THEaiTRobot: An Interactive Tool for Generating Theatre Play Scripts

Rudolf Rosa[μ], **Patrícia Schmidtová**[μ], **Alisa Zakhtarenko**[μ],
Ondřej Dušek[μ], **Tomáš Musil**[μ], **David Mareček**[μ], **Saad Obaid**[μ],
Marie Nováková[σμ], **Klára Vosecká**[δ], **Daniel Hrbek**[σδ] and **David Košťák**[σ]
[μ]Charles University, Faculty of Mathematics and Physics, Prague, Czechia
[σ]The Švanda Theatre in Smíchov, Prague, Czechia
[δ]The Academy of Performing Arts in Prague, Theatre Faculty (DAMU), Prague, Czechia
rosa@ufal.mff.cuni.cz

Abstract

We present a free online demo of THEaiTRobot, an open-source bilingual tool for interactively generating theatre play scripts, in two versions. THEaiTRobot 1.0 uses the GPT-2 language model with minimal adjustments. THEaiTRobot 2.0 uses two models created by fine-tuning GPT-2 on purposefully collected and processed datasets and several other components, generating play scripts in a hierarchical fashion (title → synopsis → script). The underlying tool is used in the THEaiTRE project to generate scripts for plays, which are then performed on stage by a professional theatre.

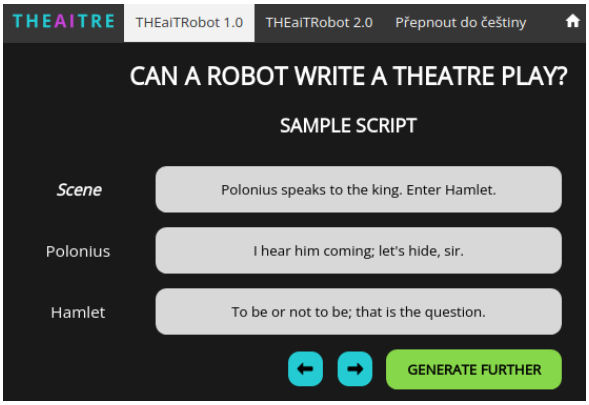

Figure 1: THEaiTRobot 1.0 input screen.

1 Introduction

We present a demo version of THEaiTRobot, a tool for interactively generating theatre play scripts. THEaiTRobot 1.0 is a vanilla GPT-2 model (Radford et al., 2019) with several adjustments for the theatrical domain (Rosa et al., 2021a); THEaiTRobot 2.0 features two fine-tuned GPT-2 models operating in a two-step hierarchical fashion. Machine translation allows the tool to operate both in English and in Czech.[1]

The tool was used within the THEaiTRE project to generate scripts of theatre plays, which were then staged by a professional theatre. The script of the 60-minute-long first play, *AI: When a Robot Writes a Play*, consists from 90% of automatically generated texts, with only 10% human contributions and edits (THEaiTRobot 1.0 et al., 2021), unprecedented for such a long play. The online premiere of the play was viewed by thousands of spectators worldwide (Moutinho, 2021).

One of our goals in the THEaiTRE project is making the public more familiar with the actual state and operation of artificial intelligence and natural language generation in particular. In addition to theatrical shows complemented by follow-up discussions with the audience, we want to further support this goal by making the demo of the THEaiTRobot tool freely available online for anyone to experiment with.[2] A short video showing the usage of the demo is available on YouTube.[3]

2 Related Work

A number of GPT-based language generation tools is available online, such as news generators (Zellers et al., 2019; Geitgey, 2019),[4,5] text adventure games,[6] code completion tools,[7,8] or chatbots.[9] However, to the best of our knowledge, no script generation tool has been released so far.

Script generation has been explored in several other projects, none of which however make their tools publicly available, often not even sharing quite enough details about the design of the tool, and also either using extensive human curation

[1]Adapting to a different language is very simple, as the translation system is external to the generator. It can easily be replaced by changing the respective call to the external API.

