The Theater Stage as Laboratory:
Review of Real-Time Comedy LLM Systems for Live Performance

Piotr Mirowski
Improbotics
London, UK
piotr.mirowski@computer.org

Boyd Branch
Improbotics
Coventry, UK

Kory Mathewson
Improbotics
Montréal, Canada

Abstract

In this position paper, we review the eclectic recent history of academic and artistic works involving computational systems for humor generation, and focus specifically on live performance. We make the case that AI comedy should be evaluated in live conditions, in front of audiences sharing either physical or online spaces, and under real-time constraints. We further suggest that improvised comedy is therefore the perfect substrate for deploying and assessing computational humor systems. Using examples of successful AI-infused shows, we demonstrate that live performance raises three sets of challenges for computational humor generation: 1) questions around robotic embodiment, anthropomorphism and competition between humans and machines, 2) questions around comedic timing and the nature of audience interaction, and 3) questions about the human interpretation of seemingly absurd AI-generated humor. We argue that these questions impact the choice of methodologies for evaluating computational humor, as any such method needs to work around the constraints of live audiences and performance spaces. These interrogations also highlight different types of collaborative relationship of human comedians towards AI tools.

1 Introduction

Attempting to combine humor with machine intelligence is a long-standing subject of scientific enquiry and it is perceived as a fundamental challenge (Raskin, 1979). Amin and Burghardt (2020), Veale (2021) and Sharples and y Pérez (2022) provide authoritative reviews of this nascent field, which can be supplemented with examples of recent works that rely on large language models (Winters et al., 2018; Toplyn, 2023; Chen et al., 2023; Jentzsch and Kersting, 2023; Tikhonov and Shtykovskiy, 2024).

According to several computational humor researchers including Winters (2021), *"humans are the only known species that use humor for making others laugh. Furthermore, every known human civilization also has had at least some form of humor for making others laugh"* (Caron, 2002; Gervais and Wilson, 2005). This observation is often extrapolated into the assertion that humor remains an elusive goal for AI (in the same vein, researchers in computational storytelling have defined improvisational storytelling as a *grand challenge* for AI (Martin et al., 2016)). According to recent surveys (Mirowski et al., 2024), this skeptical view about AI's comedic potential is a strongly-held opinion shared by the wider comedy community, from actors and audiences to reviewers and journalists writing about comedy.

For this reason, we posit that comedy audiences and performance spaces are the ultimate environments to critically evaluate the quality of systems for the computational generation of humor. Even though some studies like (Gorenz and Schwarz, 2024) have evaluated AI-generated humor using crowd-sourced workers, human-computer interaction researchers have raised concerns about the poor quality of crowd-sourced evaluation of open-ended text generation (Karpinska et al., 2021), which can be attributed to lack of buy-in from evaluators, and to missing context. The setup of live professional comedy invites **paying audiences** expecting to have a good time, and **critical reviewers** judging the overall performance: it thus provides with a realistic and challenging testbed for computational humor systems. Moreover, the live nature of comedy performance creates rich, **interactive exchanges** between comedians and audiences, which—unlike pure online evaluation—allows comedians (and comedy generation systems) to incorporate **real-time feedback** and rich sensory and **cultural context** in the **safe** environment of theater.

At the intersection of live theater and humor sits improvisational theater and comedy (Johnstone, 1979), a complex theatrical art form that can be

traced back to Rhapsodes in Ancient Greece or to Commedia dell'Arte (Lea, 1934; Collins, 2001; Mathewson, 2019). Because it relies on natural human interaction and demands constant adaptation to an evolving context, theatrical improvisation (like jazz) has been qualified as *"real-time dynamic problem solving"* (Magerko et al., 2009; Johnson-Laird, 2002). According to Mathewson and Mirowski (2018) *"improv requires performers to exhibit acute listening to both verbal and non-verbal suggestions coming from the other improvisers, split-second reaction, rapid empathy towards the other performers and the audience, short- and long-term memory of narrative elements, and practiced storytelling skills"*, making it a highly interesting challenge for AI systems. Interestingly, theatrical improvisation encourages risk taking and experimentation, and it even "celebrates failure" thanks to a tacit agreement between improvisers and audiences who acknowledge the challenge of making up comedic material live on the stage. The improv stage thus provides a "safe" environment to test technological tools like artificial intelligence.

