

M. Taimoor Khan, Ph.D.

☎ +49-17 666 596 217 | ✉ taimoor.nlp@gmail.com | 🌐 taimoor-khan-747b5610 | 📍 Köln, Germany

I am Taimoor, a researcher, academician and data scientists having research interest in natural language processing, adaptive (continual) machine learning, sentiment analysis and more recently hate speech detection and reproducibility. currently, I am serving as Senior researcher (Post-Doc) in Knowledge Technologies for Social sciences dept. at GESIS, Köln. I contribute Digital Behavioral Data (DBD) project as technical lead for Methods Hub service. My responsibilities involve developing and facilitating the development of reproducible AI/ML models for the social science use cases. I have worked on Hadoop/Spark infrastructure for distributed data storage and processing.

Awards

- **Co-chairship** of 2nd MethodsNET conference, 2025.
- **Performer of the year award** at GESIS for the year 2023.
- **Student researcher of the year** in Ph.D., Bahria University, 2016.
- **Multiple honorariums** for publishing in high impact journals.
- **Grants** to present research work at conferences in Singapore (2016), Malaysia (2017) and Vietnam (2018).
- **Performance based** higher education scholarships and fee-weavers.

EXPERIENCE

- **PostDoc at ((KTS Dept), GESIS - Leibniz-Institut für Sozialwissenschaften, Germany (2022-10 - present)**
 - Project Lead/Service manager for reproducible AI/ML niche platform for social science.
 - Supervising a team of PostDoc, PhD students, software engineer and SHKs
 - Managing portal development in coordination with other stakeholders
 - Defining reproducible AI/ML models guidelines, checklists and documentation templates for social science
 - Developing pioneering content (computational models as methods and tutorials) as gold samples for the platform
 - project management, reporting and dissemination at relevant venues
- **Asst. Professor (CS) at (FAST-NUCES Islamabad), Pakistan (2014-08 - 2022-08)**
 - Taught, Computer Science / Data Science courses in BS, MS and Ph.D. programs
 - As Graduate program coordinator, I coordinated with faculty and students to arrange for offering courses, advise on student progress and suggest electives, enforcing rules and ensuring timely completion of student degrees
 - Supervised BS projects and MS thesis on AI/ML and NLP topics
 - Collaborated in AI/ML tool development for low-resourced languages
 - ORIC focal person, Batch Advisor and Lead of student groups
- **Lecturer (CS) at (COMSATS University Islamabad), Pakistan (2011-12 - 2014-07)**
 - Teaching Computer Science courses in BS program
 - Volunteered as campus WebMaster and helped organizing workshops, seminars/symposiums and sport events
- **Multiple Web Developer** contract-based positions in Islamabad (JinTech) and Melbourne (DesignBuildSource, NetworkVideos, StandOutVideos) (2007-2008, 2010-2011)

EDUCATION

- Ph.D. (Computer Science) from Bahria University Islamabad, Pakistan **Thesis title: Lifelong machine learning aspect extraction from large-scale data, 2012 - 2018**
- MS-Information Technology (Professional Computing) with Web development stream from Swinburne University, Australia (**TeamLead in Industry project**), 2009 - 2010
- BS (Computer and Information Sciences) from PIEAS, Pakistan **Thesis title: A comparative analysis of sequential and parallel sparse storage formats, 2003 -2007**

Research

Natural language processing, Machine learning (Continual learning) ([Google scholar profile](#))

• Journal publications:

- Wajid U, Hamza M., **Khan M.T.**, Azam N. A three-way decision approach for dynamically expandable networks. International Journal of Approximate Reasoning. 2024 Mar 1;166:109105. **3.9IF**
- **Khan, M.T.**, Azam, N., Khalid, S. and Aziz, F., Hierarchical lifelong topic modeling using rules extracted from network communities. PloS one, 17(3), p.e0264481, 2022. **3.24IF**
- Shah A., Azam N., and Ali B., **Khan M.T.**, and Yao J. "A Three-way Clustering Approach for Novelty Detection", Information Sciences, Vol-569, pages 650-668, 2021. **6.8IF**
- **Khan, M. T.**, Azam N., Khalid S., and Yao J. "A Three-way Approach for Learning Rules in Automatic Knowledge-based Topic Models", International Journal of Approximate Reasoning, Vol-82, pg 210-226, 2017. **2.845IF**
- **Khan, M. T.**, Durrani M. Y., Khalid S., and Aziz F., Online Knowledge-Based Model for Big Data Topic Extraction, Computational Intelligence and Neuroscience, vol. 2016, Article ID 6081804, 10 pages, 2016. doi:10.1155/2016/6081804, **1.215IF**

