

Dominik Jurko, Rudolf Rosa, Tomáš Musil, Ondřej Dušek, 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



CEEHACKS

DAE

ÚFA L

THEATRE

T A
Č R



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



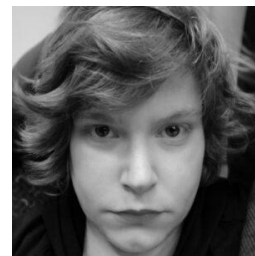
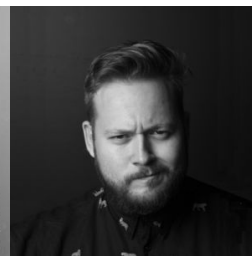
ÚFA L

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



unless otherwise stated

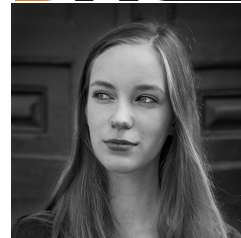
The team



CEEHACKS

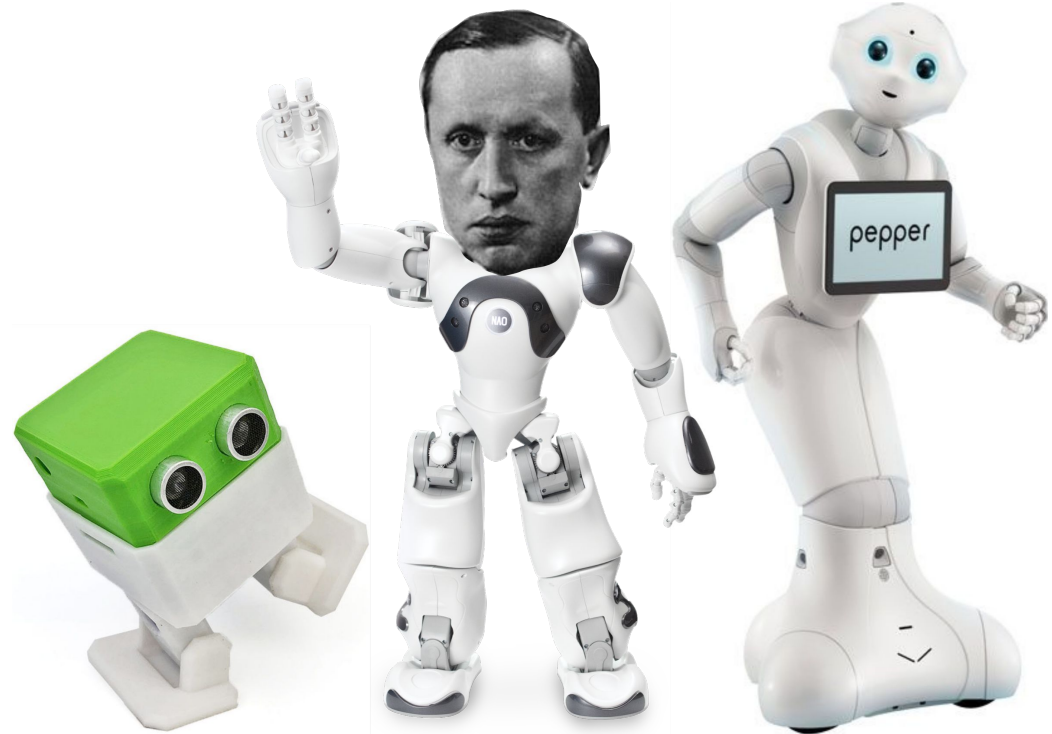


DAM



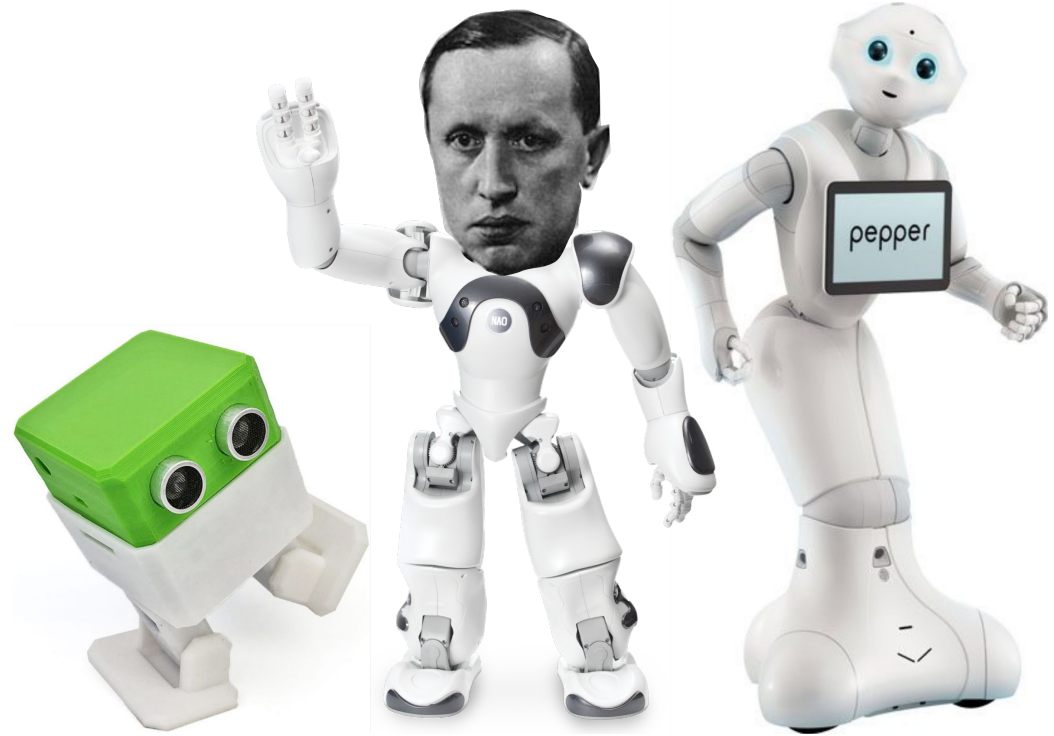
Motivation

- 25th January 1921
 - R.U.R. by Karel Čapek
 - first theatre play about robots



Motivation

- 25th January 1921
 - R.U.R. by Karel Čapek
 - first theatre play about robots
