Course Project

Overview

Project Title: Gourmet GPT: A Bespoke & Compact AI Recipe Advisor

Project Objective:

The objective of the project is to create a highly efficient, personalized recipe recommendation chatbot specifically for Indian cuisine. This project's core innovation is the development of a **Custom and Compact Small Language Model (SLM)**, built from the ground up using a **Transformer-based decoder-only architecture**. This domain-specific approach avoids the computational overhead of large, general-purpose models, making the system highly efficient and scalable. The SLM will be optimized for the food and nutrition domain, aiming for superior contextual accuracy in recipe matching and conversational interaction by understanding the nuances of Indian ingredients, cooking methods, and dietary preferences.

Dataset

The dataset will consist of a high-quality collection of **Indian recipes**, including ingredients, step-by-step instructions, and nutritional information. I'll also incorporate culinary articles and conversational data related to Indian cuisine to enhance the model's understanding of context and conversational flow.

Role of Neural Networks

The project relies heavily on **neural networks** for two key functions:

1. **Transfer Learning:** I'll leverage a pre-trained model like Gemma as a foundational architecture. This allows the model to transfer its general language understanding to the specific domain of Indian cuisine, accelerating training and improving performance.
2. **Stochastic Next Word Prediction:** The **Transformer-based decoder-only architecture** at the heart of the SLM uses a mechanism called **stochastic next word prediction**. This allows the model to generate a sequence of words (a recipe or a conversational response) by predicting the most probable next word based on the preceding sequence. The model's "stochastic" nature means it introduces a

degree of randomness, preventing repetitive or bland responses and making the conversation feel more natural.

Transformer Architecture:

a. **nanoGPT**: A minimalist, educational, decoder-only Transformer model that employs **masked multi-head self-attention** and a **two-layer feedforward neural network** with a ReLU activation function.
b. **Gemma 3**: A decoder-only Transformer that uses **Grouped-Query Attention (GQA)** with a mix of **local and global attention layers** to efficiently handle a long context, and a **Gated Linear Unit (GeGLU)** as its activation function in the feedforward network.

Measures of Success

Success will be measured by several key metrics:

- **Model Compactness:** The primary goal is to build a small model, with a target size of around a **million parameters**, to demonstrate efficiency and reduced computational footprint.
- **Training and Validation Loss:** I will monitor the loss curves during training to ensure the model is learning effectively and generalizing well to unseen data. A continuously decreasing loss on both the training and validation sets will indicate success.
- **Perplexity:** This metric will be used to evaluate the model's language generation quality. A lower perplexity score indicates that the model is more confident in its predictions and that its generated text is more similar to human-written text.

Breakdown of Specific Tasks

1. Dataset Preparation

The first step is to curate a high-quality, domain-specific dataset. This will involve scraping and sanitizing a vast collection of Indian recipes, culinary articles, and food-related conversational data from various online sources. The data will be cleaned, structured, and formatted for optimal model training.

2. Pre-training with NanoGPT and Gemma 3 Architecture

The foundational SLM will be pre-trained on the prepared dataset. I will adapt and explore modern pre-training techniques, leveraging methodologies from **nanoGPT** for efficiency and **Gemma** for its lightweight, state-of-the-art architecture. This hybrid approach will allow me to build a powerful yet compact model from scratch.

3. Instruction Fine-tuning for an AI Recipe Advisor

The final and most critical stage is **instruction fine-tuning**. I'll train the pre-trained model on a specialized set of prompts and responses tailored for a recipe advisor. This will teach the model to handle specific user requests, provide step-by-step instructions, suggest substitutions, and answer questions about Indian cuisine with high accuracy and relevance. This stage will teach the model to go beyond mere text generation and become a truly helpful and knowledgeable conversational agent.