

# Kreiranje fiktivnih priča koristeći GPT-Neo

---

**Kanjovsky, Tin**

**Undergraduate thesis / Završni rad**

**2023**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Pula / Sveučilište Jurja Dobrile u Puli**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:137:110695>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-08-05**



*Repository / Repozitorij:*

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli

**Tin Kanjovsky**

**Kreiranje fiktivnih priča koristeći GPT-Neo**

**(Making a fiction story generator using a GPT-Neo)**

Završni rad

**Pula, 2023.**

Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli

**Tin Kanjovsky**

**Kreiranje fiktivnih priča koristeći GPT-Neo**

**(Making a fiction story generator using a GPT-Neo)**

Završni rad

**Ime Prezime studenta/studentice, JMBAG: Tin Kanjovsky, 0303061089**

**Studijski smjer: preddiplomski sveučilišni studij informatika**

**Znanstveno područje: Društvene znanosti**

**Znanstveno polje: Informacijske i komunikacijske znanosti**

**Znanstvena grana: Informacijski sustavi i informatologija**

**Mentor: izv.prof.dr.sc. Darko Etinger**

**Pula, rujan 2023.**

## IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani \_\_\_\_\_, kandidat za prvostupnika  
\_\_\_\_\_ ovime izjavljujem da je ovaj  
Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim  
istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene  
bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na  
nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio  
rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten  
za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

\_\_\_\_\_

U Puli, \_\_\_\_\_, \_\_\_\_\_ godine

## IZJAVA

o korištenju autorskog djela

Ja, \_\_\_\_\_ dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom

\_\_\_\_\_

\_\_\_\_\_ koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu

U Puli, \_\_\_\_\_, \_\_\_\_\_ godine

Potpis

\_\_\_\_\_

# SADRŽAJ

<b>SAŽETAK (ABSTRACT)</b> .....	1
<b>1. UVOD</b> .....	1
<b>2. METODOLOGIJA</b> .....	5
<b>2.1. MODEL</b> .....	5
<b>2.2. PODACI</b> .....	7
<b>2.3. METODE</b> .....	8
<b>3. REZULTATI</b> .....	13
<b>4. RASPRAVA</b> .....	16
<b>5. ZAKLJUČAK</b> .....	19
<b>Popis literature i drugih izvora podataka koji su upotrijebljeni u izradi završnog rada</b> .....	20

## SAŽETAK (ABSTRACT)

Ovaj rad istražuje koncept generiranja izmišljenih priča koristeći GPT-Neo, besplatan i open-source jezični model. Generator bi imao za cilj olakšati stvaranje zanimljivih i originalnih izmišljenih priča. GPT-Neo-ova sposobnost učenja iz velikih skupova podataka i generiranja koherentnog i prirodnog jezika čini ga alatom koji bi mogao uspješno obaviti ovaj zadatak. Model je fino podešen koristeći veliki skupa podataka od 300 tisuća napisanih priča uparenih s upitima s online foruma. Suočavajući se s izazovom fino podešavanja GPT-Neo modela na posebno odabranom skupu podataka bogatom fikcijom, ostvareno je značajno poboljšanje u razumijevanju i kvaliteti generiranih tekstova. Fino podešeni model pokazuje značajna poboljšanja, kako i u metrikama poput perpleksnost, tako i u generiranim pričama u odnosu na osnovni model.

**Ključne riječi:** GPT-Neo, prirodna obrada jezika (NLP), Transformers, fine-tuning, generiranje teksta, kreativno pisanje

This paper explores the concept of generating fictional stories using GPT-Neo, a free and open-source language model. The fiction story generator aims to facilitate the creation of interesting and original fictional stories. GPT-Neo's ability to learn from large datasets and generate coherent and natural language makes it a tool that could successfully perform this task. The model is fine-tuned using a large dataset of 300,000 written stories paired with prompts from online forums. Faced with the challenge of fine-tuning the GPT-Neo model on a specially selected dataset rich in fiction, significant improvements in the understanding and quality of generated texts have been achieved. The finely tuned model shows significant improvements both in metrics like perplexity and in the generated stories compared to the baseline model.

**Keywords:** GPT-Neo, Natural Language Processing (NLP), Transformers, fine-tuning, text generation, creative writing

# 1. UVOD

Pričanje priča u generiranju teksta je zanimljivo područje u tehnologiji obrade prirodnog jezika (NLP). Predstavlja umjetnost i znanost izrade narativa putem automatiziranih sredstava, gdje je cilj stvoriti koherentne i zanimljive priče koje zahvaćaju biti ljudske kreativnosti i komunikacije. Ovo polje u nastajanju pomiče granice trenutnih mogućnosti generiranja teksta, zahtijevajući delikatnu ravnotežu između tematske dosljednosti, kreativnosti i strukturalnog integriteta uvjerljive radnje.