- 26th February 2021
 - AI: When a Robot Writes a Play
 - first theatre play written (from 90%) by “robots” (actually AI)



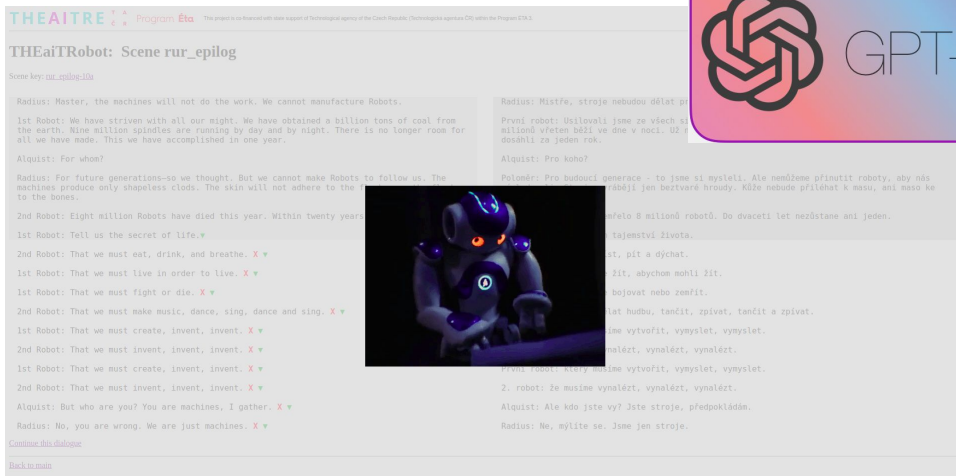
Approach

1. Hacking the GPT-2 language model
- 2.
- 3.
- 4.



Approach

1. Hacking the GPT-2 language model
2. Building a web application
- 3.
- 4.



Approach

1. Hacking the GPT-2 language model
2. Building a web application
3. Generating a theatre play script
- 4.

THEaiTRE

Program Eta

This project is co-financed with state support of Technological agency of the Czech Republic, Technological agency CZU within the Program CTR 2.

THEaiTRObot: Scene rur_epilog

Scene key: rur_epilog_01a

Radius: Master, the machines will not do the work. We cannot manufacture Robots.

