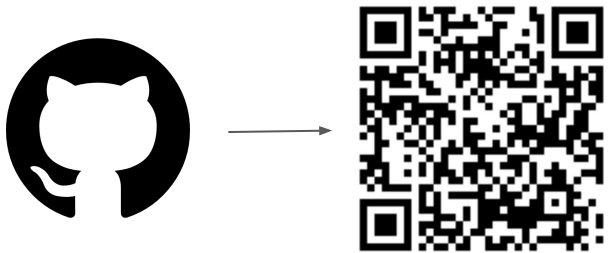
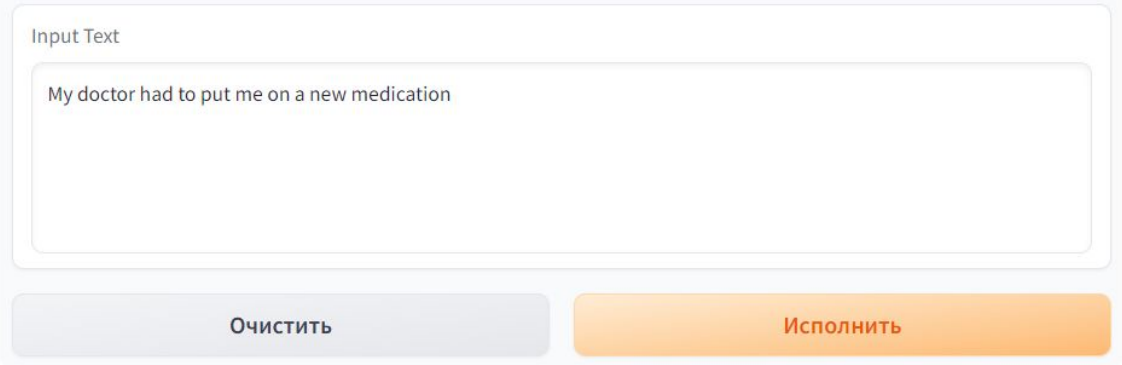
Jokes Generation
(Decoder-based transformer)