Natural Language Processing (NLP) je područje umjetne inteligencije koje se usredotočuje na interakciju između računala i ljudskog jezika (Brown i ostali, 2020). Njegov glavni cilj je omogućiti računalima razumijevanje, interpretaciju i generiranje ljudskog jezika na način koji je smislen i koristan. NLP kombinira principe računalne znanosti, lingvistike i strojnog učenja kako bi razvio algoritme i modele koji mogu obraditi i analizirati tekstualne podatke. Među brojnim otkrićima u ovom području, pojava velikih jezičnih modela istaknula se kao značajna prekretnica. Primjer za NLP u kontekstu generiranja teksta je GPT-Neo, veliki jezični model temeljen na tehnologiji Transformer. GPT-Neo može razumjeti i generirati ljudski jezik na visokoj razini složenosti.

Ovaj rad istražuje pokušaj stvaranja generatora fiktivnih priča pomoću GPT-Neo, jednog od najsuvremenijih jezičnih modela. GPT-Neo je vrsta jezičnog modela temeljena na arhitekturi Transformer razvijenih od strane EleutherAI, posebno dizajnirana za zadatke obrade prirodnog jezika (NLP). To je produžetak i varijacija GPT (Generative Pre-trained Transformer) obitelji modela, čiji je cilj ponuditi ravnotežu između veličine modela i performansi. Naziv "GPT" označava "Generative Pre-trained Transformer," što se odnosi na osnovnu arhitekturu, dok "Neo" označava novu ili unaprijeđenu verziju ovih modela (*GPT Neo*, 2020/2023). Kao i drugi GPT modeli, GPT-Neo je istreniran na ogromnoj količini tekstualnih podataka s interneta, što mu omogućava da nauči obrasce i strukture prirodnog jezika. Ovo treniranje omogućava GPT-Neu da izvršava širok spektar zadataka vezanih za razumijevanje i generiranje prirodnog jezika, kao što su dopunjavanje teksta, prijevod, odgovaranje na pitanja i mnoge druge, s impresivnom



preciznošću i točnošću (Brown i ostali, 2020). Različite verzije GPT-Nea se mogu razlikovati po veličini modela, podacima za treniranje i fino podešavanje za specifične zadatke. Nekoliko ključnih karakteristika i značajki GPT-Neo su:

- **Transformer arhitektura:** GPT-Neo je izgrađen na Transformer arhitekturi, koja je postala temelj za mnoge najsuvremenije NLP modele. Ova arhitektura omogućuje hvatanje složenih jezičnih obrazaca i odnosa u tekstualnim podacima.
- **Prethodno uvježbavanje(Pre-training):** Kao i njegovi prethodnici, GPT-Neo prolazi pre-training fazu u kojoj uči predvidjeti sljedeću riječ u rečenici na temelju golemog korpusa tekstualnih podataka. Ova faza prije obuke pomaže modelu da stekne široko razumijevanje jezika i znanja o svijetu.
- **Fino podešavanje (Fine-tuning):** Nakon prethodne obuke, GPT-Neo se može fino prilagoditi određenim NLP zadacima. Fino podešavanje je proces prilagodbe modela za dobru izvedbu određenih zadataka, kao što je klasifikacija teksta, analiza osjećaja ili generiranje teksta. Fino podešavanje čini GPT-Neo svestranim i sposobnim za različite zadatke vezane uz jezik.
- **Varijante veličina:** GPT-Neo dolazi u različitim varijantama veličina, omogućujući korisnicima odabir modela koji odgovaraju njihovim specifičnim zahtjevima u pogledu veličine modela i računalnih resursa. Manje varijante su pristupačnije za zadatke s ograničenim računalnim resursima, dok veće varijante nude poboljšane performanse.
- **Razumijevanje prirodnog jezika:** GPT-Neo može razumjeti i generirati tekst sličan ljudskom, što ga čini vrijednim za zadatke poput generiranja teksta, sažimanja, prijevoda jezika i razvoja chatbota.
- **Kreativno pisanje i generiranje sadržaja:** GPT-Neo se često koristi za zadatke kreativnog pisanja, kao što je generiranje priča, pjesama ili sadržaja za razne aplikacije. Može pomoći piscima pružajući inspiraciju i generirajući tekst na temelju zadanih upita.

U ovom radu korištena je varijanta GPT-Neo 125M (*EleutherAI/gpt-neo-125m · Hugging Face*, 2023) zbog toga što je to model EleutherAI-a s najmanje parametara, i moguće ga je stabilno pokretati na Paperspace platformi. Također, GPT-Neo je opširniji i bolje se ponaša u određenim slučajevima od GPT-3, koji ujedno nije ni open-source. Za fino podešavanje korišten je skup podataka koji sadrži 300.000 priča i promptova za te priče. Tim iz Facebook-a (Fan i ostali, 2018) je sakupio priče i promptove s Reddit-ovog WritingPrompts foruma (*Prompts and motivation to create something out of nothing*, bez dat.), koje su korisnici pisali tijekom 3 godine, i kombinirao ih u ovaj skup podataka. WritingPrompts je Reddit zajednica gdje korisnici inspiriraju jedni druge da pišu tako što postavljaju promptove ili upite, a drugi korisnici slobodno odgovaraju u obliku priča. Svaki upit može imati više priča. Promptovi pokrivaju širok spektar tema, dužina i detalja. Priče moraju sadržavati najmanje 30 riječi, izbjegavati vulgarnost i neprikladan sadržaj, i trebaju biti inspirirane odgovarajućim promptom.

