Evaluating Human Perception and Bias in AI-Generated Humor

Narendra Nath Joshi
Adobe Inc
San Jose, CA, USA
joshinarendranath@gmail.com

Abstract

This paper explores human perception of AI-generated humor, examining biases and the ability to distinguish between human and AI-created jokes. Through a between-subjects user study involving 174 participants, we tested hypotheses on quality perception, source identification, and demographic influences. Our findings reveal that AI-generated jokes are rated comparably to human-generated ones, with source blindness improving AI humor ratings. Participants struggled to identify AI-generated jokes accurately, and repeated exposure led to increased appreciation. Younger participants showed more favorable perceptions, while technical background had no significant impact. These results challenge preconceptions about AI's humor capabilities and highlight the importance of addressing biases in AI content evaluation. We also suggest pathways for enhancing human-AI creative collaboration and underscore the need for transparency and ethical considerations in AI-generated content.

1 Introduction

Advancements in generative artificial intelligence have opened up new avenues in creative expression. These powerful language and content generation models produce remarkably human-like text, images, audio, and code. One particularly intriguing application is its ability to generate humorous content.

Humor is a fundamental aspect of human communication and interaction. It serves various social and psychological functions, from facilitating bonding and group cohesion to reducing stress and diffusing tense situations (Martin and Ford, 2018). Psychologists have long studied the role of humor in human development, cognition, and emotional expression (Berger, 2014). Humor improves mood, enhances creativity, and fosters feelings of empathy and trust between individuals (Kuiper and Nicholl, 2004).

Humor is a uniquely human trait that has been beyond the capabilities of machines until recently. However, the latest breakthroughs in natural language processing, neural networks, and large language models challenge this assumption. Generative artificial intelligence systems can now be trained on vast repositories of human-created humor, from witty one-liners to elaborate comedic sketches. By identifying patterns, analyzing the structure of humor, and learning to mimic the creative processes of human comedians, these models can generate original humorous content that often surprises and delights its audience.

Generating, understanding, and appreciating humor requires complex cognitive processes, including pattern recognition, perspective-taking, and juxtaposing incongruous concepts (Veale, 2004). Previous research has highlighted humor's nuanced and context-dependent nature, with cultural norms, personal experiences, and social dynamics all playing a role in an individual's humorous sensibilities (Polimeni and Reiss, 2006). Exploring how generative artificial intelligence systems can capture and replicate these multifaceted elements of human humor is a fascinating and challenging area of inquiry.

However, generating high-quality humor remains a significant challenge for current systems. Humor is a complex and subjective phenomenon, often relying on cultural references, contextual understanding, and the ability to surprise and delight the audience (Ritchie, 2009). Existing generative artificial intelligence models may need help to capture the full depth and nuance of human humor, leading to humor perceived as generic or lacking in authenticity (Augello et al., 2008).

Furthermore, it is crucial to understand how human perception and biases influence the evaluation of AI-generated humor. Humans may have preconceived notions or skepticism about machines' ability to generate genuinely humorous content,

Proceedings of the 1st Workshop on Computational Humor (CHum), pages 79–87
January 19, 2025.

which could lead to biased assessments.

As generative artificial intelligence advances, it becomes increasingly important to understand its impact on various domains, including the creative arts and human-computer interaction. In this work, we explore the current state of generative artificial intelligence and its humor applications. We examine the human perception of humor created by generative artificial intelligence.

Building upon these foundational concepts and challenges, this study aims to investigate specific hypotheses and research questions regarding human perception of AI-generated humor. Our research is guided by the following hypotheses and corresponding research questions:

H1 Humans believe they can reasonably identify if humor is AI-generated. **(Reasonable Identification Hypotheses)**

- **H1a** Participants' accuracy in identifying AI-generated jokes is higher than chance.