• Conference publications:

- ...

• Talks

- Reproducibility of AI/ML methods in computational social science at Meta-Rep, Munich, 2024-10
- Methods Hub: A platform for sharing computational social science methods at KONVENS, Vienna, 2024-09
- HybridDetox - Multilingual text detoxification CLEF, France, 2024-09
- 2020 - Guest speaker at Microsoft event, Microfest++ on the topic "Lifelong Machine learning and the open-world assumption"
- 2019 - Invited speaker session on Lifelong Machine Learning at ICET-2019
- 2019 Workshop on Building Google Chrome Extensions 2014 Workshop on Web development with Wordpress
- 2013 - Web Designing
- 2012 - MySQL and PHP development
- 2011 - Model-View-Controller: Separating business logic from application

PROJECTS and Collaborations

- **TIER2 (Enhancing Trust, Integrity and Efficiency in Research through next-level Reproducibility)**
 - **Horizon EU project** focusing on reproducibility in computer, life and social sciences
 - **Contribution:** reproducibility checklists and the accompanying templates and guidelines for computational social sciences
 - Contributed to **TIER2, EU Horizon project**, representing GESIS as partner institute *2023-06 - 2024-03*

Digital Behavioral Data (DBD)

- GESIS long-term project on collecting, analyzing developing offers from DBD
- **Contribution:** Methods Hub (A DBD service) as PostDoc/ Project lead/Service manager assisted in content development/curation/review, developing guidelines, checklists and templates, and portal development
- Supervisor and lead developer in establishing GESIS first **in-house Big Data Hadoop infrastructure**. Managing, developing and curating research procedures for Digital behavioral data analysis, *2022-10 - 2024-06*. **Managing a team of 6 people** Project Lead/Service Manager, assisting with reporting, managing and supervision responsibilities *2024-08 - present*
- Collaborations
 - **Focal person** in MoU with Pashto academy, to support with **low-resourced language tools**, *2020 - 2022*
 - Earned multiple **HEC Pakistan Grants**, along with institutional scholarships and awards

Teaching and Supervision

- **Teaching**
 - **Online courses** on text mining, sentiment analysis and data cleaning, with 2000+ students from 150+ countries.
 - Grad courses: Text Mining, Social Networks Analysis, Machine learning
 - Undergrad courses: Introduction to Programming (C++, Java, PHP), Object Oriented Programming, Object Oriented Analysis and Design, Web Technologies, Database Systems, Software Engineering, Artificial Intelligence, Data mining and warehousing
 - Online Three courses on text and data analysis with around 2000 students world-wide
- **Supervision**
 - **Supervised** more than 20 BS/MS thesis including pre-doctoral **Voluntary Mentorship** at ([FAPFP](#)), *2014 - 2024*
 - **Graduate Research**
 - * Identifying Optimal Thresholds for Dynamically Expanding Deep Neural Networks, Fatima Al-Fihri Fellowship (Voluntary mentorship) (2021 - Present)
 - * Automatic Thesaurus Generation for Low Resourced Language, Fatima Al-Fihri Fellowship (Voluntary mentorship) (2021 - Present)
 - * MS(CS) Optimizing Knowledge Transfer to Address the Problem of Catastrophic Forgetting in Continual Learning, FAST-NUCES (2021 - Present)
 - * MS(CS) research in Classification of Research Paper Through Language and Style based Features, FAST-NUCES (2019-2020)
 - **Undergrad projects**