What follows is a literature and performance review of the state of the art of computational humor systems deployed in real-time in front of live audiences, whether in physical or in virtual spaces. We group these according to the type of scientific or artistic questions that they raise, starting with questions around robotic **embodiment** of chatbots, anthropomorphism and **competition** between humans and machines (Section 2), and questions around **liveness**, **timing** and utility in the artistic process (Section 3). We address the human **interpretation** and **justification** of seemingly absurd AI-generated humor (Section 4) and finish with a discussion on how the constraints of live audiences and performance spaces impact the choice of methodologies for evaluating computational humor (Section 5). We therefore suggest that the setting of live performance allows to define collaborative relationships between human comedians and AI tools.

2   Robot comedy as a test of humanity

As introduced in the previous section, a commonly held belief is that humor is seen as the last frontier of intelligence (Winters, 2021). Robot comedy can then be seen as a challenge to humanity itself[1].

---

[1] This prompted comedy critic Logan (2023) to unwittingly center comedy over other art forms: *"unlike music and visual art, comedy can't be easily reduced to an algorithm." (sic)*

2.1   Can robots deliver comedy on stage?

Robot embodiment presents with unique challenges of audience reception. Some robotics and theater practitioners like Hiroshi Ishiguro and Oriza Hirata took the route of anthropomorphism (Pluta, 2016) to make the robot presence as human-like as possible, whereas others like Tom Sgouros built a custom robotic arm[2] or even, like Annie Dorsen, simply staged two laptops "talking" philosophy[3].

In 2010, social roboticist Prof. Heather Knight[4] pioneered staged comedy performances with an Aldebaran Nao[5] robot delivering human-written comedy and gathering audience feedback thanks to camera sensors that track audience sentiment following each line delivered by the robot, and used this information to modify next joke selection based on audience feedback (Knight et al., 2011). Starting in 2014, Austin, Texas-based multidisciplinary artist and engineer Arthur Simone staged toy-like humanoid robots to be his partners in improvised theater performances: *Bot Party: Improv Comedy with Robots*[6], thus investigating how to improvise with a robot.

In 2016, theater improvisers and robotics researchers Dr. Piotr Mirowski and Dr. Kory Mathewson from duo *HumanMachine*[7] developed large language models (Sutskever et al., 2014) for improvisational comedy (Mathewson and Mirowski, 2017a). Unlike previous, rule-based AI methods geared at generating comedy, they trained general conversational models (Vinyals and Le, 2015) trained on OpenSubtitles (Tiedemann, 2009). The language model was coupled with speech recognition to listen to their human partner, text-to-speech and text-dependent robotic control to operate a small scale robot such as the Nao or EZ-Robot JD Humanoid[8]. Comedy derived from the human actor attempting to justify whatever the robot said.

Some of those robot performances incorporated implicit audience feedback (Knight et al., 2011; Mathewson and Mirowski, 2017a), but we hypothesize that audiences may have evaluated the novelty of the premises of those shows in addition to their comedic quality.

---

[2] https://sgouros.com/judy/
[3] https://anniedorsen.com/projects/hello-hi-there/
[4] https://www.ted.com/talks/heather_knight_silicon_based_comedy
[5] https://www.aldebaran.com/en/nao
[6] https://www.botparty.org/
[7] https://humanmachine.live
[8] https://www.ez-robot.com/

2.2 Computational humor presented as a competition between humans and machine

The recent rapid deployment of AI in the creative fields has raised ethical issues around the cannibalization of creative economies (Frosio, 2023) and the lack of consent in how training data for AI was obtained (Zhong et al., 2023). As a consequence, the public debate around AI is currently driven by the fear of replacement; as illustrated below, performance artists engaging AI ask the question whether AI-generated humor can ever *match* human level.