H2 Humans have reasonable doubt in AI's abilities to generate quality humor. **(Reasonable Doubt Hypotheses)**

- **H2a** Humans rate AI-generated jokes lower in quality compared to human-generated jokes.
- **H2b** The perceived quality of AI-generated jokes improves when participants are unaware of the source.

H3 Bias towards AI-generated humor changes with exposure. **(Repeated Exposure Hypotheses)**

- **H3a** Participants' ratings of AI-generated jokes improve after repeated exposure.

H4 Demographic factors influence perception of AI-generated humor. **(Demographic Hypotheses)**

- **H4a** Younger participants rate AI-generated jokes higher than older participants.
- **H4b** Participants with a background in technology or AI are more accepting of AI-generated humor.

To investigate these hypotheses, we formulated the following research questions:

RQ1: Are participants able to accurately identify AI-generated jokes more often than by chance? *(tests Hypothesis H1a)*

RQ2: Do humans rate AI-generated jokes lower in quality compared to human-generated jokes? *(tests Hypothesis H2a)*

RQ3: Does the perceived quality of AI-generated jokes improve when participants are unaware of the source? *(tests Hypothesis H2b)*

RQ4: Do participants' ratings of AI-generated jokes improve after repeated exposure? *(tests Hypothesis H3a)*

RQ5: Do younger participants rate AI-generated jokes higher than older participants? *(tests Hypothesis H4a)*

RQ6: Are participants with a background in technology or AI more accepting of AI-generated humor? *(tests Hypothesis H4b)*

Through a carefully designed user study, we aim to address these research questions and test our hypotheses, contributing to the understanding of human perception and bias in the context of AI-generated humor.

2 Related Work

The field of computational humor has evolved from early rule-based systems to more sophisticated data-driven approaches powered by modern machine learning techniques. Researchers have increasingly focused on understanding human perception and biases towards AI-generated humor.

Previously, researchers primarily focused on rule-based systems that attempted to capture the logical structures and linguistic patterns underlying humorous expressions (Binsted, 1996 and Ritchie, 2001). These systems relied on pre-defined rules and templates to generate puns, jokes, and other forms of humor. However, they were often limited in their ability to adapt to the nuances and complexities of human humor, which can be highly context-dependent and subjective.

As the field of artificial intelligence advanced, researchers began exploring the use of machine learning algorithms to generate humor in a more data-driven manner. Mihalcea and Strapparava, 2005 developed one of the early data-driven systems, which utilized semantic relationships and linguistic features to identify and generate humorous one-liners. This approach showed promise but needed higher quality and semantically diverse training data. Valitutti et al., 2016 developed a system that could produce puns and other forms of wordplay by exploiting linguistic patterns and semantic relationships. Similarly, Winters et al., 2019 presented a general framework for computational humor that learns joke structures and parametrization from rated example jokes by learning from datasets of human-created humor. These findings suggest that modern machine learning systems can recognize and replicate the nuances of humor.

The recent advancements in large language models have further expanded the capabilities of computational humor generation. LLM-based approaches, such as those leveraging GPT-3 (Brown, 2020) or other transformer-based models, have demonstrated impressive performance in generating coherent and contextually relevant humor. These models are trained on vast amounts of text data, allowing them to capture more nuanced linguistic patterns and common-sense knowledge that can be leveraged for humor generation (Hossain, 2020).

Alongside these advancements in computational humor generation, researchers have also begun to explore the importance of understanding human perception and biases towards AI-generated humor. Humor is a highly subjective and complex phenomenon, often relying on cultural references, contextual understanding, and the ability to surprise and delight the audience (Ritchie, 2009).

3 Methodology

3.1 Study Design

We conducted a between-subjects experimental study to investigate human perception of AI-generated humor and potential biases in evaluation. The study was designed to examine how knowledge of a joke's source (human or AI) influences perception, and how different presentation contexts affect evaluation accuracy and bias.

