Raghav Jain

+919034043285
raghavjain106@gmail.com

Education

Delhi, India **Delhi Technological University** **August 2017 – May 2021**

- Bachelor of Technology in Software Engineering, GPA: 8.62/10

Research Experience

Research Assistant **Oxford TVG** **June 2023 - Present**

Supervisor: **Prof. Philip Torr**

- Engaged in a collaborative effort with researchers at Meta to analyze the phenomenon of disinformation on social network platforms.
- Actively researching the application of graph neural network techniques to investigate instances of coordinated inauthentic behavior on social media platforms.

Research Collaborator **AI-NLP-ML Lab, IIT Patna** **June 2021- Present**

Supervisor: **Prof. Sriparna Saha**

- Led an industry-funded project on automated English training software and worked on all NLP aspects of the software including an automated essay grading system, grammar, and spelling checking components and translation of the English language to other regional languages.
- Proposed an 'extractor-abstractor' framework to outperform its predecessors by a margin of 0.5 ROUGE-1, 0.4 ROUGE-2, 1 METEOR, and 0.9 WMS (Word Mover Similarity) scores.
- Proposed a new evaluation metric (WIDAR) for summarization systems in an attempt to overcome the limitations of ROUGE and it is able to outperform its predecessor and various state-of-the-art metrics.
- Worked extensively on cyberbully detection in code-mixed language and its explainability. Also developed a system which can intervene Cyberbully and suggest the corresponding normalized sentence
- Proposed a GPT-2 based Domain aware Italian Language tutoring conversational Agent. Currently working on proposing a Shannon theory-based knowledge selection system in conversational agents.
- Currently engaged in a collaboration with researchers at Microsoft to investigate the effectiveness of various LLMs and prompting techniques in relation to temporal reasoning tasks.

Research Collaborator **University of Innsbruck** **Jan 2022- Present**

Supervisor: **Prof. Adam Jatowt**

- Conducted a thorough survey of techniques and models in the biomedical summarization domain. Reviewed more than 40 papers in the biomedical summarization domain and wrote a survey paper on the findings.
- Currently working on understanding Sentence fusion and information combination mechanisms in NLG tasks.
- Proposing an evaluation metric capable of capturing the information fusion process of Transformers.

Research Publication

Accepted Works [1]

- Apoorva Singh, **Raghav Jain***, Sriparna Saha, "Reimagining Complaint Analysis: Adopting Seq2Path for a Generative Text-to-Text Framework" **AACL-IJCNLP 2023**
- Apoorva Singh, Apoorv Verma, **Raghav Jain**, Sriparna Saha, "Investigating the Impact of Multimodality and External Knowledge in Aspect-level Complaint and Sentiment Analysis" **CIKM 2023**
- **Raghav Jain***, Apoorva Singh*, Vivek Gangwar, Sriparna Saha, "AbCoRD: Exploiting multimodal generative approach for Aspect-based Complaint and Rationale Detection" **ACM MultiMedia 2023**
- **Raghav Jain**, Tulika Saha, Sriparna Saha, "T-VAKS: A Tutoring-based Multimodal Dialog System via Knowledge Selection" **ECAI 2023**

