CHENNAI MATHEMATICAL INSTITUTE

# AI-POWERED RESUME STANDARDIZATION SYSTEM

ANIKET TIWARI | MDS202308
DEEPANSHU MITTAL | MDS202320
KIRONMOY ROY | MDS202332

APPLIED MACHINE LEARNING PROJECT | PROF. RAGHAV KULKARNI

# Problem Statement

Companies often receive resumes in varied formats, making it difficult to evaluate candidates consistently. There is a need for an automated system that can extract relevant details from any resume and transform it into the company's predefined CV format.

# Project Objective

To develop a GenAI application that takes any candidate's resume (PDF, DOCX, etc.) and converts it into a standardized company CV template by extracting all required fields using intelligent parsing and language modeling.

# Workflow Diagram

## User Journey

User enters Gemini API key one time, after that uploads the resume file (PDF, Doc, Image).

↓

User select the company standard template for the conversion from the left sidebar.

↓

In next couple of seconds, user's resume converted to company's standardized template.

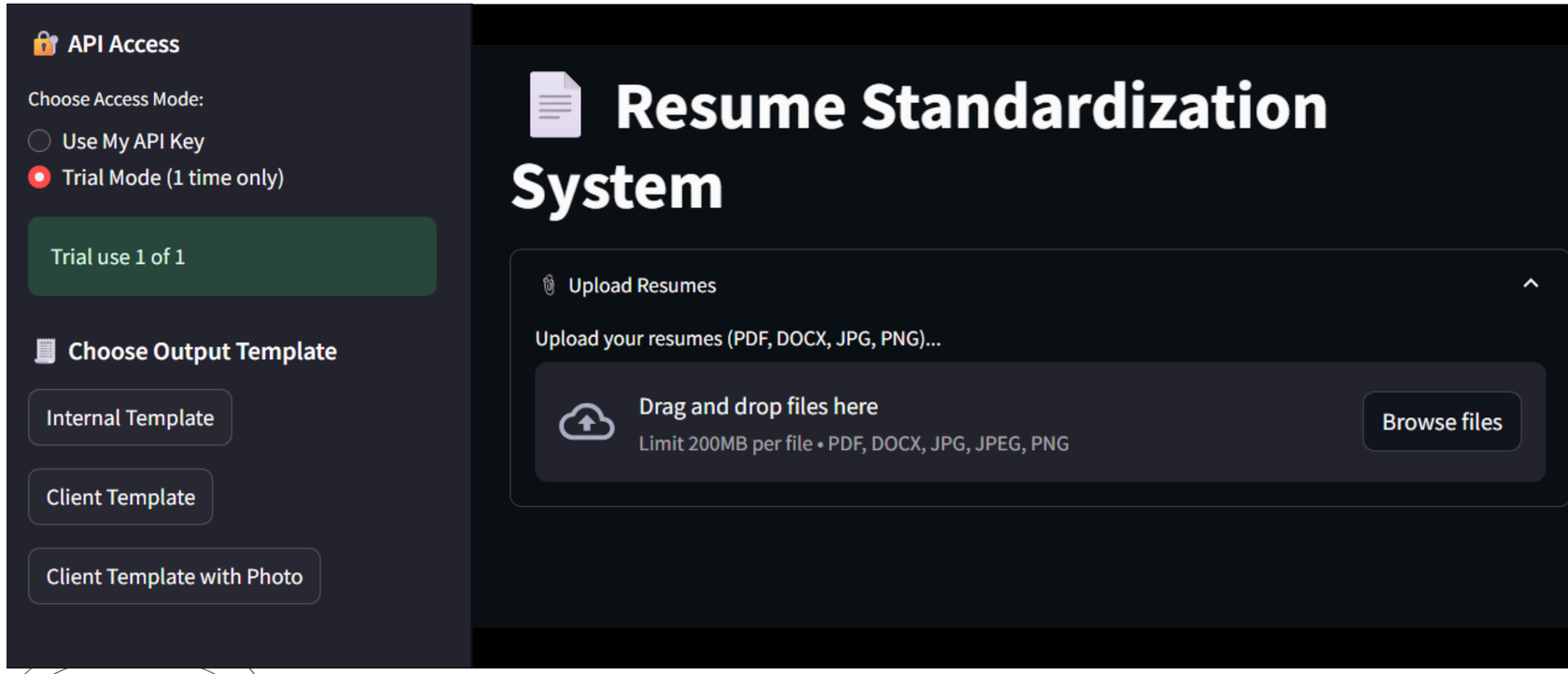


*Fig: Streamlit page for the application.*
*Two Options:*
*(a) Use your API key*
*(b) Trial Mode*

# Tech Stack

Programming language

NLP & Parsing

PDF & Doc Parsing:

- pdfplumber
- PyMuPDF
- docx

LLM Used

**Gemini-2.0-Flash**

Why ?

- Open Source
- Can handle image input
- Good performance
- Low latency & lightweight

Image Processing

- cv2, PIL

PDF Generation

- reportlab

Others

- BytesIO
- zipfile
- re

Frontend hosting

Streamlit

Custom made functions for field extraction and template filler

## BACKEND JOURNEY

**Input:** Resume file (PDF, Doc, Image), if the input is a Docx file it is first converted to a pdf file before further processing.

Extract the complete text from the PDF to create a raw input for the LLM. Then, apply custom-built field extractor functions to retrieve relevant information from the raw text according to a standardized CV format.

Once the relevant information has been extracted, use a custom-built formatter or template filler to populate the company's standardized CV template. Finally, download the newly tailored CV

# PROJECT DEMO

# Key Challenges

- Selecting a multimodal **open-source LLM** capable of handling both textual content and embedded images (e.g., graphs, profile photos, logos) from resumes. We chose Gemini-2.0-Flash, but it has its own limitation of **low tokens per unit of time**. So many iterations at a time cannot be done.
- **Unstructured & Varied Formats:** Resumes come in different formats (PDF, DOCX, TXT) and styles, making parsing difficult.
- **Section Identification:** Extracting and segmenting resume into sections (Experience, Education, Skills) is non-trivial due to inconsistent headings.
- **Context Understanding:** Differentiating between job responsibilities vs. project descriptions.

# Future Improvements

- **Plug-and-Play Integration:** Develop REST APIs or SDKs for easy integration into ATS, HRMS, Workday and other HR softwares.
- Use **semantic similarity models** to compare standardized resumes against job descriptions & add a scoring or **ranking mechanism** to prioritize resumes for specific job roles.
- **Scalability:** Upgrade to faster and more accurate AI models that can handle large numbers of resumes quickly and reliably as the system grows.
- **Multilingual Support:** Support resume parsing and standardization in multiple languages using multilingual transformers like mBERT or XLM-R.

# Primary Work Allocation

### Aniket Tiwari (MDS202308)

- Integrated all custom-built functions into the main app.py file, designed the Streamlit user interface, and deployed the application.

### Deepanshu Mittal (MDS202320)

- Developed support functions (e.g., docx2pdf, raw text extraction from PDFs) and implemented field extraction logic for structured data retrieval.

### Kironmoy Roy (MDS202332)

- Developed formatter and template-filling functions (e.g., fill_table_education_years) and designed company-standard CV templates.

# THANK YOU

*"Alone we can do so little; together we can do so much."*
*-Helen Keller*