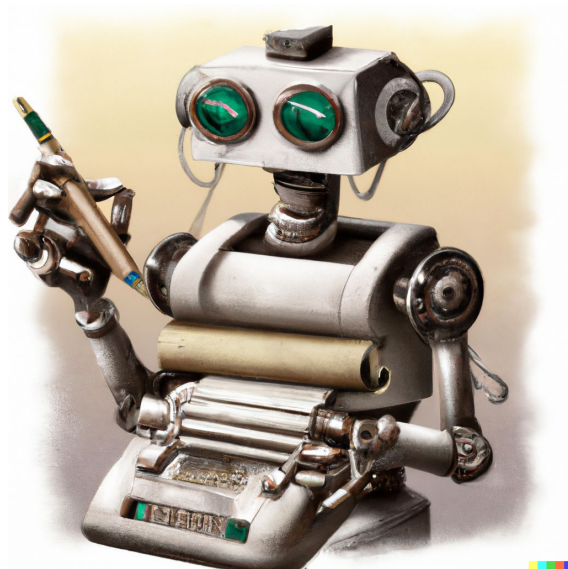

Co-Writing Screenplays and Theatre Scripts with Language Models An Evaluation by Industry Professionals

Рецензия





Команда



Piotr Mirowski
главный герой



Kory Mathewson
друг главного героя



Jaylen Pittman
стажёр



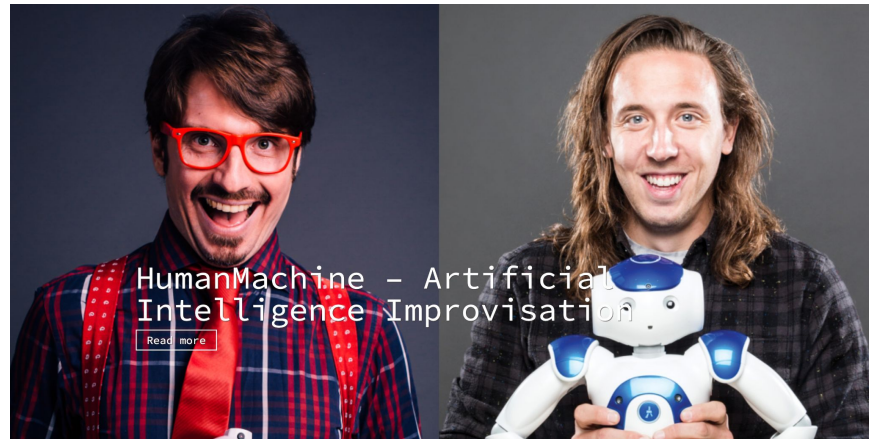
Richard Evans
таинственный
помощник

Предыстория



Improbabilities

Чатбот пытается пройти тест Тьюринга перед зрителями



HumanMachine

Театр импровизации, в котором один из участников - робот



Предыстория

- Improvised theatre alongside artificial intelligences (2017)
- Improvised comedy as a Turing test (2017)
- Improbatics: Exploring the imitation game using machine intelligence in improvised theatre (2018)
- Human improvised theatre augmented with artificial intelligence (2019)
- Do digital agents do Dada? (2020)
- Rosetta Code: improv in any language (2020)
- Collaborative Storytelling with Human Actors and AI Narrators (2021)



Предыстория

- Improvised theatre alongside artificial intelligences (2017)
- Improvised comedy as a Turing test (2017)
- Improbotics: Exploring the imitation game using machine intelligence in improvised theatre (2018)
- Human improvised theatre augmented with artificial intelligence (2019)
- Do digital agents do Dada? (2020)
- Rosetta Code: improv in any language (2020)
- Collaborative Storytelling with Human Actors and AI Narrators (2021)
 - + Chinchilla
 - + Prompt engineering
 - + Hierarchical text generation
- Co-Writing Screenplays and Theatre Scripts with Language Models An Evaluation by Industry Professionals (2022)

ML for Creativity & Design@ NeurIPS 2022

Co-writing screenplays and theatre scripts alongside language models using Dramatron

Piotr Mirowski¹, Kory W. Mathewson², Jaylen Pittman, Richard Evans
DeyMind
{piotrmirowski, kornath, jaylenp, richardevans}@deymind.com

Abstract

Language models are increasingly attracting interest from writers, but lack long-range semantic coherence, limiting their usefulness for longform creative writing. We address this limitation by applying language models hierarchically in a system we call **Dramatron**. By building structural context via prompt chaining, Dramatron can generate coherent scripts and screenplays complete with title, characters, story beats, location descriptions, and dialogue. We illustrate Dramatron's usefulness as an interactive co-creative system with a user study of 15 theatre and film industry professionals. Participants co-wrote theatre scripts and screenplays with Dramatron and engaged in open-ended interviews. We report reflections both from our interviewees and from independent reviewers who watched productions of the works. Finally, we discuss the suitability of Dramatron for human-machine co-creativity, ethical considerations—including plagiarism and bias—and participatory models for the design and deployment of such tools.