- * Semantic Search Engine, FAST-NUCES (2021)
- * Urdu Poetry Classification, FAST-NUCES (2021)
- * Legal Judgment Prediction and Analysis, FAST-NUCES (2020)
- * Optimized load predictions for solar hybrid system at residential level, FAST-NUCES (2020)
- * Fake reviews detection, FAST-NUCES (2019)
- * Narrow folksonomy for social research groups, FAST-NUCES (2018)
- * Exploring hierarchy in topics using formal concept analysis, FAST-NUCES (2017)
- * Context-aware YouTube video recommender system, FAST-NUCES (2017)
- * Analyzing reasons for online social conflicts, FAST-NUCES (2017)
- * Personal analytics for activities management, FAST-NUCES (2016)
- * Government Policy Monitoring through Semi-supervised Aspect-based Sentiment Analysis, FAST-NUCES (2016)
- * Lifelong Machine Learning for Aspect Extraction, FAST-NUCES (2016)
- * Unsupervised Network Analysis of Textual Data, FAST-NUCES (2015)
- * Contextual Categorization of Implicit Bookmarking for Revisitation, FAST-NUCES (2015)
- * Sentiment Analysis using Hadoop, FAST-NUCES (2014)
- * Automatic timetable generation through controlled random approach, COMSATS IIT (2013)
- * Health Management System, COMSATS IIT (2012)

Organizational

- Research events
 - 2019 - 15th IEEE International Conference on Emerging Technologies(ICET-19), Peshawar, Pakistan
 - 2017 - 6th IEEE International Conference on Electrical Engineering and Informatics (ICEEI-17), Langkawi, Malaysia
 - 2016 - 6th IEEE International Conference on Computing and Innovative Technology (INMIC-16), Islamabad
 - 2014 - Research Innovations in Information Technology and Engineering (RIITE-14), Attock
 - 2013 - Research Innovations in Information Technology and Engineering (RIITE-13), Attock
 - 2013 - 11th IEEE international conference on Frontiers of Information Technology (FIT-13), Islamabad
 - 2012 10th IEEE international conference Frontiers of Information Technology (FIT-12), Islamabad
- Review
 - Reviewer at prestigious journals and conferences including Elsevier journals (**KBS**) and (**RESS**)
 - Knowledge-based Systems, Elsevier
 - Journal of Intelligent Systems, De Gruyter
 - International Journal of Privacy and Health Information Management (IJPHIM)
 - International Journal (IEEE Wireless Communications Magazine)
 - International Journal (Elsevier Cognitive Systems Research)
 - (ICET-2019) IEEE International Conference ICET-2019 (IEEE ICET-15)
 - (HICSS-2020) 53rd Hawaii International Conference on System Sciences, Maui, HI, USA (HICSS-53)
 - (HICSS-2019) 52nd Hawaii International Conference on System Sciences, Maui, HI, USA (HICSS-52)
 - (C CODE-2019) 2nd International Conference on Communication, Computing and Digital Systems, Bahria University, Pakistan
 - (ComTech-2017) International Conference on Communication Technologies, NUST, Pakistan

Tools and Technologies)

- First (**Pashto spell checker**) using morphological operations for detecting and fixing out of vocabulary words.
- (**Pashto syllable analysis**) to analyze syllable structures in Pashto words. We are currently using it to develop a Pashto TTS model that will support multiple dialects of the language.
- (**Pashto thesaurus**) offered as browser extension for client-side support.

SKILLS

Programming: Python, PHP, C++, Java

Web programming: PHP, Javascript, SQL/noSQL databases

Markup languages: MD, HTML, CSS, \LaTeX , XML, JSON

Data science tools: Sci-Kit-Learn, NLTK, WordNet, Pandas, Numpy, NetworkX, Gephi

Large Language Models: Tensorflow, keras, PyTorch, Huggingface

Big Data: Hadoop, Spark, MapReduce, MongoDB, Prometheus and Grafana

OS: Windows, Linux (Ubuntu/CentOS)

Project Management: ASANA, MS-Office, Git

languages: English, Urdu, Pashto, Deutsch (A1)