Participants were randomly assigned to one of six experimental groups, each designed to test specific aspects of humor perception and source identification:

- **Group A (Human Baseline)** Participants evaluated only human-generated jokes, establishing a baseline for humor quality ratings and identification accuracy.
- **Group B (AI Baseline)** Participants evaluated only AI-generated jokes, allowing assessment of perceived quality and identification accuracy 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.

3.2 Participant Selection and Demographics

We recruited 193 total participants from Amazon Mechanical Turk (Paolacci et al., 2010). Participating workers received a \$5.00 compensation based on an estimated work of 30 minutes for a projected wage of \$10 (US federal minimum wage is \$7.25). The workers provided informed consent before completing the study. After completing the task, participants also answered questions about demographics and prior experience with Mechanical Turk.

We performed several integrity checks for our participants. Similar to prior studies deployed on Mechanical Turk (Ashktorab et al., 2021), we excluded workers whose mean rating time was less than 3 seconds and removed workers who had uniform responses ($\sigma < 5$) in the rating responses. We also removed individuals from the study who fell outside of the mean $\pm$2SD statistic for each of the dependent variables. This process left us with 174 participants.

Demographic data was collected for:

- **Gender** (Male, Female, Non-binary, Prefer not to say, Other)
- **Age Range** (18-24, 25-34, 35-44, 45-54, 55+)
- **Experience with AI technologies** (5-point scale from "Never use it" to "Deep understanding")

3.3 Stimulus Selection and Preparation

3.3.1 Human-Generated Jokes

Human-generated jokes were sourced from (Phillips, 2013), curated by a panel of three independent raters to ensure consistent quality and appropriate content. Selected jokes represented various humor styles (wordplay, observational, situational) while controlling for potentially confounding variables such as length and complexity. The jokes were completely text-based.

3.3.2 AI-Generated Jokes

AI-generated jokes were created using Claude 3.5 Sonnet (Anthropic, 2024), with consistent prompting techniques to ensure comparable quality and style variety. The jokes underwent the same rating and filtering process as human-generated jokes to maintain experimental control. The jokes were completely text-based.

3.4 Experimental Procedure

3.4.1 Joke Presentation

Each participant evaluated twenty-five jokes in their assigned condition. In trials, we found twenty-five jokes to be appropriate as an increased number of jokes could lead to potential cognitive fatigue and a drop in study experience and quality of results. Jokes were presented individually in randomized order (except for Group C's alternating sequence) to control for order effects.

3.4.2 Rating Sessions

The experiment consisted of two phases:

1. **Initial Rating Phase**: All participants rated jokes on a 5-point scale (Very funny to Not funny at all).
2. **Source Assessment Phase**: Group E performed additional tasks:
   - Source identification (Human/AI)
   - Confidence ratings (5-point scale)

4 Results

4.1 Overall Analysis

Our analysis reveals several significant patterns in how humans perceive and evaluate AI-generated humor. We present our findings organized by research questions, incorporating both quantitative metrics and qualitative observations. The results challenge several preconceptions about AI-generated humor while confirming others, particularly regarding demographic influences and exposure effects.

4.2 Source Assessment Performance (RQ1)

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

- Accuracy: 0.43034
- Mean Confidence: 3.892

The near-chance accuracy rates suggest that distinguishing between human and AI-generated humor has become increasingly challenging. High-quality AI-generated jokes, in particular, were frequently misattributed to human authors with high confidence, indicating significant advancement in AI's ability to generate natural-seeming humor.

4.3 Quality Perception and Source Bias

4.3.1 Comparative Quality Ratings (RQ2)

Contrary to initial expectations, our analysis shows that participants do not rate AI-generated jokes significantly lower in quality compared to human-generated jokes (μ_{AI} = 2.97393, σ = 1.4137; μ_{human} = 2.94769, σ = 1.4046).

This finding is particularly noteworthy given the common assumption that AI-generated content would be perceived as inferior to human-created content. The ratings distribution shows considerable overlap between the two sources, with AI-generated jokes occasionally receiving higher scores in categories such as wordplay and situational humor.