In 2023, and in the context of public releases of generative AI tools, and of their subsequent short-term impact on creative industries (contributing to the Writers Guild of America (WGA) labour action), Los Angeles-based comedians Allisson Goldberg and Brad Einstein created *Comedians vs. AI: Stage Against the Machine*[9]. Their show featured two teams of comedians, one "human" relying only on their skills, another one supported by Gen AI software like ChatGPT and DALL-E. The show evaluated AI in an adversarial context, pitting one team against another, and promising the audiences reassurance about limited capabilities of the machines; to quote one comedian, "We have the benefit of having trauma and life experience to pull from that AI doesn't have integrated yet, and that makes us more dynamic and sensitive and hilarious for now." (Jamerson, 2023).

In that same year of 2023, New York-based Comedy Bytes[10] refined this concept to focus on improvised *rap battles* and *comedy roasts* between a small cast of human performers, and cartoonish virtual avatars puppeteered by actors or text-to-speech, reading AI-generated jokes (Tett, 2023).

Other improv performances built around adversarial human-AI relationships include *The AI Improv Show* (2023) by London improv school The Free Association[11] (featuring ChatGPT-generated jokes) and Amsterdam-based Boom Chicago who produced *The Future Is Here... And It Is Slightly Annoying*[12] (2019) with improv sketches involving a tele-presence robot on wheels connected to a chatbot developed by Botnik Studios[13].

2.3 Can an AI Pass the Comedy Turing Test?

Building upon the idea of human-AI comparison, improv duo *HumanMachine* adapted in 2017 the Turing test (Turing, 1950) and introduced its comedic counterpart (Mathewson and Mirowski, 2017b). They assembled in 2018 a large-cast improv troupe, *Improbotics*[14], featuring human actors, some of whom (called *Cyborgs*) get lines from AI via headphones connected to a portable FM radio receiving lines transmitted from the AI chatbot's text-to-speech. Over hundreds of performances, the troupe has devised diverse short-form and long-form improv games featuring the Cyborg in disclosed or concealed identity. In addition to evaluating audiences' perception towards AI, the troupe evaluates audiences' perception of its language capabilities: they devised a comedy Turing test by staging non-Cyborg actors who pretend to be controlled by AI alongside the Cyborg actors. One would expect the comedy Turing test to become harder as LLM technology develops, but the comedians invented ways to "sound like an AI" to confuse the audiences, thereby demonstrating the limitations of the Turing test.

Computational humor researcher Dr. Thomas Winters designed, with comedian Lieven Schiere, a more formal Turing test performed on the stage, and aimed at evaluating advances in large language models for writing comedy ahead of the performance (Winters, 2024).

2.4 Comedic deception of audiences by AI

The idea of deception has been explored in game contexts beyond the Turing test. In 2023, filmmaker Dr. Manuel Hendry designed a dark comedic installation *The Feeling Machine*[15], where a chatbot-powered, ELIZA-inspired "psychotherapist" is embodied by an animated mask: once that "psychotherapist" establishes a rapport with an individual spectator (Hendry et al., 2023), the system provocatively shows a deep fake of that spectator making up false memories, to raise questions about misuses of technology. *The Feeling Machine* targets art museum audiences acquainted with ethical discussions around AI; at the opposite side of the spectrum sits a general audience show made by TV company Endemol Italy and presented in 2023: *Fake Show: Diffidate delle imitazioni*[16], an impro-

---

[9] https://www.comediansvsai.com/
[10] https://www.comedybytes.io/
[11] https://www.thefreeassociation.co.uk/
[12] https://boomchicago.nl/shows/the-future-is-here/
[13] https://botnik.org/

[14] https://improbotics.org
[15] https://www.hendry.me/
[16] https://www.raiplay.it/programmi/fakeshowdiffidatedelleimitazioni

visational game show featuring deep fakes.

Company *Improbotics* adapted that concept in 2024: they comically explore alternative life choices of a consenting audience member, acted out by different improvisers who drive live-generated deep fakes (Glennon, 2024).

3 Live performance and real-time interaction as a test for generative AI

The commonality behind the shows presented in Section 2 was that they addressed ethical interrogations about the role of AI in comedy. In this section, we review shows that investigate how to effectively deliver computational humor on stage.