Team 13

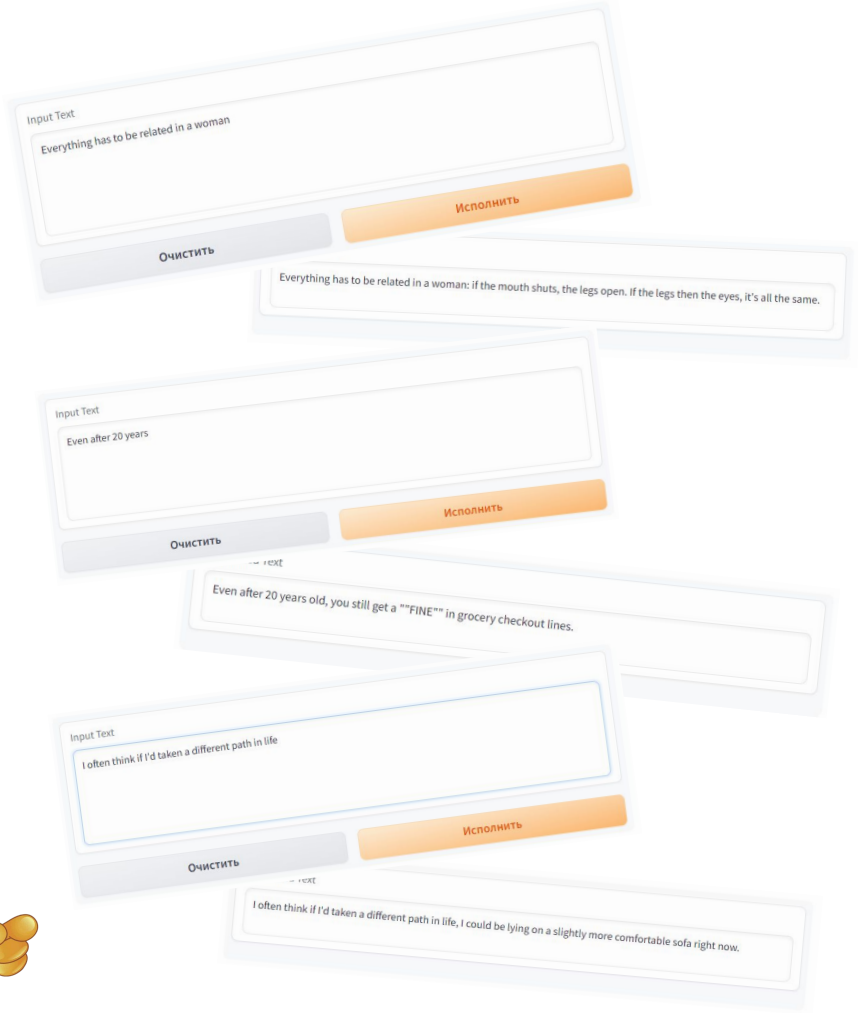
Rafail Venediktov
Aliaksei Korshuk

Project description

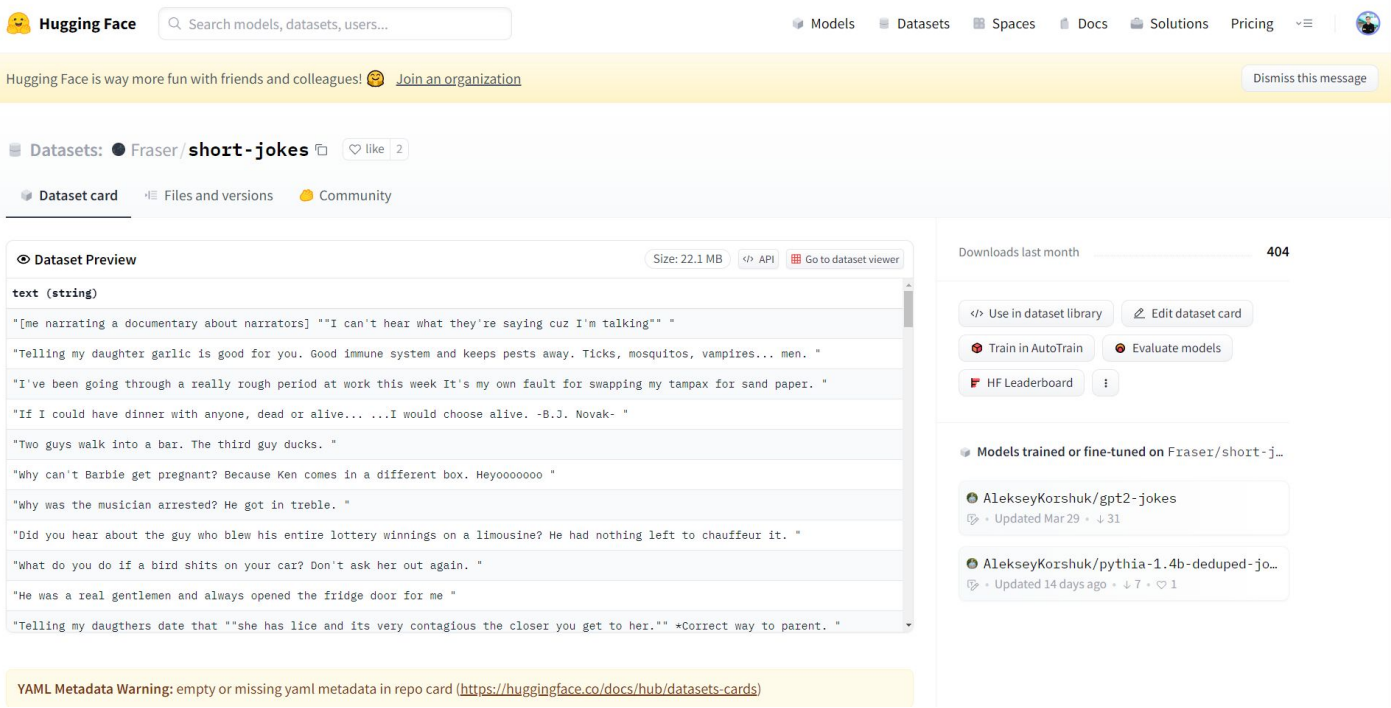
The J**oke Generation** (Decoder-based transformer) is a natural language processing tool as a part of an NLP course at the university. The bot generates hilarious and witty jokes from a random seed, making use of natural language processing techniques to analyze the semantics of the input and produce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