Prompt: [WP] You are cooking lunch in the kitchen when you suddenly realize that you have misplaced your kitchen knife. Luckily, when you turn around, your teddy bear is holding a knife and approaching you.
Priča: Ted was such a good little bear. Even with that scary expression he just looked so cute. "Now Ted I know you want to help me but you need to be careful swinging that thing around, somebody could get hurt" I said It almost seemed like Ted had smiled after hearing that, but maybe I was just seeing things. Leaving that aside Ted was really helpful whenever a tool or a knife had gone missing I would see Ted holding onto it just a few steps behind me. I know he's a stuffed animal but how does he always know that I'm looking for a certain tool. But he's also very clumsy. The other day when I returned home I noticed there were shards of glass scattered all around the house. Ted was in the distance trying to grab a glass cup but he kept on dropping them...

Tablica 1. Primjer prompta i priče s Reddit WritingPrompts foruma

Ovaj će rad istražiti koliko je učinkovit i efikasan fino podešeni GPT-Neo u kreiranju fiktivnih priča i dali bi mogao poslužiti kao pomoć u kreativnom pisanju. Link na model se može pronaći na Hugging Face stranici:

[https://huggingface.co/Tincando/fiction\\_story\\_generator](https://huggingface.co/Tincando/fiction_story_generator). Link na izvorni kod se nalazi ovdje: [https://github.com/Tincando/GPT-Neo\\_fiction\\_story\\_generator/](https://github.com/Tincando/GPT-Neo_fiction_story_generator/)

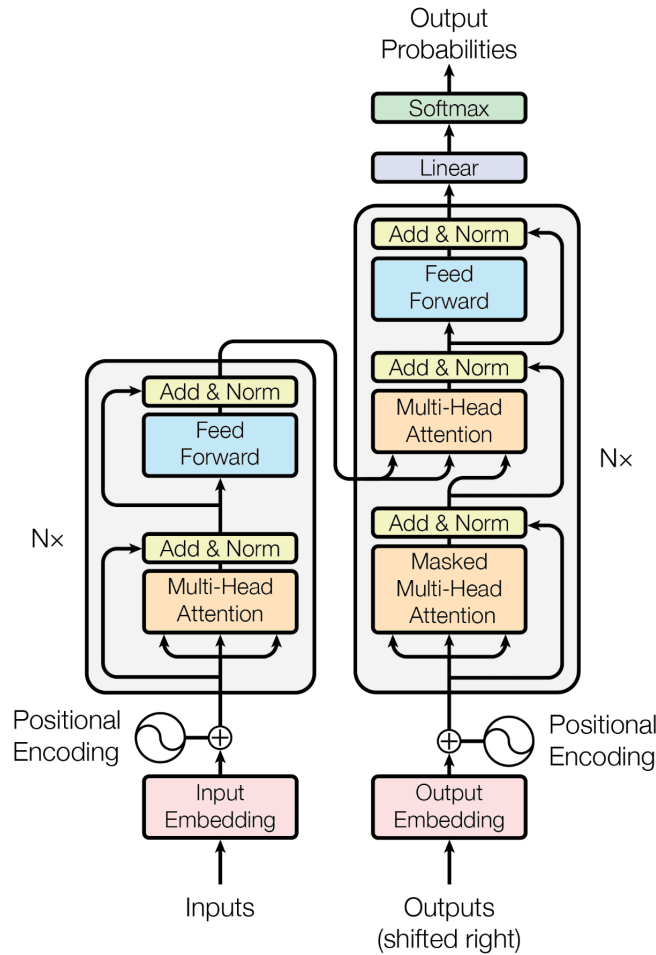
## 2. METODOLOGIJA

### 2.1. MODEL

GPT-Neo, koji je razvio EleutherAI, dio je GPT (Generative Pre-trained Transformer) obitelji modela. Dizajniran je kao učinkovitija i skalabilnija varijanta GPT-a. Dok je GPT-3 tvrtke OpenAI jedan od najpoznatijih modela u ovoj obitelji, GPT-Neo nudi slične mogućnosti sa smanjenim računalnim zahtjevima (*GPT Neo*, bez dat.).

Arhitektura GPT-Neo temelji se na arhitekturi Transformer, koja se pokazala vrlo učinkovitom za različite zadatke obrade prirodnog jezika. Kao i drugi modeli u obitelji GPT, GPT-Neo koristi Transformer arhitekturu. Ova arhitektura uključuje slojeve mehanizama samopozornosti i neuronske mreže s povratnom vezom, omogućavajući modelu da uhvati kontekstualne informacije iz ulaznog teksta (Vaswani i ostali, 2023). Radi se o modelu jezika sličnom GPT-2, obučenom na velikom skupu podataka zvanom the Pile (Gao i ostali, 2020). Arhitektura je slična GPT-2, osim što GPT Neo koristi lokalnu pažnju u svakom drugom sloju s prozorom veličine 256 tokena.