[2]https://theaitre.com/demo
[3]https://youtu.be/B3U38UgeZ9w
[4]https://rowanzellers.com/grover/
[5]https://newsyoucantuse.com/
[6]https://play.aidungeon.io/
[7]https://copilot.github.com/
[8]https://www.tabnine.com/
[9]https://projectdecember.net/

Proceedings of the 15th International Conference on Natural Language Generation: System Demonstrations, pages 10 - 13
July 18-22, 2022 ©2022 Association for Computational Linguistics

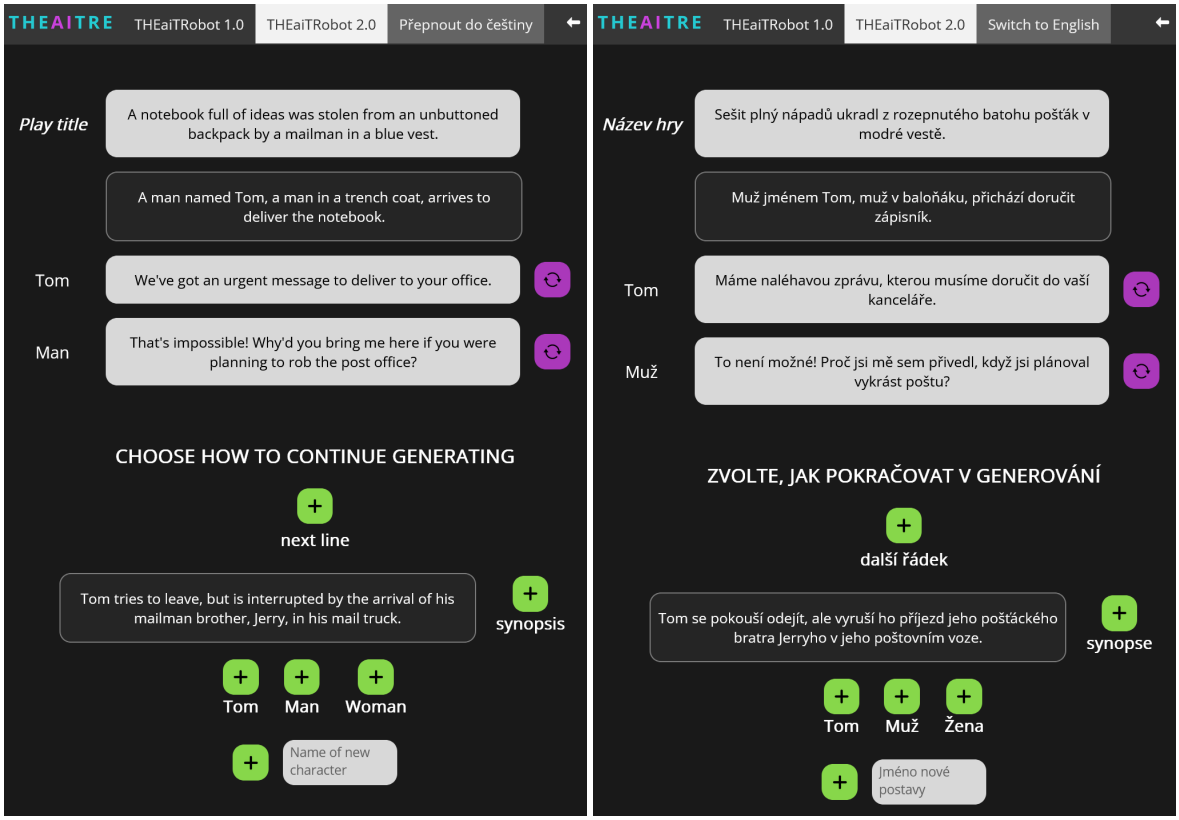

Figure 2: THEaiTRobot 2.0 synopsis → script interface (left: English, right: Czech with the same script).

and/or generating only very short scripts (Colton et al., 2016; Benjamin et al., 2016; Helper and Gillies, 2018; Mathewson and Mirowski, 2017).

3 THEaiTRobot 1.0

The first version of our tool is based on a vanilla GPT-2 XL language model with several adjustments, mainly employing TextRank-based (Mihalcea and Tarau, 2004) extractive text summarization to deal with GPT-2's limited window of 1,024 tokens, so longer scripts can be generated without losing context (Rosa et al., 2021b). The tool uses CUBBITT (Popel et al., 2020) for on-the-fly machine translation of the outputs into Czech.

In the demo version of the tool, the user can input a scene setting, character names and their first lines, or use one of the predefined inputs (see Figure 1), from which we construct the input prompt for the GPT-2 model in the following format:

```
Scene setting.

Character Name: Character line.

Character Name: Character line.
```

The tool then generates a continuation of the script line by line. At each step, the user can choose to continue generating or to regenerate a previously generated line (i.e. generate a different continuation from that position onward).

An early version of this demo was presented to the public in an exhibition at Goethe-Institut in Prague,[10] where about 100 users interacted with it, mostly finding it amusing and intriguing. We also gained feedback that helped us improve the demo.

4 THEaiTRobot 2.0

The second version of the tool uses a two-step hierarchical generation approach, first generating a play synopsis and then expanding that synopsis into a full play script, with specific models trained on our datasets for each of the steps.

