Proceedings, The Thirteenth AAAI Conference on
Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17)

Improvised Theatre Alongside Artificial Intelligences

Kory W. Mathewson,[1,2] Piotr Mirowski[2]
[1]University of Alberta Edmonton, Alberta, Canada
[2]HumanMachine, London, UK
korymath@gmail.com, piotr.mirowski@computer.org

Abstract

This study presents the first report of Artificial Improvisation, or improvisational theatre performed live, on-stage, alongside an artificial intelligence-based improvisational performer. The Artificial Improvisor is a form of artificial conversational agent, or chatbot, focused on open domain dialogue and collaborative narrative generation. Using state-of-the-art machine learning techniques spanning from natural language processing and speech recognition to reinforcement and deep learning, these chatbots have become more lifelike and harder to discern from humans. Recent work in conversational agents has been focused on goal-directed dialogue focused on closed domains such as appointment setting, bank information requests, question-answering, and movie discussion. Natural human conversations are seldom limited in scope and jump from topic to topic, they are laced with metaphor and subtext and face-to-face communication is supplemented with non-verbal cues. Live improvised performance takes natural conversation one step further with multiple actors performing in front of an audience. In improvisation the topic of the conversation is often given by the audience several times during the performance. These suggestions inspire actors to perform novel, unique, and engaging scenes. During each scene, actors must make rapid fire decisions to collaboratively generate coherent narratives. We have embarked on a journey to perform live improvised comedy alongside artificial intelligence systems. We introduce Pyggy and A.L.Ex. (Artificial Language Experiment), the first two Artificial Improvisors, each with a unique composition and embodiment. This work highlights research and development, successes and failures along the way, celebrates collaborations enabling progress, and presents discussions for future work in the space of artificial improvisation.

Introduction

Improvisational theatre, or improv, is the spontaneous creation of unplanned theatrics, often performed live on-stage in front of an audience. Improv is a form of collaborative interactive storytelling, where two or more people work together to generate novel narratives. It is grounded in the connections between the performer(s) and the audience. Improv requires the performers to work as a team. The actors must rapidly adapt, empathize, and connect with each other to achieve natural, fluid collaboration. To truly excel at the art form, performers must think and react to audiences reactions quickly, and work together to accept and amplify each other's offers—an act that can be seen as *real-time dynamic problem solving* (Magerko and others 2009). Improv demands human performers handle novel subject matter through multiple perspectives ensuring the audience is engaged while progressing narrative and story. Due to the incredible difficulty, improvisors must embrace failure surrender to spontaneity (Johnstone 1979).

Improvised theatre has been a platform for digital storytelling and video game research for more than 20 years (Perlin and Goldberg 1996; Hayes-Roth and Van Gent 1996). Past research has explored several knowledge-based methods for collaborative storytelling and digital improvisation (O'Neill et al. 2011; Si, Marsella, and Pynadath ; Zhang et al. 2007; Magerko, Dohogne, and DeLeon 2011). Similar work explores how humans interact with system which improvise music and dance (Hoffman and Weinberg 2011; Thomaz et al. 2016). Computer aided interactive storytelling has been considered for applications in video games with an aim to create endless narrative possibilities in video game universes for user engagement (Riedl and Stern 2006).

Robotic performances have been explored previously. In 2000, Tom Sgorous performed *Judy, or What is it Like to Be A Robot?* In 2010, the realistic humanoid robot Gemenoid F performed *Sayonara*, which was later turned into a movie. In 2014, Carnegie Mellon University's Personal Robotics Lab collaborated with their School of Drama to produce *Sure Thing* (Zeglin and others 2014). In these performances, the robots were precisely choreographed, deterministic, or piloted on stage. These shows required the audience to suspend disbelief and embrace the mirage of autonomy. These performances verge ever closer to the deep cliffs surrounding the uncanny valley—the idea that as the appearance of a humanlike robot approaches, but fails to attain, human likeness, a person's response would abruptly shift from empathy to revulsion (Mori, MacDorman, and Kageki 2012).

