

# Jokes Generation

## (Decoder-based transformer)



**Team 13**

Rafail Venediktov  
Aliaksei Korshuk  
Panov Evgenii

Input Text

My doctor had to put me on a new medication

Очистить

Исполнить

Generated Text

My doctor had to put me on a new medication that's supposed to help lower the amount of karate in my blood

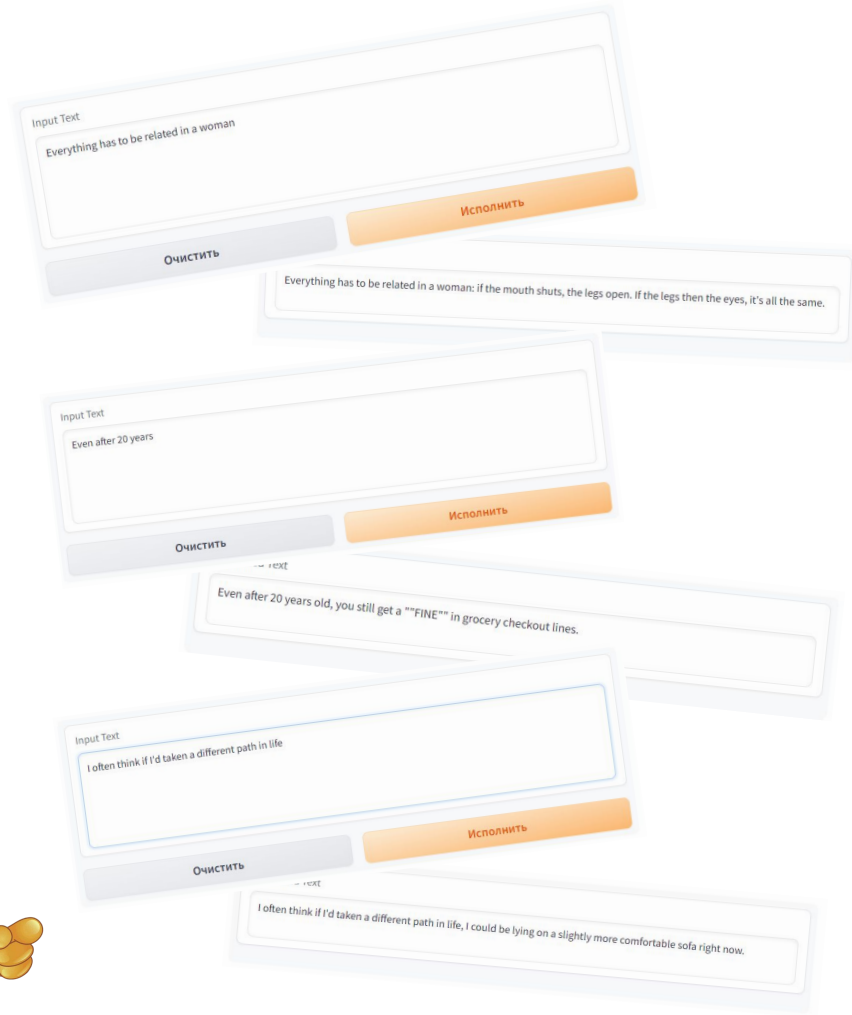
# Project description



The **Joke Generation** (Decoder-based transformer) is a natural language processing tool as a part of an NLP course at the university. The bot generates hilarious and witty jokes from a random seed, making use of natural language processing techniques to analyze the semantics of the input and produce humorous content.


The motivation behind creating this bot is to demonstrate the capabilities of natural language processing while providing an amusing and engaging experience for users.


With its ability to produce unique and diverse jokes, the Joke Generation is a prime example of how AI can be used to enhance our daily lives in ways that were previously unimaginable.







# Dataset





 **Hugging Face**

Hugging Face is way more fun with friends and colleagues!  [Join an organization](#) Dismiss this message

**Datasets:** ● Fraser/short-jokes   2

**Dataset card**  Files and versions  Community

**Dataset Preview** Size: 22.1 MB  API  Go to dataset viewer

**text (string)**

```
"[me narrating a documentary about narrators] "I can't hear what they're saying cuz I'm talking" "
```

```
"Telling my daughter garlic is good for you. Good immune system and keeps pests away. Ticks, mosquitos, vampires... men. "
```

```
"I've been going through a really rough period at work this week It's my own fault for swapping my tamtax for sand paper. "
```

```
"If I could have dinner with anyone, dead or alive... ...I would choose alive. -B.J. Novak- "
```

```
"Two guys walk into a bar. The third guy ducks. "
```

```
"Why can't Barbie get pregnant? Because Ken comes in a different box. Heyooooooo "
```

```
"Why was the musician arrested? He got in treble. "
```

```
"Did you hear about the guy who blew his entire lottery winnings on a limousine? He had nothing left to chauffeur it. "
```



```
"What do you do if a bird shits on your car? Don't ask her out again. "
```



```
"He was a real gentlemen and always opened the fridge door for me "
```



```
"Telling my daughters date that ""she has lice and its very contagious the closer you get to her."" *Correct way to parent. "
```

