SRM
INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
DELHI-NCR CAMPUS, MODINAGAR, U.P

AI-Powered Sanskrit NLP Framework for Morphological and Semantic Analysis

Project -I

B.Tech. IIIrd Year/ 6th Semester

SUBJECT NAME : IT PROJECT MANAGEMENT
SUBJECT CODE : 21MGH303P

Problem Statement

Develop an AI-powered NLP framework for morphological analysis, semantic parsing, and knowledge extraction from classical Sanskrit texts by computationally modeling Paninian grammar, enabling digital preservation, linguistic research, and intelligent language applications.

1. Introduction

1.1 Background

Classical Sanskrit is one of the most structured and grammatically rigorous languages in the world. Its linguistic framework, formally defined by the ancient grammarian Panini through the Ashtadhyayi, represents a rule-based system that precisely governs word formation, morphology, and syntax. Despite its systematic nature, much of this linguistic knowledge remains confined to traditional texts and manual interpretation, limiting its accessibility and application in modern digital systems.

In recent years, advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have enabled the automated analysis of human languages. However, most NLP systems are optimized for modern, less rule-intensive languages and often fail to capture the depth and precision required for classical languages such as Sanskrit. This gap highlights the need for a specialized computational framework that can effectively model Paninian grammar while leveraging modern AI techniques.

1.2 Problem Definition

Although Sanskrit possesses a highly formalized grammatical system, the absence of scalable and accurate computational tools has restricted its digital preservation, semantic understanding, and large-scale linguistic analysis. Existing Sanskrit NLP tools either rely heavily on rigid rule-based systems or purely data-driven models, both of which have limitations in handling complex morphological and semantic structures.

The core problem addressed in this project is the lack of an integrated AI-driven NLP framework that can perform morphological analysis, semantic parsing, and knowledge extraction from classical Sanskrit texts by computationally aligning Panini's grammatical rules with modern machine learning approaches.

1.3 Project Objectives

The primary objectives of this project are:

- To design and develop an AI-powered NLP framework for analyzing classical Sanskrit texts.
- To computationally model Paninian grammar for accurate morphological and semantic interpretation.
- To enable structured knowledge extraction for academic research and digital preservation.
- To bridge traditional linguistic theory with modern computational linguistics.
- To provide a scalable and extensible system for future language technology applications.

2. Project Identification Methodology & Stakeholder Description

2.1 Project Identification Methodology

Phase	Description
Problem Identification	Loss of classical Sanskrit knowledge due to non-digitized grammar
Opportunity Analysis	AI + NLP can preserve and operationalize ancient grammar
Feasibility Study	Technical (ML + rules), Economic (moderate), Operational (high scholarly demand)
Project Justification	Cultural preservation + academic + AI research value
Approval Gate	Academic institutions / Govt / Digital heritage bodies

2.2 Stakeholder Analysis

Stakeholder	Role	Interest Level
Project Sponsor (Govt / University)	Funding & oversight	High
Project Manager	Planning, execution	High
Sanskrit Linguists	Grammar modeling	Very High
NLP Engineers	Model development	High
Data Scientists	Corpus training	Medium
Cultural Institutions	Preservation	Medium
End Users (Researchers, Students)	Usage & feedback	High
Cloud / Infra Providers	Hosting	Low

3. Project Cost Estimation & Capital Budgeting

3.1 Cost Components

Cost Category	Description
Development	NLP models, grammar engines
Infrastructure	Cloud compute, storage
Data	Corpus creation & annotation
Tools	NLP frameworks, ML libraries
Testing	Accuracy validation
Maintenance	Updates, model retraining

3.2 Software Cost Estimation Models Applied

a) COCOMO (Intermediate – Organic Mode)

- Project size: ~25–30 KLOC
- Effort ≈ 70–90 Person-Months
- Cost Driver: High analyst capability, high complexity

b) Function Point Analysis

- Inputs: Sanskrit texts
- Outputs: Parsed grammar trees, semantic frames
- Complexity: High
- Estimated FP: ~450–500

c) Expert Judgment

- Sanskrit NLP is niche → higher specialist cost
- Rule-engine development inflates effort

3.3 Estimated Project Cost (Indicative)

Component	Cost (INR)
Development	₹18–22 Lakhs
Data Annotation	₹6–8 Lakhs
Infrastructure	₹3–4 Lakhs
Testing & QA	₹2 Lakhs
Contingency (15%)	₹4–5 Lakhs
Total	₹33–41 Lakhs

3.4 Capital Budgeting Technique

Preferred Method:
Net Present Value (NPV) – Positive due to long-term academic & licensing value
Payback Period – ~3–4 years (via grants, licensing, APIs)

4. Market Demand Analysis & Demand Planning

4.1 Market Demand Drivers

- Growing interest in digital humanities
- AI-driven linguistic research
- Government-funded language preservation programs
- EdTech platforms integrating classical languages
- UNESCO & cultural digitization initiatives

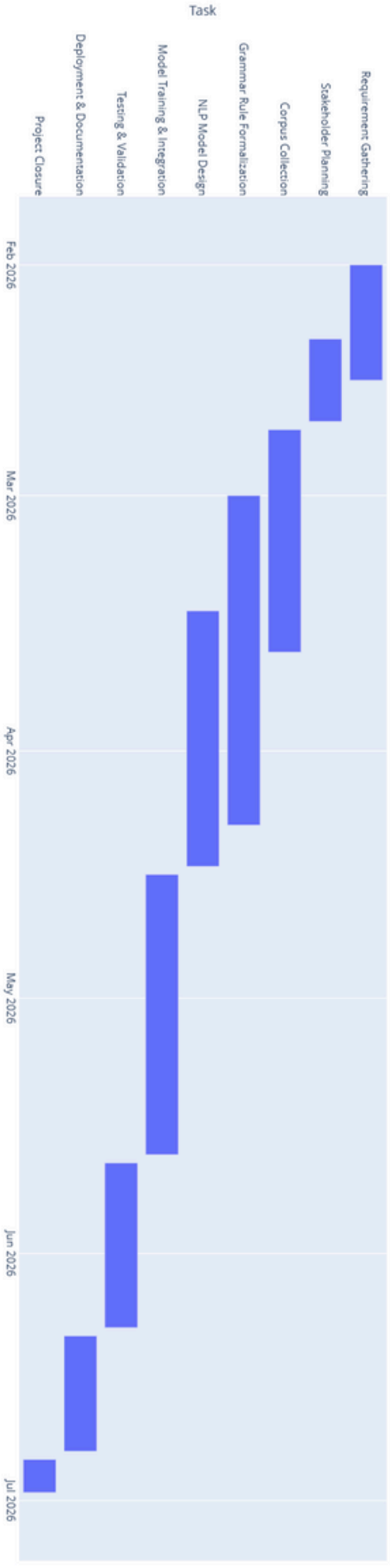
4.2 Target Market Segments

Segment	Demand
Universities & Research Institutes	High
Digital Libraries	Medium
Language Learning Platforms	Medium
Govt Cultural Bodies	High
AI Research Labs	Medium

4.3 Demand Planning

Phase 1: Academic adoption
Phase 2: API & research tools
Phase 3: EdTech & multilingual NLP expansion

GANTT CHART


Task
Requirement Gathering
Stakeholder Planning
Corpus Collection
Grammar Rule Formalization
NLP Model Design
Model Training & Integration
Testing & Validation
Deployment & Documentation
Project Closure
Feb 2026
Mar 2026
Apr 2026
May 2026
Jun 2026
Jul 2026