Sanya Bathla Taneja

Email: sanyabt11@gmail.com
Website: sanyabt.github.io | LinkedIn | Google Scholar

EXPERIENCE

Johnson & Johnson Innovative Medicine
Postdoctoral Scientist | *November 2024 – October 2026*
- Data science and advanced analytics to support clinical trials with natural language processing and machine learning.

University of Pittsburgh, Intelligent Systems Program
Graduate Student Researcher | *February 2020 – November 2024*
- Led knowledge graph (KG) research and development with 1M nodes and 9M edges, combining biomedical literature and ontologies for generation of mechanistic hypotheses for natural product-drug interactions and safety surveillance using pharmacovigilance signals from adverse event reporting systems. Presented related work at 7 conferences, with 1 Best Poster Award and 5 peer-reviewed publications.
- Implemented pipelines for end-to-end named entity recognition, entity linking, text mining, KG inferences and embeddings with Python.
- Led development and evaluation of prototype system for evidence synthesis with databases, KG, large language models for retrieval augmented generation.
- Developed custom OMOP vocabulary and OBO ontology representations for 300 natural products and constituents for KG and pharmacovigilance using natural language processing and embedding-based models to capture natural products in spontaneous reports.
- Responsible for EHR data extraction, data analysis, and technical development of machine learning and case-control epidemiological analyses OMOP Common Data Model for Alzheimer's disease risk factors using OHDSI methods in R and Python, with 1 peer-reviewed publication and 2 conference presentations.

National Library of Medicine, National Institutes of Health (NIH)
Research Intern | *May – July 2023*
- Developed entity linking methods for diseases in PubMed articles using natural language processing with biomedical ontologies, BERT-based models, and GPT.

University of Pittsburgh, School of Medicine
Research Assistant | *Sept 2018 – Feb 2020*
- Developed natural language processing and machine learning pipelines for twitter surveillance of vaping. Responsible for RITHM software framework maintenance, documentation, and upkeep of the GitHub repository for real-time Twitter data mining, with 3 peer reviewed publications.

University of Pittsburgh and Malawi, Africa | Research Intern | *June – August 2019*
- Developed Bayesian networks and machine learning models to diagnose childhood malaria in Malawi in collaboration with Global Health Informatics Institute.

EDUCATION

PhD Intelligent Systems (Major: Artificial Intelligence) | University of Pittsburgh | *2024*
MS Intelligent Systems | University of Pittsburgh | *2020*
B.Tech. Computer Science and Engineering | Indira Gandhi Delhi Technical University | *2018*

SUMMARY

AI researcher and computer scientist with experience in natural language processing, machine learning, and knowledge representation and their applications in healthcare. Proficient in Python and SQL, with strong communication and writing skills.

SKILLS AND INTERESTS

Skills and Interests: Natural Language Processing, Machine Learning, Knowledge Graphs, Graph Representation Learning, Biomedical Ontologies, EHR Data Analysis, Bayesian Networks, Scientific Writing, Large Language Models
Technologies: Python, SQL, R, Git, C++, RDF, OWL, Neo4j, GPT
Libraries: NLTK, Spacy, Pandas, Scikit-learn, Keras, Networkx, Tensorflow, Langchain

PROFESSIONAL ACTIVITIES

- Student Editorial Board Member, Journal of the American Medical Informatics Association (JAMIA) | *2022-2023*
- Awarded Provost Fellowship | *2023*
- Peer Review (Journal of Biomedical Informatics, Bioinformatics, BMC Bioinformatics, Intelligent Systems for Molecular Biology, AMIA Informatics Summit)
- Co-organized & coordinated discussion group on Symbolic AI and Knowledge Graphs with 20 participants and 6 guest speakers | *2022*