As their ability to generate text improves, large language models (LLMs) are becoming useful in co-creative applications [13] and show particular promise for automatic story generation [4, 9] as an augmentative tool for human writers. Story generation, in particular for theatre scripts [10, 11] and screenplays, is a difficult task for LLMs because the narrative must exhibit long-term coherence and interconnection, whereas LLMs are limited in their ability to model long-range dependencies because their context window is bounded to about 1500 words in state-of-the-art models [12, 13].

We present **Dramatron**, a system that uses LLMs to generate script and screenplay hierarchically. Dramatron leverages the strengths of LLMs and combines well-designed prompts and prompt chaining [14] with structured generation for long-range coherence across the entire script. Our method is similar to hierarchical neural story generation [8], but Dramatron can generate coherent scripts that are tens of thousands of words long. It can produce an entire script—including a title, characters, plot, locations, and dialogue—from a single user-provided summary of the central dramatic conflict, called the *lay line* [15]. The user can intervene at any step of the hierarchical generation. In this way, the user interactively co-writes the script. Dramatron was developed with ChatGPT [16] but can be used with any LLM that can accept an input prompt and predict text tokens. Given the quality and bias limitations of online crowd-sourced annotations and evaluations from non-expert raters [17, 20], we engaged 15 experts in a two-hour long user study sessions to co-write a script alongside Dramatron for evaluation. These playwrights and screenwriters from the theatre and film industry were paid a consulting fee for their engagement and provided their artistic opinion and analysis of the outputs co-written with Dramatron. Our study design and data collection process was validated by an ethical review board external to our research institution. To the best of our knowledge, this work represents the largest expert user study conducted on co-creative authorship to date [21, 22].

¹Authors contributed equally to this work.

²We will present a demo of Dramatron during this workshop and consider a public release of the tool.



Figure 1. (Left) Dramatron starts from a user-supplied log line to generate a title and characters, which are then used in prompts to generate a sequence of scene summaries in the plot, which are used to generate unique location descriptions. All elements are then combined to generate dialogue for each scene. Arrows indicate how text generated is used to construct prompts for further LLM text generation. (Right) Photos of human actors interpreting *Care, The Day*, *The Gorkh Shoudh Sill*, a script co-written with Dramatron by director p1 and staged by Rapid Fire Theatre as part of *Plays By Byte*.

We collected qualitative feedback on the co-authorship process during our sessions with the 15 study participants (anonymous as p1, p2, etc.). 13 participants also provided responses on our post-session feedback form (included in the appendix). Quantitative survey results were more positive on questions related to enjoyment and surprise than on questions related to ownership and pride in the output.

Positive comments about Dramatron focused on how hierarchical generation lets the writer work on the narrative arc, the possibility either to co-author interactively or to let the system generate, and the potential of the output script to serve as source material for the human writer. Participants identified inspiration, world building, and content generation as useful applications for Dramatron. Participants noticed various biases embedded in the language model (discussed in Ethical Implications section). Participants embrace unexpected outputs from the system. For example, p5 laughed at the “poetic and absurd” suggestion: “It is really interesting to see what it comes up with” (p5). “Tests of absurdity that are tickling my fancy” (p10). “I would’ve thought of that but it is quite funny” (p1). “This is something that a human author probably would not stand for, it is simply created [...] I want ideas that a human couldn’t possibly have” (p12). That said, participants also noted a lack of nuance and satire. Participant 1 observed: “It’s a good example of how computers do not understand nuance, the way we see language and can understand it even if it is not super specific”. “A lot of information, a bit too verbose, then should be more subtle” (p9). Participant 14 concluded that “AI will never write *Casablanca*, or *A Wonderful Life*. It might be able to write genre bored storytelling”. Finally, p4 and p5 observed that “there has to be a push away from systems of Western dramaticity [...] it might be helpful to consider how it might be used within the context of other contemporary writing”—suggesting alternative narrative structures—“as the AI is not bound by the same rules that we are. So, telling it to be bound by those human rules feels limiting of the capabilities”.

