FAIZEL KHAN

Minneapolis, MN | faizelkhan11@gmail.com | 573-200-4012 | www.linkedin.com/in/faizelkhan/ | Google Scholar Page

SUMMARY

Experienced Software Engineer with a track record of 5+ years in delivering software solutions. Strong expertise in Machine Learning Model Management, developing full-stack software and ETL pipelines, and hands-on experience deploying applications in production. Proven track record of utilizing complex data sets to drive decisions and provide valuable insights.

SKILLS

Languages: Python, SQL, Javascript, Git
Framework: Django, Vue.js, Node.js, Bootstrap, REST API
AI Tools: Sklearn, Pandas, Pytorch, Tensorflow
Dev Tools: Github, Unittest, Airflow, Kafka, Zookeeper
DB/Cloud: PostgreSQL, NoSQL, AWS (EC2 & S3)
Analytics: Tableau, Matplotlib, W&B, JIRA, Jupyterlab

CERTIFICATIONS

Introduction to TensorFlow, Coursera, Jun 2022
Good Clinical Practice & Research Protection, CITI, Jan 2022
FERPA Training, UMN, Jan 2022
NLP with TensorFlow, Udemy, Sep 2022
HIPAA Training, UMN, Jan 2022
Statistical Learning, Stanford University, Sep 2017

EXPERIENCE

Data Scientist Dec 2021 – May 2023
University of Minnesota Medical School (Minneapolis, MN) **Key Skills**: Tableau, Pandas, API, SQL, NoSQL

- Partnered with cross-functional teams to develop data-driven solutions for various medical research studies.
- Optimized data pipelines by analyzing electronic health records (EHR) data, revealing key patterns and trends.
- Co-authored two research papers, contributing to the analysis, interpretation, and visualization of data using Tableau.
- Provided data analysis consultation to diverse medical research teams, facilitating effective survey execution, data collection, and performing analysis.

Data Science Intern May – Aug 2022
Boston Scientific (Minneapolis, MN) **Key Skills**: NLP, Pytorch, AWS (EC2 & S3), Vue.js

- Fostered a cross-functional collaboration with Regulatory Affairs to devise a solution for expedited document retrieval.
- Developed a semantic search engine app for document retrieval, significantly reducing response time to federal agencies.
- Executed an efficient ETL pipeline for refining the text corpus used in the NLP model, utilizing pandas, NLTK, and Spacy.
- Built and deployed a full-stack web application on AWS, ensuring scalability and reliability in a production environment.

Lead Software Engineer Jun 2018 – May 2021
BioKyowa Inc (Cape Girardeau, MO) **Key Skills**: Django, JIRA, API, Airflow, Kafka, PostgreSQL

- Architected a Django-based full-stack web application, enhancing data workflows for manufacturing operations.
- Enhanced analytics capabilities and saved 3,000 annual work hours by designing an ETL pipeline with Airflow and Kafka.
- Directed a team of junior software engineers, delivering high-quality and scalable applications via Agile methodologies.
- Devised and implemented technical solutions tailored to customer requirements and product objectives using JIRA.
- Designed an AI-powered chatbot as a proof-of-concept using NLP techniques to respond to employee queries effectively
- Saved $200k annually in overtime pay by leading initiatives to introduce Learning and Work Management systems.
- Negotiated contracts, securing up to 30% cost reduction in annual agreements with vendors like Monday.com.

RESEARCH PUBLICATIONS

Fritsma T, et al. Factors Associated With Health Care Professionals' Choice to Practice in Rural Minnesota. JAMA Netw Open. 2023;6(5):e2310332. https://doi.org/10.1001/jamanetworkopen.2023.10332

- Partnered with the Minnesota Department of Health to execute a cross-sectional study of 32,086 healthcare professionals.
- Instrumental role in data acquisition, analysis, and interpretation, contributing significantly to manuscript revisions.
- Unveiled key findings highlighting the urban-centric work preference of most healthcare professionals in Minnesota.
- The study underscored the necessity for strategic attention to factors influencing the choice to practice in rural areas.