Our work serves as a bridge between the artificial intelligence labs and improvisational theatre stages. We aim to build a bridge over the uncanny valley, toward a future where humans and autonomous agents converse naturally together. Our work is partially inspired by the narratives behind George Bernad Shaw's Pygmalion (Shaw 1913), Mary

Shelly's Frankenstein (Shelley 1818), and Alan Jay Lerner's My Fair Lady (Cukor 1964). In these stories, creators attempt to design and build reflections of themselves, fabricating their respective ideal images of perfection.

We present our methods and details on the systems which power the first two Artificial Improvisors. We concisely report on findings, and discuss future work at the intersection between artificial intelligence and improvisational theatre.

Methods

Over the last year we iterated from *Version 1: Pyggy* using classic machine learning and deterministic rules to *Version 2: A.L.Ex.* which uses deep neural networks, advanced natural language processing, and a much larger training dataset. While improvisational theatre is a complex artform mixing dialog, movement, and stagecraft, and there exist many improvisational rules for the novice improvisor, in this study we focus on a single component: training the dialog system. An Artificial Improvisor dialog system is composed of three major building blocks: 1) speech recognition, 2) speech generation, and 3) a dialogue management system. The three modules comprise a simplified framework, inspired by the *General Architecture of Spoken Dialogue Systems*, for extemporaneous dialog systems (Pietquin 2005). We detail these components for both Pyggy and A.L.Ex.

Version 1: Chatbot-based Improvisor Pyggy

Pyggy, short for Pygmalion, was an early version of an Artificial Improvisor. Pyggy was built using speech recognition powered by Google Cloud Speech API. Speech recognition translated sound waves from human voice, to text through a network-dependent API. Pyggy used Apple Speech Synthesis for translated output text to sound. Dialogue management was handled with Pandorabots and Chatterbot.

Pandorabots handled hard-coded rules and deterministic responses. For example, when the human said: "Let's start improvising", the system would *always* respond: "Ok". Pandorabots also handled saving simple named entities. For example, if the human said: "My name is Lana" then the system could answer the simple recall question: "What is my name?" Chatterbot was introduced to handle open dialogue and add a sense of randomness to the system. Chatterbot was pre-trained on a set of dialogue, as described below, and then "learned" based on responses the human gave back to the system. For a given human improvisor statement, each of these systems would generate response, which were concatenated and output to the user.

Pre-training of Pyggy was done through an interactive website where individuals could directly interact in basic chit-chat. Unfortunately, when the general public had the ability to interact with Pyggy many started to act adversarially and mischievously, training the system to say rude and inappropriate things. Once the compiled training set was cleaned and filtered, it was quite small. Additional clean training data was appended from the Cornell Movie Dialogue Corpus (Danescu-Niculescu-Mizil and Lee 2011). The dataset is composed of 220579 conversational exchanges from 617 movies and provided the system plenty of novel, interesting, and unique dialogue to pull from.

Pyggy is embodied by a visualization as seen in Fig. 1. The dynamic image-based visualization of Pyggy was accomplished with Magic Music Visualizer. The simple animation system shifted the mouth on an image of a face slightly when sound was being played by the speech generation system. This effect gave Pyggy an animated life, face, and physical embodiment on stage.

Version 2: Neural Network-based A.L.Ex. (Artificial Language Experiment)

There were limitations to the dialogue which Pyggy could produce, as it was restricted to the set of sentences present in the training data. The system was crude in this sense, recalling the most likely response to any input from the human. As well, Pyggy had no means by which to understand or track the topic of a scene. These limitations prompted us to explore a word-by-word generation approach.

