Project Proposal: Regional Dialect Synthesis Pipeline (Haryanvi)

Project Title: Bridging the Linguistic Divide: A Generative Pipeline for Hindi-to-Haryanvi Translation and Speech Synthesis

Focus Dialect: Haryanvi (Bangru/Desari)

1. Problem Statement & Motivation

While major Indian languages like Hindi are well-supported by digital tools, widely spoken dialects like **Haryanvi** (approx. 10 million speakers) remain "low-resource." Current translation systems (e.g., Google Translate) output standard Hindi, failing to capture the unique vocabulary (e.g., *"Kade"* vs *"Kab"*), grammar, and the distinct, aggressive tonal quality of the region. This creates a digital divide for native speakers.

Objective:

We propose a strictly scoped Deep Learning pipeline to bridge this gap. The system will accept Standard Hindi text and generate high-fidelity **Haryanvi speech**.

- **Scope Refinement:** we have removed lip-sync and multi-dialect requirements. The project is exclusively focused on **Hindi to Haryanvi** to ensure feasibility and high-quality evaluation.

2. Proposed Methodology & Architecture

The pipeline consists of two modular components leveraging State-of-the-Art (SOTA) open-source models:

Module A: Hindi-to-Haryanvi Translation (Text-to-Text)

- **Goal:** Convert Standard Hindi text into grammatically and lexically correct Haryanvi.
- **Model: Llama-3.1-8B-Instruct (Quantized)**.
- **Technique:** We will use **QLoRA (Quantized Low-Rank Adaptation)** to fine-tune the model. Llama-3.1 is chosen for its superior instruction-following capabilities and "in-context learning," which allows it to grasp dialect rules (like gender marker shifts from *raha* to *rya*) better than traditional LSTM-based models.

Module B: Haryanvi Text-to-Speech (Text-to-Audio)

- **Goal:** Generate natural-sounding speech from the translated Haryanvi text.
- **Model: VITS (Conditional Variational Autoencoder)**.
- **Technique:** Transfer Learning. We will initialize the model using **AI4Bharat's Indic-TTS** checkpoints (pre-trained on Hindi/Rajasthani). This provides a strong baseline for

Indo-Aryan phonetics, allowing us to converge faster on Haryanvi prosody compared to training from scratch.

3. Dataset Strategy (Data Sources)

To overcome the "low-resource" challenge, we will use a hybrid data strategy:

- **Translation Data (Hindi to Haryanvi):**
  - **Synthetic Generation:** We will generate a parallel corpus of 2,000+ sentence pairs using **Gemini**. We will prompt the model with specific linguistic rules (e.g., "Replace 'hum' with 'mhaare'") to create an initial training set.
  - **Validation:** These synthetic pairs will be manually validated by native speakers in the group to ensure accuracy before training.
- **Audio Data (TTS):**
  - **Primary Source:** The **ankitdhiman/haryanvi-tts** dataset (Hugging Face), containing ~5 hours of aligned Bangru dialect audio.
  - **Secondary Source:** A subset of the **IndicGenBench** (bgc dialect code).

4. Evaluation Plan

We have defined a concrete evaluation framework balancing automated metrics and human judgment:

Component	Metric	Description
Translation	**chrF++**	Character n-gram F-score; more suitable for morphologically rich Indian languages than BLEU.
Translation	**Human Eval**	A random sample of 50 translations will be reviewed for **Lexical Accuracy** (correct vocabulary usage) and **Grammatical Consistency**.
TTS	**MCD**	**Mel-Cepstral Distortion** to measure the spectral distance between generated audio and ground-truth validation audio.

TTS	**MOS**	**Mean Opinion Score**. We will conduct a blind listening test with 5 native speakers (available within our network) rating **Naturalness** and **Accent Authenticity** on a 1-5 scale.

5. Novelty & Differentiation

1. **Hyper-Local Focus:** Unlike Bhashini or Google, which focus on scheduled languages, we are building specifically for an *unscheduled* dialect (Haryanvi).
2. **SOTA Adaptation:** We are demonstrating how general-purpose SOTA models (Llama 3.1, VITS) can be adapted for hyper-local tasks using synthetic data augmentation, a novel approach for low-resource dialect engineering.