4.3.2 Source Awareness Effects (RQ3)

The blind testing condition (Group E) demonstrates significantly different ratings compared to the informed condition (Group F):

- Blind condition: μ = 3.34064, σ = 1.1056
- Informed condition: μ = 2.92737, σ = 1.4164
- Statistical significance: [t(df) = 6.04, p = $1.78e-09$]

The data reveals a clear pattern of bias when participants are informed about the source. In the blind condition, participants evaluated jokes primarily on their inherent humor value, leading to more favorable ratings for AI-generated content. This suggests that preconceptions about AI capabilities may influence judgment more than actual content quality.

Group	Mean Rating (First 7 Jokes)	Mean Rating (Last 7 Jokes)
A	3.02721	2.84354
B	2.88095	3.15476
C	2.90043	3.06494
D	3.21693	3.56085
E	2.99078	3.47926
F	2.91353	2.96617

Table 1: Mean Humor Ratings for First 7 and Last 7 Jokes by Group

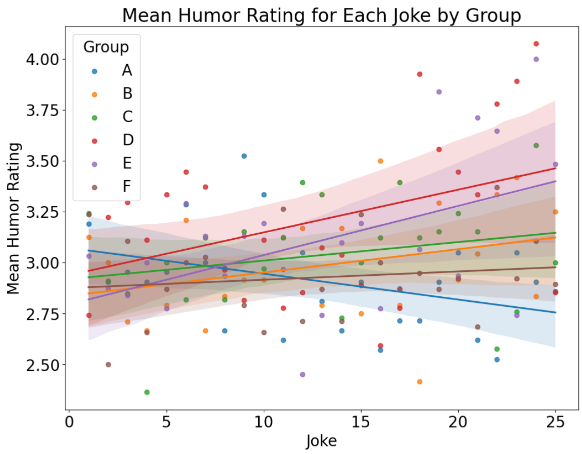

Figure 1: Mean Humor Rating for each Joke for each Group. Note that Group A is only human-generated jokes

4.4 Exposure and Learning Effects (RQ4)

Analysis of rating progression shows a significant upward linear trend in groups with AI-generated jokes (Groups B, C, D, E and F) as shown in Figure 1. The improvement in ratings over time as shown in Table 1 suggests a familiarization effect, where initial skepticism gives way to increased appreciation of AI-generated humor. This trend is particularly evident in the mixed presentation sequence group D and the blind AI test group E.

4.5 Demographic Influences

4.5.1 Age-Related Effects (RQ5)

Results reveal significant age-group differences as shown in Figure 2. The age-related differences in ratings show a clear generational pattern across different groups, with younger participants demonstrating more openness to AI-generated humor. This trend remains consistent across different joke types and presentation formats.

4.5.2 Technical Background Impact (RQ6)

Analysis of variance indicates no significant differences based on AI expertise across groups as shown in Figure 3. Participants with technical backgrounds did not provide more favorable ratings but showed a more nuanced appreciation for AI-generated humor, often mentioning technical aspects in their qualitative feedback.

4.6 Qualitative Observations

Participant feedback revealed several recurring themes. Many expressed initial skepticism, with some noting, *"It's strange to think of AI 'trying' to be funny when it doesn't actually experience humor. I'm not sure it will ever truly understand what makes people laugh"* (Participant 36) and *"The jokes were clever enough, but there's something unsettling about humor coming from a machine. It's missing the human touch"* (Participant 71). However, others were surprised by the quality of the AI-generated humor, with comments such as, *"It's impressive how well the AI captured the timing and wit usually found in human jokes. I wouldn't have guessed it was machine-generated"* (Participant 64) and *"I was genuinely surprised that an AI could come up with something this funny! I didn't expect it to pick up on such subtle humor"* (Participant 146). Finally, some participants noted recognizing patterns in the AI's humor, with one stating, *"I noticed a lot of the humor felt very structured, almost too perfect. I think the AI relies on patterns that work, but it doesn't quite get the unpredictability that human humor has"* (Participant 21), and another remarking, *"When I started paying attention, I could tell the AI was using patterns that were too precise. It didn't have the imperfections or unpredictable elements that make human humor feel fresh"* (Participant 167).