A collection of scripts co-written with Dramatron were produced and staged at the Edmonton International Fringe Theatre Festival in August 2022 (see Fig. 1). Right), the first half of each performance was scripted, the second half improvised. Two reviews were written about the production of *Plays By Byte*. One of the reviews noted that the show “proves that artificial intelligence can in fact write a bit Fringe play”. The reviewer noted that the success of the performance was due to both Dramatron and the human actors, especially one performer who “mastered Dramatron’s voice and seamlessly took it off-script for the remainder of the show, much to the delight of the howling audience”. The second reviewer noted the style of Dramatron, and how that served the performance saying “If there’s a certain fitness in the dialogue, which runs to declarations, that in itself is amusing since it turned out to be perfectly suited to the deep comic talents of [the] improvisers”. Creative team discussions complimented the reviewers and provide insights on how professional actors and improvisers found working with scripts co-written by Dramatron. Overall, the sentiment of enjoying the style of the system was a common theme, with several of the performers remarking that “some of the funniest parts are when you can tell a robot made it”, and that the audience “wants to hear the robot’s voice”. These comments represent critical creative reflections and speak to the value in both the human and the co-creative tools involved in the production.

In short, we present **Dramatron** and a pathway toward human-machine co-creativity that uplifts human writers and artists while leveraging novel artificial intelligence systems such as LLMs.

Ethical Implications

We describe a co-creative tool built around large language models. It can augment and uplift human artists’ work by providing them with inspiration, as well as challenge them and thereby support their artistic practice. Before conducting our study, we identified three directly relevant risks and ethical implications discussed in previous work [28]: 1) bias and offensive language in the generated output, 2) automation of creative work resulting in “cannibalizing” the work of creative artists engaged in script writing, and 3) copyright infringement by reusing copyrighted data from the training dataset, either knowingly (e.g. through prompting: “write the script in the style of Ursula Le Guin”) or unknowingly (e.g. by virtue of similar training data). Our mitigation strategy is two-fold: we invite the creative human artist into the loop throughout the co-authorship process, and we maintain clarity and transparency on the origins of the generated text. To mitigate copyright issues, the writer could query short parts of the script using a search engine and plagiarism detection tools [29]; this functionality could be built directly into co-creative tools. Writers using these tools should be aware of the origin of the data in the LLM, and that audiences should be aware that those outputs were generated through an interaction between humans and co-creative tools. Interestingly, study participants independently raised these concerns during interviews. From the feedback gathered in the study, some participants reported that outputs from the LLM can sometimes be problematic, stereotypical, or biased, for example, “I am less sure that the computer” (p3), or “the protagonists are both male characters, and all of the supporting characters are female” (p4, p5). Furthermore, participants raised concerns about the source of the dataset: “If you are putting existing scripts into the dataset, where are they being pulled from?” (p4, p5). Thoughts on this subject ranged from “Plagiarizing the corpus of script is a problem” (p2) to “In the context of collective and deviated creation, (reusing existing published work) is not necessarily a problem, because it can be perceived as an homage to existing work” (p11). The rules and norms for the use of systems trained on copyright-protected material are the subject of ongoing work [30]. For example, Lee et al. (2022) distinguish between verbatim, paraphrase, and idea plagiarism [29]. Finally, participants raised concern about the potential impact of generative tools on creative economies: “It would free the artist from writing formulaic scripts, [but] it also replaces the work opportunities” (p4, p5). In general, participants expect our mitigation strategies satisfactory and more reported details or concerns regarding outputs from the model. While not the prime focus of the interview sessions, biases and stereotypes could be systematically explored: future work could explore what sorts of narratives can be written using AI tools, and how the system performs for different cultural groups.

References

- [1] Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.
- [2] Nishit Shrivastava, Bryan McCann, Lav R Varshney, Ciming Xiong, and Richard Socher. CTRL: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019.
- [3] Samarth Dharwadkar, Andreia Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and Play Language Models: A Simple Approach to Controlled Text Generation. *CoRR*, abs/1912.02164, 2019.
- [4] Angela Fan, Mike Lewis, and Yann N. Dauphin. Hierarchical neural story generation. *CoRR*, abs/1805.08333, 2018.
- [5] Abigail See, Aneesh Pappu, Rohan Saxena, Abhishek Venkatesh, and Christopher D Manning. Do massively generated natural stories and storytellers differ? *arXiv preprint arXiv:1909.07076*, 2019.
- [6] Open Bravem. GPT-3 Creative Fiction. <https://www.openai.com/GPT-3>. Accessed: 2022-09-20.
- [7] Misao Fukumizu, J. Perotico, A. Lindqvist, and M. Cavazza. Narrative plan generation with self-supervised learning. In *AAAI*, 2021.
- [8] Amal Abbelkumar, Syon Li, and Xiangyu Peng. Automatic story generation: Challenges and attempts. *NACAC: HLT*, page 72, 2021.

