# Dr. Mehwish Fatima

[Mehwish Fatima - Linkedin Profile](#)      [Mehwish Fatima - Google Scholar](#)

Contact: +92-329-5443291      Email:    [mehwishfatima.raja@gmail.com](mailto:mehwishfatima.raja@gmail.com)

---

## Summary

AI researcher and educator with a PhD in Computational Linguistics from Heidelberg University, currently serving as Assistant Professor in the Department of Artificial Intelligence & Data Science at SEECS—School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Pakistan's top-ranked university. Specialized in Generative AI, NLP, and multimodal systems, with hands-on experience deploying LLMs using PyTorch, HuggingFace, and DeepSpeed. Published in top-tier venues (ACL, EMNLP) and actively contributes as a reviewer for major NLP conferences. Currently supervising research in federated learning, speech, and math reasoning.

---

## Technical Skills

**Languages & Programming:** Python, Java, C++, C,
**ML and DL:** PyTorch, TensorFlow, Scikit-learn, HuggingFace Transformers, DeepSpeed, LangChain, NLTK, OpenAI APIs, Vector DBs
**Specializations:** Generative AI, LLMs, Cross-lingual NLP, Multimodal AI, Federated Learning
**Tools & Platforms:** Google Colab, AWS, GitHub, WEKA
**Research & Development:** Applied AI Prototyping, Technical Writing, Team Collaboration, Curriculum Design, Academic Supervision

---

## Education

**PhD** in Computational Linguistics      [2018—2024]
Ruprecht-Karls-Universit¨at Heidelberg, Heidelberg, Germany

**MS(CS)**      [2010—2012]
CGPA: 3.17/ 4.00
National University of Computing and Emerging Science (FAST-NU), Lahore, Pakistan

**BS(CS)**      [2003—2007]
CGPA: 3.83/ 4.00  (**Gold medalist**)
University of Agriculture Faisalabad, Faisalabad, Pakistan

---

## Publications

1. Maham Khan, Hajra Binte Naeem, Faiza Khan, Laiba Rana, Seemab Latif, **Mehwish Fatima**, and Raja Khurram Shahzad, "*NUST Titans at SemEval-2025 Task 11: AfroEmo: Enhanced Multilingual Emotion Detection with Afro-XLMR*" (Accepted in SemEval-2025)

2. Mariam Babar Khan, Huma Ameer, Seemab Latif, and **Mehwish Fatima**, "*NUST Nova at RIRAG 2025: A Hybrid Framework for Regulatory Information Retrieval and Question Answering*", In Proceedings of the 1st Regulatory NLP Workshop (RegNLP 2025), pp. 73-78. 2025.

3. Muhammad Rouhan Faisal, Faizyab Ali Shah, Muhammad Abdullah, Shalina Riaz, Huma Ameer, Seemab Latif, and **Mehwish Fatima**, "*NUST Alpha at RIRAG 2025: Fusion RAG for Bridging Lexical and Semantic Retrieval and Question Answering*", In Proceedings of the 1st Regulatory NLP Workshop (RegNLP 2025), pp. 79-84. 2025.

4. Huma Ameer, Muhammad Hannan Akram, Seemab Latif, and **Mehwish Fatima**, "*NUST Omega at RIRAG 2025: Investigating Context-Aware Retrieval and Answer Generation-Lessons Learned and Challenges*", In Proceedings of the 1st Regulatory NLP Workshop (RegNLP 2025), pp. 85-90. 2025.

5. **Mehwish Fatima**, Tim Kolber, Katja Markert, Michael Strube(2023), "*SimCSum: Joint Learning of Simplification and Cross-lingual Summarization for Cross-lingual Science Journalism*" Proceedings of the 4th New Frontiers in Summarization Workshop (EMNLP2023), Association for Computational Linguistics, Online and Singapore, 24-40.

6. **Mehwish Fatima**, Michael Strube (2023), "*Cross-lingual Science Journalism: Select, Simplify and Rewrite Summaries for Non-expert Readers*" Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL 2023), Association for Computational Linguistics, Online and in Toronto, Canada, 1843–1861.

7. **Mehwish Fatima**, Michael Strube (2021), "*A novel Wikipedia based dataset for monolingual and cross-lingual summarization*", Proceedings of the Third Workshop on New Frontiers in Summarization

(EMNLP 2021), Association for Computational Linguistics, Online and in Dominican Republic, edition. 39-50.

8. **Mehwish Fatima**, Mark-Christoph Muller (2019), "*HITS-SBD at the FinSBD Task: Machine Learning vs. Rule-based Sentence Boundary Detection*", Proceedings of the First Workshop on Financial Technology and Natural Language Processing (IJCAI 2019), Macao, China, edition. 115-121.

9. Muhammad Sharjeel, **Mehwish Fatima**, Saba Anwar and Rao Muhammad Adeel Nawab (2018) "*MAPonSMS-Overview of the Multilingual SMS-based Author Profiling Task at FIRE'18*", Forum for Information Retrieval Evaluation, 202-212.