Temelj GPT modela je Transformer arhitektura, kako su izvorno predložili Vaswani i ostali u radu "Pažnja je sve što trebate". Ova arhitektura koristi mehanizme samo-pažnje (self-attention) za obradu ulaznih podataka i hvatanje odnosa između elemenata u podacima. GPT modeli sastoje se od niza identičnih ili sličnih Transformer blokova (Vaswani i ostali, 2023). Svaki blok se sastoji od dvije glavne komponente: mehanizma samo-pažnje s više glava i neuronske mreže za pozicijsko zračenje. Ovi blokovi se slažu jedan na drugi kako bi stvorili duboku mrežu.



Slika 1. arhitektura Transformer modela (Vaswani i ostali, 2023)

GPT-Neo dolazi u različitim veličinama, kao što su GPT-Neo 125M, GPT-Neo 350M i više. Te varijante imaju različite veličine modela i računalne zahtjeve, što ih čini prikladnima za različite primjene i ograničenja hardvera. U svom radu koristio sam GPT-Neo 125M kojeg sam preuzeo preko Hugging Face knjižnice Transformerera (*EleutherAI/gpt-neo-125m · Hugging Face, 2023*).

Hugging Face je popularna platforma za rad s transformer modelima i pruža korisnicima jednostavan pristup i korištenje različitih modela, uključujući GPT-Neo 125M, putem svoje knjižnice Transformerera. Omogućuje jednostavno i opsežno sučelje za rad s

najsuvremenijim modelima obrade prirodnog jezika (NLP), posebno onima koji se temelje na arhitekturi Transformera (Wolf i ostali, 2020).

## 2.2. PODACI

GPT-Neo je obučen na velikom skupu podataka zvanom the Pile. Pile je engleski korpus teksta od 825 GiB napravljen za obuku velikih jezičnih modela. Pile je izgrađen od 22 različitih pod skupova podataka visoke kvalitete—i postojećih i novo konstruiranih—od kojih mnogi potječu iz akademskih ili profesionalnih izvora (Gao i ostali, 2020).

Za fino podešavanje svog model u cilju generiranja fiktivnih priča korišten je skup podataka iz Facebook Research Fairseq repozitorija (*Fairseq/Examples/Stories at Main · Facebookresearch/Fairseq*, bez dat.). Taj skup podataka prikupljen je s Redditovog foruma WRITINGPROMPTS. Ova internet zajednica potiče korisnike da se međusobno potiču na pisanje putem podnošenja ideja za priče, ili promptova, dok drugi korisnici odgovaraju na njih u obliku kratkih priča. Promptovi koji korisnici postavljaju na forum WRITINGPROMPTS (*Prompts and motivation to create something out of nothing*, bez dat.) pokrivaju širok spektar tema, duljina i razine detalja, osiguravajući raznolikost u skupu podataka. Uklonjene su automatizirane objave botova, izbrisane objave, posebne obavijesti, komentari moderatora i priče koje su bile kraće od 30 riječi. Ovaj proces čišćenja pomaže u osiguravanju kvalitete i relevantnosti podataka. 5% promptova rezervirano je za skup za validaciju, dok je još 5% odvojeno za skup za testiranje. Ovi skupovi često se koriste za procjenu performansi modela tijekom treninga. Zbog prirode podataka koji su izvučeni s internetskog foruma, u skupu podataka postoji znatan broj rijetkih riječi i pogrešaka u pisanju.

<b>Train Stories</b>	<b>272,600</b>
<b>Test Stories</b>	<b>15,138</b>
<b>Validation Stories</b>	<b>15,620</b>
<b>Prompt Words</b>	7.7M
<b>Story Words</b>	200M
<b>Average Length of Prompts</b>	28.4
<b>Average Length of Stories</b>	734.5

Tablica 2. Statistika WritingPrompts skupa podataka (Fan i ostali, 2018)

Sve svemu, ovaj skup podataka pruža raznovrsnu i očišćenu kolekciju promptova i odgovarajućih priča s foruma WRITINGPROMPTS, što ga čini pogodnim za fino podešavanje i evaluaciju modela za generaciju fiktivnih priča.

### 2.3. METODE

Nakon preuzimanja tekstualnih datoteka s Facebook Research Fairseq repozitorija. Trebamo povezati Promptove i priče pošto su spremljene u odvojene datoteke. Datoteka valid.wp\_source sadrži navedene promptove, dok valid.wp\_target sadrži odgovarajuće priče. Također, treba napomenuti da je skup podataka za obuku prilično velik. Kako bi se istovremeno unijeli i upiti i priče u GPT-Neo, upiti i priče se spajaju u jedinstvenu cjelinu. Također, izrađen je manji postupak čišćenja teksta kako bi se osiguralo da se razmaci između riječi i interpunkcijskih znakova ispravno postave. Ovaj postupak čišćenja pomaže u obradi teksta kako bi se osigurala njegova pravilna formatiranost.

```

DATAPATH='writingPrompts'
def combinetext(prompt, story):
    fp=open(os.path.join(DATAPATH,prompt),encoding='utf8')
    fs=open(os.path.join(DATAPATH,story),encoding='utf8')
    prompts=fp.readlines()
    stories=fs.readlines()
    assert len(prompts)==len(stories)
    combine=[]
    for i in range(len(prompts)):
        combine.append(prompts[i].rstrip()+ ' <sep> '+' ".join(stories[i].split()::300))
    return combine