Automatic Language Generation in Improvised Theatre The very nature of improvised theatre relies on spontaneous generative conversational abilities. Improvised theatre training relies on teaching the actors games which force them to perform fast-paced word associations (e.g., "electric ... car ... company") or sentence completion (Johnstone 1979) without over-thinking any of their decisions. During these word generation games, spontaneity is encouraged and failure (e.g., a non-grammatical choice of word, an onomatopoeia instead of a word or simply a made-up, garbled word suggestion) is tolerated and celebrated. By celebrating failure, improvisors actively reinforce spontaneity and liberate the creative process. Some of the games directly draw on the Surrealists' *Cadavres Exquis* idea of taking turns in collaborative art generation and require the players to build coherent narratives. Surprisingly, even the apparently most challenging exercises, such as singing and songwriting, as practiced in musical improvisation, still rely on the faculty of spontaneous text generation. In the latter, the performers follow the rhythm and tune of an accompanist while singing rhyming text. Many musical improv teachers, and freestyle rap artists recommend not to prepare rhymes in advance. Rather, they encourage starting lines without predetermined ideas of what rhyme can be found, and let the rhymes arise organically in the mind of the improvisor.

While the word generation process is destined to be spontaneous, it is not intrinsically random. Improvisors use their cultural background, their literary and pop-culture knowledge, eloquence skills, and vernacular, to generate sequences of words which seem most *obvious* to them. Each line is *statistically likely* to occur given the context of the improvisation.

Neural Language Model-based Text Generation We decided to imitate the creative process of improvisation using a statistical language model that can generate text as a sequence of words. While building an open-domain conversational agent is AI-hard, a generative dialogue system that is conditioned on previous text and that mimics collaborative writing could give to the audience an illusion of sensible dialogue (Higashinaka and others 2014). The need for generative dialogue and language models required

Green Can Use

shifting from the rule-based, deterministic learning systems of Pyggy to deep neural network-based language model which would generate sentences word by word. This new model, called A.L.Ex (Artificial Language Experiment) was built using recurrent neural networks (RNN) with long-short term memory (LSTM) (Mikolov and others 2010; Hochreiter and Schmidhuber 1997). Contrary to much similar work in text generation (Sutskever, Martens, and Hinton 2011; Graves 2013), we decided not to rely on character-based RNNs. This facilitates curating the vocabulary produced by the dialogue system and thus immediately replace or remove offensive words generated by the LSTM.

We experimented with multiple LSTM architectures with the goal of building a dialogue model that can handle the topics within an improvised scene over dozens of exchanges between the human and the AI. Starting from a first version consisting of 100,000 linear input word embeddings and a two-layer LSTM with 256 hidden units followed by a softmax over 100,000 output words. The second version contained 4 layers of 512-dimensional LSTMs and extra 64 inputs to the first LSTM, coming from a Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) topic model, that enables the language model to integrate long-range dependencies in the generated text and capture the general *theme* of the dialogue (Mirowski et al. 2010), following the implementation from (Mikolov and Zweig 2012). The third version relied on pre-trained word embeddings (GloVe, Global Vectors) (Pennington, Socher, and Manning 2014) as inputs, resulting in a larger vocabulary of 250,000 input words (the GloVe word embedding matrix was considered as pre-trained and stayed fixed over the training) and only 50,000 output words. The fourth version cloned the 4-layer LSTM into a query embedding module and a response generating module in a seq2seq architecture (Sutskever, Vinyals, and Le 2014; Kiros et al. 2015) with an attention model over the query embedding vectors (Shang, Lu, and Li 2015).

Dataset The language model of A.L.Ex was trained on transcribed subtitles from 102,916 movies from OpenSubtitles.org, going from 1902 to early 2016. This user-contributed subtitles dataset for dialogue model training contains multiple languages and versions for each movie (Vinyals and Le 2015). The data were available as XML files, with precise timestamps for each line of dialogue. We kept one English subtitle version per movie. As we noticed that subtitles tend to be split over time and that each change of interlocutor is marked by a dash sign, we processed the XML files to adjoin lines of dialogue separated by 1 sec, starting with lowercased words and without an initial dash, into single lines of dialogue. Further processing involved correcting common spelling mistakes to account for the often erroneous subtitle input (e.g., substitutions of "l" by "I" or vice-versa, extra spaces between an apostrophe and the contracted word or repetitions of letters, using a painstakingly hand-crafted set of about a thousand of regular expressions) and removal of such as information as "subtitles by ...". The resulting files were lowercased. After text clean-up, we calculated that the top 50,000 words accounted for about 99.4% of the total words appearing in the corpus. The text contained about 880 million tokens (including dashes).

