

# When a Robot Writes a Play: Automatically Generating a Theatre Play Script

Rudolf Rosa<sup>μ</sup>, Tomáš Musil<sup>μ</sup>, Ondřej Dušek<sup>μ</sup>, Dominik Jurko<sup>μ</sup>, Patrícia Schmidtová<sup>μ</sup>,  
David Mareček<sup>μ</sup>, Ondřej Bojar<sup>μ</sup>, Tom Kocmi<sup>μ</sup>, Daniel Hrbek<sup>σ,δ</sup>, David Košťák<sup>σ</sup>, Martina Kinská<sup>σ</sup>,  
Marie Nováková<sup>σ</sup>, Josef Doležal<sup>δ</sup>, Klára Vosecká<sup>δ</sup>, Tomáš Studeník<sup>χ</sup> and Petr Žabka<sup>χ</sup>

<sup>μ</sup>Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Prague, Czechia

<sup>σ</sup>The Švanda Theatre in Smíchov, Prague, Czechia

<sup>δ</sup>The Academy of Performing Arts in Prague, Theatre Faculty (DAMU), Prague, Czechia

<sup>χ</sup>CEE Hacks, Prague, Czechia

uru@ufal.mff.cuni.cz

## Abstract

We inform about *AI: When a Robot Writes a Play*, a theatre play with a mostly artificially generated script. We describe the THEaiTRobot 1.0 tool, which was used to generate the script. We discuss various issues encountered in the process, including those that we solved to some extent as well as those which we plan to solve in a future version of the system.

## The Play

On 26th February 2021, we premiered a theatre play with the title *AI: When a Robot Writes a Play*<sup>1</sup> in the Švanda theatre in Prague, Czechia. The 60-minute play, followed by a 75-minute discussion, was streamed online via the webpage of the THEaiTRE project (Rosa et al., 2020), within which it was created.<sup>2</sup> The play was performed in Czech and accompanied by English subtitles.

The script (i.e., the dialogues) for the play was generated by our THEaiTRobot 1.0 system (Rosa et al., 2021) based on the GPT-2 language model (Radford et al., 2019) with very little human intervention. Specifically, our analysis (to be published soon) shows that 92% of the words in the script were generated by GPT-2. We describe some aspects of the generation process and some issues that we had to resolve. The subsequent staging of the play, on the other hand, was completely in the hands of human theatre professionals.

The THEaiTRobot 1.0 system has several limitations, such as being capable of only generating scripts for individual scenes, not for a full play. We discuss some of the most important issues and our intended approach at solving them. This will eventually lead to a THEaiTRobot 2.0 system, which will be used to generate a script for a new theatre play, with a premiere planned for January 2022.

The text of this extended abstract is largely based on the THEaiTRobot 1.0 short paper by the same authors (Rosa et al., 2021).

<sup>1</sup><https://www.svandovodivadlo.cz/en/inscenations/673/ai-kdyz-robot-pise-hru/3445>

<sup>2</sup><https://www.theaitre.com/>

## Generative Art

There already is a range of works of art created with the help of artificial intelligence. Among the most notable are the short sci-fi movie *Sunspring* (Benjamin et al., 2016), the musical *Beyond the Fence* (Colton et al., 2016), and the theatre play *Lifestyle of the Richard and Family* (Helper, 2018).

However, to the best of our knowledge, our project is the first to produce a full-length theatre play with such a small amount of human intervention into the script.

## The Generation Process

The process of generating a theatre play scene script starts by the user (a theatre dramaturge in our case) defining the start of the scene: a scene setting and two initial lines of dialogue. For the first play, we defined a set of inputs revolving around a common topic to ensure some basic coherence of the whole play. The THEaiTRobot tool then uses the vanilla GPT-2 XL model to generate continuing lines.

For the play, 727 lines of script were generated. The user had the option to discard any generated line (together with all subsequent lines), prompting the tool to generate a different continuation (used 46×). The user could also manually enter a line into the script, which became part of the input for GPT-2 (used 8×). The script was then post-edited by deleting 214 lines, and changing 362 words (8%) on 146 lines.

The tool itself is implemented as a web application with a server backend, using the Huggingface Transformers library (Wolf et al., 2020). We have published a video showing the operation of THEaiTRobot 1.0,<sup>3</sup> a sample of its outputs,<sup>4</sup> and its source codes.<sup>5</sup> We will also soon publish the full script of the theatre play with marked human interventions.