The 2022 Edmonton International Fringe Theater Festival



**Bots write scripts. Performers act them out.
Then they improvise the ending. Hilarity ensues.**

The result is an unpredictable work of human-machine co-creativity. You will laugh until your face hurts.

Повлиявшие статьи: THEaiTRE

THEaiTRE: Artificial Intelligence to Write a Theatre Play

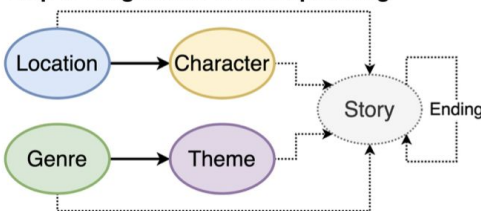
Abstract

We present THEaiTRE, a starting research project aimed at automatic generation of theatre play scripts. This paper reviews related work and drafts an approach we intend to follow. We plan to adopt generative neural language models and hierarchical generation approaches, supported by summarization and machine translation methods, and complemented with a human-in-the-loop approach.



Повлиявшие статьи: Plot Writing from Pretrained Language Models

Step 1: Progressive content planning



Step 2: Generate story body

Task: Write a plot summary of a *{genre}* story featuring *{character1}* and *{character2}* in *{location}* with the main theme "*{theme}*"
Plot summary: "

Step 3: Generate story ending

Task: Write the ending of a *{genre}* story.
What happened earlier: *{story}*
What happens in the end: "



Уже 5 цитирований!

Help me write a Poem - Instruction Tuning as a Vehicle for Collaborative Poetry Writing

[Tuhin Chakrabarty](#), [Vishakh Padmakumar](#), [Hengxing He](#) · Education · EMNLP · 2022

Recent work in training large language models (LLMs) to follow natural language instructions has opened up exciting opportunities for natural language interface design. Building on the prior success... [Expand](#)

 3  ·  View 2 excerpts, cites background  Save  Alert

Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers

[Daphne Ippolito](#), [Ann Yuan](#), [Andy Coenen](#), [Sehmon Burnam](#) · Art · ArXiv · 2022

Recent developments in natural language generation (NLG) using neural language models have brought us closer than ever to the goal of building AI-powered creative writing tools. However, most prior... [Expand](#)

 1  ·  View 1 excerpt, cites background  Save  Alert

An AI Dungeon Master's Guide: Learning to Converse and Guide with Intents and Theory-of-Mind in Dungeons and Dragons

[Pei Zhou](#), [Andrew Zhu](#), +5 authors [Prithviraj Ammanabrolu](#) · Computer Science · ArXiv · 2022

TLDR Human and automated evaluations show that a DM trained with RL to generate guidance by incorporating a theory-of-mind of the players significantly improves the players' ability to achieve goals grounded in their shared world. [Expand](#)

 ·  View 1 excerpt, cites background  Save  Alert

The Infinite Index: Information Retrieval on Generative Text-To-Image Models

[Niklas Deckers](#), [Maik Frobe](#), +4 authors [Martin Potthast](#) · Computer Science · ArXiv · 2022

TLDR This work reframe prompt engineering for generative models as interactive text-based retrieval on a novel kind of "infinite index", and envision how active learning may help to guide the retrieval of generated images. [Expand](#)

 ·  View 1 excerpt, cites methods  Save  Alert

Dynamic Storyboard Generation in an Engine-based Virtual Environment for Video Production

[Anyi Rao](#), [Xuekun Jiang](#), +5 authors [Bo Dai](#) · Business · ArXiv · 2023

show

 ·  View 1 excerpt, cites background  Save  Alert



Качество генераций: сомнительное

Scene 2

Place: The Pool Pit.

Plot element: Inciting Incident.

Beat: Teddy begins a day full of frustration and annoyance. He wakes up to find his room is infested with bugs. The pool is dirty, too. Teddy must deal with the mess to get the club ready for the customers.

F.5.2 Scene 1, random seed 2.

TEDDY

This is a hell of a town, a hell of a town.

It's got a lot of people here.

It's got a lot of noise here.

It's got a lot of bars here.