Rricha Jalota
ML Researcher

+49 1768 0231609
Berlin, Germany
rrichajalota@gmail.com
linkedin.com/in/rrichajalota

EXPERIENCE

AppTek   **July 2023 — Present**
Research Scientist   *Aachen, Germany*

- Conducting research on LLMs and assessing their capabilities as a part of the LLM team.
- Contributed to the benchmarking and implementation of the automatic evaluation pipeline for ClimateGPT [2].
- Conducted comprehensive analysis on serving ClimateGPT variants (using vLLM); assessed variations in latency & throughput with increasing model size and requests.
- Developing and assessing LLM-based summarizer tailored for call center data.

German Research Center for Artificial Intelligence (DFKI)   **June 2021 — May 2022, Oct 2022 – July 2023**
Research Assistant   *Saarbrücken, Germany*

- Conducted research on identifying, quantifying, and mitigating translation artifacts (a.k.a. Translationese) in human- or machine-translated text.
- Extended an established bias-removal technique to reduce the "bias" induced by translation at word-embedding and sentence-embedding levels [4].
- Developed a monolingual translation based style-transfer approach to mitigate translationese in human translations, notably, in the absence of parallel translated and original data in the same language [3].

Apple Inc.   **Jun 2022 — Sept 2022**
Research Intern at Apple Siri   *Aachen, Germany*

- Conducted research on multi-dialect Neural Network Language Modeling (NNLM) for on-device hybrid Siri ASR pipeline.
- Examined different data sampling techniques and the application of dialect-specific adapters in (non-transformer based) production NNLMs [1].
- Developed a multi-dialect NNLM that could meet the accuracy, latency and memory constraints of on-device single-dialect models [1].

MorningSun Technology (EdTech-Startup)   **May 2021 — Apr 2022**
Research Assistant   *Saarbrücken, Germany*

- Enhanced rule-based Grammatical Error Correction pipeline by integrating transformer-based framework.
- Based on the Translationese Mitigation approach [4], devised a method for classifying essays written by language learners of diverse native backgrounds [5].

Paderborn University, Data Science Group   **Sept 2017 — May 2021**
Research Engineer   *Paderborn, Germany*

- Conducted research on Knowledge Graph-based QA [6,8,10] and dialogue systems [9], information retrieval [11] and text classification [7].

EDUCATION

Saarland University   **Oct 2020 - Aug 2023**
Master of Science (M. Sc.) in Language Science and Technology   *Saarbrücken, Germany*

Indira Gandhi Delhi Technical University for Women (IGDTUW)   **Sept 2013 - May 2017**
Bachelor of Technology (B. Tech.) in Electronics and Communication Engineering   *New Delhi, India*

PUBLICATIONS

[1] **Rricha Jalota**, Lyan Verwimp, Markus Nussbaum-Thom, Amr Mousa, Arturo Argueta, Youssef Oualil. 2024. "Towards A World-English Language Model for on-Device Virtual Assistants." *In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 10811-10815). IEEE.*

[2] David Thulke, Yingbo Gao, Petrus Pelser, Rein Brune, **Rricha Jalota**, ... 2024. "ClimateGPT: Towards AI Synthesizing Interdisciplinary Research on Climate Change." *arXiv preprint*

[3] **Rricha Jalota**, Koel Dutta Chowdhury, Cristina España-Bonet, Josef van Genabith. 2023. "Translating away Translationese without Parallel Data." *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 7086-7100).*

[4] Koel Dutta Chowdhury, **Rricha Jalota**, Cristina España-Bonet, Josef Genabith. 2022. "Towards Debiasing Translation Artifacts." *In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3983-3991).*

[5] **Rricha Jalota**, Peter Bourgonje, Jan Van Sas, Huiyan Huang. 2022. "Mitigating learnerese effects for CEFR classification." *In Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022) (pp. 14-21).*

[6] Daniel Vollmers, **Rricha Jalota**, Diego Moussallem, Hardik Topiwala, Axel-Cyrille Ngonga Ngomo, Ricardo Usbeck. 2021. "Knowledge graph question answering using graph-pattern isomorphism." *In Further with Knowledge Graphs: Proceedings of the 17th International Conference on Semantic Systems.*

[7] Hamada M. Zahera, **Rricha Jalota**, Mohamed Ahmed Sherif, Axel-Cyrille Ngonga Ngomo. 2021. "I-AID: identifying actionable information from disaster-related tweets." *IEEE Access, 9, pp.118861-118870.*

[8] **Rricha Jalota**, Daniel Vollmers, Diego Moussallem and Axel-Cyrille Ngonga Ngomo. 2021. "LAUREN-Knowledge Graph Summarization for Question Answering." *In 2021 IEEE 15th International Conference on Semantic Computing (ICSC) (pp. 221-226). IEEE.*

[9] **Rricha Jalota**, Priyansh Trivedi, Gaurav Maheshwari, Axel-Cyrille Ngonga Ngomo and Ricardo Usbeck. 2019. "An approach for ex-post-facto analysis of knowledge graph-driven chatbots–the dbpedia chatbot." *In International Workshop on Chatbot Research and Design (pp. 19-33).*

[10] Ria Hari Gusmita, **Rricha Jalota**, ... 2019. "QUANT-Question Answering Benchmark Curator." *In Semantic Systems. The Power of AI and Knowledge Graphs: 15th International Conference, SEMANTiCS 2019, Proceedings 15 (pp. 343-358).*

[11] **Rricha Jalota**, Nikit Srivastava ... 2019. "Finding Datasets in Publications: The University of Paderborn Approach." *'Rich Search and Discovery for Research Datasets', chapter 9, SAGE Publications.*

Selected Projects

Interpretation of Text Classification Models   **2021**
- Studied two interpretability methods, Integrated Gradients and Layer-wise Relevance Propagation, to understand the behavior of CNN-based sentiment classifier. Link: github.com/rrichajalota/Interpreting-CNN-BiLSTM

Bilingual Word-Embeddings using Word Aligner   **2021**
- Trained bilingual EN-DE word embeddings using co-occurrence information obtained from the word alignment tool, fast_align. Link: github.com/rrichajalota/WinterSem-2020-21/blob/main/Computational-Linguistics/BiLingual_Word_Embeddings/

Achievements and Awards

Hasso Plattner Institute Career Grant   **Oct 2020**
- Received the career grant to participate in European Women in Technology conference.

DBpedia Delegate at Google Summer of Code Mentor Summit   **Oct 2019**

Rich Context Competition Finalist   **Feb 2019**
- Represented my team (from Paderborn) in the Rich Context Competition. The competition focused on identifying and extracting dataset mentions and research methods from full-text social science publications [11].

Volunteer Experience

Google Summer of Code (GSoC) Mentor   **2018, 2019 Editions**
- As a mentor in GSoC (an initiative encouraging student developers to contribute to open-source software development), I supervised two under-graduate students for the project "Neural-QA model for DBpedia" (github.com/dbpedia/neural-qa). Supported them in improving the model architecture and automating the template generation process of creating training datasets for the model.

Skills

Tools and Languages	Python, Git, Fairseq, PyTorch, TensorFlow, SPARQL, LaTeX
Concepts	Natural Language Processing, Machine Translation, Machine Learning, Deep Learning, Knowledge Graphs
Communication	English *(fluent)*, Hindi *(native)*, German *(beginner)*