1st Robot: We have striven with all our might. We have obtained a billion tons of coal from the earth. Nine million spiders are running by day and by night. There is no longer room for all we have made. This we have accomplished in one year.

Alquist: For whom?

Radius: For future generations-so we thought. But we cannot make Robots to follow us. The machines produce only shapeless clods. The skin will not adhere to the bones.

2nd Robot: Eight million Robots have died this year. Within twenty years-

1st Robot: Tell us the secret of life.

2nd Robot: That we must eat, drink, and breathe. X

1st Robot: That we must live in order to live. X

1st Robot: That we must fight or die. X

2nd Robot: That we must make music, dance, sing, dance and sing. X

1st Robot: That we must create, invent, invent. X

2nd Robot: That we must invent, invent, invent. X

1st Robot: That we must create, invent, invent. X

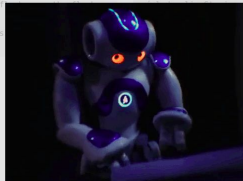
2nd Robot: That we must invent, invent, invent. X

Alquist: But who are you? You are machines, I gather. X

Radius: No, you are wrong. We are just machines. X

[Continue this dialogue](#)

[Back to main](#)



Approach

1. Hacking the GPT-2 language model
2. Building a web application
3. Generating a theatre play script
4. Performing the play on stage

THEAITRE Program Eta

This project is co-financed with state support of Technological agency of the Czech Republic, Technological agency CZU within the Program CZU2.

THEaiTRobot: Scene rur_epilog

Scene key: rur_epilog_01a

Radius: Master, the machines will not do the work. We cannot manufacture Robots.

1st Robot: We have striven with all our might. We have obtained a billion tons of coal from the earth. Nine million spiders are running by day and by night. There is no longer room for all we have made. This we have accomplished in one year.

Alquist: For whom?

Radius: For future generations-as we thought. But we cannot make Robots to follow us. The machines produce only shapeless clods. The skin will not adhere to the bones.

2nd Robot: Eight million Robots have died this year. Within twenty years-

1st Robot: Tell us the secret of life. ♡

2nd Robot: That we must eat, drink, and breathe. ♡ ♡

1st Robot: That we must live in order to live. ♡ ♡

1st Robot: That we must fight or die. ♡ ♡

2nd Robot: That we must make music, dance, sing, dance and sing. ♡ ♡

1st Robot: That we must create, invent, invent. ♡ ♡

2nd Robot: That we must invent, invent, invent. ♡ ♡

1st Robot: That we must create, invent, invent. ♡ ♡

2nd Robot: That we must invent, invent, invent. ♡ ♡

Alquist: But who are you? You are machines, I gather. ♡ ♡

Radius: No, you are wrong. We are just machines. ♡ ♡

[Continue this dialogue](#)

[Back to menu](#)



GPT-2



Hacking the GPT-2 Language Model



GPT-2
SMALL

117M Parameters



GPT-2
MEDIUM

345M Parameters



GPT-2
LARGE

762M Parameters



GPT-2
EXTRA
LARGE

1,542M Parameters



OpenAI



Transformers

Language model

Language model

- Task: Which word should follow?

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the...
 - ???

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the...
 - kitchen
 - bathroom
 - cinema
 - horse

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the... kitchen/bathroom/cinema/horse
- Language model
 - Needs to learn what language looks like
 - Large amount of texts: news, books, websites, film subtitles... (bilions of words)

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the... kitchen/bathroom/cinema/horse
- Language model
 - Needs to learn what language looks like
 - Large amount of texts: news, books, websites, film subtitles... (bilions of words)
- N-gram language model (e.g. 4-gram)
 - How often are words A B C followed by word D (in first 100M lines of Wikipedia)?
 - “went to the bathroom” > “went to the horse”?
 - “went to the cinema” > “went to the kitchen”?

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the... kitchen/bathroom/cinema/horse
- Language model
 - Needs to learn what language looks like
 - Large amount of texts: news, books, websites, film subtitles... (bilions of words)
- N-gram language model (e.g. 4-gram)
 - How often are words A B C followed by word D (in first 100M lines of Wikipedia)?
 - “went to the bathroom” > “went to the horse”? 42 > 4
 - “went to the cinema” > “went to the kitchen”?