```

Slika 2. funkcija za spajanje prompta s odgovarajućom pričom

```

def cleanpunctuation(s):
    for p in '!,:;?':
        s=s.replace(' '+p,p)
    s=s.replace(' '+n\t', 'n\t')
    s=s.replace(' '+\s', '\s')
    s=s.replace(' '+\re', '\re')
    s=s.replace(' '+\ve', '\ve')
    s=s.replace(' '+\ll', '\ll')
    s=s.replace(' '+\am', '\am')
    s=s.replace(' '+\m', '\m')
    s=s.replace(' '+\ m', '\ m')
    s=s.replace(' '+\ m', '\ m')
    s=s.replace(' '+\ ve', '\ve')
    s=s.replace(' '+\ s', '\s')
    s=s.replace('<newline>', '\n')
    return s

```

Slika 3. funkcija za čišćenje teksta

Kao rezultat ovih funkcija, svaka linija unutar ove ujedinjene datoteke sadrži i prompt i odgovarajuću priču.



```
1 train_text[10]
```

```
"[ WP ] Season 30 of Game of Thrones <sep> Note: I just watched Episode 1 of Season 4 and haven't read any of the books, so I don't know what happens after that. Please no spoilers! This is also my first WP post, but constructive feedback is welcome. Also, just for fun I tried to do an unofficial screenplay format, hence the weird capitalization. \n \n -- - \n \n Fade in to a mound of swords, tips facing us. We hear footsteps on stone, slowly getting closer. Meanwhile, the camera zooms out until we recognize this as The Iron Throne. Camera sweeps around the throne, until we are in front of middle-aged DAENARYS sitting on the throne, face blank. \n \n After several seconds, in addition to the footsteps we now also hear the jangle of a maester's chain. Both sounds then stop, and we hear an aged but familiar voice, `` Your Grace... `` \n \n The camera turns to face SAM, in full maester garb. We see him approaching the camera for a few seconds, and then switches to a side view as he stops at the base of the stairs. DAENARYS nods and SAM climbs the first set of stairs towards the throne and stops. \n \n Looking from behind SAM we see DAENARYS in the throne which towers above her head. To her right stands SIR JORAH, after all these years, he's still protectively looking over her. The Hand of the King's medallion hangs from his armor. \n \n SAM: I've just received word from The Wall. It's still holding, but"
```

Slika 4. primjer kombinirane priče i prompta

GPT-Neo koristi isti tokenizer kao i GPT-2, Byte-Pair Encoding (BPE) za tokenizaciju niza teksta. BPE počinje s vokabularom pojedinačnih znakova i iterativno spaja najčešće parove znakova u nove podtokene, efikasno kodirajući strukturu teksta (Shibata i ostali, 1999).

Kako bi se osigurala konzistentna duljina nizova u istom skupu podataka, postavlja se maksimalna duljina niza na 512 znakova, skraćuju se dulji nizovi i nadopunjavaju kraći. Budući da funkcija tokenizer vraća samo `input_ids` i `attention_mask`, za potrebe obuke modela moraju se pružiti oznake. Stoga se za svaki niz `input_ids` stvara odgovarajući niz oznaka.

```
def create_labels(inputs):
    labels=[]
    for ids,attention_mask in zip(inputs['input_ids'],inputs['attention_mask']):
        label=ids.copy()
        real_len=sum(attention_mask)
        padding_len=len(attention_mask)-sum(attention_mask)
        label[:]=label[:real_len]+[-100]*padding_len
        labels.append(label)
    inputs['labels']=labels
```

Slika br.5 funkcija za kreiranje oznaka

Broj uzoraka za obuku iznosio je 170.375. Korištenjem jednog GPU-a na Paperspace-u za obuku modela, trebalo je nekoliko dana da se model potpuno obuči zbog automatskog gašenja Paperspace-a nakon 6 sati. Za fino podešavanje modela korišten je Hugging Face Trainer API.

Hugging Face Trainer API (Sučelje za treniranje Hugging Face) sastoji se od moćnog alata za fino podešavanje i treniranje modela strojnog učenja, posebno za obradu prirodnog jezika (NLP), koristeći Transformers biblioteku (*Trainer*, bez dat.).

```
from transformers import TrainingArguments

training_args = TrainingArguments(
    output_dir="fiction_story_generator",
    evaluation_strategy="epoch",
    learning_rate=5e-5,
    num_train_epochs=5,
    push_to_hub=True,
    resume_from_checkpoint=True
)
```

Slika br.6 konfiguracija za treniranje

Prije fino podešavanja modela moramo postaviti neke osnovne parametre.