Wothe, J. K., et al. Academic and Wellness Outcomes Associated with use of Anki Spaced Repetition Software in Medical School. Journal of Medical Education and Curricular Development. https://doi.org/10.1177/23821205231173289

- Partnered with a medical research group to assess the correlation between the Anki app utilization and learner outcomes.
- Played a crucial role in the analysis, and interpretation of data, along with the revision of the manuscript.
- Uncovered findings that daily Anki users among medical students showed a notable increase in the USMLE Step 1 score.
- Revealed a correlation between Anki usage and improved sleep quality, while other wellness or extracurricular metrics remained unaffected.

EDUCATION

M.S. Computer Science (***Focus:*** *Artificial Intelligence*) **Graduated**: May 2023
University of Minnesota, Minneapolis, MN **GPA**: 3.87

B.S. Computer Science (**Minor**: Mathematics) **Graduated**: May 2018
Southeast Missouri State University, Cape Girardeau, MO **GPA**: 3.77

RESEARCH PROJECTS

Meta-Analysis of Differentially Private Fine-tuned Language Model Feb – May 2023
NLP Research **Key Skills**: Differential Privacy, W&B, Pytorch

- Investigated privacy in Large Language Models, preserving accuracy and cost reduction with parameter-efficient methods
- Successfully showcased that employing parameter-efficient techniques like BitFit for private finetuning of LLMs could match the accuracy level of non-private finetuned LLMs, thus overcoming the privacy-accuracy tradeoff.

Measuring and improving supervision in the clinical learning environment Jan 2022– May 2023
Medical Research **Key Skills**: Pandas, SQL, Matplotlib

- Pioneered a research initiative on the influence of resident-doctor interactions in medical education, utilizing EHR data where previously no validated research existed.
- Devised and deployed a statistical learning model, generating pivotal insights from data, such as interaction patterns, feedback quality, and supervision levels, effectively improving the clinical learning environment.

Sep – Dec 2022
Choosing the right experts for BioMedical Question Answering (BioQA) tasks
NLP Research **Key Skills**: LLM, W&B, Pytorch, NVIDIA A100 GPU

- Build a sparsely gated Mixture of Expert based QA methods aimed to simplify the diversity in biomedical questions.
- Exploited data augmentation and parameter tuning techniques to compete with the current state-of-the-art model.

Sep – Dec 2022
ML-driven Short Video Streaming
AI Research **Key Skills**: Pytorch, Reinforcement Learning, PPO

- Build a Reinforcement Learning (RL) approach to decide which video to download, at what quality, and when to pause.
- Improved scores by 87% - 164% compared to the baselines, using an actor-critic RL algorithm, PPO, with action masking.

KEY GRADUATE CLASSES

Artificial Intelligence
- Concepts Learned: Adversarial Search and Optimization, Markov Chain, Ensemble Learning, Reinforcement Learning
- Project: Implemented Symmetric Delete Spelling Correction (SDSC) algorithm to create a quicker spell checker.

Deep Learning
- Concepts Learned: Deep Neural Networks, Autoencoders, Vision Detection and Segmentation, GAN, Transformer model
- Project: Incorporated actor-critic reinforcement learning to improve short video streaming.

Natural Language Processing
- Concepts Learned: Sep2Seq, RNN, LSTM, Evaluation Metrics, Natural Language Generation, Prompt Engineering
- Project: Improved Q&A systems for biomedical tasks exploiting data augmentation and parameter tuning techniques.

Dirty, Private, and Federated Data
- Concepts Learned: Data Cleaning, Deep Entity Matching, Differential Privacy, Federated Systems, Data Error Detection
- Project: Implemented Differential Privacy for Large Language Models using parameter-efficient techniques.

Software Engineering
- Concepts Learned: Software Design and Architecture, System Testing, Data flow coverage, Regression testing
- Project: Build YAML workflow configurations and system tests to perform code coverage on an Android app.