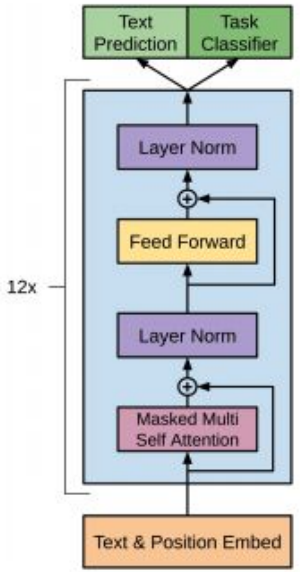


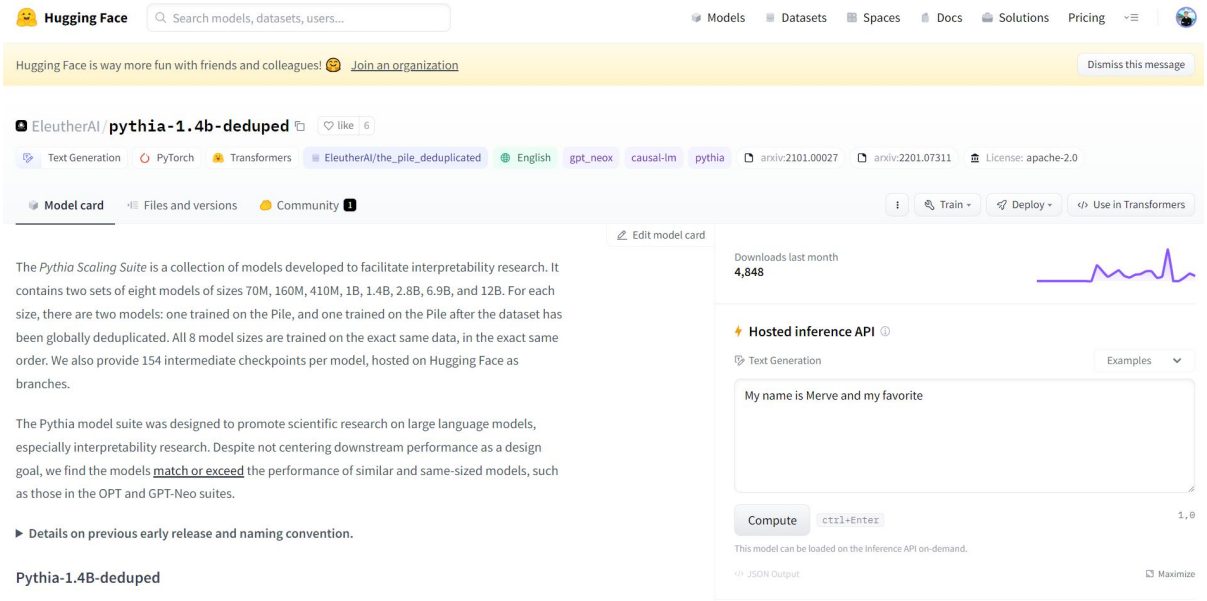
Dataset



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



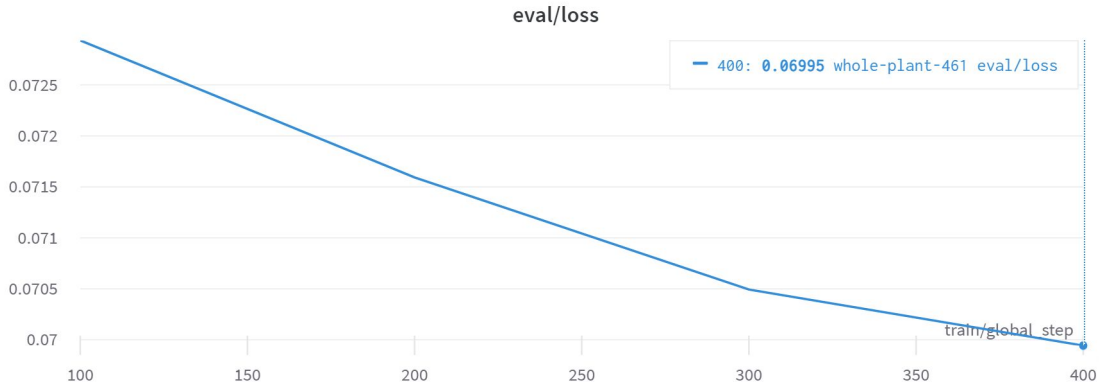


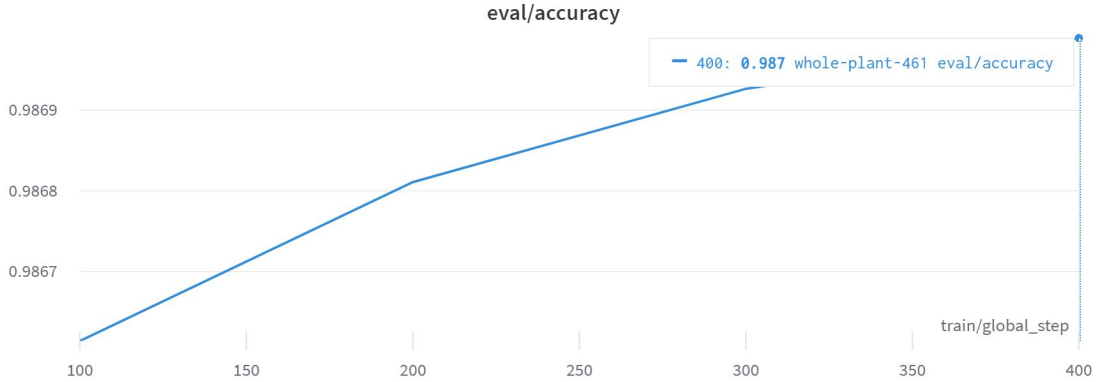
The **Pythia Scaling Suite** is a collection of models developed to facilitate interpretability research. Trained on the Pile after the dataset has been globally deduplicated.

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.





TEAM 13

Dataset

GitHub

Playground

Aliaksei Korshuk

Training model and
making a playground

Rafil Venediktov

Search, analysis and preparation
of dataset and model