# NADIA SAEED

LinkedIn:
linkedin.com/in/nadia-saeed-408980245/
GitHub: github.com/NadiaSaeed
Email: n.anwar040@gmail.com
+92 3025817294 & +92 3015538448
Rawalpindi-Punjab, 44000 Pakistan

## PROFESSIONAL SUMMARY

Experienced data scientist with a robust background in machine learning, specializing in Natural Language Processing (NLP). I possess a solid track record in creating, retraining, and deploying machine learning models, and transforming prototypes into production solutions. My Ph.D. in Computer Science, combined with over six years of software research experience, underscores my proficiency in software engineering, deep learning, and MLOps. Dedicated to advancing the field of NLP, I have published multiple peer-reviewed papers and led international research collaborations.

## SKILLS

**Programming Languages:** Python, R, Java, PHP

**Machine Learning:** Deep Learning, Ensemble Learning, SVM, Logistic Regression, Random Forests, LSTMs, CNNs

**NLP:** Embedding Models, Abstractive Summarization, Entity Recognition

**Software Engineering:** System Design, Databases, Prototyping

**MLOps:** ZenML, Deployment of ML Models to Production

**Tools & Frameworks:** Scikit-Learn, Keras, TensorFlow, SQL

**Data Visualization:** Polling Pie Charts, Customizable Surveys

## EDUCATION

**Ph.D. in Computer Science**
National University of Computer & Emerging Sciences, Islamabad, Pakistan
Graduation Date: November 2023
GPA: 3.89/4.0

**MS in Computer Science**
National University of Computer & Emerging Sciences, Islamabad, Pakistan
Graduation Date: May 2017
GPA: 3.12/4.0

**B.Ed. in Computer Science**
AIOU, Pakistan
Graduation Date: 2014

**MCS (Master of Computer Science)**
VU, Pakistan
Graduation Date: 2013
GPA: 3.12/4.0

## RESEARCH WORK

### Research Project | Published Work

*As First Author*

- Medical Embedding Module address OOV problem: https://doi.org/10.3389/fmolb.2022.928530 (2022)
- EHR Abstractive Summarization addressing Entity Hallucination Problem: https://doi.org/10.1007/s10115-023-02055-6 (2023)
- Build an AI model for the Prediction and Correction of the wrong Diagnostic Disease based on Patient details: https://aclanthology.org/2024.clinicalnlp-1.32/ (2024)
- Used LLM for Hybrid Visual and Textual Dermatology dataset to generate Response for Treatment of a Disease in Three Languages (Chinese, English, and Spanish): https://aclanthology.org/2024.clinicalnlp-1.31/ (2024)
- NLP and Generative AI: A Roadmap for Enhanced CRISPR Design (Submitted:2024)

### Graduate Research Assistant | CBRL-NUCES

*Islamabad - Aug 2019 - Current*

- Designed and implemented a novel stacking ensemble model using meta-learning of the ant colony optimization (ACO) system.
- Published two first-authored peer-reviewed research papers:
  - "Medical terminology-based computing system: a lightweight post-processing solution for out-of-vocabulary multi-word terms" (Frontiers in Molecular Biosciences, 2022)
  - "Transforming Abstractive Summarization of Scientific Articles with Linguistic Analysis and Concept Reinforcement" (Knowledge and Information Systems, 2023)
- Team Leader for the Clinical NLP Workshop 2024, leading an international coding competition focused on Error Detection and Correction in Pathogen Records, and contributing to a forthcoming research publication in the field.

## PROFESSIONAL WORK

### Machine Learning NLP Engineer | AdMaxim. Inc

*Lahore-Punjab - Oct 2017 - Current (Remote)*

- Successfully researched, designed, developed, and deployed machine learning models for client projects, utilizing advanced techniques like SVM, Logistic Regression, Random Forests, LSTMs, and CNNs.
- Proficient in scikit-learn, Keras, and TensorFlow on both CPU and GPU environments.
- Developed and deployed a production-ready demographic classification model predicting user demographics like gender, age, and income level based on internet log data.

### Computer Science Visiting Faculty | Fatima Jinnah Women's University

*Rawalpindi-Punjab- Oct 2023 - Feb 2024*

- Delivers engaging lectures and hands-on labs in Microsoft Office suite (Word, Excel, and PowerPoint) for undergraduate students.
- Shares expertise in Windows 10 OS and relevant software, fostering technological literacy among students.

## ACCOMPLISHMENTS

- Conducted research funded by the Higher Education Commission of Pakistan and the Ministry of Planning Development and Reforms under the National Center in Big Data and Cloud Computing (NCBC).
- MSCS Teaching Assistant for Machine Learning (2017).
- Received a two-year research stipend from the CBRL lab for significant contributions to the project (2020-2022).
- Achieved #1 Rank on the EurekaAlert Dataset for abstractive summarization of scientific research articles:

  https://paperswithcode.com/paper/medtss-transforming-abstractive-summarization (2023).
- Acted as Research Paper's Reviewer of Clinical NLP Workshop 2024.
- MediFact-CoRR 2024 achieved the second-best score in the Clinical NLP Workshop 2024 (https://aclanthology.org/2024.clinicalnlp-1.57.pdf).

## RESEARCH PROJECTS AND INITIATIVES

**MCS Final Year Project (2011-2013)**

- Developed a survey-based application using PHP, Java, and SQL.
- Enabled users to create surveys with various templates, customized questions/answers, and multiple-choice questions (MCQs).
- Implemented login functionality for users to view detailed responses in the form of polling pie charts for each question.

**MSCS Project (2014-2017)**

- Conducted an ensemble-based classification of 20 different UCI datasets with binary and multi-class classification.
- Published the project locally, utilizing Java language.
- Completed Face Recognition using MatLab language.

**Ph.D. Research (2018-2023)**

- Studied Advanced NLP, genomics bioinformatics, and data science.
- Published peer-reviewed papers addressing critical issues in NLP and healthcare.

**Clinical NLP Workshop 2024**

- Participated in solo projects:
  - Built an AI model for the Prediction and Correction of wrong Diagnostic Diseases based on patient details: arXiv, 2024.
  - Developed responses for disease treatment in three languages using LLM for a hybrid visual and textual dermatology dataset: arXiv, 2024.

## CURRENT PROJECTS

- Research on RAG-Llama indexing to address hallucination and falsification problems using guided, rule-based, and verification-based modules.
- Gene editing project exploring sources specifically profluent, with a focus on completing a review paper on CRISPR9.

## TEACHING INTERESTS

- Core subjects: Data Structures, Object Oriented Programming
- Applied courses: Image Processing, Machine Learning, Bioinformatics, Natural Language Processing, Statistics, Data Science

*Teaching is the best way to reinforce my knowledge and I am dedicated to fostering a deep understanding of these subjects in my students.*