10. **Mehwish Fatima**, Saba Anwar, Amna Naveed, Waqas Arshad Cheema, Rao Muhammad Adeel Nawab, Muntaha Iqbal, Alia Masood (2018), "*Multilingual SMS-based Author Profiling: Data and Methods*", Natural Language Engineering, Cambridge University Press, *24*(5), 695-724, **IF**: **2.36.**

11. **Mehwish Fatima**, Komal Hasan, Saba Anwar, Rao Muhammad Adeel Nawab (2017), "*Multilingual author profiling on Facebook*", Information Processing & Management, Elsevier, *53*(4), 886-904, **IF**: **8.083.**

12. Syed Asad Hussain, **Mehwish Fatima**, Atif Saeed, Imran Raza, and Raja Khurram Shahzad (2017), "*Multilevel classification of security concerns in cloud computing*", Applied Computing and Informatics *13*(1), 57-65.

13. Raja Khurram Shahzad, **Mehwish Fatima**, Niklas Lavesson, and Martin Boldt (2015), "*Consensus Decision Making in Random Forests*", Book Chapter in Machine Learning, Optimization and Big Data, Springer, 347-358.

14. Adnan Noor Mian, **Mehwish Fatima**, Raees Khan, and Ravi Prakash (2014), "*An empirical evaluation of lightweight random walk-based routing protocol in duty cycle aware wireless sensor networks*", The Scientific World Journal (Thesis outcome).

| Experience | |
|---|---|
| | **Assistant Professor** [08.2024—Present] |

**Assistant Professor** [08.2024—Present]
AI & DS Department at SEECS-NUST, Pakistan
Courses: NLP (UG), DSA (UG) ML(UG), LLMs (PG), Generative AI & Its Applications (PG)

**GenAI Solution & Content Architect** [07.2023—05.2024]
Educative Inc, Washington, USA
Solution & content strategist and developer for GenAI, DS, ML and DL courses, learning paths, cheat sheets and blogs.

**Guest NLP Scientist** [10.2018—03.2023]
Heidelberg Institute for Theoretical Studies (HITS), Germany
Topic: Single Document Cross-lingual Abstractive Summarization for Science Journalism

**Lecturer** [08.2012—08.2018]
COMSATS University Islamabad (CUI), Lahore Campus, Pakistan
Courses: DSA (Java/ C++) | OOP (Java) | FCP (C++)

**Visiting Lecturer** [02.2012—06.2012]
University of Education, Lahore, Pakistan
Course: Distributed Databases

**Lab Instructor** [08.2011—12.2011]
National University of Computer & Emerging Sciences (FAST-NU), Lahore, Pakistan
Course: FCP (C++ & graphics)

**Research Projects & Supervision**

**Ongoing Supervised Research @ NUST (MS/PhD Level)**
1. Federated learning for reducing the carbon footprinting of LLMs
2. Solving maths word problems with reasoning based models
3. Hybrid agentic caller agents with LLMs
4. Bad smells identification and code refactoring with AI
5. Thermal comfort and energy optimization in office buildings with AI
6. Telecommunication and LLMs: Regulatory documents to user support

**Completed Theses Supervision @ CUI (MS Level)**
1. Gender Identification for Urdu
2. Author Identification for Urdu
3. Author Profiling for Multilingual Text
4. Cross-Genre Author Profiling Using Semantic Tagging

**Internships of Universität Heidelberg @ HITS**

| | |
|---|---|
| **Projects/ Grants** | **Commercial project**—KisaanCare for early disease prediction and generating alerts, action items, recommendations of pesticides and fertilizers and reports (English) for Waseela, alert dissemination via sms in local languages (Urdu and Punjabi) [April 2025]<br><br>The 5th **Forum for Women in Research** 2025—Multi-Task Learning for Legal Judgment Prediction. Among the shortlisted finalists. [February 2025]<br><br>Investigating the **impact of climate change** in women and children of the rural area in South Asia with AI. [in progress] |
| **Certifications** | NVIDIA DLI Certificate—Building Transformer-Based Natural Language Processing Applications—NVIDIA Deep Learning Institute [April 2021]<br>Lisbon Machine Learning Summer School [July 2019] |
| **Other Activities** | Co-Instructor in LLM BootCamp 2024 @ NUST.<br>Co-organizer of the MAPonSMS'18 competition on digital text forensics—**FIRE'18** @ CUI.<br>Reviewer @ Natural Language Engineering \| Information Processing & Management \| ICIC-2021 \| LREC-2022 \| EMNLP-2022 \| EACL-2023 \| ACL-2023 \| EMNLP-2023 \| LREC-COLING-2024-2025 \| SemEval2025 |
| **Awards** | HEC-DAAD Scholarship ["HRDI-UESTPs/UETs, Phase-1" Batch-VI] (2018 - 2022)<br>HITS Scholarship (10.2022 - 03.2023) |
| **References** | Prof. Dr. Michael Strube<br>Heidelberg Institute for Theoretical Studies (HITS), Heidelberg<br>michael.strube@h-its.org<br><br>Dr. Raja Khurram Shahzad<br>Mid Sweden University, Sweden<br>Raja-Khurram.Shahzad@miun.se<br><br>Dr. Muhammad Saqib Ilyas<br>Lahore University of Management Sciences<br>saqibm@lums.edu.pk |