DEPARTMENT OF APEX INSTITUTE OF TECHNOLOGY

# PROJECT PROPOSAL

## 1. Project Title: -

AI Powered Legal Document Summarizer

## 2. Project Scope: - (Max 500 words)

1. **Document Analysis**
   * **Input Types:** Support for a variety of document formats, including PDF, DOCX, and scanned documents (OCR integration).
   * **Content Parsing:** Extract text from structured and unstructured formats.
   * **Legal Language Understanding:** Train the AI to identify legal terminologies, clauses, and critical information.
2. **AI Model Development**
   * **Natural Language Processing (NLP):** Use advanced models like OpenAI’s GPT, BERT, or domain-specific legal NLP models.
   * **Pre-Training and Fine-Tuning:**
     + Pre-train on large datasets of legal documents.
     + Fine-tune the model for specific legal domains (e.g., contracts, wills, NDAs, court judgments).
   * **Context Retention:** Ensure the summarization process retains the original intent and meaning of the document.
3. **Summarization Features**
   * **Summary Length Control:** Provide options for generating brief, detailed, or custom-length summaries.
   * **Key Clause Extraction:** Highlight critical clauses, deadlines, obligations, and exceptions.
   * **Language Simplification:** Offer simplified summaries for non-legal users while retaining legal integrity.
4. **User Interface**
   * **Web Application:**
     + Drag-and-drop upload for documents.
     + User-friendly dashboard displaying summaries and extracted key points.
   * **Customization Options:**
     + Select specific sections to summarize.
     + Adjust legal jargon level for different audiences.
   * **API Integration:** Provide APIs for seamless integration with law firm workflows or enterprise systems.
5. **Performance Metrics**
   * **Accuracy:** Measure the quality of the summaries compared to human-curated ones.
   * **Speed:** Optimize for real-time or near-real-time summarization.
   * **Compliance:** Ensure adherence to local and international data privacy laws, such as GDPR.
6. **Use Cases**
   * **Law Firms:** Efficient review of contracts, case files, and legal agreements.
   * **Corporate Legal Teams:** Rapid summarization of regulatory updates and compliance documents.
   * **General Public:** Simplifying legal documents for personal understanding.

## 3. Requirements: -

* Hardware Requirements

**Development and Training Environment**

1. High-Performance Computing (HPC) Setup or Cloud-Based GPU
2. RAM: Minimum 64 GB
3. Storage: 500 GB SSD(Minimum)

* Software Requirements

1. **Operating System**

* **Linux:** Ubuntu 20.04 or later (preferred for development and server environments).
* **Windows Server:** For certain enterprise environments.

2. **Programming Languages:**

* Python (primary language for AI and NLP development).

3. **IDEs:**

* PyCharm, VS Code, or Jupyter Notebook for Python development.

**STUDENTS DETAILS**

|  |  |  |
| --- | --- | --- |
| **Name** | **UID** | **Signature** |
| SNEHA KUMARI | 21BCS5019 |  |
| PAIDI HEMANTHI NAIDU | 21BCS5239 |  |

**APPROVAL AND AUTHORITY TO PROCEED**

We approve the project as described above, and authorize the team to proceed.

|  |  |  |
| --- | --- | --- |
| **Name** | **Title** | **Signature**  **(With Date)** |
| DR. PRIYANKA KAUSHIK | SUPERVISOR |  |