Značenje postavljenih parametra sa slike br.6 su:

- `output_dir`: Određuje direktorij u koji će se spremiti rezultati treniranja.
- `evaluation_strategy`: Postavlja strategiju evaluacije, u ovom slučaju, nakon svake epohe (jedne potpune iteracije kroz trening podatke).
- `learning_rate`: Određuje stopu učenja, tj. koliko brzo model uči iz podataka.
- `num_train_epochs`: Postavlja broj epoha, odnosno koliko puta će model proći kroz sve trening podatke tijekom treniranja.
- `push_to_hub`: Omogućuje automatsko postavljanje rezultata treniranja na Hugging Face Hub, platformu za dijeljenje modela i podataka.
- `resume_from_checkpoint`: Omogućuje nastavak treniranja iz prethodno spremljene kontrolne točke, korisno za ponovno pokretanje treniranja nakon prekida.

```
from transformers import Trainer

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=traindata,
    eval_dataset=validdata,
    data_collator=data_collator,
    tokenizer=tokenizer,
)
```

Slika br.7 definiranje Trainer objekta

Nakon stvaranja Trainer objekta i njegove konfiguracije, koristi se metoda `trainer.train()` kako bi se pokrenuo proces fino podešavanja modela. Trainer će automatski izvršavati treniranje modela kroz zadani broj epoha, evaluaciju i spremanje kontrolnih točaka.



Nakon fino podešavanja možemo primijetiti da je perpleksnost za skup podataka za validaciju 23.01, što je znatno bolje nego rezultat prije fino podešavanja. Ovo poboljšanje u perpleksnosti nakon fino podešavanja ukazuje na to da je model postao bolji u razumijevanju i generiranju teksta na skupu podataka za validaciju. Perpleksnost je mjera koja ukazuje na to koliko je model "zbunjen" ili nesiguran pri predviđanju sljedeće riječi ili sekvence u tekstu (Miaschi i ostali, 2021). Niža perpleksnost obično znači da je model postao precizniji i pouzdaniji u svojim predviđanjima. Zatim prompt iz validacijskog skupa podataka je ponovno odabran, unesen u model, i od modela je zatraženo da generira priču duljine od 300 riječi.

```
=== GENERATED STORY 1 ===
[ WP ] Every person in the world develops a weird mutation/power the day they turn 16. Everyone's powers are always different, some more insignificant than others. You turn 16, and watch as all your friends discover their newfound ability's. That is, until you discover the severity of your own.
<sep> I was born with a power. It was a strange one, I don't know what it was or why it happened, but it was something that I had never really thought about before. The first time it happened, it was a little scary. My parents had been worried about me growing up with a power like this, but they had no idea how to handle it. They had to keep it hidden from everyone else. But I knew that I could be a hero. And so, I did what any good parent would do: I became a superhero.

When my parents were diagnosed with cancer, they had to go through a lot of chemo. At first, they tried everything. A lot of them tried drugs. Some of them tried to stop me from killing myself. Others tried to make me stronger. But I was always stronger than them. So, I was given a power that I never thought I could use, and it made me stronger than anyone else.
=== GENERATED STORY 2 ===
[ WP ] Every person in the world develops a weird mutation/power the day they turn 16. Everyone's powers are always different, some more insignificant than others. You turn 16, and watch as all your friends discover their newfound ability's. That is, until you discover the severity of your own.
<sep> `` I'm not sure what to do about it `` I thought. The last time we met was the day before my 16th birthday. We were walking down the street, and I was looking at a man who looked like he was about to die. He was wearing a suit, and a black tie. His face was covered by a black hood, and his eyes were red.

`` Hey, buddy! How's it hanging? `` I asked. He looked up at me, and I could see that his face was a mix of anger and sadness.

`` It's fine. Just a little bit worried. What's wrong with you? `` He asked. I didn't answer him.

`` Well, I don't know. My power isn't working anymore, but it's still working. `` I said.

`` Oh... `` He said, and I could feel the tears starting to form in his eyes.

`` Yeah, well, I guess I'll just have to try it sometime. Maybe tomorrow? `` I asked.

`` Sur
```

Slika br.9 Primjeri priča fino podešenog modela

Kao što možemo primijetiti na slici br.9, primjeri priča generiranih od fino podešenog modela znatno su bolje kvalitete u usporedbi s onima generiranim prije fino podešavanja. Ovi primjeri priča imaju veću dosljednost u vezi s temama i kontekstom

koji su postavljeni putem promptova. Fino podešavanje modela igra ključnu ulogu u ovom poboljšanju.

Kroz proces fino podešavanja, model se prilagođava specifičnom zadatku ili skupu podataka, čime se omogućava bolje razumijevanje i bolja usklađenost s temama i smjernicama koje se pružaju putem promptova. To rezultira generiranim pričama koje su više relevantne, suptilnije i bolje prate tematski kontekst.

Training Loss	Epoch	Step	Validation Loss
3.0842	1.0	34075	3.1408
3.0026	2.0	68150	3.1275
2.9344	3.0	102225	3.1270
2.8932	4.0	136300	3.1306
2.8517	5.0	170375	3.1357

Slika br.10 prikaz rezultata treniranja

Iz ovih rezultata sa slike br.10 možemo vidjeti da se training loss postupno smanjuje kako model prolazi kroz epohe i korake, što je pozitivan znak. To znači da se model uči na temelju treninga podataka. Međutim, također je važno pogledati validation loss. Ako validation loss počne rasti dok se training loss smanjuje, to može ukazivati na overfitting.

