Rohan Charudatt Salvi

rohancsalvi@gmail.com | Personal Website

217-974-5025 | Google Scholar

About Me

I am Rohan Charudatt Salvi, a recent master's graduate from the University of Illinois Urbana-Champaign. My primary research interest lies in Natural Language Processing (NLP). I have prior experience, and my ongoing research involves applying NLP techniques to tackle diverse problems. Additionally, I am keenly interested in working with language models, particularly in information extraction, conversational agents, and biases. I am enthusiastic about pursuing a Ph.D. to further my learning and engage in research that advances the frontiers of NLP. Proficient in Python, R, C++, NLP, Deep Learning, and Computer Vision.

Education

Aug 2021 – July 2023 **University of Illinois Urbana-Champaign**
MS in Information Science (GPA: 4.0/ 4.0)

Aug 2016 – May 2020 **Visvesvaraya National Institute of Technology, Nagpur, India**
Bachelor of Technology in Computer Science & Engineering (CGPA: 8.56/10)

Employment

Sept 2023 – Present **University of Illinois Urbana Champaign-** *Graduate Research Assistant*
- Advisor: Catherine Blake
- Research focus: **Utilizing Large Language Models for Precision Medicine through Non-Genetic Information**

Sept 2023 – Present **Ashoka Trust for Research in Ecology and Environment (ATREE) -** *NLP Engineer*
- Spearheading the development of a chatbot for the Plants of India Portal at ATREE to enhance access to information on plant taxonomy, distribution, and discovery.
- Implementing retrieval ranking and chain of thought prompting techniques in the chatbot to improve user query understanding and response accuracy.
- Utilizing retrieval-augmented generation in the chatbot to provide more precise and contextually relevant information from ATREE's plant database.
- Evaluating various domain adaptation strategies for large language models to understand plant taxonomies effectively.

Mar 2023 – Jul 2023 **University of Illinois Urbana Champaign-** *Graduate Research Assistant*
- Compared performance of varied information extraction techniques to extract GDPR facets around the use of personal data in Privacy policies.
- Employing prompting methods to extract GDPR facets through large language models

May 2022 – Dec 2022 **Bayer R&D Services-** *NLP Intern*
- Created a Natural Language Interface for Databases (NLIDB) for the company's Data Warehouse.
- Designed an NLP pipeline consisting of text preprocessing, parsers, and pattern recognition to extract columns, tables, and filtering conditions from the user's query.
- Developed a dialog system-based NLIDB solution capable of generating questions to resolve ambiguity or get more information from the user.

Aug 2021 – Dec 2021 **Center for Health Informatics-WHO Collaboration Center, UIUC -** NLP Research Intern
- Data collection related to COVID-19 from different social media sources.
- Cleaning the raw data and transforming it into features appropriate for Neural Network models.

- Train and evaluate deep text classification models such as CNN, BERT, and Bi-LSTM to classify information into six categories.

Jul 2020 – Jul 2021 **Tata Research Development and Design Center, India -** *Research Intern*
- Developed a system dynamics model for COVID-19 in India using stock and flow models
- Simulated and analyzed the effects of interventions such as vaccines, testing, contact tracing, etc.

May 2019 - Jul 2019 **Credit Suisse, India -** *Intern Technology*
- Built a Content Management System (CMS) for FAQs associated with Compliance.

Awards

iSchool Research Showcase 2022- Second runner up:
Poster: Stereotype Detection in LLM using Computational Grounded Theory.

Publications

- Zheng Qingxiao, Shengyang Xu, Lingqing Wang, Yiliu Tang, **Rohan Salvi**, Guo Freeman, and Yun Huang. 2023. Understanding Safety Risks and Safety Design in Social VR Environments. Proceedings of the ACM on Human-Computer Interaction, CSCW.

- Souvik Barat, Aditya Paranjape, Anwesha Basu, **Rohan Salvi**, Supratim Gosh, Vinay Kulkarni. Modeling and Simulation for the spread of Covid-19 in an Indian City. Proceedings of the 2022 Winter Simulation Conference.

- **Rohan Charudatt Salvi**, Catherine Blake, Masooda Bashir. PrivacyChat: Utilizing Large Language Models for Fine-Grained Information Extraction over Privacy Policies. Iconference, 2024. (*Under Print*)

- Day, K., Christl, D., **Salvi, R.**, & Sriram, P. (2023, March 24). *Video Pre-trained Transformer: A Multimodal Mixture of Pre-trained Experts*. ArXiv.org. https://arxiv.org/abs/2304.10505

Publications (In Review)

- **Rohan Charudatt Salvi**, Catherine Blake. Finding the Devil in the Details: Large Language Models for High Precision Information Extraction. Journal of the American Medical Informatics Association, 2023.

- **Rohan Charudatt Salvi**, Nigel Bosch. Investigating Perception of Stereotypes in Large Language Models: A Computational Grounded Theory Approach. Big Data & Society, 2023.

Leadership and Teaching Experience

Jan 2022 - May 2022 **Mentored a Junior Student, Neha Mathews, University of Illinois Urbana Champaign**
Poster: Portrayal of COVID-19 by the American News Media, Ischool student showcase 2022

Aug 2019 - Dec 2019 **CSL304: Neuro-Fuzzy Techniques, Visvesvaraya National Institute of Technology, Nagpur**
Undergraduate Teaching Assistant - Professor Poonam Sharma
Created Jupyter notebooks as practice assignments and references related to neural networks.

Aug 2018 - May 2020 **Student Mentor, Visvesvaraya National Institute of Technology, Nagpur**
Planned mentorship programs, development seminars, and community bonding activities.

Service

Speaker at the following events:

- **NLP and Computer Vision** at NLP Applications Workshop in National Center for Supercomputing Applications Student Research Conference, 2023
- **Introduction to Artificial Intelligence and AI for Ecology**, ATREE
- **Emerging Technologies for Environment Conservation Workshop**, ATREE

Student Volunteer for the following conferences:
Computer-Supported Cooperative Work and Social Computing 2022, Winter Simulation Conference 2022

Ongoing Research Projects

Sept 2023 **Professor Catherine Blake -** University of Illinois Urbana-Champaign (UIUC)

Project: Utilizing Large Language Models for Precision Medicine through Non-Genetic Information

- Fine-grained extraction: Utilizing language models to extract granular information and relations from randomized clinical trials.
- Working with full text: Assessing LLMs' capability to understand and extract information from the full text of randomized control trials.
- Personalized intervention recommendation: Evaluating the use of LLMs as conversational agents to recommend interventions. Specifically focused on their ability to answer questions, maintain conversation, and provide grounded information.

Sept 2023 **Professor Nigel Bosch -** University of Illinois Urbana-Champaign (UIUC)

Project: Robust framework for measuring and mitigating stereotypical biases in LLM

- Methodologies for bias evaluation and mitigation: Developing automatic prompt generation and innovative prompt-based tasks to measure and identify stereotypical biases in LLMs across varied contextual settings. Comparing the model generation on handcrafted prompts against automated prompts.
- Evaluation metrics: Creating a novel evaluation metric that analyses biases based on prompt-based generation by LLMs. We aim to create a metric that can view the bias from multiple lenses and have begun initial research by working on intersectional biases. Our evaluation will also consider the noise in prompts, model parameters, and their impact. This comprehensive approach will allow us to gauge and understand biases more accurately.