Internship Report: GeminiDecode - Multilanguage Document Extraction Using Gemini Pro

1. Introduction
The *GeminiDecode* project is designed to leverage Google's Generative AI models to facilitate the extraction of multilingual document content. It specifically aims to process various document formats, such as invoices and legal papers, and translate them into actionable insights, breaking down language barriers with the help of AI. The project's front end is developed using Streamlit to provide an interactive user experience.

2. Objectives
- To create a tool that extracts key information from documents in various languages.
- To enable seamless translation and processing of multilingual data.
- To ensure the project can handle a variety of document formats, especially invoices.

3. Technologies Used
- Streamlit: To build the web interface for user interaction.
- Google Generative AI (Gemini Pro): To process and extract data from documents.
- Pillow (PIL): For image processing and handling of uploaded files.
- dotenv: For secure handling of environment variables, such as API keys.
- Python: Primary programming language used for building the application.

4. Key Components
1. Image Uploader:
   - Users can upload images of documents in formats like `.jpg`, `.jpeg`, and `.png`. This functionality is powered by Streamlit's file uploader widget.

2. AI Integration:
   - The project uses Google's Generative AI models, particularly `gemini-1.5-flash`, to process the document images. The models extract relevant content based on the specified prompt, which focuses on understanding and answering questions related to invoices and other documents.

3. Multilingual Capabilities:
   - The AI model's ability to comprehend and respond in multiple languages is key to handling diverse document types. This is crucial for processing global invoices and legal documents.

4. User Interface:
  - The web app includes a user-friendly interface where users can upload documents, view their images, and extract information at the click of a button.

5. Code Overview
5.1 Main Application (app.py)
The code starts by importing necessary libraries such as Streamlit, dotenv, and Google's Generative AI SDK. The page configuration is set at the beginning to ensure proper rendering.

- Main Functionality:
  - The user uploads a document, which is processed and passed to the AI model. The model generates a response based on the input prompt and the uploaded image.
  - Errors were addressed by switching to the latest generative AI model (`gemini-1.5-flash`) after the deprecation of the older version.

5.2 Dependencies (requirements.txt)
The project relies on several Python packages as described below:
- Streamlit: To build the UI and manage interactivity.
- Google-GenerativeAI: To connect with and utilize Google's advanced AI models.
- python-dotenv: For managing environment variables.
- Pillow (PIL): For handling image uploads and display.
- Additional Libraries: langchain, PyPDF2, chromadb, faiss-cpu.

(For more details, refer to the full list in `requirements.txt`【24†source】).

6. Challenges Faced
- Model Deprecation: The initial AI model, `gemini-pro-vision`, was deprecated in July 2024, necessitating a switch to `gemini-1.5-flash`. This required adjustments in code and testing to ensure compatibility with the new model.
- Handling Multilingual Texts: Ensuring that the AI could process and accurately extract information from documents in multiple languages was challenging but was addressed through careful prompt engineering and model selection.

7. Future Work
- Enhancing AI Capabilities: Incorporating more advanced AI models to improve accuracy, especially with complex legal and financial documents.
- Multiformat Support: Expanding the project to handle PDFs and other document

formats using libraries such as PyPDF2.
- Security Enhancements: Improving the secure handling of uploaded documents, especially in cloud-based environments.

8. Conclusion
GeminiDecode successfully demonstrates how generative AI models can be applied to real-world document extraction challenges. With its multilingual capabilities and intuitive interface, it can be scaled further to cater to a broader range of use cases, from financial to legal document processing.
