GiggleLab: AI-Powered Turkish Joke Generator

Purpose of the Project:

The aim of this project is to develop an artificial intelligence model that understands and generates short humorous narratives known as “fıkra”, which are an integral part of Turkish culture. By combining traditional humor with modern technologies, the project seeks to offer an entertaining and creative experience.

Project Steps

Preparing the Dataset

The jokes (Fıkra) were compiled, cleaned, and structured.

Choosing a Model and Fine-Tuning

Chose the Trendyol LLM and fine-tuned it.

Developing the Website

The code was written, and the interface and model were integrated.

Testing and Reporting

The model outputs were evaluated and the documentation was prepared.

Generating Jokes with Artificial Intelligence

Culture and Entertainment

Artificial intelligence supports the preservation and digitization of cultural elements by being used in humorous content generation, such as the creation of fıkra (short jokes).

AI in Turkish Culture

Artificial intelligence trained with figures like Nasreddin Hoca helps keep local humor alive in digital environments.

Education and Language

Artificial intelligence captures students' interest through humorous stories while also contributing to language development.

Traditional Humor with Artificial Intelligence

AI models trained with fıkra support the preservation of cultural heritage and its transmission to future generations.

Project Development Process

Used Model

Within the scope of the project, the Turkish language model named Trendyol-LLM-7b-chat-dpo-v1.0 was used. This model is a large language model trained on Turkish data and possesses a broad understanding of the language.

Fine-Tuning Process

The model was fine-tuned using a specially compiled fıkra dataset. This dataset included hundreds of traditional and creative jokes, presented in a "prompt" and "story" format. As a result, the model not only learned the structure of fıkra but also successfully generated new ones.

Datasets Used

- fikra_dataset.json: An original dataset consisting of Nasreddin Hoca jokes.
- clean_jokes.json: A cleaned, AI-compatible supplementary humor dataset.
- All data was structured in the "prompt" and "story" format to ensure compatibility with the model.

Technical Infrastructure

- Python: Used for data preparation, model fine-tuning, and interface development.
- GiggleLab Application: The integration of the interface and the model was achieved through the main.py and fikra_llm.py files.

Testing Process

The jokes generated by the model were evaluated based on naturalness, humor level, and grammatical accuracy.

Reading tests were conducted with test users, and interactive presentations were prepared with laughter sound effects.

The test scenarios are documented in the test case report.pdf file.

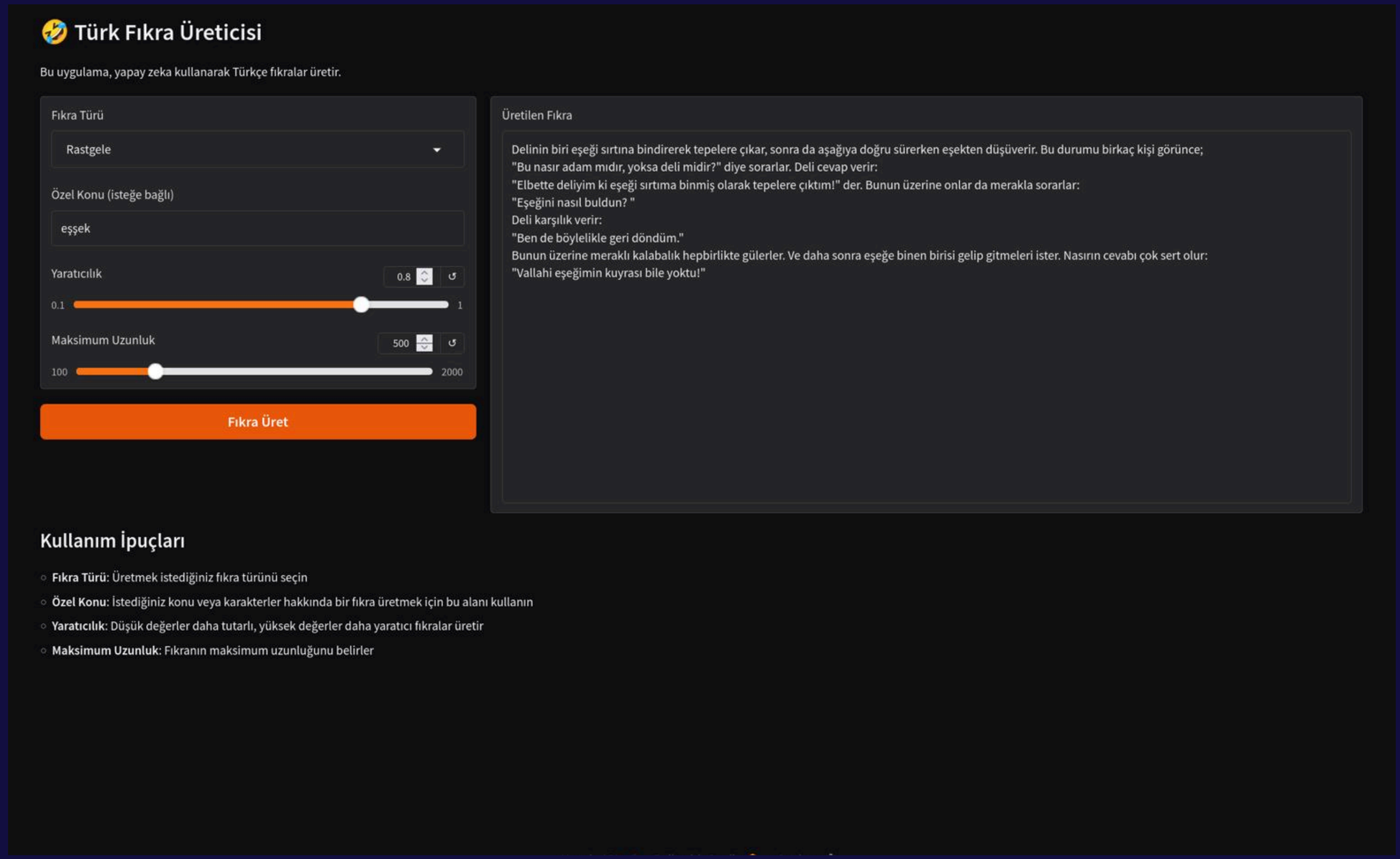
Risks and Precautions

Data quality and diversity play a significant role in humor generation.

To prevent the model from misinterpreting Turkish idioms and cultural elements, the diversity of examples was increased.

The details are presented in the risk_report.pdf document.

Output Sample:



Conclusion

As part of this project, a comprehensive dataset of Nasreddin Hoca jokes was created, and the Trendyol LLM model was fine-tuned with a focus on cultural humor.

The resulting model:

- Understands and generates local humor,
- Has learned language patterns and comedic styles unique to Turkish culture,
- Produces entertaining, creative, and culturally rich content.

This project demonstrates that artificial intelligence can be a powerful tool not only in technical fields but also in preserving and revitalizing cultural values in digital environments.

Thank you for listening!