

# Rohit Saxena

rohitsaxena.uoe@gmail.com | saxenarohit.github.io | +44-7424710731

 LinkedIn |  GitHub |  Google Scholar

## SUMMARY

My research interests focus on natural language understanding and multimodal learning, specifically on long-context modeling and improving reasoning in vision and language models. Prior to my PhD, I had 7 years of industry research experience at TCS Research, where I worked on problems in the media and entertainment domain.

## EDUCATION

- **University of Edinburgh** Sep. 2019 – Present  
*PhD, Natural Language Processing*  
◦ Advisor: Frank Keller, Pasquale Minervini  
◦ Topic: Long Document and Multimodal Summarization
- **Uttar Pradesh Technical University** Aug. 2007 – Jun. 2011  
*Bachelor of Technology in Computer Science and Engineering*  
India

## EXPERIENCE

- **TCS Research, Tata Research Development and Design Center (TRDDC)** Mar 2012 - Aug 2019  
*Researcher* Pune, India  
◦ Part of the CTO Group, where I conducted research in the Media and Entertainment Research Group, focusing on NLP and multimodal machine learning.  
◦ Mentored a team of researchers to successfully complete multiple high-impact research projects, several of which were adopted into production systems for major clients in the media and entertainment sector.  
◦ Developed POCs for new models and have experience presenting research work to CxO-level stakeholders.
- **Amazon Web Services (AWS-AI)** Sep 2023 - Dec 2023  
*Applied Science Intern* Seattle, USA  
◦ Proposed and implemented a method for mitigating hallucination in large language models by generating attribution for generated text, enabling correction of hallucinated responses and enhancing model reliability

## SELECTED PUBLICATIONS

1. **Rohit Saxena**, Pasquale Minervini, Frank Keller. *PosterSum: A Multimodal Benchmark for Scientific Poster Summarization*. Under review.
2. **Rohit Saxena**, Aryo Pradipta Gema, Pasquale Minervini. *Lost in Time: Clock and Calendar Understanding Challenges in Multimodal LLMs*. Reasoning and Planning for LLMs at ICLR 2025.
3. Dongqi Liu, Chenxi Whitehouse, Xi Yu, Louis Mahon, **Rohit Saxena** et al. *What Is That Talk About? A Video-to-Text Summarization Dataset for Scientific Presentations*. Under review.
4. **Rohit Saxena**, Hao Tang, Frank Keller. *End-to-End Long Document Summarization using Gradient Caching*. Under review.
5. Aryo Pradipta Gema, Joshua Ong, Giwon Hong, Alessio Devoto, Alberto Mancino, **Rohit Saxena**, et al. *Are We Done with MMLU?* Association for Computational Linguistics: NAACL 2025
6. **Rohit Saxena**, Frank Keller. *Select and Summarize: Scene Saliency for Movie Script Summarization*. Findings of the Association for Computational Linguistics: NAACL 2024.
7. **Rohit Saxena**, Frank Keller. *MovieSum: An Abstractive Summarization Dataset for Movie Screenplays*. Findings of the Association for Computational Linguistics: ACL 2024.
8. Giwon Hong\*, Aryo Pradipta Gema\*, **Rohit Saxena\***, et al. *The Hallucinations Leaderboard—An Open Effort to Measure Hallucinations in Large Language Models*. Under review \*equal contribution
9. **Rohit Saxena**, Savita Bhat, Niranjana Pedanekar. *EmotionX-Area66: Predicting Emotions in Dialogues*. Association for Computational Linguistics Workshop on Natural Language Processing for Social Media: ACL 2018.
10. **Rohit Saxena**, Savita Bhat, Niranjana Pedanekar. *Live on TV, Alive on Twitter: Quantifying Continuous Partial Attention of Viewers During Live Television Telecasts*. 17th IEEE International Conference on Data Mining series Workshop on Data Science for Human Performance in Social Networks: ICDM 2017
11. **Rohit Saxena**, Niranjana Pedanekar. *I Know What You Coded Last Summer: Mining Candidate Expertise from GitHub Repositories*. 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing: CSCW 2017

## PATENTS

---

1. [US20160269417A1](#) - Dynamic data masking for mainframe application.
2. [US20150381703A1](#) - Automating a process for web-based software.
3. [US10599864B2](#) - Sensitive audio zone rearrangement for customer verification.
4. [US10198322B2](#) - Method and system for efficient selective backup strategy in an enterprise.
5. [US10296523B2](#) - System and method for estimating temporal importance of data
6. [201621003887](#) - Systems and methods for estimating skill-sets of users in a distributed environment.