[1] * denotes Equal Contribution

- Krishanu maity, Nilabja Ghosh, **Raghav Jain**, Sriparna Saha, Pushpak Bhattacharyya, "StereoHate: Towards Identifying Stereotypical Bias and Target Group in Hate Speech Detection" **Natural Language Engineering journal, 2023**
- John Nay, David Karamardian, Sarah Lawsky, Wenting Tao, Meghana Bhat, **Raghav Jain**, Aaron Travis Lee, Jonathan H. Choi, Jungo Kasai, "Large Language Models As Tax Attorneys: A Case Study in Legal Capabilities Emergence" **Phil. Trans. of the Royal Society A 2023**
- **Raghav Jain***, Apoorv Verma*, Apoorva Singh, Vivek Gangwar, Sriparna Saha, "Aspect-based Complaint and Cause Detection: A Multimodal Generative Framework with External Knowledge Infusion" **ECML 2023**
- Apoorva Singh*, **Raghav Jain***, Prince Jha, Sriparna Saha, "Peeking inside the black box: A Commonsense-aware Generative Framework for Explainable Complaint Detection" **ACL 2023**
- Prince Jha*, Krishanu Maity*, **Raghav Jain**, Sriparna Saha, Pushpak Bhattacharyya, "Explain Thyself Bully': Sentiment Aided Cyberbullying Detection with Explanation" **ICDAR 2023**
- **Raghav Jain,** Krishanu Maity, Prince Jha, Sriparna Saha, "Generative models vs Discriminative models: Which performs better in detecting cyberbullying in memes?" **IJCNN 2023**
- Sharmistha Das, Apoorva Singh, **Raghav Jain**, Sriparna Saha, Alka Maurya, "Let the model make financial senses: a Text2Text generative approach for financial complaint identification", **PAKDD 2023**
- Deep Gandhi*, **Raghav Jain***, Jay Gala*, Jhagrut Lalwani*, Swapneel S Mehta*, "Expanding Access to ML Research through Student-led Collaboratives", **NeurIPS 2022 Workshop WBRC**
- **Raghav Jain**, Tulika Saha, Souhitya Chakraborty, Sriparna Saha, "Domain Infused Conversational Response Generation for Tutoring based Virtual Agent", **IJCNN 2022**
- **Raghav Jain***, Vaibhav Mavi*, Anubhav Jangra*, Sriparna Saha, "WIDAR - Weighted Input Document Augmented Rouge", **ECIR 2022**
- **Raghav Jain***, Prashant Kumar*, Shivam Chaudhary*, Sanjay Kumar, "Solving Community Detection in Social Networks: A comprehensive study", **ICCMC 2021**
- Anubhav Jangra*, **Raghav Jain***, Vaibhav Mavi*, Sriparna Saha, Pushpak Bhattacharyya, "Semantic Extractor Paraphraser based Abstractive Summarization", **ICON 2020**

Under Submission

- **Raghav Jain**, Anubhav Jangra, Sriparna Saha, Adam Jatowt, A Survey on Medical Document Summarization, **ACM Computing Surveys** (**Under Review**)

Work Experience

NLP Researcher — **Nintee** — **Sept 2022- February 2022**

Manager: **Mr. Paras Chopra**

- Developed a GPT-3 powered Health Search Engine that returns responses to a user query based on openly available podcasts, journals, and subreddits.
- Developed an end-to-end GPT-3 based search engine for providing nutritional information for different food items.
- Worked extensively with OpenAI GPT-3 and different prompt engineering tools.

ML Engineer — **TelerApp** — **April 2022- August 2022**

Manager: **Dr. Pracheta Sahoo**

- Curated a dataset of medical reports (X-ray reports, radiology reports, and CT reports) for the speech-to-text task.
- Fine-tune various speech-to-text models (such as Facebook Wav2vec 2.0) for generating transcripts from medical report audio.
- Designed a Django-based Web Interface for hosting the developed Speech-to-Text model.

Open Source Project

Founding Engineer — **SimPPL** — **June 2022- Present**

- **SimPPL** is a non-profit online platform that allows anyone external to a platform (such as Twitter or Reddit) to evaluate the spread of online news examining which accounts spread it, when they spread it, and how much interest it garnered–all at the click of a button.
- We received a previous grant to create a dashboard to study what information is spread online and how it spreads and partnered with The Sunday Times (UK) to verify if there was any coordinated activity surrounding the spread of news from two Russian state-backed media outlets resulting in the first iteration of our system under the 'Networks and Topics' tab of our website.

Reviewer

- Reviewer: EMNLP 2022, Expert Systems With Applications, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)
- Secondary Reviewer: WebConf 2022, KDD 2022