U ovom slučaju, čini se da se training loss smanjuje tijekom epohe, što je dobar znak. Validation loss također je stabilan, što sugerira da model nije overfitting i vjerojatno uči korisne obrasce iz podataka. Cilj u treniranju modela je postići nizak training loss uz održavanje niskog validation lossa kako bismo osigurali dobru generalizaciju na nove podatke.

## 4. RASPRAVA

Početni rezultati bili su obećavajući. Priče generirane fino podešenim GPT-Neo 125M bile su intrigantne, pokazujući sposobnost modela za oblikovanje uzbudljivih priča. Međutim, kao i u svakom kreativnom pothvatu, bilo je nijansi koje treba uzeti u obzir. Jedno značajno zapažanje bilo je da generirane priče, iako solidne kvalitete, ponekad odstupaju od unesenih promptova.

Ovo odstupanje između promptova i generiranih priča ukazivalo je na složenost zadatka. Ostaje pitanje hoće li model s više parametara, kao što je GPT-Neo 2.7B, možda donijeti još bolje rezultate. Ideja o korištenju masivnije neuronske mreže s unaprijeđenim sposobnostima i dubljim razumijevanjem jezika i konteksta zvuči zanimljivo. GPT-Neo 2.7B potencijalno bi mogao nadmašiti GPT-Neo 125M kao generator fikcijskih priča, posebno s dovoljno podataka i računalnih resursa. Međutim, moramo biti svjesni rizika od overfittinga, osobito s većim modelom. Manji modeli poput GPT-Neo 125M mogu biti manje skloni overfittingu.

Kako bi se ocijenila učinkovitost modela, fino podešeni model je testiran korištenjem novih promptova preuzetih s WritingPrompts foruma. Pristupilo se novim temama i promptovima kako bi se procijenila reakcija modela i kakve će priče generirati na temelju tih novih podataka..

Rezultati su predstavljeni u obliku generiranih priča, a svaka priča odražava sposobnost modela da interpretira i proširuje zadane upute u smislene i intrigantne naracije. Ove priče su proizvod modelovog razumijevanja jezika i kreativnosti, i pružaju uvid u to kako se model nosi s različitim temama i kontekstima.

**Primjer prompta 1:** You traveled a year into the future and saw the Earth devoid of life. You wanted to find out what happened and went back day by day, over a 1000 days and there are still no signs of life.

“ What is this? ” I looked up from my book. It was a book I had been reading for years now. A book that had been my home for the past 3 years. My father had always told me that it was a book he had read when he was a child, but I never understood. The pages were black and the words were black. “ This is the future. We have been here for 1000 days and nothing has changed. There is no sign of life in the world. But we have to go forward. If we don't we will be destroyed. So we must go forward. ” I stood there for a few seconds, confused as to why I was still here. I looked around. Everything was black. No sign of life. Nothing. Just a blank white room with nothing. Then I saw it. A small, black object. It was a large, black box. I could not tell where the box ended or what it was made of. I looked closer. It was a large, black cube. I looked again. It was made of metal, and it was covered in dust.

**Primjer prompta 2:** 'Need a hand?', you nod, and watch in horror as your friend lends you a still-warm, still-twitching, perfectly amputated hand from their pocket.

I'm not sure if this is what I wanted to write, or if it's the only thing I've been able to get out of my life for. It was a normal day at work when I received an email that said “ Need help with your new prosthetic? ” I looked up from the computer screen to see my boss, Mr. Smith, sitting on his couch, reading a book about prosthetics and how they could be used to replace people who have lost their limbs. He had a clipboard with him and a pen, and he was scribbling away. “ What do you mean, 'need a hand'? ” I asked, trying to sound as inconspicuous as possible. “ Well, we're all about prosthetics right now. We can't use them to replace people who have lost their limbs, because there's no other way around it. So we need to make sure we don't lose our hands. And we need to keep them warm so they can't die. That's why we're going to make sure that they stay warm, and that they're still alive. You'll have to find a replacement. ” The man on the couch looked at me

**Primjer prompta 3:** You are cooking lunch in the kitchen when you suddenly realize that you have misplaced your kitchen knife. Luckily, when you turn around, your teddy bear is holding a knife and approaching you.

“ I'm sorry, but I don't think this is what I wanted. ” The knife was still in his hand, but it had already begun to move. It had been a long day at work, and he was tired of working for himself. He was tired of the boss, and the constant reminders that he was a failure. His boss had been the same as always, but he was different. He was different from everyone else, and he had to make sure that everyone was right. But he was not. He had to make sure that he was right. And so, he began to make his way towards the kitchen. He had to make sure that the teddy bear was holding the knife. That was the first thing he had to do. As he approached, he noticed something that he had never seen before. A small, round object. It was a small, wooden box. It was made out of a plastic, and it had been placed on the



floor. There was a large hole on the bottom of the box, and it was covered in a thick layer of dirt. He looked around, and found that he was in a large, wooden room with a larg

**Primjer prompta 4:** After being eaten by a monstrous whale, you sing to pass the time; unknowingly, your voice becomes a Siren, drawing in doomed sailors.