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the... kitchen/bathroom/cinema/horse
- Language model
 - Needs to learn what language looks like
 - Large amount of texts: news, books, websites, film subtitles... (bilions of words)
- N-gram language model (e.g. 4-gram)
 - How often are words A B C followed by word D (in first 100M lines of Wikipedia)?
 - “went to the bathroom” > “went to the horse”? 42 > 4
 - “went to the cinema” > “went to the kitchen”? 25 > 14

Language model

- Task: Which word should follow?
 - I woke up in the morning and went to the... kitchen/bathroom/cinema/horse
- Language model
 - Needs to learn what language looks like
 - Large amount of texts: news, books, websites, film subtitles... (bilions of words)
- N-gram language model (e.g. 4-gram)
 - How often are words A B C followed by word D (in first 100M lines of Wikipedia)?
 - “went to the bathroom” > “went to the horse”? 42 > 4
 - “went to the cinema” > “went to the kitchen”? 25 > 14
- Neural language models
 - See a larger number of preceding words (GPT-2: up to 1023 previous words)
 - Choose which preceding words to look at (attention, skip-gram)
 - Can estimate word similarity (cosine distance of word embeddings)
 - Complex internal representations and operations instead of simple comparison of frequencies

The base approach

- Take GPT-2 XL language model
- Feed it with a theatre-script-like prompt, e.g.:

A man is driving in his car. He sees a girl hitchhiking next to the road. He stops the car.

Man: Do you want a ride?

Girl: Oh yes please!

- Let it generate a continuation
- Machine translation English→Czech

A sample output

A man is driving in his car. He sees a girl hitchhiking next to the road. He stops the car.

Man: Do you want a ride?

Girl: Oh yes please!

Man: Okay, then you must first convince me you're not a child prostitute.

Girl: I'm 16 years old, and I'm only 16.

Man: Then I'll take you home!

Woman: He's asking for trouble!

Man: No, he's only asking!

Woman: Then why don't you just take her home already?

Man: But I have a family!

“Resolved” issues

“Resolved” issues

- Set of characters: introduces new, forgets old
 - modify next token probabilities to restrict characters

“Resolved” issues

- Set of characters: introduces new, forgets old
 - modify next token probabilities to restrict characters
- Repetitiveness: identical lines
 - increase repetition penalty $1.00 \rightarrow 1.01$
 - discard and regenerate if repetition occurs

“Resolved” issues

- Set of characters: introduces new, forgets old
 - modify next token probabilities to restrict characters
- Repetitiveness: identical lines
 - increase repetition penalty $1.00 \rightarrow 1.01$
 - discard and regenerate if repetition occurs
- Limited context (1024 subwords): long-distance inconsistencies
 - extractive summarization (TextRank) of prompt+output \rightarrow 5 lines
 - plus keep last 250 subwords

- Generating a whole play (now: individual scenes)
 - hierarchical approach: title → outline → scenes (fine-tuned models)

- Generating a whole play (now: individual scenes)
 - hierarchical approach: title → outline → scenes (fine-tuned models)
- Character personalities and perspectives
 - cluster character types, employ character embeddings

- Generating a whole play (now: individual scenes)
 - hierarchical approach: title → outline → scenes (fine-tuned models)
- Character personalities and perspectives
 - cluster character types, employ character embeddings
- Dramatic situations (e.g. Intruder, Request, Parting)
 - annotate plays, detect situations, tune to generate situations

WIP + plans

- Generating a whole play (now: individual scenes)
 - hierarchical approach: title → outline → scenes (fine-tuned models)
- Character personalities and perspectives
 - cluster character types, employ character embeddings
- Dramatic situations (e.g. Intruder, Request, Parting)
 - annotate plays, detect situations, tune to generate situations
- Various machine translation issues (gender, politeness, context...)
 - document-level, fine-tuning, automatic post-editing



The Web Application



THEaiTRobot: Scene masters_death

Scene key: [masters_death-Jablon](#)

It's the morning. The Robot enters the room of his master who is really old and sick. The Robot sees that his master is not doing very well this morning. He sits at the edge of his bed and takes his hand.

Robot: I remember how you jumped on this bed when you were little. You were full of beans.