The choice of a movie dialogue corpus, derived from movie scripts, is fitting. Often improv comedy actors draw on previous experience, personal culture and practice in their spontaneous creative process (Martin, Harrison, and Riedl 2016). Future work will explore a variety of text-based datasources including plays, short stories, transcripts of improvised performances, and symbolic plot points (Cook 1928).

System Architecture A.L.Ex. was designed to subvert the multiplicity of connected services which formed the architecture of Pyggy. A.L.Ex. aimed to be an offline, standalone Artificial Improvisor. While, similarly to Pyggy, speech recognition and generation are still performed by ready-made tools, respectively Apple Enhanced Dictation and Apple Speech Synthesis, these tools are run offline, on the same computer. Moreover, the entire text-based dialogue system (coded in Lua and Torch), was encapsulated into a single program which then makes system calls to speech recognition and text-to-speech, and was controlled through a graphical user interface which visualizes results (both the recognized and generated sentences in the dialogue). The core system is then extended with additional modules; it also runs a fault-resilient server which accepts incoming HTTP GET requests from client applications. These applications include software controlling a humanoid robot with pre-programmed motions that are activated when A.L.Ex speaks (see Figure 2). Applications have been written for controlling both the EZ-Robot JD Humanoid and the Aldebaran Nao.

Results

There are challenges associated with testing, and quantitatively evaluating, open-domain dialogue systems (Glass 1999; Higashinaka and others 2014). An obvious and reasonable first measure for qualitative assessment would be similar to that of a human improvisor; that is, the audience-perceived performance level of an Artificial Improvisor during an improvisational performance. Thus, each of these systems has been tested live in front of audiences between 5 and 100 people, for a total of 25, 7-60 minute performances between 8 April 2016 and 1 June 2017. As is common in improvisation, show structure and order remained largely consistent, while content varied based on audience suggestion [1]. Through audience feedback, the system has been iteratively improved, through enhancement to the neural network-based dialogue system, the addition of performance *props* (audio and video user interface, robotic avatar), novel improv comedy games involving the AI and the humans (with optional audience participation), and scripted narrative elements performed by the humans.

While human audience evaluation is realistic and comparable with human improvisor evaluation, there are limitations. First, while an audience is evaluating the performance of the AI in the scene, they may be biased by the fact that an artificial intelligence is performing these lines. Due to this bias, it will be worthwhile to compare human and AI-

[1] https://youtube.com/watch?v=FjxBdMRAZlE

Figure 1: Human performing on-stage with Pyggy. The avatar of the AI is projected behind the human performer.

based responses and scenes offline where the rater is blind to the source of the sides of the dialog. As well, the performance of these systems can be limited at several points in the dialog. Speech recognition is not solved, and environmental noise (e.g. cheering/laughing audience) can influence the quality of input audio for speech-to-text. Finally, because of the high variability of live theatrical performance, the performance spaces (e.g. dark pub basements, or bright conference venues), the narrative elements, the technical difficulties encountered during performance, and iterative experimentation, it is difficult for us to provide quantitative assessment of the AI-Improvisor in this study.

Even with these limitations, we believe it is important to report on the success and failure thus far with qualitative examples, and leave subjective assessment to the reader. Important future work will include large scale qualitative assessment from trained improvisors for qualities of improvisation and from the general public.

Version 1: Pyggy

The first version of AI-based improv, Pyggy, performed in the show *Artificial Improvisation* live and in front of an audience on April 8, 2016, as part of an invited performance during Rapid Fire Theatre's 2016 Bonfire Festival. While technically demanding, it was generally well received by the audience (Figure 1). Here is a short excerpt:

HUMAN: We're having a picnic! Would you like cheese and bread?

PYGGY: Seen one human, you've seen them all.

HUMAN: Promise me, right now, we're here listening to each other.

PYGGY: For fun? Where are all your queer party friends?

HUMAN: Why don't you have any feelings for me? That's why I had to invite you on a picnic. I wanted to get some time to talk to each other.

