Shafiuddin Rehan Ahmed

ahmeshaf.github.io linkedin.com/in/ahmeshaf github.com/ahmeshaf shah7567@colorado.edu

Looking for full-time NLP/ML positions post December 2023

RESEARCH INTERESTS

Natural Language Processing · Information Extraction and Retrieval · Knowledge Graphs · Machine Learning · Deep Learning · Neuro-Symbolic NLP · Generative Models

EDUCATION

University of Colorado, Boulder, USA *Aug 2017 - Dec 2023*
MS and *Ph.D.*, Computer Science and Engineering

Indian Institute of Technology, Hyderabad, India *Aug 2008 - May 2012*
Bachelor of Technology, Computer Science and Engineering

TECHNICAL SKILLS

Programming:	Python, Java, C++, OCaML, Bash, SQL, SPARQL/RDF, LaTeX
Software & Tools:	Emacs, Eclipse, Visual Studio, Git, MySQL
Deep Learning:	PyTorch, TensorFlow, Keras, GPT, AllenNLP, HuggingFace, spaCy

RESEARCH EXPERIENCE

ExplosionAI GmbH (makers of spaCy), Berlin, Germany (100% Remote) *May 2023 - Aug 2023*
Machine Learning Engineer Intern (`Python`)

- Design and Development of Semantic Role Labeling system using spacy-llms

Papers:

- A project concerning usage of LLMs for Semantic Role Labeling (in-review)
- A project concerning usage of LLMs for Event Coreference Resolution (in-review)

ExplosionAI GmbH (makers of spaCy), Berlin, Germany (100% Remote) *May 2022 - Aug 2022*
Machine Learning Engineer Intern (`Python`)

- Developed annotation guidelines and interface for model-in-the-loop Event Extraction and Cross Document Event Coreference Resolution using prodi.gy annotation tool.

Papers:

- https://aclanthology.org/2023.law-1.14/

University of Colorado, Boulder, USA *Jan 2018 - May 2022*
Research Assistant (`Java, Python`)

- Led the technical development of the Entity and Event Coreference Resolution pipeline of the Colorado team in DARPA's AIDA program[1]
- Collaborated with multiple universities to develop a cross-lingual and cross-modal coreferencing system run on documents represented as knowledge graphs
- Participated in Streaming Multimedia Knowledge Base Population (SM-KBP[2]) 2018, 2019, and 2020 with consistently the best results in Event Coreference Resolution task
- Worked in collaboration with Anschutz medical center in creating clinical annotation tools for colon cancer bio-marker relations detection and classification.

Sopris Health, USA *June 2018 - Aug 2018*
Software Engineering Intern II (`Python`)

[1] https://www.darpa.mil/program/active-interpretation-of-disparate-alternatives
[2] https://tac.nist.gov/2020/KBP/SM-KBP/

- Implemented a CNN classifier that detects clinically irrelevant utterances while transcribing clinical conversations using Sopris Health app, achieving an accuracy of about 95% surpassing previous best

HP Inc. R&D, Bangalore, India *Aug 2012 - Aug 2017*
Senior R&D Engineer (`C#, Python`)

- Formed and led the team working on NLP solutions for HP printer's customer care service
- Designed and developed an unsupervised model that learns troubleshooting steps for printer issues from customer care service records.
- Developed a solution that converts unstructured customer care text records into Markov graphs of troubleshooting steps.
- Developed a data-driven chat-bot that guides the agents in finding the best solution for printer issues.

TEACHING EXPERIENCE

Natural Language Processing, University of Colorado, Boulder *Aug 2023 - Dec 2023*
Teaching Assistant

- Assisted a class of 90 graduate students on Professor James Martin's NLP, which involved designing five class assignments and mentoring the students in their final projects.

Fundamentals of Software Engineering, University of Colorado, Boulder *Jan 2023 - Present*
Teaching Assistant

- Currently teaching a class of 100 graduate students on Professor Michael Barinek's Software Engineering, which involves designing 5 assignments, 5 quizzes, and mentoring student's final project.

Natural Language Processing, University of Colorado, Boulder *Aug 2022 - Dec 2022*
Teaching Assistant

- Assisted a class of 104 graduate students on Professor James Martin's NLP, which involved creating and completing five class assignments.
- Mentored students in publishing their project results for SemEval 2023 Tasks 4 and 6[3]

PATENTS

1. Shameed Sait M A, Shafiuddin Rehan Ahmed, Niranjan Damera Venkata. *Providing Solutions Using Stochastic Modelling*. en. 2018. URL: `https://patents.google.com/patent/US20210049489A1/`

AWARDS

1. Best Frame Recall for Cross Document Event Coreference Resolution in Text Analysis Conference, 2019
2. 3rd Place Outstanding Poster - In-Progress Research, Graduate Students' Research expo., 2018-2019
3. Merit-cum-Means Scholarship for Undergraduate Studies, 2009-2012

PAPERS

1. Shafiuddin Rehan Ahmed, Abhijnan Nath, Michael Regan, Adam Pollins, Nikhil Krishnaswamy, and James H. Martin. "How Good Is the Model in Model-in-the-loop Event Coreference Resolution Annotation?" In: *Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 136–145. URL: `https://aclanthology.org/2023.law-1.14`
2. Shafiuddin Rehan Ahmed, Abhijnan Nath, James H. Martin, and Nikhil Krishnaswamy. "$2 * n$ is better than n^2: Decomposing Event Coreference Resolution into Two Tractable Problems". In: *Findings of the Association for Computational Linguistics: ACL 2023*. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 1569–1583. URL: `https://aclanthology.org/2023.findings-acl.100`

[3] https://semeval.github.io/SemEval2023/tasks.html

3. Shafiuddin Rehan Ahmed and James H. Martin. *Within-Document Event Coreference with BERT-Based Contextualized Representations*. 2020. arXiv: 2102.09600 [cs.CL]. URL: https://arxiv.org/abs/2102.09600

4. Shafiuddin Rehan Ahmed. "CharTransE: An Extension of TransE on Character n-grams". en. In: (2019). URL: http://rgdoi.net/10.13140/RG.2.2.23249.86886

5. Shafiuddin Rehan Ahmed and Dhanendra Soni. "Wikification via Binary and Ranking Techniques". In: (2018). URL: https://rgdoi.net/10.13140/RG.2.2.26121.93282

6. Adam Wiemerslage and Shafiuddin Rehan Ahmed. *From Algebraic Word Problem to Program: A Formalized Approach*. 2018. arXiv: 2003.11517 [cs.CL]. URL: https://arxiv.org/abs/2003.11517

7. Cecilia Mauceri, Shafiuddin Rehan Ahmed, and Timothy O'Gorman. "RAMFIS System Report TAC 2018." In: *Proceedings of the 2018 Text Analysis Conference, TAC 2018, Gaithersburg, Maryland, USA, November 13-14, 2018*. NIST, 2018