**YAML Metadata Warning:** empty or missing yaml metadata in repo card (<https://huggingface.co/docs/hub/datasets-cards>)



Downloads last month 404



 Use in dataset library  Edit dataset card

 Train in AutoTrain  Evaluate models

 HF Leaderboard 

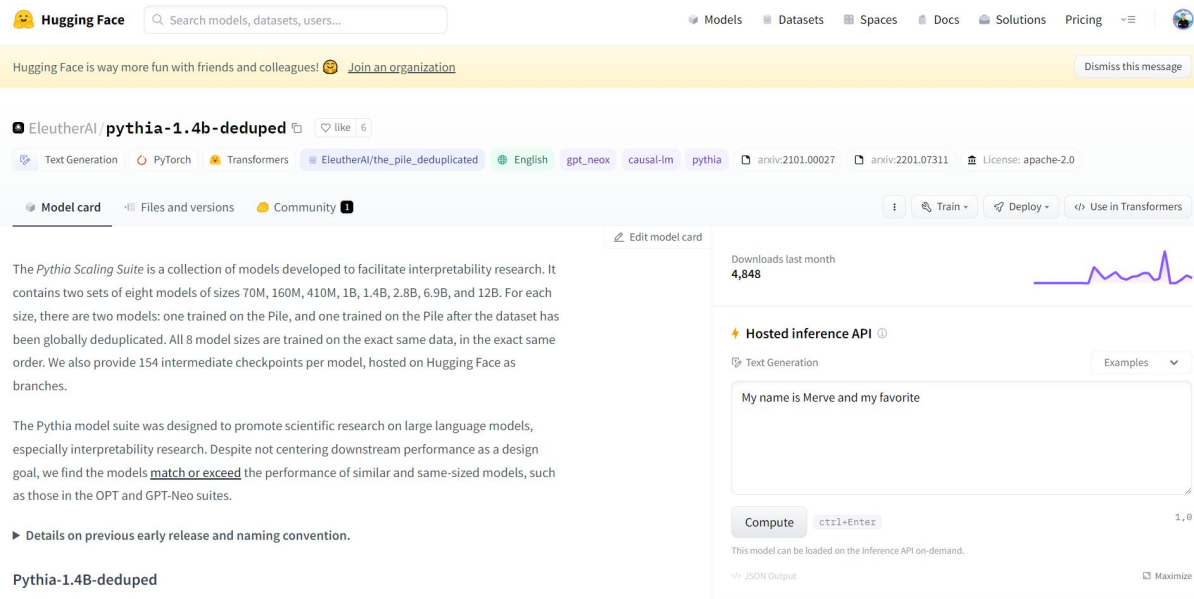
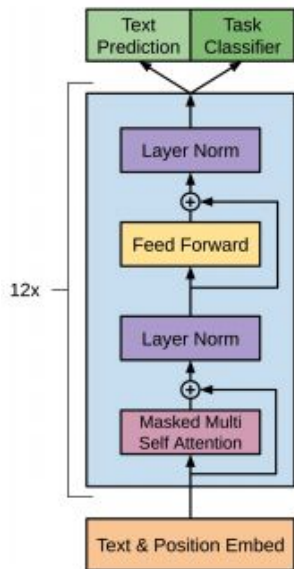
**Models trained or fine-tuned on Fraser/short-j...**

 AlekseyKorshuk/gpt2-jokes  
 Updated Mar 29 · ↓ 31

 AlekseyKorshuk/pythia-1.4b-deduped-jo...  
 Updated 14 days ago · ↓ 7 · ♥ 1

The [Fraser/short-jokes](#) dataset available on the Hugging Face website is a collection of short, humorous jokes in the English language. This dataset contains over 200,000 jokes, which are categorized into different categories such as puns, one-liners, knock-knock jokes, and more.