PYGGY: I miss your vivid sense of humour.

Figure 2: Human performing onstage with A.L.Ex., personified as a robot. The user interface, showing speech recognition results and machine-generated response, is projected behind the performers. Image credits anonymised.

Version 2: A.L.Ex. (Artificial Language Experiment)

Independently developed A.L.Ex. performed for the first time in front of a live audience on July 24, 2016, during an improv drop-in in London. It has been performed 24 times and on two continents since then (Figure 2). We provide below a short transcript from an early performance with A.L.Ex. As can be noticed, the dialogue system somewhat manages to keep track of the general theme of the conversation (dramatic, with words related to navigation and combat).

HUMAN: Captain, the ship is under attack. The frigate is arriving...

A.L.EX: You see, I don't feel the same way. I just don't want to stay in the position to get burnt.

HUMAN: We're all going to get burned by a cannon fire...

A.L.EX: While you're in there, tell me. When I get dressed, I'll go up there and see what happens.

HUMAN: Captain, you need to get on the deck, the crew is expecting you to speak, to take leadership. We are in the middle of a war and this is an invasion of foreign ships. Captain, show us your leadership.

A.L.EX: We won't be able to navigate without them as we are, serving for us to act like a life of violence and to help us die from the most powerful life of our time.

The subsequent experiments detailed in this section were all performed using iteratively improved versions of A.L.Ex.

Human-AI Collaboration Through Games

With the base system in place, we have explored several alternative formats, often called structures or games in improvisation (Riedl and Stern 2006; Magerko and others 2009). We explored a variety of formats as the context within which the AI-Improvisor performs is critical toward the measure of audience perception of quality. In improvisation, a major tenant is to make one's fellow improvisors *shine*. We created a show built on structures where both A.L.Ex. and human improvisors could shine.

Figure 3: Two humans performing on-stage with A.L.Ex. One of the humans is remotely connecting, adding to the complexity of the show setup.

Justification Game

The most extreme case of enhancing the stature of a human improvisor arises from games where the actor is confronted with ridiculously difficult challenges that he or she successfully overcomes (Johnstone 1979). One such game is called *pick-a-line* or *lines from a hat* and consists in the player intermittently and randomly picking a line of dialogue (typically unrelated to the current improvisation), reading it aloud, and seamlessly integrating it into the scene. The humour generally arises from the improvisor's skill in justifying that line of dialogue or from the line being coincidentally appropriate. We found that, because of the limitations of speech recognition and of the dialogue system in A.L.Ex, many of the human-AI interactions ended up following the paradigm of justification games.

Multiple-choice Human-mediated Dialogue

A multiple-choice game was the first format that we explored outside of the basic structure of two improvisors engaging in a basic dialog in a scenic setting. In this format, the AI visually presented several candidate responses on a screen, but did not say any of the responses. Instead, an audience volunteer would select their preferred response and read it aloud. In this way, we were able to directly engage an audience member in the performance (this demolition of the fourth wall is common in improvisation). When the audience is invited to directly interact with the AI on stage, an additional tension is introduced in the room: how will an untrained human react if the AI offers multiple interesting candidates, and what if there are no interesting candidates generated? We observed that these games presented the challenge of the audience member having to share attention between screen and human improvisor and suffered from low energy and audience engagement.

Multiple-person Games

Additionally, we explored dynamics where the AI played a single character in a scene with multiple humans. First, we introduced multiple humans in the same physical space. In this situation, A.L.Ex. plays alongside two human performers. We noticed that there is often a tendency for the two humans to form a 'side', acting together 'against' the AI. Much more interesting scene dynamics emerged when we challenged one of the human performers to align with A.L.Ex's character in the scene. Extending from this work, we then tried including the second human through a remote connection (Google Hangout, see Fig. 3). A.L.Ex. was then able to interact with the physical human and the remote human. Network latency continues to prove challenging and we are still exploring means by which to overcome these challenges. We then instantiated multiple versions of A.L.Ex. in scenes. In this way, we could balance the two humans on stage with two robotic improvisors. This presented opportunities for interesting connections and relationships, as well as challenges as the timing of the two AI-based improvisors can be noticeably different.