These qualitative insights provide context for the quantitative findings and highlight the complex nature of human perception of AI-generated content.

4.7 Summary of Key Findings

Our results reveal several important patterns. AI-generated jokes receive comparable ratings to human-generated ones, challenging preconceptions about AI's humor capabilities. Source blindness significantly improves perception of AI-generated humor, indicating the presence of implicit bias. Participants demonstrate poor ability to distinguish between AI and human-generated jokes despite high confidence. Repeated exposure leads to improved

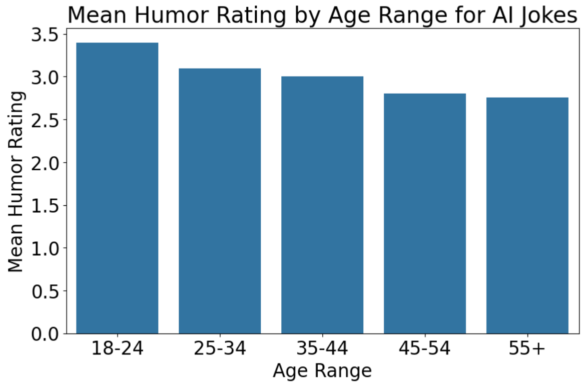

Figure 2: Mean Humor Rating by Age Range for AI-Generated Jokes

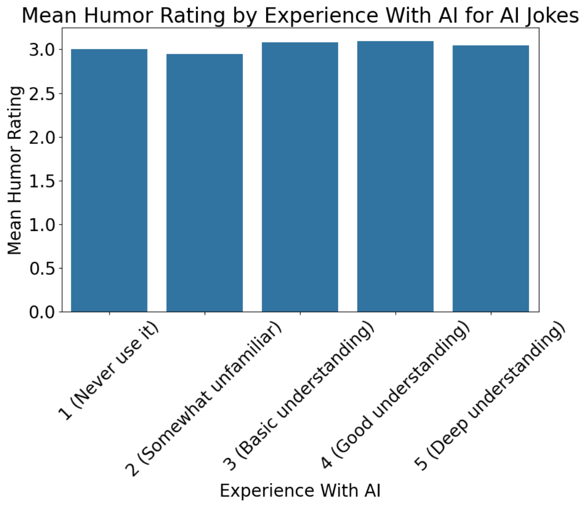

Figure 3: Mean Humor Rating by Experience with AI for AI-Generated Jokes

ratings and reduced bias. Younger participants show more favorable perceptions of AI-generated humor. Participants with technical backgrounds do not demonstrate more favorable perceptions of AI-generated humor.

5 Discussion

5.1 Human Ability to Distinguish AI Content

The inability of participants to accurately identify AI-generated jokes beyond chance levels (RQ1) has significant implications for the sophistication of current AI humor generation systems, especially when reported confidence is high. It also raises the potential need for AI disclosure in creative content, especially as the line between human and AI-generated works continues to blur. Furthermore, this finding highlights the future of human-AI creative collaboration and the ethical considerations surrounding the attribution of AI-generated content.

5.2 Evolution of AI Humor Perception

Our findings challenge the prevalent assumption that humans inherently prefer human-generated humor over AI-generated content. The comparable ratings between AI and human-generated jokes (as shown in RQ2) suggest that AI systems have reached a significant milestone in generating contextually appropriate and genuinely amusing content. This advancement reflects the sophisticated natural language processing capabilities of modern AI systems, particularly in understanding and replicating the nuanced patterns that make content humorous.

