Anmol Singhal

[in] | [O] | [g] | [X]

Email: anmolsinghal@cmu.edu
Website: anmolsinghal98.github.io

Research Interests

Document Processing, Legal Language Understanding, LLM-guided Requirements Engineering

Education

Carnegie Mellon University, Software and Societal Systems Department   August 2024 – Present
Ph.D. in Societal Computing   *Pittsburgh, PA*
Indraprastha Institute of Information Technology, Delhi   August 2017 – June 2021
B.Tech - Computer Science Engineering; CGPA: 9.10/10.00   *New Delhi, India*

Publications and Preprints

- **Anmol Singhal**, Preethu Anish, Chirag Jain, Arkajyoti Chakraborty, Smita Ghaisas. Clarification Question Generation for Disambiguating Contracts. *In LREC-COLING 2024* **(Paper) (Code)**
- **Anmol Singhal**, Preethu Anish, Shirish Karande, Smita Ghaisas. Towards Mitigating Unfairness in Contracts from a Non-legal Stakeholder's Perspective. In *Natural Legal Language Processing Workshop at EMNLP 2023* **(Paper)**
- **Anmol Singhal**, Preethu Anish, Pratik Sonar and Smita Ghaisas. Data is about Detail - An Empirical Investigation for Software Systems with NLP at Core. In *IEEE/ACM International Conference on AI Engineering (CAIN), 2022* **(Paper)**
- Smita Ghaisas, **Anmol Singhal**. Data Management in NLP4RE. In *Natural Language Processing for Requirements Engineering Book, Springer-Verlag* **(Chapter)**
- **Anmol Singhal***, Mihir Goyal*, Jainendra Shukla, Raghava Mutharaju. Feature Fused Human Activity Recognition Network (FFHAR-Net). In *HCI International 2021 - Posters (HCII 2021).* **(Paper) (Code)**

*Equal Contribution

Technical Skills

Languages: Python, Java, Javascript, HTML, CSS, C++, C, SQL

Tools and Technologies: Langchain, OpenAI, PyTorch, TensorFlow, NLTK, SpaCy, Git, Django, Vue.js, Bootstrap

Relevant Courses: Machine Learning, Deep Learning, Natural Language Processing, Software Development using Open Source, Trustworthy AI Systems, Statistical ML, Semantic Web, Probability & Statistics, Linear Algebra

Research Experience

TCS Research   September 2021 – July 2024
Researcher, Data and Decision Sciences Area   *New Delhi, India*

- **Towards Mitigating Unfairness in Contracts (Paper)**
  Advised by: Dr. Smita Ghaisas and Dr. Shirish Karande
  - Conducted a questionnaire-based study to define fairness in contracts from a non-legal stakeholder's perspective.
  - Incorporated a semi-supervised fine-tuning method to detect potentially unfair clauses using limited labeled data.
  - The method accurately classified **84%** sentences and outperformed Chain of Thought prompting by a margin of **9%**.
  - **Keywords:** *Fairness, Legal Text Processing, Semi-Supervised Learning, Prompting*

- **Clarification Question Generation for Disambiguating Contracts**
  Advised by: Dr. Smita Ghaisas
  - Devised a retrieval-augmented prompting approach to generate questions for clarifying ambiguities in contracts.
  - Manual and automated evaluation showed that our approach performs ambiguity detection on a document-wide scale with an F2 score of **0.87**. It surpasses open-source Large Language Model baselines by a margin of **20%**.
  - **Keywords:** *Ambiguity, Requirements, Legal Text Processing, Retrieval, Prompting*

- **Analyzing Data Requirements for Document-Processing Software Systems (Paper)**
  Advised by: Dr. Smita Ghaisas
  - Interviewed experts to determine the data-related challenges faced when developing systems with NLP at their core.
  - Empirically reviewed the mitigation strategies and best practices employed by practitioners to address the challenges.
  - **Keywords:** *Data, Document Processing, SE4NLP, Requirements*

Multimodal Digital Media Analysis Lab (MIDAS) January 2020 – June 2021
Undergraduate Researcher *New Delhi, India*
- **English Grammatical Error Detection and Correction (Code) (Report)**
  Advised by: Dr. Rajiv Ratn Shah

  * Developed a hybrid method using rule-based and Transformer-based models to correct English grammatical errors.
  * The solution was deployed as a Grammatical Error Correction (GEC) tool for **i-Saksham NGO** to aid the education of about 2000 underprivileged kids living in Bihar, India.
  * **Keywords:** *Transfer Learning, Hybrid Methods*

- **Grammatical Error Annotation Tool for Hindi (Report)**
  Advised by: Dr. Rajiv Ratn Shah

  * Proposed an artificial data generation method for Hindi GEC by directly infusing lexical and syntactic errors in data.
  * Built a data-agnostic tool to evaluate the performance of Hindi GEC models against different linguistic error types.
  * **Keywords:** *Low-resource Natural Language Processing, Artificial Data Generation*

Indraprastha Institute of Information Technology, Delhi August 2019 – May 2020
Undergraduate Researcher *New Delhi, India*
- **Smart Human Activity Recognition (Paper)**
  Advised by: Dr. Jainendra Shukla and Dr. Raghava Mutharaju

  * Developed a method using the early fusion of minimalistic features such as time and location to detect the daily activities of senior citizens and medical patients residing in smart homes.
  * Our method showed extremely accurate results with an F1 score of 0.98.
  * **Keywords:** *Activity Recognition, Sequential Modeling, Human-Centered Computing*

Teaching Experience

CSE343 Machine Learning IIIT Delhi
Teaching Assistant for a class of 150 senior undergraduate students *August 2020 – December 2020*

CSE508 Information Retrieval IIIT Delhi
Teaching Assistant for a class of 190 senior undergraduate and graduate students *January 2021 – May 2021*

Relevant Projects

Covid-19 Peak Prediction June 2020 - July 2020
Advised by: Dr. Pankaj Jalote *New Delhi, India*
- Analysed the effectiveness of state-of-the-art ML models to predict the peak of Covid-19 cases in India.
- Conducted extensive research on various time-series-based architectures, optimisers and methods for feature extraction.
- **Keywords-** *Machine Learning for Healthcare, Sequential Modeling*

Calibration of ML Systems January 2021 - May 2021
Advised by: Dr. C. Anantaram *New Delhi, India*
- Explored the role played by calibration in ML systems to ensure the trustworthiness of model predictions.
- Experimented with custom loss functions to reduce calibration error in neural networks trained on image classification.
- **Keywords-** *Trustworthy AI, Calibration*

Awards and Achievements

- **ACM SIGSOFT Distinguished Paper Runner-up at CAIN'22**: Work on data requirements nominated among top 3 papers

- **Innovation Spark Award at TCS Research**: For receiving international acclaim for research conducted

- **Deans Award for Excellence in Academics**: For excellent performance in 2019-20 and 2020-21 academic sessions

Services and Leadership

- **Project Lead at TCS Research**: Supervised two interns and one full-time associate

- **Program Committee Member** for NLPP4RE book published by Springer: Responsible for reviewing submitted manuscripts

- Delivered a talk at **ICON 2022** on data-related challenges encountered in industry-scale AI projects

- **Student Mentor at IIITD**: Responsible for mentoring 6 freshman students to help them navigate college life

- **Member, Cultural Committee, IIITD**: Organized events to bolster student life in campus

- **Member, Organizing Team** of Odyssey, Cultural Fest of IIITD