Vision & Scope Document:

Turkish Joke-Telling AI Model

1- Introduction:

1.1 Project Background:

While large pre-trained language models have shown impressive capabilities in understanding and generating text across multiple domains, they often lack domain-specific expertise, especially in humor, which requires not only linguistic knowledge but also cultural context. This project aims to address this gap by fine-tuning a pre-existing language model to generate jokes in Turkish.

1.2 Project Objectives:

The primary objective of this project is to fine-tune an existing large language model (LLM) so that it can specifically generate Turkish jokes. By adapting the model to the nuances of Turkish humor, cultural references, and language structure, the system will be able to produce content that is both entertaining and culturally relevant. This will be accomplished through a carefully curated dataset of Turkish jokes, followed by training the model to understand and replicate the joke-telling process in Turkish.

The key goals of the project include:

- Fine-tuning a pre-trained language model: We will leverage a pre-trained model, such as LLaMa 3.3 or similar, to fine-tune it specifically for Turkish joke generation.

- Generating culturally relevant content: The model will be tailored to ensure that it generates humor that is contextually appropriate for Turkish-speaking audiences.

2. Vision Statement:

The vision of this project is to create an AI system capable of generating high-quality, contextually appropriate, and culturally relevant jokes in Turkish. By fine-tuning the LLaMa 3.3 model, the goal is to provide an engaging and entertaining experience for Turkish speakers, learners of the language, and those interested in Turkish humor. This AI-driven system will bridge the gap between traditional joke-telling and modern AI technologies, making it a fun, innovative tool for both entertainment and education.

3. Scope of the Project:

In-Scope:

- **Model Fine-Tuning:** Fine-tuning the LLaMa 3.3 model specifically for Turkish joke generation using a curated dataset of Turkish jokes.

- **User Interaction:** Developing a user-friendly interface through which users can request jokes. This will include a simple web-based application for input and output.

- **Cultural Relevance:** Ensure that all jokes generated by the model are culturally sensitive and appropriate for the Turkish-speaking audience.

- **Deployment:** The AI model will be deployed as a web-based application for easy access by users.

Out of Scope:

- **Multilingual Joke Generation:** This project is focused solely on generating Turkish jokes and will not be able to move beyond the basic LLaMa model in other languages at this stage.

- **Voice or Sound Generation:** The project will not include voice or audio-based joke delivery; it will focus entirely on text-based jokes.

- **Large-Scale Production:** Our focus is on creating a prototype for Turkish joke generation, not on delivering the solution to a large number of users at this time.

4. Objectives

The main objectives of this project are focused on successfully fine-tuning and launching the model. These objectives guide the development and completion of the project.

Fine-Tuning Success: Successfully fine-tune the LLaMa 3.3 model to produce high-quality Turkish jokes. This will be measured by the relevance, humor, and cultural relevance of the jokes.

Joke Generation: Jokes should be unique, not presenting the data they were trained on to the user.

User-Friendly Interface: Develop a user interface that allows users to interact with the system and request jokes.

Cultural Sensitivity: Ensure that jokes are appropriate for a wide audience in Turkish-speaking communities, avoiding culturally insensitive or inappropriate content.

5. Target Audience

The target audience for this project is individuals and groups interested in Turkish culture, language, and humor. The system is designed to provide an

engaging and entertaining experience, making Turkish jokes accessible and enjoyable to a variety of users. Specific target audiences include:

Turkish Speakers: Native Turkish speakers who love humor and want to experience a modern, AI-generated form of entertainment in their own language.

AI Enthusiasts: Individuals interested in the development and implementation of AI, particularly natural language processing (NLP) and its use in cultural and humorous content production.

Cultural Explorers: Individuals interested in Turkish culture and humor, who want to discover and experience jokes that reflect Türkiye's cultural context.

6. Key Features and Requirements

Key Features:

Fine-Tuning: The LLaMa 3.3 model will be fine-tuned using a dataset of Turkish jokes. Training will include understanding the nuances of Turkish humor, puns, and cultural references.

User Interface: A simple and intuitive web-based interface will allow users to request jokes. This interface will display jokes generated by the model in response to user input.

Content Filtering: Data will be carefully selected to ensure that the generated jokes are culturally appropriate and free of offensive content.

Technical Requirements:

Model: The LLaMa 3.3 70B pre-trained model will be used as the base model for fine-tuning.

Programming Languages and Libraries: It will be done using the Lora method in the Finetunin Unslost library. Python and Pytorch will be used.

Dataset: A carefully selected dataset of Turkish jokes taken from general joke collections or found on the internet.

Hosting and Distribution: The application will be a web application running locally.

7. Success Criteria

The following criteria will be used to measure the success of the project:

Model Quality: The fine-tuned model should produce jokes that are entertaining, contextually relevant, and culturally sensitive. The jokes should be unique.

User Experience: The user interface should be easy to navigate and users should be able to interact with the system without technical issues.