Suyog Y. Joshi

(+91) 9168167911 | suyogyj@gmail.com | LinkedIn | GitHub

Education

- **Ashoka University** *August 2021 - May 2024*
  Bachelor of Science (Honours) in Computer Science Sonipat, India
- **Ashoka University** *August 2024 - May 2025*
  Postgraduate Diploma in Advanced Studies and Research, Computer Science Sonipat, India

Skills

Languages: Python, C, Java, JavaScript, HTML, CSS
Frameworks: SpaCy, NLTK, LangChain, Tensorflow, PyTorch
Tools and Platforms: Git, Neo4j, Windows

Publications

Saurav Chowdhury, **Suyog Joshi**, and Lipika Dey. 2024. Cross Examine: An Ensemble-based approach to leverage Large Language Models for Legal Text Analytics. In *Proceedings of the Natural Legal Language Processing Workshop 2024*, pages 194–204, Miami, FL, USA. Association for Computational Linguistics.

Projects

- **LLMs for Legal Text Analysis** *August 2024 - Present*
  Capstone Thesis/Project [GitHub]
  - Built an **LLM-driven pipeline** that automates **entity extraction** (e.g., statutes, provisions, case details) from **Indian food safety legal texts**, reducing manual analysis of lengthy documents.
  - Engineered a **knowledge graph** in **Neo4j** to map relationships between cases, statutes, and legal provisions, enabling deeper insights into legal trends and case similarities.
  - Leveraged **transformer models** and **zero-shot LLMs** (GPT-4o-mini) to structure and de-duplicate extracted entities, improving accuracy and facilitating data-driven legal reform analysis.
- **Applied Machine Learning in Football** *January 2024 - May 2024*
  Independent Study Module [GitHub]
  - Developed an end-to-end pipeline to collate and process football events data and **train predictive models such as gradient-boosted decision trees** to determine the likelihood of a shot converting to a goal.
  - **Engineered new features from the data** to improve prediction performance, **achieving the industry-standard**.
  - Performed **hypothesis testing** to determine whether there were inherent differences between leagues, and how these differences affected the model.
- **AcademAI Chatbot** *April 2024 - May 2024*
  Course Project (Natural Language Processing) [GitHub]
  - Developed a **retrieval-augmented generation (RAG)-based LLM chatbot** using the LLaMA API to answer queries about Ashoka University's academic policies and course handbooks.
  - Built a complete LangChain-based pipeline allowing for **question-answering** on the policy documents, including a **history-aware document retriever** to allow for conversational chatting.
  - **Conducted evaluation** on various document splitting and retrieval methods to identify the **best-performing algorithms**.

Research Experience

- **Ashoka University** *June 2024 - August 2024*
  Research Intern [GitHub]
  - Interned under the guidance of Dr. Lipika Dey to study the feasibility of using **LLMs for Legal Text Analytics**.
  - Developed **pipelines for the summarization of documents**, along with **NER for legal entities**.
  - Created a **legal knowledge graph to represent relationships** between cases, statutes, and provisions.
  - Wrote a paper which was accepted to the *Natural Legal Language Processing* workshop at *EMNLP* 2024.
- **Mphasis Makerspace at Ashoka University** *June 2023 - August 2023*
  Summer Research Intern Sonipat, India
  - Conducted research to develop methods for **object detection through walls using CNNs** and low-frequency software defined radios.
  - **Performed literature review** to identify the best existing solutions and find gaps in the research.
  - Implemented the **data collection and cleaning pipeline**.

Volunteer Experience

- **Peer Mentor and Events Team Member** *September 2023 - Present*
  Computer Science Society, Ashoka University
  - As a Peer Mentor, I guided a small group of students, providing support and clarification on any computer science-related doubts they faced.
  - Helping to create a collaborative and open space for Computer Science at Ashoka by planning and organising events such as mixers, research reading groups, and panel discussions.
- **Organisational Assistant** *September 2024*
  1^{st} *Annual Workshop of the Centre for Digitalisation, AI and Society*
  - Worked to ensure the smooth functioning of the workshop, handling logistics before and during panel discussions.
  - Took charge of ensuring an exemplary experience for the speakers, from travel to meals and refreshments.

Relevant Coursework

Artificial Intelligence, Natural Language Processing, Applied Machine Learning in Football (Independent Study Module), AI Alignment (Independent Study Module), Computer Vision, Introduction to Machine Learning, Algorithm Design and Analysis.