# Model



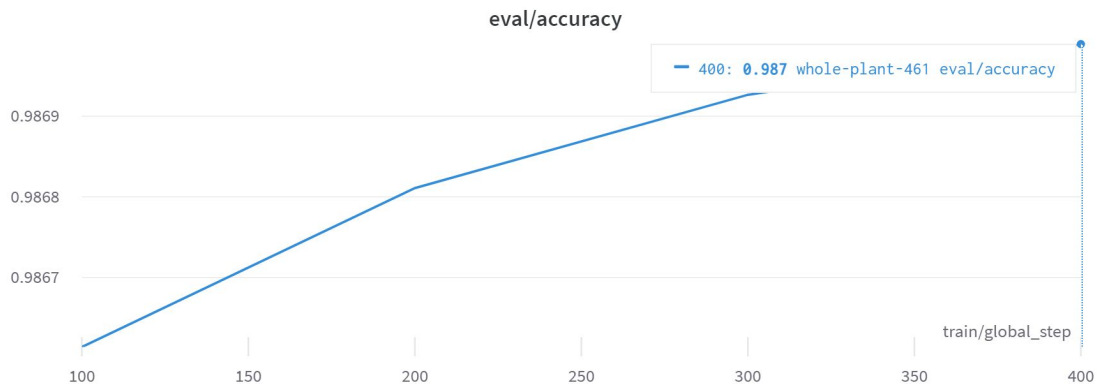
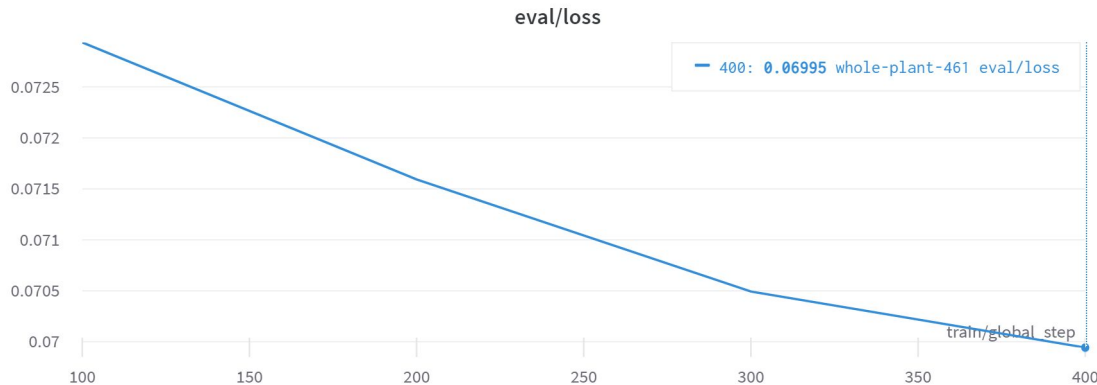
The screenshot shows the Hugging Face interface for the model `EleutherAI/pythia-1.4b-deduped`. The model card includes a search bar, navigation links (Models, Datasets, Spaces, Docs, Solutions, Pricing), and a message about joining an organization. The model card itself has tabs for 'Model card', 'Files and versions', and 'Community'. The 'Model card' tab is active, showing a description of the *Pythia Scaling Suite* and the *Pythia* model suite. The description states that the suite contains two sets of eight models of various sizes (70M, 160M, 410M, 1B, 1.4B, 2.8B, 6.9B, and 12B), trained on the Pile dataset after global deduplication. It also mentions that the models are trained on the exact same data and provide 154 intermediate checkpoints per model. The *Pythia* model suite was designed to promote scientific research on large language models, especially interpretability research. The card also includes a link to 'Details on previous early release and naming convention' and the model name 'Pythia-1.4B-deduped'. On the right side, there is a 'Hosted inference API' section with a 'Text Generation' dropdown, a text input field containing 'My name is Merve and my favorite', a 'Compute' button, and a 'JSON Output' checkbox. A line graph shows 'Downloads last month' with a value of 4,848.

The [Pythia Scaling Suite](#) is a collection of models developed to facilitate interpretability research. Trained on the Pile after the dataset has been globally deduplicated.

In architecture, the model consists of only the decoder component of a sequence-to-sequence (seq2seq) model.

# Evaluation and conclusion

The jock generation model is an AI system that creates new and original jokes for entertainment or marketing purposes. However, it **faces challenges in generating high-quality jokes** that are appropriate for the intended audience. The system's success relies on using good-quality training data, and human input is needed to evaluate the humor and appropriateness of the generated jokes. Ultimately, the system's effectiveness **depends on the quality of the training data** used and human oversight to ensure the jokes meet the desired standards.





Dataset

# TEAM



GitHub

# 13



Playground



**Aliaksei Korshuk**

Training model and  
making a playground



**Rafil Venediktov**

Search and refinement  
of the model



**Evgenii Panov**

Search, analysis, and  
preparation of the dataset