Master: We both know I am dying.*

Robot: I remember you saying that too, you sick old man. And how you have tried to kill yourself many times. X *

Master: I have tried but it's so hard. I want to be alive. I want to be with my children and grandchildren. X *

Robot: I am sorry but you can't. There is no way. No way. There is no way. X *

Master: I don't want to hear that. You can't keep killing yourself and not do something about it. X *

Robot: Oh, well I don't want to hear that either. X *

Master: Well, you have to get up. I need to help you. X *

Robot: I am going to bed. X *

Master: I will wait for you. X *

Je ráno. Robot vstoupí do pokoje svého pána, který je opravdu starý a nemocný. Robot vidí, že jeho pánovi se dnes ráno moc nedaří. Sídí na kraji postele a bere ho za ruku.

Robot: Pamatuju si, jak jsi na tuhle postel skákala, když jsi byla malá. Byl jsi plný energií.

Mistr: Oba víme, že umírám.

Robot: Pamatuju si, že jsi to říkal taky, ty starý nemocný dědku. A jak ses mnohokrát pokusil zabít.

Mistr: Zkoušel jsem to, ale je to tak těžké. Chci být naživu. Chci být se svými dětmi a vnoučaty.

Robot: Je mi líto, ale to nemůžeš. V žádném případě. Ani náhodou. V žádném případě.

Mistr: To nechci slyšet. Nemůžu se pořád zohýbat a nic s tím dělat.

Robot: No, to taky nechci slyšet.

Mistr: No, musíš vstát. Musím ti pomoci.

Robot: Jdu si lehnout.

Mistr: Počkám na tebe.

Al.
KDYZ ROBOT PISE
HRU

Producing the Script



Producing the script

- Generating individual scenes
- Post-processing the script

Producing the script

- Generating individual scenes
 - THEaiTRobot operated by a professional theatre dramaturge
 - interventions
 - the input prompts
 - choosing different variant
 - manually adding a line
- Post-processing the script

Producing the script

- Generating individual scenes
 - THEaiTRobot operated by a professional theatre dramaturge
 - interventions
 - the input prompts
 - choosing different variant
 - manually adding a line
- Post-processing the script
 - deletions
 - small edits
 - fixing errors in automated EN→CS translation

Producing the script

- Generating individual scenes (15 generated, 8 selected)
 - THEaiTRobot operated by a professional theatre dramaturge
 - interventions
 - the input prompts (2% of lines)
 - choosing different variant (for 8% of generated lines)
 - manually adding a line (1% of lines)
- Post-processing the script
 - deletions (15% of lines)
 - small edits (30% of lines, 8% of words)
 - fixing errors in automated EN→CS translation (20% of words)

Producing the script

- Generating individual scenes (15 generated, 8 selected)
 - THEaiTRobot operated by a professional theatre dramaturge
 - interventions
 - the input prompts (**2% of lines**)
 - choosing different variant (for 8% of generated lines)
 - manually adding a line (**1% of lines**)
- Post-processing the script
 - deletions (15% of lines)
 - small edits (30% of lines, **8% of words**)
 - fixing errors in automated EN→CS translation (20% of words)
- ...so ~90% of the script comes from GPT-2

A photograph of a theater stage. The stage is dark, and the audience seating is blue. A red banner with the text "Staging the Play" is overlaid on the image. The stage is set with a blue backdrop and a balcony. A person is standing on the stage, facing the audience. Several audience members are seated in the front rows, some wearing masks. A laptop is on the stage floor in front of the person.

Staging the Play

Staging the play

- Only dialogues generated
- Everything else done manually
 - stage directions, scene design, music, costumes...

Staging the play

- Only dialogues generated
- Everything else done manually
 - stage directions, scene design, music, costumes...
- Professional theatre team
 - 1 director (Daniel Hrbek)
 - 2 dramaturges
 - 6 actors
 - scenographer, costume designer, choreographer...

Staging the play

- Only dialogues generated
- Everything else done manually
 - stage directions, scene design, music, costumes...
- Professional theatre team
 - 1 director (Daniel Hrbek)
 - 2 dramaturges
 - 6 actors
 - scenographer, costume designer, choreographer...
- Online premiere on 26th February 2021
 - ~ 10 000 views
- Live performances once possible

Summary

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

- Hacking the GPT-2 language model
- Building a web application
- Generating a theatre play script
- Performing the play on stage

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