5.3 The Role of Source Bias

The improvement in perceived quality when the source is unknown (RQ3) reveals a crucial insight into human cognitive bias. This "source bias" effect demonstrates how preconceived notions about AI capabilities can influence humor appreciation. The disconnect between blind and informed ratings suggests that humans may hold unconscious biases against AI-generated content. These biases can significantly impact their evaluation of creative content, with the quality of AI-generated humor often being systematically undervalued when its source is known.

5.4 Demographic and Exposure Effects

The improvement in ratings with repeated exposure (RQ4) indicates a learning effect that could have important implications for AI content integration strategies. It suggests that public acceptance of AI-generated creative works may increase over time, as audiences become more familiar with the technology. This improvement also points to the potential for long-term shifts in perception, as repeated exposure helps individuals to better understand and appreciate AI-generated content.

5.5 Generational Differences

The observed age-related differences in humor appreciation (RQ5) reflect broader patterns in technology adoption and acceptance. Younger participants' higher ratings of AI-generated humor suggest a generational shift in attitudes toward AI-generated content. These differences may indicate potential future trends in AI acceptance, with younger generations being more open to embracing AI-driven creative works. Moreover, the role of early exposure to technology in shaping these perceptions cannot be overlooked.

5.6 Technical Literacy Impact

While technical background does not show a direct correlation with humor appreciation (RQ6), this finding still provides valuable insights into how knowledge may influence perception. It suggests that understanding AI systems could potentially reduce skepticism toward AI-generated content, even though such a relationship is not evident in the data. Additionally, education about AI capabilities may play a role in influencing public reception of AI-generated works. While no clear connection between technical literacy and humor appreciation is found, it remains a potential area for future exploration, especially in terms of how technical knowledge might shape bias against AI-generated content.

5.7 Technical Implications

Our findings suggest several key implications for the development of humor-generating AI systems. One crucial factor is the importance of context awareness in humor generation, ensuring that AI systems can understand and adapt to the nuances of different situations. The need for diverse training data is also highlighted, as it could help AI systems better appeal to various demographics, accounting for differences in humor preferences. Additionally, incorporating user feedback mechanisms into the design of AI systems may enhance their ability to tailor humor more effectively. Lastly, maintaining stylistic consistency in AI-generated humor is valuable, as it can create a more coherent and relatable experience for the audience.

5.8 Societal Implications

The study's findings also have broader societal implications, particularly in the context of AI's role in everyday life. As AI continues to integrate into various aspects of our society, it will influence not only the entertainment and media sectors but also broader cultural perceptions of creativity and originality. Understanding these dynamics can help shape policies and frameworks that govern the use and development of AI technologies in socially sensitive areas.

5.9 Creative Industry Impact

The implications of our findings extend to the future of AI in creative industries. One key aspect is the evolving role of AI in content creation, which could redefine how creative works are produced. The potential for human-AI collaborative content creation also stands out, with AI augmenting human creativity in new and innovative ways. The future of entertainment and media production will likely see a blending of human and AI-generated content, which could change how audiences consume creative works. Additionally, this shift may influence employment and skill requirements in creative fields, as new roles emerge to manage and work alongside AI systems.

5.10 Design Considerations

The study suggests several design principles that could guide future AI humor systems. Transparency in source attribution is important, ensuring users are aware of whether content is AI-generated. Adapting to user preferences and feedback is another key principle, allowing AI systems to evolve and better align with individual tastes over time. Incorporating cultural and contextual awareness will also be crucial in making AI-generated humor more relevant and relatable. Finally, a balance between novelty and familiarity is essential, as AI humor should be fresh and surprising without straying too far from what audiences find familiar and enjoyable.

5.11 Ethical Considerations

Our study raises several important ethical questions regarding AI-generated content. One of the main concerns is the need for transparency, particularly in the disclosure of AI-generated content. This disclosure is vital in maintaining trust and ensuring that audiences are informed about the nature of the material they engage with. The potential impact of AI on human creativity and expression also warrants attention, as it may alter how people perceive and value human-created art. Furthermore, the role of AI in shaping cultural narratives is a significant ethical consideration, as AI could influence societal values and perceptions. Lastly, preserving human agency in the creative process is essential, ensuring that AI serves as a tool to augment human creativity, not replace it.