Comparison with ELIZA

Finally, we built an audience interaction game in homage to one of the earliest chatbot systems, ELIZA, by Joseph Weizenbaum (Weizenbaum 1966). In this format, an audience member is invited to the stage to discuss an ailment with an AI therapist, that being A.L.Ex. in ELIZA mode. Interestingly, while ELIZA is powered by relatively simple deterministic response rules given certain decompositions of the human's input statement, this is an audience favorite and often well received during shows. It is important to pay special attention to this note, as the holistic performance of an AI-improvisor should be evaluated based not only on how well it is received, but also on the novelty and uniqueness of the scenes it performs.

Many games were selected to allow for clear, distinct trade-off between multiple improvisors within consistent settings. Often our systems fail through mis-understanding speech-to-text input, or robot/human interruption, due to lack of social cueing and perception. By embracing and learning from these failures, we will continue to innovate and experiment to better understand and showcase the strengths of A.L.Ex.

Discussion

Future work will incorporate recent advances in deep reinforcement learning for dialogue generation (Ranzato et al. 2015). Through design of reward functions, more interesting dialogue may be encouraged. Three useful conversational properties recently shown to improve long-term success of dialogue training are: informativity, coherence, and ease of answering (Li and others 2016). Additional reward schemes may improve, or tune, the trained deep neural network based dialogue managers. Recent work has shown that reinforcement learning can be used to tune music generation architectures (Jaques and others 2016). Rewarding linguistic features (i.e. humor, novelty, alliteration) may prove useful in dialog generation (Hollis, Westbury, and Lefsrud 2016).

This study focused on building a dialog system for improvisational performance. Improv theatre is a young practice, but there exists several books of rules for novice improvisational training which could be useful for future studies (Napier 2004). Future iterations of these systems could include common improvisational rules, such as 'status contrast' the 'Yes, and...' theory of accept and expand, and comedic rules, such as the 'rule-of-three' (Johnstone 1979).

Adversarial methods for natural language are another means of recent exploration (Li et al. 2017; Rajeswar et al. 2017). While the results are interesting and informative, these works are still limited in the objective functions and evaluation criteria used often relying on log-likelihood scores, BLEU (Papineni et al. 2002) or ROUGE (Lin 2004) scores. Additional evaluation metrics must be devised to score these open-domain dialog systems (Liu et al. 2016; Lipton, Berkowitz, and Elkan 2015; Ranzato et al. 2015).

Natural human conversations are seldom limited in scope, jump from topic to topic, and are laced with metaphor and subtext. Artificial Improvisors of the future should make use of recent advances in artificial memory and attention models. As well, humans often make use of non-verbal cues during dialog. By incorporating this additional information, human(s) could both consciously and subconsciously inform the learning system (Mathewson and Pilarski 2016). Additionally, if the Artificial Improvisor is modeled as a goal-seeking agent, then shared agency could be quantified and communicative capacity could be learned and optimized for during the performance (Pilarski and Mathewson 2015).

While the system is trained to perform dialog, it is not trained to tell a cohesive story with a narrative arc. The addition of memory network advancements may improve call-back; additional engineering and training will be necessary to collaboratively build a narrative arc. In 1928, William Cook published a book on algorithmic plot development which may serve this purpose, and implementations and connections have yet to be explored (Cook 1928).

Thought must be given to the interface through which humans and artificial performers interact (Yee-King and dInverno 2016; McCormack and d'Inverno 2012; 2016). The embodiment of the Artificial Improvisor has been investigated with Pyggy and A.L.Ex. using on-screen visualizations and robotics. Stage presence is critical to ensure that an Artificial Improvisation show is enjoyable and engaging. Improvisational performances are not strictly conversational and often demand physicality from performers. Collaboration between scientists and creatives will lead to innovative interactions and immersive art. With the growing popularity of interactive mixed-reality experiences, as well as recent advances in natural language processing, speech, and music generation, there are exciting avenues of future investigation (Oord and others 2016; Arik and others 2017).