3.1 AI co-creating real-time comedy dialogue

The development of large language models and conversational AI applications mostly focuses on single-user text-based dialogue. Speech recognition and dialogue systems struggle with Multi-Party Chat (MPC). Branch et al. (2024) describe how they approached this problem in *Improbotics* performances, where multiple actors interact with an AI *Cyborg* stage partner, just like in a traditional improv comedy show featuring a large cast in a lively performance. Instead of simple turn-taking in human-chatbot dialogue, the troupe resorts to human-in-the-loop curation of continuously AI-generated lines, where the most comedic or relevant lines are sent to the Cyborg performer via an earpiece; introducing a second performer who takes responsibility for selecting the AI-generated line creates a "writer's room" setup and introduces two levels of human interpretation of AI-generated material. In their 2024 performances at the Edinburgh Festival Fringe, the troupe replaced earpieces by augmented-reality glasses, delegating the role of AI line curation to the Cyborg performer, who would simultaneously read some of the AI-generated lines and try to maintain eye contact with stage partners.

Timing! The most important rule of comedy is... *Improbotics* needed to design both technology (fast speech recognition and language models, and asynchronous generation) to accelerate the robot's responses (Branch et al., 2024), and dramaturgy ("slow-burn" improv scenes relying on non-language communication to fill the time lags) (Mathewson and Mirowski, 2018). Improviser Cyborgs expressed they had struggled with AI generated lines because of slow timing and delays; their perception was that the audience preferred timely responses to higher quality but delayed responses.

In Oregon, Prof. Naomi Fitter focused on comedy timing as she has been running since 2019 regular comedy nights where her robot comedian Jon relies on audience laughter to control the timing and delivery of jokes (Srivastava and Fitter, 2021).

3.2 AI for inspiration and world building

Liveness in AI improv shows is not limited to dialogue: human actors can leverage AI-generated ideas for real-time performance. Notably, San Francisco-based Alexa Improvise[17] has used an AI assistant for game ideas since 2018; *Yes, Android* by Toronto company *Bad Dog*[18] featured actors reading LLM-generated lines in 2017; Nouméa-based *La Claque*[19] incorporated French-language AI for short-form improv suggestions in 2023; and India-UK troupe *ClimateProv* leveraged Gen AI to inspire climate-themed improvisation in 2022. Winters and Mathewson (2019) designed automatic slide generators[20] for *Powerpoint karaoke* games.

Several projects explored LLMs for long-form improvisation by supporting storytelling. Notably *Plays by Bots*, staged since 2022 by Edmonton-based *Rapid Fire Theatre*[21], rely on scripts co-written with *Dramatron* (Mirowski et al., 2023) to build the world for improvisers; and in 2021, *Improbotics* used an AI as narrator for long-form scenes (Branch et al., 2021b).

Finally, and while they do not use AI in real time during their performance, many comedians have presented material co-written with AI in front of live audiences, including Darren Walsh[22] in 2023 and Anesti Danelis[23] in 2024 at Edinburgh Fringe.

The commonality between all those shows is to employ computational humor systems as mere writing tools to support live human performance; as a consequence, audience evaluation is focused primarily on the human performers and how they engage with their audiences.

3.3 Live performance with AI in digital spaces

The development of computer-mediated communication technology has introduced a new way for humans to congregate and redefined the notion of liveness and audience interaction. Live performance

---

[17]https://ai.nickradford.dev/
[18]https://baddogtheatre.com/
[19]https://laclaqueimpro.com/
[20]https://talkgenerator.com/
[21]https://rapidfiretheatre.com/
[22]https://darrenwalsh.co.uk/
[23]https://www.anestidanelis.com/

no longer requires a physical space, as performers and audience can congregate virtually via teleconference and chat, overcoming long geographical distances, as proved by *Failed to Render*[24], a comedy club in virtual reality, or most improv teams performing and rehearsing online during Covid-19.

Branch et al. (2023) analysed how shared VR environments and telepresence enhance improvisational flow more than traditional teleconference; a tele-immersive environment was used in 2020-2021 for VR rehearsals and performances[25] of *Improbotics*, where the AI agent was represented by an avatar (Branch et al., 2021a). Jacob et al. (2019) used computer vision models for physical improv games in *Robot Improv Circus VR*[26]. *PORTAGING* was a humorous prompt battle with Gen AI performed on a Discord channel at NeurIPS 2022[27]. In these shows, audience engagement could be measured in chat interactions during streaming and, in some cases, laughter on live audio channels.