5.12 Limitations and Future Research

5.12.1 Study Limitations

There are some limitations to consider in this study. First, the specific context and timing of the research may influence the findings, as perceptions of AI-generated content can evolve over time. Additionally, the limited range of humor styles tested in the study means that the findings may not fully reflect the diversity of humor types that exist. Sampling biases in participant selection could also affect the generalizability of the results. Finally, the rapid evolution of AI capabilities means that the study's conclusions may need to be revisited as new advancements emerge.

5.12.2 Future Research Directions

Future research should address several areas to further explore the impact of AI on humor appreciation and content creation. One potential direction is studying long-term changes in perception over extended exposure to AI-generated content, which could reveal how familiarity affects audience reception. Cross-cultural variations in AI humor appreciation should also be examined, as different cultures may have distinct humor preferences. Additionally, the impact of various AI disclosure methods on audience reactions warrants further investigation. Future studies could also explore the role of personalization in AI humor generation, examining how tailored content influences user satisfaction. Finally, the influence of different humor styles and contexts on AI-generated content should be explored to better understand how AI systems can cater to diverse comedic tastes.

6 Conclusion

Our findings suggest that the relationship between humans and AI-generated humor is more complex and nuanced than previously understood. As AI systems continue to evolve, the distinction between human and AI-generated content becomes increasingly subtle, challenging our preconceptions about creativity and humor. The study highlights the importance of understanding and addressing human biases in the development and deployment of AI creative systems, while also suggesting potential pathways for improving human-AI creative collaboration.

References

Anthropic. 2024. Claude. https://www.anthropic.com. Version 3.5 Sonnet.

Zahra Ashktorab, Michael Desmond, Josh Andres, Michael Muller, Narendra Nath Joshi, Michelle Brachman, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Christine T Wolf, et al. 2021. Ai-assisted human labeling: Batching for efficiency without overreliance. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–27.

Agnese Augello, Gaetano Saccone, Salvatore Gaglio, and Giovanni Pilato. 2008. Humorist bot: Bringing computational humour in a chat-bot system. In *2008 international conference on complex, intelligent and software intensive systems*, pages 703–708. IEEE.

Peter L Berger. 2014. *Redeeming laughter: The comic dimension of human experience*. Walter de Gruyter GmbH & Co KG.

Kim Binsted. 1996. Machine humour: An implemented model of puns.

Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Nabil Tarique Hossain. 2020. *Creative natural language generation: humor and beyond*. University of Rochester.

Nicholas A Kuiper and Sorrel Nicholl. 2004. Thoughts of feeling better? sense of humor and physical health.

Rod A Martin and Thomas Ford. 2018. *The psychology of humor: An integrative approach*. Academic press.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538.

Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419.

Bob Phillips. 2013. *The World's Greatest Collection of Clean Jokes*. Harvest House Publishers.

Joseph Polimeni and Jeffrey P Reiss. 2006. The first joke: Exploring the evolutionary origins of humor. *Evolutionary psychology*, 4(1):147470490600400129.

Graeme Ritchie. 2001. Current directions in computational humour. *Artificial Intelligence Review*, 16:119–135.

Graeme Ritchie. 2009. Can computers create humor? *AI Magazine*, 30(3):71–71.

Alessandro Valitutti, Antoine Doucet, Jukka M Toivanen, and Hannu Toivonen. 2016. Computational generation and dissection of lexical replacement humor. *Natural Language Engineering*, 22(5):727–749.

Tony Veale. 2004. Incongruity in humor: Root cause or epiphenomenon?

Thomas Winters, Vincent Nys, and Danny De Schreye. 2019. Towards a general framework for humor generation from rated examples. In *ICCC*, pages 274–281.