

# M. Taimoor Khan, Ph.D.

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

I am Taimoor, currently serving as Postdoc in Knowledge Technologies for Social sciences dept. at GESIS Köln. I am contributing as project/service manager (for dept.) to the development of a niche platform for reproducible AI/ML social science research models. In this capacity, my responsibilities include supervising portal development, defining guidelines for documenting reproducible models, developing pioneering content for the platform, supervising my team. Additionally, managing project and resources, progress reporting on the project, dissemination at relevant venues, student supervision. We are also collaborating with other 3rd party projects to standardize, synergize, and co-create artifacts.

I bring on over 10 years of experience in academia and have published on ML/AI and NLP topics at reputed venues.

## Highlights

---

- 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*
- Contributed to **TIER2, EU Horizon project**, representing GESIS as partner institute *2023-06 - 2024-03*
- **Supervised** more than 20 BS/MS thesis including pre-doctoral **Voluntary Mentorships** at (FAPFP), *2014 - 2024*
- **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
- Reviewer at prestigious journals and conferences including Elsevier journals (**KBS**) and (**RESS**)
- **Earned student Researcher of the year** at Bahria University Islamabad, Pakistan, *2016*

## 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)

---

- 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**

### Recent 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

## PROJECTS

---

### 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

### 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

## NLP Tools Developed (for Low-Resourced Language)

---

- 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,  $\text{\LaTeX}$ , 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, Deutsch (A1)