## SELECTED INDUSTRY & RESEARCH PROJECTS

---

- **Retrieval Augmented Generation for Question Answering** *May 2024 - Jun. 2024*  
*University of Edinburgh*
  - Developed a retrieval-augmented prompting (RAG) approach to incorporate external knowledge sources into question answering (QA) systems.
  - Investigated the impact of external knowledge on QA performance, focusing on complex, multi-disciplinary benchmarks.
  - Tested the RAG approach on the MMLU dataset, analyzing improvements in accuracy and robustness on knowledge-intensive questions.
- **Mitigation of Hallucination in Large Language Models** *Sep. 2023 - Dec. 2023*  
*Amazon Science*
  - Proposed a method to generate attribution for generated text, using it to correct hallucinations in large language models.
  - Improved performance on the Natural Question Answering and XSum datasets.
- **Multimodal Content Analysis of Television Shows** *Sep. 2018 - Aug. 2019*  
*TCS Research*
  - Developed methods for emotion detection in dialogues using subtitles, speaker identification with audio features, and actor face tracking with FaceNet and OpenCV.
  - Implemented a contrastive loss-based approach to enhance speaker identification accuracy in movies.
- **Style Transfer for Advertisements** *Mar. 2017 - Aug. 2018*  
*TCS Research*
  - Created a photo-realistic style transfer network to render stylized advertisements, minimizing distraction during video playback.
  - Evaluated this approach on 5 advertisements across 10 popular Hollywood movie scenes.
- **Analysis of TV Viewer's Attention and Character Popularity** *Mar. 2015 - Aug. 2018*  
*TCS Research*
  - Developed a model to quantify viewer attention using live tweeting behavior during TV broadcasts.
  - Ranked characters based on features such as audio-visual cues (energy, aesthetics, memorability) and social attributes (mentions, dialogue popularity).

## SKILLS

---

- **Programming Languages:** Python (Advanced), Java-SE (Advance), Java-EE (Advance), C++, C
- **Deep Learning Frameworks:** PyTorch (Advance), Keras, Tensorflow, Caffe
- **Libraries:** Transformers, Pandas, Numpy, scikit-learn
- **Web Skills** JavaScript, HTML, CSS, D3.js, JQuery
- **Development Tools:** Visual Studio, PyCharm, Eclipse, Spyder, IBM Rational Application Developer,  $\LaTeX$ , Git
- **Databases:** IBM DB2, Oracle, MySQL, PostgreSQL
- **Servers:** IBM WebSphere Application Server, JBoss Application Server, Tomcat
- **Operating System:** Ubuntu, Windows

## AWARDS AND CERTIFICATIONS

---

- Awarded full scholarship for PhD from UK Research and Innovation, valued at approximately £152,836 (covering tuition, stipend, and program costs).
- Awarded scholarship "Airtel Scholar Hunt" by Middlesex University, UK, for undergraduate studies (not pursued).
- Awarded "ILP Kudos" for ranking in the top 1% in TCS Initial Learning Program (among 40,000 peers).
- 2nd place in the SocialNLP 2018 Data Challenge.
- Semi-finalist in The Great Mind Challenge (TGMC) 2009 and 2010, organized by IBM.
- Oracle Certified Associate, Java SE 7 Programmer.
- Verified Coursera certificates: Deep Learning Specialization, Machine Learning by Stanford Online, Classification by University of Washington.

## ACADEMIC EXPERIENCE

---

- **Reviewer:** ACL, EMNLP, ACL Rolling Review
- **Teaching Assistant, University of Edinburgh:** Tutor for Natural Language Understanding, Generation, and Machine Translation
- **Teaching Assistant, University of Edinburgh:** Marker for Accelerated Natural Language Processing

## ADDITIONAL INFORMATION

---

**Presented work:** CIKM 2016, CSCW 2017, ICDM 2017, ACL 2018, NAACL 2024, ACL 2024.

**Interests:** [Photography](#), Travelling.