4   AI language and human interpretation

The remaining question about computational humor systems for live performance is how they help communicate, or how they challenge human actors to make sense of AI-generated output.

On one hand, AI can be used for meaning making: multilingual improv in *Rosetta Code* is mediated by speech recognition, machine translation, and in-ear text-to-speech (Mirowski et al., 2020). Incidently, these three tools are applications that underlie the development of language models.

On the other hand, we alluded in Section 2 to human actors trying to justify "seemingly absurd" AI-generated text. Improvisers can leverage LLMs as a creative and acting challenge (Mathewson and Mirowski, 2018), and *THEaiTRE*'s scripted production of *AI: When a Robot Writes a Play* exemplifies the glitch aesthetic of involuntarily funny absurdist LLMs (Rosa et al., 2021). Absurdist theatre, however, requires supportive audiences. The *Dramatron* system (Mirowski et al., 2023) was an attempt at making AI-generated theatrical scripts sound less "absurdist", and it aimed at supporting actors by generating more coherent narratives.

More than Human, produced in 2019 by Dr. Gunter Lösel, went in the opposite direction. Its human cast (one of whom was taking lines from an LLM) did not attempt to justify those AI-generated suggestions at all. Instead, and following the principles of Dadaism, they used AI to explore and celebrate their own "inner machine" (Loesel et al., 2020; Lösel, 2024).

5   Discussion: evaluation of live AI humor

This position paper claims that audience feedback from live performances enables a challenging testbed for computational humor systems. Arguably, some comedy material is not amenable to live or improvised formats (e.g., memes, comedic videos and films) as they are pre-written and with no live audience interaction. Nevertheless, these can be assessed by measuring audience engagement on social media, in ratings or at the box office.

Human-Computer Interaction literature provides many toolboxes for assessing live audience engagement and the creative process. Branch et al. (2024) and Mathewson and Mirowski (2018) rely on audience and performer surveys after performances. Srivastava and Fitter (2021) measure audience laughter and engagement using microphones. Mirowski et al. (2024) proposed focus groups with comedians engaging in writing tasks with LLMs and assessing AI using Creativity Support Tool metrics like (Cherry and Latulipe, 2014; Chakrabarty et al., 2024): these metrics can be applied to live and improvisational contexts as well.

The fundamental advantage of framing the evaluation of computational humor in the wider context of audience reception and feedback, is that it simultaneously defines the role that AI tools can play in the wider comedy ecosystem–as creativity support tools. Such a framing encourages a collaborative relationship between human comedians and AI tools instead of an adversarial one, and helps approach the various ethical questions around AI art (and comedy in particular) on artists (and comedians) (Epstein et al., 2023; Jiang et al., 2023).

Humans have used the technologies of their time to tell stories, from cave paintings to the internet. Generative AI is one such technology, and this paper gave examples of storytellers trying to adopt it as a writing tool for performance. Humor and comedy writers can evaluate those tools through real-time human feedback, which can be uniquely provided by live theater—an ideal experimental substrate for creative technology for storytelling.

Acknowledgments

References

Miriam Amin and Manuel Burghardt. 2020. A survey on approaches to computational humor generation. In *Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 29–41.

Boyd Branch, Christos Efstratiou, Piotr Mirowski, Kory W Mathewson, and Paul Allain. 2021a. Tele-immersive improv: Effects of immersive visualisations on rehearsing and performing theatre online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Boyd Branch, Piotr Mirowski, Kory Mathewson, Sophia Ppali, and Alexandra Covaci. 2024. Designing and evaluating dialogue llms for co-creative improvised theatre. In *Proceedings of the 15th International Conference on Computational Creativity*. Association for Computational Creativity.

Boyd Branch, Piotr Mirowski, and Kory W Mathewson. 2021b. Collaborative storytelling with human actors and ai narrators. In *Proceedings of the 12th International Conference on Computational Creativity*. Association for Computational Creativity.