`` I'm sorry, " I said. She looked up at me, her eyes wide and her face pale. `` It's okay. We're not gonna be able to get back to the ship again, " she said, her voice cracking. I looked down at my feet, and I saw the ocean. The waves crashed against the rocks, crashing into the shoreline. A few of the sailors were still alive, and I could hear the screams of the sailors. They were screaming for help. `` What are you doing? " she asked. Her voice was soft and soothing. `` Just singing. You can't hear us. " I looked at her, and I saw the sea. It was a sea that was not mine. It was a sea of fire and pain. `` Why did you do this to us? Do we have to die? " `` No, no, it doesn't matter. This is the only thing that matters. " I looked at the ocean again, and I saw the ocean. The waves crashed against the rocks, crashing into the shoreline. A few of the sailors were still aliv

Tablica 3. primjeri generiranih priča

## 5. ZAKLJUČAK

U ovom radu, preuzet je uzbudljiv izazov finog podešavanja GPT-Neo modela koristeći posebno odabrani skup podataka ispunjen zanimljivim pričama iz fikcije. Rezultati su bili značajni, jer su pokazali jasno poboljšanje perpleksnosti, što ukazuje na značajno unapređenje razumijevanja teksta modela. Dodatno je zanimljivo, kada su procijenjene priče generirane fino podešenim modelom, primijećeno je značajno poboljšanje i u kvaliteti.

Međutim, kao student, teško je ne uvidjeti da generativno modeliranje jezika predstavlja izuzetno kompleksan posao. Rastuća razlika između ljudske kreativnosti i sposobnosti umjetne inteligencije postaje sve očitija. Naša vještina oblikovanja priča, prenošenje emocija i prilagođavanje različitim stilovima pisanja dokaz su kompleksnosti ljudskog jezika.

Iako postizemo impresivan napredak, moramo shvatiti da smo i dalje daleko od kopiranja dubine i bogatstva ljudskog jezika. GPT-Neo je koristan alat, ali nije čarobnjak kada je riječ o pisanju cjelovitih priča od 300 riječi. Može biti koristan za prevladavanje blokade pisanja i poticanje kreativnosti, ali nedostaje mu potpuna automatizacija generiranja fiktivnih priča. GPT-Neo može nuditi prijedloge i ideje kako biste započeli pisanje. Međutim, ne treba očekivati da će GPT-Neo obaviti sve teške zadatke. Još nije sposoban stvoriti potpunu priču od početka do kraja s dubinom i emocijama koje može pružiti ljudski pisac. Umjesto toga, koristite GPT-Neo kao pomoćnika u pisanju. To je koristan alat za pisce, ali nije zamjena za ljudsku maštovitost i vještine pripovijedanja. Dakle, iako može biti koristan dio vašeg procesa pisanja, nije prečac do potpunog automatiziranog pisanja fikcije.

## Popis literature i drugih izvora podataka koji su upotrijebljeni u izradi završnog rada

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodi, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. <http://arxiv.org/abs/2005.14165>

*EleutherAI/gpt-neo-125m* · Hugging Face. (2023, svibanj 3). <https://huggingface.co/EleutherAI/gpt-neo-125m>

*Fairseq/examples/stories at main* · facebookresearch/fairseq. (bez dat.). GitHub. Preuzeto 11. rujan 2023., od <https://github.com/facebookresearch/fairseq/tree/main/examples/stories>

Fan, A., Lewis, M., & Dauphin, Y. (2018). *Hierarchical Neural Story Generation* (arXiv:1805.04833). arXiv. <http://arxiv.org/abs/1805.04833>

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., & Leahy, C. (2020). *The Pile: An 800GB Dataset of Diverse Text for Language Modeling* (arXiv:2101.00027). arXiv. <http://arxiv.org/abs/2101.00027>

*GPT Neo*. (bez dat.). Preuzeto 11. rujan 2023., od [https://huggingface.co/docs/transformers/model\\_doc/gpt\\_neo](https://huggingface.co/docs/transformers/model_doc/gpt_neo)

*GPT Neo*. (2023). [Python]. EleutherAI. <https://github.com/EleutherAI/gpt-neo> (Original work published 2020)

Miaschi, A., Brunato, D., Dell'Orletta, F., & Venturi, G. (2021). What Makes My Model Perplexed? A Linguistic Investigation on Neural Language Models Perplexity. *Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, 40–47. <https://doi.org/10.18653/v1/2021.deelio-1.5>

*Prompts and motivation to create something out of nothing*. (bez dat.). Preuzeto 11. rujan 2023., od <https://www.reddit.com/r/WritingPrompts/>

Shibata, Y., Kida, T., Fukamachi, S., Takeda, M., Shinohara, A., & Shinohara, T. (1999). *Byte Pair Encoding: A Text Compression Scheme That Accelerates Pattern Matching*.

*Trainer*. (bez dat.). Preuzeto 11. rujan 2023., od [https://huggingface.co/docs/transformers/main\\_classes/trainer](https://huggingface.co/docs/transformers/main_classes/trainer)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. <http://arxiv.org/abs/1706.03762>

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., ... Rush, A. M. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing* (arXiv:1910.03771). arXiv. <http://arxiv.org/abs/1910.03771>