Project Scope Document: Fine-Tuning and Deployment of Turkcell-LLM-7b-v1 with Turkish Jokes

1. Project Overview

This project aims to fine-tune the Turkcell-LLM-7b-v1 language model to specialize in Turkish joke generation and deploy it through a Streamlit web application. The model will serve as a Turkish joke telling assistant that delivers contextually appropriate humorous content on request.

2. Project Objectives

The primary objective is to create and deploy a specialized version of Turkcell-LLM-7b-v1 that excels at Turkish joke generation, supported by a comprehensive dataset and efficient fine-tuning techniques. The model will maintain its general language capabilities while developing expertise in humor, all accessible through a user-friendly Streamlit interface.

3. Deliverables

The project will deliver a curated dataset of Turkish jokes formatted for training, necessary preprocessing scripts, fine-tuned model weights, performance reports, and a complete Streamlit web application with deployment documentation.

4. In Scope

This project encompasses joke data collection and preparation, model fine-tuning using parameter-efficient methods, performance evaluation, and Streamlit application development. The Streamlit app will include user input fields, response display, and basic styling with deployment instructions for various environments.

5. Out of Scope

Enterprise-level production features, commercial development, content from copyrighted sources, architectural modifications to the base model, languages beyond Turkish, advanced content moderation, and third-party service integrations fall outside this project's boundaries.

6. Constraints

The project faces constraints in computing resources (minimum 16GB VRAM GPU for training), deployment server requirements, a 5-9 week timeline, limited budget, dependency on publicly available joke data, and model size considerations for deployment.

7. Assumptions

We assume the base model possesses sufficient knowledge of Turkish language and culture, adequate joke data availability, fine-tuning feasibility on available hardware, appropriate licensing, and Streamlit's capability to serve the model with reasonable performance.

8. Stakeholders

Key stakeholders include the Project Lead, Data Scientists, Data Collectors, Web Developers, Technical Support staff, and End Users who will interact with the application.

9. Timeline

The 9-week timeline covers data collection and preparation (weeks 1-2), environment setup (week 3), model fine-tuning (weeks 4-5), evaluation (weeks 6-7), Streamlit development (week 8), and deployment with documentation (week 9).

10. Resources

Required resources include GPU computing power, deployment server capacity, storage space, technical tools (Python, PyTorch, Transformers, PEFT, Streamlit), personnel with NLP and web development experience, and a dataset of 1,000-5,000 Turkish jokes.

11. Success Criteria

Success will be measured by the model's ability to generate appropriate and humorous Turkish jokes, outperform the base model, understand various request phrasings, complete training within constraints, achieve dataset diversity, and deliver a functional, user-friendly Streamlit application with stable performance.

12. Change Management

All scope changes require formal review and approval from the Project Lead, with impact assessments on timeline, resources, and outcomes. Approved changes will be documented and communicated to all stakeholders.

13. Streamlit Deployment Specifications

The Streamlit application will feature a project description, user input capabilities, response display, sample prompts, and model information. Technical requirements include Python 3.8+, Streamlit 1.10.0+, and optional model quantization. Deployment guidance will cover both local and server implementations with optimization recommendations.