## Resolved Issues

**Set of characters** The model does not work with a limited set of characters naturally and tends to forget characters and invent new characters too often. We resolve this by modifying the next token probability distribution within the GPT-2

<sup>3</sup><https://youtu.be/ksrZouM7Wyg>

<sup>4</sup><http://bit.ly/theaitre-samples>

<sup>5</sup><http://hdl.handle.net/11234/1-3507>

model, so that at the start of a new line, only tokens corresponding to character names present in the input prompt are allowed. We also boost probabilities of characters that have not spoken for a long time.

**Repetitiveness** GPT-2's generation may get stuck in a loop, generating one or several lines again and again. We managed to resolve this by modifying the hyperparameters of GPT-2, changing repetition penalty from 1.00 to 1.01. As a backup, we also automatically discard any generated repeated lines and prompt the model to generate another continuing line.

**Limited context** The GPT-2 model has a limit of 1024 subword tokens, within which both the input prompt and the generated output must fit. The typical solution is to crop the input at the beginning so that it fits into the window with sufficient space for generating the output. However, this means forgetting potentially important information from the input prompt and the previously generated text, which can lead to an unwanted continual topic drift and also to generating contradictory text; the text is still locally consistent, but as a whole it may be inconsistent. This severely limits the ability of GPT-2 to generate longer texts.

To handle this issue, we introduce automated extractive summarization into the process, hoping that the summarization algorithm will identify the most important pieces of information to remember. Whenever the already generated script does not fit into the limited window any more, we use the TextRank algorithm by Mihalcea and Tarau (2004) to summarize a large part of the script into only 5 lines. We concatenate those with the last 250 tokens of the script, and use that as the input for the generation process. In this way, both global and local consistency of the script can be achieved.

**Machine translation** The GPT-2 model operates on English, while we want to generate a Czech script. We therefore automatically translate the generated script using the CUBBITT neural translation model by Popel et al. (2020). As the translation tends to discard character names from the lines, we add them by identifying them in the input and translating them independently.

### Unresolved issues and future plans

**Generating a whole play** The model is not able to generate a long and complex text such as a full theatre play script. To resolve this, we intend to generate the script hierarchically, first generating a synopsis for the whole play, then expanding it into synopses for individual scenes, and finally generating each scene individually based on its synopsis. This approach is inspired by the work of Fan et al. (2018, 2019), who take a similar coarse-to-fine approach to story generation. Our situation is, however, more complex, as we plan to use one more step of the hierarchy.

**Character personalities** The characters in the play do not seem to have independent personalities in the generated script; the model seems to simply ensure consistency with already generated text, not taking the character names into account. The character personalities thus appear to switch and merge. We intend to resolve this by learning theatre character embeddings and using them to condition the language model. We plan to resolve this by clustering our data into several basic character personality types (Azab et al., 2019), then train separate character-aware language models, either by finetuning the GPT-2 model, or by using adapter models (Madotto et al., 2020; Wang et al., 2020).

**Dramatic situations** The text is generated word by word and line by line, whereas human authors of theatre plays typically operate on a more abstract level, such as dramatic situations (Politi, 1921).<sup>6</sup> While there is some work on identifying dramatic turning points (Papalampidi et al., 2019, 2020), it is too coarse-grained for our application. We are thus currently annotating a corpus of theatre play scripts with a modified set of dramatic situations, and plan to enhance the tool with this abstraction, either by adding one more layer in the hierarchical setup, or by using special tokens or embeddings to mark dramatic situations in the generated text.

**Machine translation issues** The MT model we use is tuned for news text, not theatre scripts, and translates each sentence independently. This leads to various issues, including errors in morphological gender (which should pertain to the character), variance in the honorific T-V distinction (which may vary but should be consistent for each pair of characters), and erroneous sentence splitting. We intend to tackle these issues by using a document-level translation system which takes larger context into account, fine-tuning the model on a corpus of theatre play scripts, and adding various heuristic modifications where necessary.

## Conclusion

We have described the process behind generating the script for the play *AI: When a Robot Writes a Play*, which was created by THEaiTRobot 1.0 with minimal human intervention. We have also discussed numerous issues, some of which we managed to resolve, while others are waiting to be tackled in THEaiTRobot 2.0.