Improvisational theatre is a domain where experimentation is encouraged, where interaction is paramount, and where failure flourishes. It allows artificial intelligence agents to be effectively tested, and audience reaction can provide a subjective measure of improvement and cognizance. While this may feel similar to the Turing Test, an early attempt to separate mind from machine through a game of imitation, deception and fraud, we believe it is much more (Turing 1950). Success will be measured by audience preference to engage in shows incorporating Artificial Improvisation and human desire to participate.

Games, like chess and Go, are complex, but computational solutions can be approximated. Improvisational theatre demands creativity, rapid artistic generation, and natural language processing. Improvisation is not a zero-sum game, especially as these systems learn to converse open-domain settings (Glass 1999; Higashinaka and others 2014).

Future work will continue to explore the evaluation of performance in such an open domain. Performances with Artificial Improvisors continue to spur questions and insights from other performers and audiences alike. A formal qualitative evaluation, in the lab, with expert improvisors interacting with the system, is planned to explore system quality with additional rigour. We look forward to the distant goal of the human observer, as a fly on the wall, watching AI-improvisors on-stage in front of a full audience of AI-observers.

References

Arik, S. O., et al. 2017. Deep Voice: Real-time Neural Text-to-Speech. *ArXiv e-prints*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR* 3(Jan):993–1022.

Cook, W. W. 1928. *Plotto: a new method of plot suggestion for writers of creative fiction*. Ellis.

Cukor, G. 1964. My fair lady.

Danescu-Niculescu-Mizil, C., and Lee, L. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proc Wkshp on CMCL, ACL*.

Glass, J. 1999. Challenges for spoken dialogue systems. In *Proc 1999 IEEE ASRU Wkshp*.

Graves, A. 2013. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Hayes-Roth, B., and Van Gent, R. 1996. Improvisational puppets, actors, and avatars. In *Proc Computer Game Developers Conf*.

Higashinaka, R., et al. 2014. Towards an open-domain conversational system fully based on natural language processing. In *COLING*, 928–939.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Hoffman, G., and Weinberg, G. 2011. Interactive improvisation with a robotic marimba player. In *Musical robots and interactive multimodal systems*. Springer. 233–251.

Hollis, G.; Westbury, C.; and Lefsrud, L. 2016. Extrapolating human judgments from skip-gram vector representations of word meaning. *Q J Exp Psycho* 1–17.

Jaques, N., et al. 2016. Tuning recurrent neural networks with reinforcement learning. *CoRR* abs/1611.02796.

Johnstone, K. 1979. *Impro. Improvisation and the theatre.* Faber and Faber Ltd.

Kiros, R.; Zhu, Y.; Salakhutdinov, R. R.; Zemel, R.; Urtasun, R.; Torralba, A.; and Fidler, S. 2015. Skip-thought vectors. In *Advances in NIPS*, 3294–3302.

Li, J., et al. 2016. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Li, J.; Monroe, W.; Shi, T.; Ritter, A.; and Jurafsky, D. 2017. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*.

Lin, C.-Y. 2004. Rouge: A package for automatic evaluation of summaries. In *Proc ACL Wkshp. Vol 8*.

Lipton, Z. C.; Berkowitz, J.; and Elkan, C. 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

Liu, C.-W.; Lowe, R.; Serban, I. V.; Noseworthy, M.; Charlin, L.; and Pineau, J. 2016. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*.

Magerko, B., et al. 2009. An empirical study of cognition and theatrical improvisation. In *Creativity and Cognition*.

Magerko, B.; Dohogne, P.; and DeLeon, C. 2011. Employing fuzzy concept for digital improvisational theatre.

Martin, L. J.; Harrison, B.; and Riedl, M. O. 2016. Improvisational computational storytelling in open worlds. In *Interactive Storytelling: 9th Intl Conf on Interactive Digital Storytelling, Los Angeles, USA, November 15–18, 2016, Proc 9*, 73–84. Springer.