The input for the synopsis generation step is the play title, in a fashion similar to script generation in THEaiTRobot 1.0 but with the underlying GPT-2 model fine-tuned on synopsis data. For the fine-tuning, we used a dataset consisting of ca. 65k synopses of theatre plays (scraped by us from Wikipedia), movies (Robischon, 2018; Kar et al., 2018), TV series (scraped by us from fan wiki pages) and books (Bamman and Smith, 2017).

[10] https://www.goethe.de/ins/cz/cs/ver.cfm?event_id=22345514

Once the user is happy with the generated synopsis, the synopsis is used as input for the second step.

In the second step, the play script is generated from the synopsis using a GPT-2 model fine-tuned for generating script sections based on synopsis sections. Here we use a near-domain ScriptBase corpus (Gorinski and Lapata, 2018), which contains movie synopses and scripts. We split the synopses into sentences and align these in a monotonic one-to-many fashion to script lines.

The user now has more options when generating (see Figure 2): generating a character line, also choosing which character should speak, or moving on to the next part of the generated synopsis.

5 Conclusion

We release an online demo of THEaiTRobot, a tool for interactive generation of theatre play scripts. The tool is free for non-commercial use, and its source code is released under the MIT licence.[11]

Acknowledgements

The project *TL03000348* is co-financed with the state support of Technological Agency of the Czech Republic within the ETA 3 Programme. It used services provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101).

References

David Bamman and Noah Smith. 2017. CMU book summary dataset. Released on Kaggle.

AI Benjamin, Oscar Sharp, and Ross Goodwin. 2016. Sunspring, a sci-fi short film starring Thomas Middleditch. Released on YouTube.

Simon Colton, Maria Teresa Llano, Rose Hepworth, John Charnley, Catherine V. Gale, Archie Baron, François Pachet, Pierre Roy, Pablo Gervás, Nick Collins, Bob Sturm, Tillman Weyde, Daniel Wolff, and James Robert Lloyd. 2016. The Beyond the Fence musical and Computer Says Show documentary. In *Proceedings of the Seventh International Conference on Computational Creativity*.

Adam Geitgey. 2019. *Machine Learning is Fun!* Self-published.

Philip John Gorinski and Mirella Lapata. 2018. What's this movie about? A joint neural network architecture for movie content analysis. In *Proceedings of NAACL-HLT*, pages 1770–1781, New Orleans, Louisiana.

Roslyn Helper and Harriet Gillies. 2018. Lifestyle of the Richard and family. Theatre play.

Sudipta Kar, Suraj Maharjan, A. Pastor López-Monroy, and Thamar Solorio. 2018. MPST: A corpus of movie plot synopses with tags. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France.

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

Rada Mihalcea and Paul Tarau. 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.

Sofia Moutinho. 2021. Kinky and absurd: The first AI-written play isn't shakespeare—but it has its moments. *Science*.

Martin Popel, Marketa Tomková, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar, and Zdeněk Žabokrtský. 2020. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11(4381):1–15.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. Technical report, OpenAI.

Justin Robischon. 2018. Wikipedia movie plots. Released on Kaggle.

Rudolf Rosa, Tomáš Musil, Ondřej Dušek, Dominik Jurko, Patrícia Schmidtová, David Mareček, Ondřej Bojar, Tom Kocmi, Daniel Hrbek, David Košťák, et al. 2021a. When a robot writes a play: Automatically generating a theatre play script. In *ALIFE 2021: The 2021 Conference on Artificial Life*. MIT Press.

Rudolf Rosa, Tomáš Musil, Ondřej Dušek, Dominik Jurko, Patrícia Schmidtová, David Mareček, Ondřej Bojar, Tom Kocmi, Daniel Hrbek, David Košťák, Martina Kinská, Marie Nováková, Josef Doležal, Klára Vosecká, Tomáš Studeník, and Petr Žabka. 2021b. THEaiTRE 1.0: Interactive generation of theatre play scripts. In *Proceedings of the Text2Story'21 Workshop*, volume 2860 of *CEUR Workshop Proceedings*, pages 71–76, Aachen, Germany. RWTH Aachen University.

THEaiTRobot 1.0, David Košťák, Daniel Hrbek, Rudolf Rosa, and Ondřej Dušek. 2021. AI: When a robot writes a play. Technical Report ÚFAL TR-2021-67, ÚFAL MFF UK, Praha, Czechia.

[11] https://github.com/ufal/theaitrobot

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 9054–9065. Curran Associates, Inc.