Boyd Branch, Piotr Mirowski, Sophia Ppali, Rocio Von Jungenfeld, Paul Allain, and Christos Efstratiou. 2023. Mirror placement matters in remote collaboration. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA. Association for Computing Machinery.

James E Caron. 2002. From ethology to aesthetics: Evolution as a theoretical paradigm for research on laughter, humor, and other comic phenomena. *Humor: International Journal of Humor Research*.

Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2024. Creativity support in the age of large language models: An empirical study involving professional writers. In *Proceedings of the 16th Conference on Creativity & Cognition*, pages 132–155.

Yuetian Chen, Bowen Shi, and Mei Si. 2023. Prompt to gpt-3: Step-by-step thinking instructions for humor generation. *arXiv preprint arXiv:2306.13195*.

Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):1–25.

Derek Collins. 2001. Improvisation in rhapsodic performance. *Helios*, 28(1):11–29.

Ziv Epstein, Aaron Hertzmann, Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. 2023. Art and the science of generative ai. *Science*, 380(6650):1110–1111.

Giancarlo Frosio. 2023. Generative ai in court. *Court (September 1, 2023). in Nikos Koutras and Niloufer Selvadurai (eds), Recreating Creativity, Reinventing Inventiveness-International Perspectives on AI and IP Governance (Routledge, 2023, Forthcoming)*.

Matthew Gervais and David Sloan Wilson. 2005. The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly review of biology*, 80(4):395–430.

Neave Glennon. 2024. Boti reviews | artificial intelligence improvisation. *Brighton on the inside*.

Drew Gorenz and Norbert Schwarz. 2024. How funny is chatgpt? a comparison of human-and ai-produced jokes. In *PLoS ONE*. OSF.

Manuel Flurin Hendry, Norbert Kottmann, Martin Fröhlich, Florian Bruggisser, Marco Quandt, Stella Speziali, Valentin Huber, and Chris Salter. 2023. Are you talking to me? a case study in emotional human-machine interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 19, pages 417–424.

Mikhail Jacob, Prabhav Chawla, Lauren Douglas, Ziming He, Jason Lee, Tanuja Sawant, and Brian Magerko. 2019. Affordance-based generation of pretend object interaction variants for human-computer improvisational theater. In *Proceedings of the 10th International Conference on Computational Creativity*. Association for Computational Creativity.

Megan Jamerson. 2023. A comedian and ai walk into a bar. who was funnier? https://www.kcrw.com/news/shows/greater-la/artificial-intel-smpd-homeless-oc/ai-comedy. KCRW.

Sophie Jentzsch and Kristian Kersting. 2023. Chatgpt is fun, but it is not funny! humor is still challenging large language models. *arXiv preprint arXiv:2306.04563*.

Harry H Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. Ai art and its impact on artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 363–374.

Philip N Johnson-Laird. 2002. How jazz musicians improvise. *Music Perception*, 19(3).

Keith Johnstone. 1979. *Impro: Improvisation and the Theatre*. Faber and Faber Ltd.

Marzena Karpinska, Nader Akoury, and Mohit Iyyer. 2021. The perils of using mechanical turk to evaluate open-ended text generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1265–1285.

Heather Knight, Scott Satkin, Varun Ramakrishna, and Santosh Divvala. 2011. A savvy robot standup comic: Online learning through audience tracking. In *Workshop paper (TEI'10)*.

Kathleen Marguerite Lea. 1934. Italian popular comedy: a study in the commedia dell'arte, 1560-1620 with special reference to the english stage. *(No Title)*.

Gunter Loesel, Piotr Mirowski, and Kory Wallace Mathewson. 2020. Do digital agents do dada? In *Proceedings of the 11th International Conference on Computational Creativity*, pages 488–491.

Brian Logan. 2023. Whose generated line is it anyway? ai tries to crack humour's dna. *The Guardian*.

Gunter Lösel. 2024. Theatre dialogues with machines. *Interdisciplinary Science Reviews*, 49(2):291–304.