Mathewson, K. W., and Pilarski, P. M. 2016. Simultaneous Control and Human Feedback in the Training of a Robotic Agent with Actor-Critic Reinforcement Learning. In *IJCAI Interactive Machine Learning Wkshp*.

McCormack, J., and d'Inverno, M. 2012. Computers and creativity: The road ahead. In *Computers and creativity*. Springer. 421–424.

McCormack, J., and d'Inverno, M. 2016. Designing improvisational interfaces.

Mikolov, T., et al. 2010. Recurrent neural network based language model. In *Interspeech*, volume 2, 3.

Mikolov, T., and Zweig, G. 2012. Context dependent recurrent neural network language model. *SLT* 12:234–239.

Mirowski, P.; Chopra, S.; Balakrishnan, S.; and Bangalore, S. 2010. Feature-rich continuous language models for speech recognition. In *Spoken Language Technology Wkshp, 2010 IEEE*, 241–246. IEEE.

Mori, M.; MacDorman, K. F.; and Kageki, N. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19(2):98–100.

Napier, M. 2004. *Improvise: Scene from the inside out*. Heinemann Drama.

O'Neill, B.; Piplica, A.; Fuller, D.; and Magerko, B. 2011. A knowledge-based framework for the collaborative improvisation of scene introductions. In *ICIDS*, 85–96. Springer.

Oord, A. v. d., et al. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proc 40th Meeting on Association for Computational Linguistics*, 311–318.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *EMNLP*, volume 14, 1532–1543.

Perlin, K., and Goldberg, A. 1996. Improv: A system for scripting interactive actors in virtual worlds. In *Proc Conf on Computer Graphics and Interactive Techniques*, 205–216. ACM.

Pietquin, O. 2005. *A framework for unsupervised learning of dialogue strategies*. Presses univ. de Louvain.

Pilarski, Patrick M., R. S. S., and Mathewson, K. W. 2015. Prosthetic devices as goal-seeking agents. In *2nd Wkshp on Present and Future of Non-Invasive Peripheral-Nervous-System Machine Interfaces, Singapore*.

Rajeswar, S.; Subramanian, S.; Dutil, F.; Pal, C.; and Courville, A. 2017. Adversarial generation of natural language. *arXiv preprint arXiv:1705.10929*.

Ranzato, M.; Chopra, S.; Auli, M.; and Zaremba, W. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.

Riedl, M. O., and Stern, A. 2006. Believable agents and intelligent story adaptation for interactive storytelling. In *Intl Conf on Technologies for Interactive Digital Storytelling and Entertainment*, 1–12. Springer.

Shang, L.; Lu, Z.; and Li, H. 2015. Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Shaw, G. B. 1913. *Pygmalion*. Simon and Schuster.

Shelley, M. W. 1818. *Frankenstein*. WW Norton.

Si, M.; Marsella, S. C.; and Pynadath, D. V. Thespian: An architecture for interactive pedagogical drama.

Sutskever, I.; Martens, J.; and Hinton, G. E. 2011. Generating text with recurrent neural networks. In *Proc 28th ICML*, 1017–1024.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *NIPS*, 3104–3112.

Thomaz, A.; Hoffman, G.; Cakmak, M.; et al. 2016. Computational human-robot interaction. *Foundations and Trends® in Robotics* 4(2-3):105–223.

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

Vinyals, O., and Le, Q. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Weizenbaum, J. 1966. ELIZA A Computer Program for the Study of Natural Language Communication Between Man and Machine. In *Communications of the ACM*.

Yee-King, M., and dInverno, M. 2016. Experience driven design of creative systems. In *Proc 7th Computational Creativity Conf. Universite Pierre et Marie Curie*.

Zeglin, G., et al. 2014. Herb's sure thing: A rapid drama system for rehearsing and performing live robot theater. In *2014 IEEE Intl Wkshp on Advanced Robotics and its Social Impacts*, 129–136.

Zhang, L.; Gillies, M.; Barnden, J. A.; Hendley, R. J.; Lee, M. G.; and Wallington, A. M. 2007. Affect detection and an automated improvisational ai actor in e-drama. In *Artifical Intelligence for Human Computing*. Springer. 339–358.