Neemesh Yadav

New Delhi, India - neemesh20529@iiitd.ac.in - +91-9810641769 - Linkedin - Github - Google Scholar

EDUCATION

Indraprastha Institute of Information Technology, Delhi, India — Jan 2021 — June 2024
B.Tech. in Computer Science and Biosciences — CGPA: 8.17/10

Ryan International School, Mayur Vihar, Delhi — 2019 — 2020
CBSE - Grade XII (Science) — Percentage: 94%

Ryan International School, Mayur Vihar, Delhi — 2017 — 2018
CBSE - Grade X — Percentage: 92.6%

PUBLICATIONS

In-Domain Conquers Commonsense Knowledge: Leveraging Domain-Specific Toxicity Attributes for Explanation Generation of Implied Hate Speech
Neemesh Yadav*, Sarah Masud*, Vikram Goyal, Md Shad Akhtar, Tanmoy Chakraborty
Under Review at WWW 2024

Beyond Negativity: Re-Analysis and Follow-Up Experiments on Hope Speech Detection
Neemesh Yadav*, Mohammad Aflah Khan*, Diksha Sethi, Raghav Sahni
ICLR 2023 - Tiny Paper Track

The Art of Embedding Fusion: Optimizing Hate Speech Detection
Mohammad Aflah Khan*, **Neemesh Yadav***, Mohit Jain, Sanyam Goyal
ICLR 2023 - Tiny Paper Track

EXPERIENCE

Mohammed Bin Zayed University of Artificial Intelligence (MBZUAI) — Abu Dhabi, UAE
Undergraduate Research Intern — May 2023 - Present
Advisor: Dr. Preslav Nakov

- Developed a tool to help users profile news media sources on the basis of their factuality and political stance on the fly
- Working on extending previous work by incorporating other dimensions such as rationale/reasoning to follow work in Explainable AI by generating rationales for the decisions

Laboratory for Computational Social Systems (LCS2) — Delhi, India
Undegraduate Student Researcher — May 2022 - Present
Guides: Dr Md. Shad Akhtar & Dr. Tanmoy Chakraborty

- Led a project on highlighting the existing retrieval and knowledge incorporation methods for generating explanations or stereotypical implications of Hate Speech (under review at WWW 2024)

* implies equal contribution

- Future work focuses on benchmarking reasoning in LLMs

LibrAI Remote

Research Collaborator August 2023 - Present

Collaborators: Dr. Xudong Han & Dr. Artem Shelmanov

- Collaborating with Postdoc and Ph.D. researchers from MBZUAI under Dr. Timothy Baldwin, for work on Fairness and Ethical AI
- Working on benchmarking debiasing techniques over multiple datasets, and proposing our own scoring methods extending from OOD detection

Fiverr Remote

Freelance Software Developer January 2021 - February 2022

- Creating chatbots, software development, and solving algorithmic puzzles.

PROJECTS

The Art of Embedding Fusion: Optimizing Hate Speech Detection Course Project

Github, **arXiv**, Accepted at ICLR 2023 Tiny Papers Track

Course Instructor: Dr. Md. Shad Akhtar

- We shed light on various combination techniques for several PLMs and comprehensively analyze their effectiveness
- We observe that the choice of embeddings for combination is inconsequential to the final results. This was shown by the marginal effects / slight improvements on the final scores, which come at a high computational expense.
-

Beyond Negativity: Re-Analysis and Follow-Up Experiments on Hope Speech Detection Course Project

Github, **arXiv**, Accepted at ICLR 2023 Tiny Papers Track

Course Instructor: Dr. Jainendra Shukla

- Our study finds computationally efficient yet comparable/superior methods to the status quo
- We show that simple ML models, when coupled with the correct strategies, including data analysis and high-quality embeddings can lead to results far better than Transformer-based models

Automatic QnA Generation from YT Videos Course Project

Github, **Report**

Course Instructor: Dr. Rajiv Ratn Shah

- We developed a streamlit webapp to help with the existing problem of loss of attention span during lectures.
- The webapp segments the videos into fixed sections, and generates questions over those segments which are then presented to the user. The goal is to help the user understand the concepts being taught in the video.

dCrypy

Github,

- A simple Python tool that incorporates multiple ciphers, and can be used to decode ciphertexts or encode plaintexts using the appropriate keys.

TLDR-Bot Hackathon Project

Github

- A Twitter bot that summarizes long threads into a single tweet, using the T5 model, for those who want to scope the concept of the thread.

ZoneIn: Mindmap Generator Hackathon Project

Github

- We developed a tool that can generate mindmaps of some query YouTube videos based on their transcription
- This is extremely useful for people who suffer from a relatively low attention span and can also be extended to real-time meetings with appropriate transcribing tools.

Hate Speech Curbing Bot Hackathon Project

Github

- A discord bot that has a hate speech filter and can be used to curb the same in real-time discord threads
- The bot also offers multiple games, such as Hangman and Guess the Word

SKILLS

- **Relevant Coursework:** Natural Language Processing†, Deep Learning†, Machine Learning, Information Retrieval, Reinforcement Learning, Linear Algebra, Probability & Statistics, Multivariate Calculus
- **Programming:** Python, Javascript, Arduino
- **Libraries, Frameworks & Other Technologies:** Huggingface, PyTorch, TensorFlow, Keras, W&B, Jupyter Notebooks, Streamlit, Flask, Scikit-Learn, FastAPI, Git, Linux, Bash

AWARDS AND ACHIEVEMENTS

- 1^{st} **place (as a solo-player) in 3 consecutive cryptic hunts organized by IEEE IIITD**: Won cryptic hunts, where participants are required to solve questions with ciphers and OSINT.
- 1^{st} **place** in **The12Rings**: An intense 144-question-long cryptic hunt spanning across 10 months and recording a participation of over 50,000 people.
- **Finalist Anveshan Hackathon**: Designed a Mindmap Generator from YouTube videos using transcripts and open source models.
- **Runner Up Byld + WiT Hackathon**: Designed a Discord Bot that detects toxic messages and alters them to become funny alternatives, thereby removing the hate and introducing humor in the conversation.

ORGANIZER, REVIEWER & VOLUNTEER WORK

- **Invited Reviewer** for The 8th Workshop on Online Abuse and Harms - To be organized with NAACL 2024
- **Reviewer** for LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources, and Evaluation
- **Volunteer at ICON 2022** held at IIITD: Helped out in the smooth conduct of keynotes, talks, and paper presentations

† Graduate Level Course

CO-CURRICULAR ACTIVITIES

- **Core Member Byld**: The Development Club at IIITD
- **Research Mentor** for the **Undergraduate Research Club**: Mentoring students interested in research
- **Research Event Organizing Team Lead** for **Esya 2023**: IIITD's Annual Technical Festival
  - Led the organization of HackCog, which was predominantly focused on solving tasks related to human cognition
  - Mentored the organizing team of PromptCraft, where participants were asked to craft their own prompts and come up with creative but unique solutions to problems using LLMs
  - Mentored the organizing team of MLWars, a kaggle competition, where participants were asked to differentiate between a set of spectrograms of human speech from other noises.
- **CompSoc Secretary** for **IEEE IIITD**: IEEE Student Branch of IIITD
  - Mentoring and organizing events that come under CompSoc, such as cryptic hunts (both online and offline), and workshops.