

GPT-2 Creating Star Trek Stories





What is GPT-2 ?

GPT : Generative Pre-trained Transformer

- It was released on Feb14th 2019 by OpenAI.

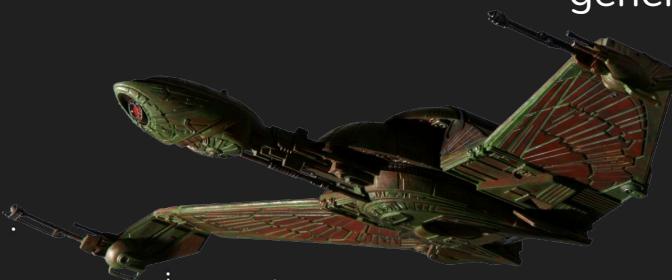
- Largest model has 1.5 billion parameters built off *WebText*.

What can it do?

GPT-2 can translate text, answers questions, summarize passages, and generates text output.

How does it work?

GPT is a general purpose learner implemented by a deep neural network. It was not created to do anything other than predict the next word in a sequence.



Let's use GPT-2



01

02

03

Data Collecting

Simple collecting & cleaning into a easy to read script format.

Setting up the environment

After many struggles Google Colab came out on top.

Training the model

Lots of playing with the parameters.

01

Data Collection & Cleaning

The Data Cleaning Process



Star Trek TOS

2641 kb



Star Trek TAS

415 kb



Star Trek TNG

5550 kb



Star Trek VOY

5905 kb



Star Trek DS9

5612 kb



Star Trek ENT

2905 kb



Star Wars ep4,5,6

174 kb

Cleaning the text

▼ root:
▼ DS9:

episode 0: " The Deep Space Nine Transcripts - Emissary Emissary Stardate: 46379.1 Original Airdate: 3 Jan, 1993 On Stardate 43997, Captain Jean-Luc Picard of the Federation Starship Enterprise was kidnapped for six days by an invading force known as the Borg. Surgically altered, he was forced to lead an assault on Starfleet at Wolf 359. [Saratoga - Bridge] LOCUTUS [on viewscreen]: Resistance is futile. You will disarm your weapons and escort us to sector zero zero one. If you attempt to intervene, we will destroy you. CAPTAIN: (a Vulcan) Red alert. Load all torpedo bays. Ready phasers. Move us to position alpha, Ensign. (The space battle begins) OPS OFFICER: (woman) They've locked on. SISKO: Rerouting auxiliary power. OPS OFFICER: Our shields are being drained. Sixty percent. Forty two. CAPTAIN: Recalibrate shield nutation. TACTICAL: (Bolian) Modulation is having no effect. OPS OFFICER: Shields have failed. SISKO: Full reverse. CAPTAIN: Maintain all Argh! (Everything goes BOOM) SISKO: Damage report. Damage report! COMPUTER: Warning. Damage to warp core. Containment failure in five minutes. TACTICAL: Direct hit decks one through four. SISKO: Let's get the civilians to the escape pods, Lieutenant. TACTICAL: Aye, sir. [Corridor] SISKO: Leave everything. Get to your assigned evacuation area now. COMPUTER: Warning. Damage to warp core. Containment failure in four minutes. (They come to a woman) SISKO: Doran. TACTICAL: I'll take care of her. You go on. SISKO: Have you seen Jennifer? [Sisko's quarters] (The room is on fire and Sisko's wife and son are trapped beneath girders and debris) SISKO: Jennifer? Jake! COMPUTER: Warning. Damage to warp core. Containment failure in three minutes. (Sisko starts throwing debris away and pulls the boy out.) SISKO: I'm going to get your mom. You're going to be okay. (More debris shifting) SISKO: Jennifer. Okay, Jake, we'll get your mom out and we'll get out of here. TACTICAL: Commander! SISKO: Help me. Jennifer, hold on. TACTICAL: (scanning her) Sir. SISKO: Just help me to get her free. TACTICAL: She's gone. There's nothing we can do. We have to leave. COMPUTER: Warning. Damage to warp core. Containment failure in two minutes. TACTICAL: Ensign, take the boy. ENSIGN: Aye, sir. TACTICAL: We've got to go now, sir. SISKO: Damn it, we just can't leave her here. (The Bolian

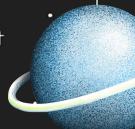
Filter...

The Deep Space Nine Transcripts - Emissary
Emissary
Stardate:
46379.1
Original Airdate: 3 Jan, 1993

On Stardate 43997, Captain Jean-Luc Picard of the Federation Starship Enterprise was kidnapped for six days by an invading force known as the Borg. Surgically altered, he was forced to lead an assault on Starfleet at Wolf 359.
[Saratoga - Bridge]

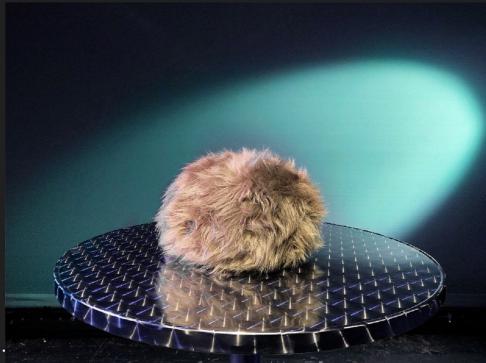
LOCUTUS [on viewscreen]: Resistance is futile. You will disarm your weapons and escort us to sector zero zero one. If you attempt to intervene, we will destroy you.

CAPTAIN: (a Vulcan) Red alert. Load all torpedo bays. Ready phasers.
Move us to position alpha, Ensign.
(The space battle begins)



02

Setting up the Environment



GPT-2 setup is a tribulation!

GPT-2 has many requirements on older libraries.

