Avneet Kaur

Cloud Engineer | Novo Nordisk

Kobenhavn Ø, Capital Region

+4531833156

avneetreen@gmail.com

avneetreen.github.io

linkedin.com/in/avneetkaur97

PROFESSIONAL SUMMARY

- Experienced in solution design and infrastructure implementation, data pipelines, managing large scale datasets, and developing scalable solutions for data processing and analysis. Skilled in collaborating with cross-functional teams to deliver high-quality, multilingual AI products. Demonstrated expertise in Python, NLP, and large-scale data engineering, with a strong publication record at top-tier venues (AAAI, ISWC, Nucleic Acids Research). Committed to advancing the capabilities of AI systems to serve global, diverse audiences
- Hold an M.Sc. in Computer Science from the University of Copenhagen, equipped with a strong foundation in machine learning, data science principles, and methodologies.

GOOGLE SCHOLAR

https://scholar.google.com/citations?user=Qoun_HcAAAAJ&hl=en&authuser=1

SKILLS

Programming Languages/ Frameworks

Python (pytorch, tensorflow, langchain, ollama, spacy, hugging face), Typescript, Java, SQL, Shell Scripting, C/C++

Data Engineering

AWS Services, Azure Cloud, Databricks, CI/CD, Kubernetes, SQL (MySQL, Oracle, SQL Server, Redshift, MSSQL)

Technical Skills

Github Actions,AWS Services ,CI/CD Pipelines,Azure Dev Ops,Azure Cloud Services ,Databricks

Machine Learning & NLP

EDUCATION

Sept 2019 - Jan 2022

Masters of Science, Computer Science

University of Copenhagen

Copenhagen, Denmark

Aug 2014 - Aug 2018

Bachelor of Technology, Computer Science and Engineering

Indraprastha Institute of Information Technology

Delhi, India

WORK EXPERIENCE

November 2024 - Present

Cloud Engineer / Data Engineer

Novo Nordisk, [Research and Early Development, Gloabal Data and AI]

- Designed and implemented platform architecture to streamline data ingestion from contract research organizations, improving integration and accessibility.
- Developed robust data pipelines leveraging Python and AWS to process large-scale datasets for analysis and machine learning model training.
- Oversaw data integration and onboarding processes within the Novo Nordisk Enterprise Data Lake, enhancing data quality and ease of access.
- Automated processes for data governance and compliance, ensuring adherence to regulatory standards and maintaining data integrity.

Aug 2021 - Jan 2022

M.Sc. Thesis Student - Development of Positron Emission Tomography pipeline

Neurobiology Research Unit

- Developed an efficient data pipeline for the pre-processing of Positron Emission Tomography (PET) images, improving the accuracy and reliability of subsequent analysis.
- Conducted in-depth research and experimentation to investigate and test methods for testing the robustness of PET image analysis algorithms.
- Contributed to the development of Nipype, an open-source Python library for analyzing MRI/PET data, by implementing new features and resolving issues.

July 2021 - August 2021

Machine Learning Intern (Natural Language Processing)

BlackRock, Inc.

- Developed a machine learning classifier using natural language processing techniques to classify bond reports as environmentally friendly, improving the efficiency and accuracy of the analysis process.
- Played an active role in enhancing the in-house NLP library for analyzing bond reports, contributing to the development of more robust and effective tools for data analysis.

July 2022 - Jan 2021

Machine Learning Intern (Deep Learning & Computer Vision)

Statumanu Aps, Copenhagen

- Developed and deployed deep learning models for detecting optic disk and retinal vein pulsations
- Documented all development activities in compliance with Medical Device Regulation, ensuring accuracy, traceability, and adherence to regulatory standards.

Data pipeline automation, NLP model development, multilingual data processing, RLHF, data quality assurance

Research & Analysis

Statistical analysis, model evaluation, bias measurement, multilingual data collection and annotation

TECHNICAL COURSEWORK

- Algorithms & Data Structures
- Design and Analysis of Algorithms
- Large Scale Data Analytics
- Machine Learning
- Data Mining
- Artificial Intelligence
- Database Management Systems
- Signal and Image Processing
- Natural Language Processing
- Probability and Statistics
- Information Retrieval
- Network Science

PUBLICATIONS

- Multidimensional Analysis of Trust in News Articles, AAAI , Paper Link
- That's Interesting, Tell Me More! Finding Descriptive Support Passages for Knowledge Graph Relationships, **ISWC, Paper Link**
- FlavorDB: a database of flavor molecules, Nucleic Acid Research, Paper Link

AREA OF INTERESTS

- Machine Learning
- Natural Language Processing
- Artificial Intelligence for social good
- Data & Cloud Engineering
- Software Engineering

AWARDS & ACHIEVEMENTS

- Grace Hopper Celebration India Scholar 2018
- Best Paper Award 2018, International Semantic Web Conference
- One of the 8 out of 200 teams selected for the Rails Girls Summer of Code Program 2018
- Gave a talk at PyCon Thailand 2019: Demystifying Conversational AI with Python
- Gave a talk at PyCon Portugal 2023: on Building Dainty Dashboards
- Gave a talk at Pycon Spain 2023: On Automated Testing of data pipelines using BDD.

PUBLICATIONS

That's interesting, tell me more! finding descriptive support passages for knowledge graph relationships

International Semantic Web Conference 2018

FlavorDB: a database of flavor molecules

Nucleic Acids Research 2018

Multidimensional Analysis of Trust in News Articles (Student Abstract)

AAAI 2020

MULTILINGUAL DATA & AI PROJECTS

The Language Effect

Cohere for AI, Expedition Aya

- Development of a multilingual framework for assessing political bias in large language models (LLMs) across multiple languages.
- Design and implementation of methodologies for language-sensitive opinion retrieval and inclusivity analysis.
- Achievement of recognition as the most innovative project during Cohere for AI's Expedition Aya program.

Multilingual climate change chatbot

Cohere for AI, Expedition Aya

- Developed a robust, multilingual chatbot utilizing Lang-graph and Reinforcement Learning with Human Feedback (RLHF) to improve accessibility and understanding of climate change issues.
- Recognized as the winning project of Expedition Aya, highlighting innovation and impact in addressing climate-related challenges.
- Contributed to raising awareness about climate change through the deployment of an inclusive and accessible conversational AI model.

Multidimensional Analysis of Trust in Indian News Articles (NLP)

IBM Research, India

- Conducted multidimensional analysis of trust in Indian news articles, utilizing advanced natural language processing techniques to assess selectivity of facts, journalistic assessment, and topic selectivity.
- Published research findings at the Association of Advancement of Artificial Intelligence (AAAI) 2020, contributing to the field of trust analysis in news articles.
- Leveraged textual content analysis to evaluate the selectivity of facts presented in news articles, providing insights into potential biases and misrepresentations.

Delineating field boundaries using sentinel-2 imagery (Remote Sensing & Computer Vision, Machine Learning)

DHI Gras, Copenhagen

- Implemented a deep learning model using convolutional networks to accurately delineate field boundaries from Sentinel-2 imagery, reducing human intervention and potential errors in agricultural records.
- Conducted extensive data pre-processing and augmentation to ensure the model's robustness and generalizability across different agricultural landscapes.
- Validated the model's performance through rigorous testing and evaluation, achieving a precision rate of 90% and significantly reducing the labor-intensive nature of field boundary delineation.