## Acknowledgements

The project TL03000348 THEaiTRE: Umělá inteligence autorem divadelní hry is co-financed with the state support of Technological Agency of the Czech Republic within the ÉTA 3 Programme.

<sup>6</sup>[https://en.wikipedia.org/wiki/The\\_Thirty-Six\\_Dramatic\\_Situations](https://en.wikipedia.org/wiki/The_Thirty-Six_Dramatic_Situations)

## References

- Azab, M., Kojima, N., Deng, J., and Mihalcea, R. (2019). Representing Movie Characters in Dialogues. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 99–109, Hong Kong.
- Benjamin, A., Sharp, O., and Goodwin, R. (2016). *Sunspring*, a sci-fi short film starring Thomas Middleditch. <https://www.youtube.com/watch?v=LY7x2Ihqjmc>.
- Colton, S., Llano, M. T., Hepworth, R., Charnley, J., Gale, C. V., Baron, A., Pachet, F., Roy, P., Gervás, P., Collins, N., Sturm, B., Weyde, T., Wolff, D., and Lloyd, J. R. (2016). The Beyond the Fence musical and Computer Says Show documentary. In *Proceedings of the Seventh International Conference on Computational Creativity*.
- Fan, A., Lewis, M., and Dauphin, Y. (2018). Hierarchical Neural Story Generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, New Orleans, LA, USA. arXiv: 1805.04833.
- Fan, A., Lewis, M., and Dauphin, Y. (2019). Strategies for Structuring Story Generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2650–2660, Florence, Italy. Association for Computational Linguistics.
- Helper, R. (2018). Lifestyle of the Richard and family. <https://www.roslynhelper.com/lifestyle-of-the-richard-and-family>.
- Madotto, A., Ishii, E., Lin, Z., Dathathri, S., and Fung, P. (2020). Plug-and-Play Conversational Models. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2422–2433, Online. Association for Computational Linguistics.
- Mihalcea, R. and Tarau, P. (2004). TextRank: Bringing Order into Text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.
- Papalampidi, P., Keller, F., Frermann, L., and Lapata, M. (2020). Screenplay Summarization Using Latent Narrative Structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1920–1933, Online.
- Papalampidi, P., Keller, F., and Lapata, M. (2019). Movie Plot Analysis via Turning Point Identification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1707–1717, Hong Kong, China. Association for Computational Linguistics.
- Polti, G. (1921). *The thirty-six dramatic situations*. JK Reeve.
- Popel, M., Tomková, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., and Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11(4381):1–15.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. Technical report, OpenAI.
- Rosa, R., Dušek, O., Kocmi, T., Mareček, D., Musil, T., Schmidtová, P., Jurko, D., Bojar, O., Hrbek, D., Košťák, D., Kinská, M., Doležal, J., and Vosecká, K. (2020). THEaiTRE: Artificial intelligence to write a theatre play. In Jorge, A., Campos, R., Jatowt, A., and Aizawa, A., editors, *Proceedings of AI4Narratives — Workshop on Artificial Intelligence for Narratives*, volume 2794 of *CEUR Workshop Proceedings*, pages 9–13, Aachen, Germany. RWTH Aachen University, RWTH Aachen University.
- Rosa, R., Musil, T., Dušek, O., Jurko, D., Schmidtová, P., Mareček, D., Bojar, O., Kocmi, T., Hrbek, D., Košťák, D., Kinská, M., Nováková, M., Doležal, J., Vosecká, K., Studeník, T., and Žabka, P. (2021). THEaiTRE 1.0: Interactive generation of theatre play scripts. In Campos, R., Jorge, A., Jatowt, A., Bhatia, S., and Finlayson, M., editors, *Proceedings of the Text2Story'21 Workshop*, volume 2860 of *CEUR Workshop Proceedings*, pages 71–76, Aachen, Germany. RWTH Aachen University.
- Wang, R., Tang, D., Duan, N., Wei, Z., Huang, X., Ji, J., Cao, G., Jiang, D., and Zhou, M. (2020). K-Adapter: Infusing Knowledge into Pre-Trained Models with Adapters. *arXiv:2002.01808 [cs]*.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.