- Create a conda environment!
- TensorFlow == 1.15
- Python == 3.7
- Install Transformers from Huggingface
- Install Pytorch

This worked well for text generation however training the model did not work on my personal computer.

+ You need a GPU to fine-tune effectively.



Google Colab saved the day

Colab has
transformers
pre-installed

You will have to
downgrade
tensorflow

Set runtime
type to **GPU**

Install **GPT-2**
model : I used the
medium : (355M)

Mounting to
Google Drive
to import data.

Train and
generate text!



03

Training the Data & Generation



Using GPT-2 Simple



A Medium Model

B Fine tuning

C Save Point

355 M is the largest model we can finetune

`gpt2.finetune(
my_data)`

Save the checkpoints, so you can load up the model later.

(The Doctor walks in)

BASHIR: Well done. Now where is it we're going?

GARAK: I'll have the replicator back within the hour.

(Garak hands over a device from the replicator.)

BASHIR: What is it?

GARAK: A small golden chalice.

BASHIR: I thought you said it was a Promenade prize.

GARAK: It's a gift.

BASHIR: How did you know?

GARAK: It's a gift.

BASHIR: Lazare broke into my datapoints, and I had to track down his source.

GARAK: It's a gift.

BASHIR: So, where are we going?

GARAK: I don't know, but the time allows.

BASHIR: You mean, we're going to Romulus?

GARAK: Yes, but not alone. The information here is on the records of

Grunts VI. Anyone can find it. A little girl from the Jadzia cluster.



How does the generated text look?

A lot of repetition.

The formatting structure came out great.

Sometimes characters talk to themselves or talk in the 3rd person.

"BEN" "I think we've got something, too. The fall of Sigma Octanus was inevitable."

"HAN" "Not since you've been around long enough to have become paranoid."

"BEN" "I'm not going to argue with that."

"THREEPIO" "Oh, I'm beginning to suspect that you're not quite so innocent after all."

"HAN" "Well, if I may say so, I've got something to tell you. And it's a very good one, kid."

"THREEPIO" "You know, I think you're going to like it here."

"THREEPIO" "I'm going to have a few words with your landlord."

"HAN" "What do you mean? I'm going to have a few words with your landlord."

"THREEPIO" "You're going to be out of here very soon, Hunchback."

"HAN" "Good. Stay here and watch after the camp, okay?"

"THREEPIO" "I have a bad feeling about this."

"HAN" "Okay, stay here and watch after the camp."

"THREEPIO" "What happened?"

"HAN" "Well, I heard some strange noises this morning. I took a few moments to investigate, but I'm afraid I'm not quite sure what you're talking about."

"THREEPIO" "What sound do you mean?"

"HAN" "Well, I'm afraid I'm not quite sure. I'm afraid the club fights are getting worse and worse every day."

"THREEPIO" "Are you kidding! How could I be so stupid? He really is a great guy."



KIRK: All right, Mister Spock, scan the level.

[Sickbay]

KIRK [OC]: Stand by. All personnel, remember.

Mister Spock, we've been dealing with a major battle
and you have it in your head to show your compassion.

[Corridor]

(A little later, Kirk is already showing compassion.)

MCCOY: You know those little physiological ticks on your skin? The thing we
had to try was to break through to the brain and get those little things
in there. But I had a theory, Jim.

KIRK: I'd love to try it.

(He reaches for his own phaser, and gets knocked out.)

MCCOY [OC]: Where's the doctor?

[Sickbay]

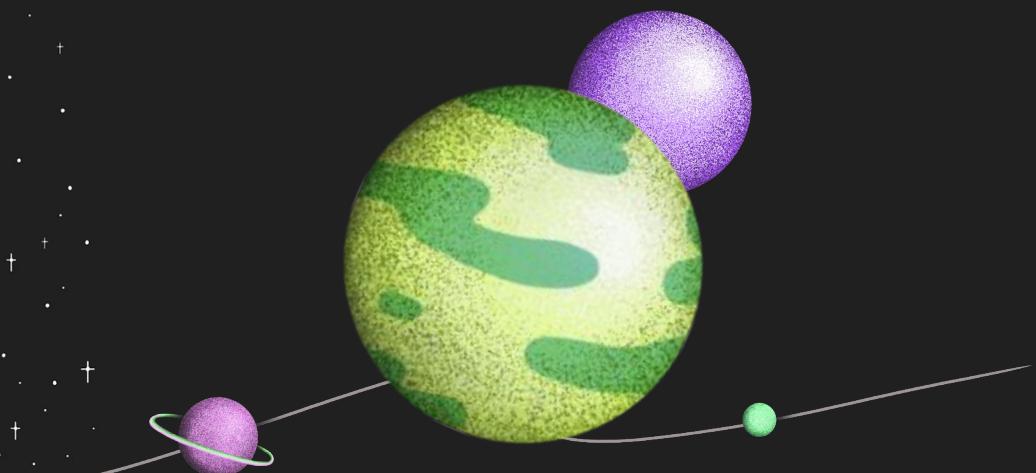
SCOTT: The Captain was very disturbed by
your behaviour. I can assure you all and all that he
would be deeply interested in learning what a Vulcan mind can

Increasing the randomness allowed
for less repetitive storytelling.

I increased the randomness
temperature
from 0.7 to 0.9 here.

Mccoy is talking to himself in the third
person here, he is the doctor.

WHAT CAN WE DO WITH THIS TEXT?



The scripts are interesting and can be edited into a coherent story. However, as is they are sprinkled with illogical statements and repetition.

I began working on using this text and running it through a multi-class classification model, to improve upon my previous project which deciphered Star Wars vs Star Trek.

For a longer term goal, a text app that could decipher which show you're talking about, and could make new stories.

