

Call to Join a Cutting-Edge AI Research Project!

Looking for UG students to contribute to a AI research tool! For more details visit the latest updates on www.scraikwar.com

Available Problems

SN	Problem	Task	Expertise Required
1	Data Acquisition	Collect & pre-process data	Python, APIs, web scraping, data cleaning
2	Summarization of data	Extract required information	NLP, Transformer models, Python
3	Clustering & Identification of trends of the data	Group similar information detect trends	ML/Unsupervised Learning, embedding, Python
4	Data Recommendation	Suggest relevant information	NLP, LLM fine-tuning
5	Data Translation	Multilingual support	NLP, Transformer-based translation
6	Data Visualization	Interactive dashboards	Python (Streamlit, Plotly, Altair), data visualization

Who Can Apply

- **1st, 2nd and 3rd year** CS/EC/BioTech/AI/Data Science/EE students pssionate about AI, NLP, research and learning
- Expertise is required, but not mandatory
- Apply as individual or as a Team (Team leader will apply, he can mention the name and expertise of team members in the other details.)

Benefits

- Certificate of Participation
- Letter of Recommendation / Mentorship
- Co-developer in the tool.
- Goodies & merchandise (T-shirts, notebooks, AI kits)
- Hands-on experience with AI, NLP & visualization
- Exposure to emerging technologies and alignment with vision and mission of the India.
- The project will have done in three phases.
- Interaction with Industry expert during phase-3 of the project.

Phase-1: Registration and the development of the module.

- Register for the participation at <https://forms.gle/uxZt6XmQ35FAKXxs6>
- **Deadline:** 05 November 2025

Phase-2: Submission of the Proof-of-Concept (POC) and Presentation

- **Submission Item:** You are required to submit a **PowerPoint Presentation (PPT)** detailing your POC (Input, Processing pipeline, Expected Output, tool and technology used, novelty, cost of maintenance, salient features, future scope).
- **Demonstration:** The POC must clearly **demonstrate the working** of the **Data Visualization** module as per the needs outlined in the original technical roadmap (attached as Annexure-A at the end of this document).

- **Template:** Please ensure your presentation strictly follows the **attached template**.
- **Deadline:** The completed PPT must be submitted on or before **07 November 2025** at [Submission Link](#).
- **POC Evaluation Criteria**
 - Your submitted PPT will be evaluated based on the following criteria, so please ensure your presentation addresses each point:
 - **Presentation Skills:** Clarity, structure, and effectiveness of the presentation.
 - **Novelty in Tool Usage with Reasoning:** Uniqueness and justification of the specific tools you select for the implementation.
 - **Sustainability:** The long-term viability and ease of maintenance of your proposed solution.
 - **Cost of Maintenance:** A clear breakdown of the estimated deployment and running costs (e.g., use of paid cloud platforms, databases, or other paid services).
 - **Preference for Indian Made Tools:** Consideration and preference given to tools developed in India.
 - **Preference for Open-Source Tools:** Emphasis on using open-source solutions to minimize costs and promote accessibility.
- Based on the evaluation of the submitted PPTs, a shortlist of candidates will be announced by 08 November 2025.

Phase-3: Presentation

- **Date:** 15 November 2025
- **Time:** Announced soon
- **Venue:** Activity Space-1

Reach us at suresh.raikwar@thapar.edu in case of any query.

Thank you and we look forward to reviewing your innovative work.

With Bests

Coordinator: Dr. Suresh Raikwar, DCSE, TIET

Roadmap of the Project

SN	Module	Details of the Task			Pre-requisite
		Input	Process	Expected Output	Steps to learn pre-requisite
1	Data Acquisition	URLs of the PDF document	- Fetch PDF using web scraping or APIs - Extract text from PDFs (PyMuPDF, GROBID) - Clean metadata (title, year) - Store structured text and metadata	Clean JSON/CSV dataset containing: title, year, full text	- Learn Python basics (loops, file handling, JSON) - Study web scraping using requests, BeautifulSoup - Practice PDF extraction using PyMuPDF or GROBID - Learn data cleaning and JSON storage - Explore metadata (DOI, year, title) structure
2	Summarization of data	Clean text and metadata	- Preprocess text (tokenization, cleaning) - Apply extractive/abstractive summarization models (T5, BART, Pegasus) to identify problem, approach, findings, limitations	JSON summaries for each pdf with 4 fields (will be given later)	- Learn NLP basics: tokenization, stopwords - Understand transformer models (HuggingFace) - Implement simple summarization model in Colab - Learn to extract key phrases from text - Test on sample PDF
3	Clustering & Identification of trends of the data	Summarized PDF	- Generate embeddings (SentenceTransformers) - Cluster related topics using K-Means/DBSCAN - Map cluster evolution year-wise - Visualize trends using Plotly or Matplotlib	JSON file of topic clusters and trend chart of the topic (given in PDF) growth over time	- Learn concept of embeddings and similarity - Study unsupervised ML (K-Means, PCA, DBSCAN) - Learn vectorization using SentenceTransformer - Understand plotting with Plotly - Try topic clustering on some PDF
4	Data Recommendation	Clustered topics	- Use LLM reasoning or RAG (Retrieval-Augmented Generation) - Extract common factors (assigned later) - Generate some questions based on topic	JSON report suggesting some questions	- Understand how RAG combines LLM + retrieval - Learn basic prompt engineering - Study LangChain architecture - Practice LLM-based question answering - Apply reasoning to suggest new questions
5	Data Translation	Summaries in English	- Translate outputs using IndicTrans2 (HuggingFace) - Post-process to ensure grammar and terminology consistency - Maintain UTF-8 format for all Indian languages	Multilingual summaries in Hindi, Tamil, Telugu, Marathi, Bengali, etc.	- Learn machine translation basics - Explore HuggingFace translation models - Implement IndicTrans2 - Understand post-processing and encoding - Test translations on small text samples
6	Data Visualization	JSON File	Design interactive dashboard using Streamlit, Dash, or React	Interactive multilingual dashboard showing new questions	- Learn basics of HTML/CSS/JS - Study Streamlit/Dash dashboard creation - Learn JSON data integration

SN	Module	Details of the Task			Pre-requisite
		Input	Process	Expected Output	Steps to learn pre-requisite
			• Display clusters, timelines, and translated outputs • Add question insights and comparison views • Integrate via simple backend API or JSON files		• Practice interactive plotting (Plotly) • Build small demo dashboard using dummy data