Brian Magerko et al. 2009. An empirical study of cognition and theatrical improvisation. In *ACM Creat. & Cog.*

Lara J Martin, Brent Harrison, and Mark O Riedl. 2016. Improvisational computational storytelling in open worlds. In *Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings 9*, pages 73–84. Springer.

Kory Mathewson and Piotr Mirowski. 2018. Improbotics: Exploring the imitation game using machine intelligence in improvised theatre. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 14, pages 59–66.

Kory W Mathewson. 2019. Humour-in-the-loop: Improvised theatre with interactive machine learning systems. *PhD Thesis, University of Alberta*.

Kory W Mathewson and Piotr Mirowski. 2017a. Improvised theatre alongside artificial intelligences. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Kory Wallace Mathewson and Piotr Mirowski. 2017b. Improvised comedy as a turing test. *arXiv preprint arXiv:1711.08819*.

Piotr Mirowski, Juliette Love, Kory Mathewson, and Shakir Mohamed. 2024. A robot walks into a bar: Can language models serve as creativity supporttools for comedy? an evaluation of llms' humour alignment with comedians. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1622–1636.

Piotr Mirowski, Kory W Mathewson, Boyd Branch, Thomas Winters, Ben Verhoeven, and Jenny Elfving. 2020. Rosetta code: Improv in any language. In *Proceedings of the 11th International Conference on Computational Creativity*, pages 115–122.

Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–34.

Izabella Pluta. 2016. Theater and robotics: Hiroshi ishiguro's androids as staged by oriza hirata. *Art Research Journal*, 3(1):65–79.

Victor Raskin. 1979. Semantic mechanisms of humor. In *Annual Meeting of the Berkeley Linguistics Society*, volume 5, pages 325–335.

Rudolf Rosa, Tomáš Musil, Ondřej Dušek, Dominik Jurko, Patrícia Schmidtová, David Mareček, Ondřej Bojar, Tom Kocmi, Daniel Hrbek, David Košť'ák, et al. 2021. When a robot writes a play: Automatically generating a theatre play script. In *Artificial Life Conference Proceedings 33*, volume 2021, page 60. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . .

Mike Sharples and Rafael Pérez y Pérez. 2022. *Story machines: How computers have become creative writers*. Routledge.

Ajitesh Srivastava and Naomi T Fitter. 2021. A robot walks into a bar: Automatic robot joke success assessment. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2710–2716. IEEE.

Ilya Sutskever, Vinyals Oriol, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112.

Gillian Tett. 2023. Can ai crack comedy? https://www.ft.com/content/818f2cab-57ff-42c3-917b-4a83f1d87802. Financial Times.

Jörg Tiedemann. 2009. News from opus-a collection of multilingual parallel corpora with tools and interfaces. In *Recent Advances in Natural Language Processing*, volume 5, pages 237–248.

Alexey Tikhonov and Pavel Shtykovskiy. 2024. Humor mechanics: Advancing humor generation with multistep reasoning. *arXiv preprint arXiv:2405.07280*.

Joe Toplyn. 2023. Witscript 3: A hybrid ai system for improvising jokes in a conversation. *arXiv preprint arXiv:2301.02695*.

Alan Turing. 1950. Computing machinery and intelligence. *Mind*, 59(236):433–460.

Tony Veale. 2021. *Your Wit is My Command: Building AIs with a Sense of Humor*. Mit Press.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Thomas Winters. 2021. Computers learning humor is no joke. *Harvard Data Science Review*, 3(2).

Thomas Winters. 2024. Evaluating humor generation in an improvisational comedy setting.

Thomas Winters and Kory W Mathewson. 2019. Automatically generating engaging presentation slide decks. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, pages 127–141. Springer.

Thomas Winters, Vincent Nys, and Daniel De Schreye. 2018. Automatic joke generation: Learning humor from examples. In *Distributed, Ambient and Pervasive Interactions: Technologies and Contexts: 6th International Conference, DAPI 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II 6*, pages 360–377. Springer.

Haonan Zhong, Jiamin Chang, Ziyue Yang, Tingmin Wu, Pathum Chamikara Mahawaga Arachchige, Chehara Pathmabandu, and Minhui Xue. 2023. Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In *Companion Proceedings of the ACM Web Conference 2023*, pages 94–98.