Dr. Shams Basir

AWS Certified ML Engineer | shamsbasir@gmail.com | (412) 608-3435 | https://www.linkedin.com/in/shamsulhaqbasir/

AWS Certified Senior Machine Learning Engineer specializing in Generative AI with 6+ years of experience in machine learning, deep learning, and AI product development. Expert in **LLMs, Generative AI (GANs, VAEs, Diffusion Models), NLP**, and building scalable AI-driven solutions. Proficient in **Python, R, MLFlow, Docker, MLOps, LLMOps, CI/CD, Git**, and **data science frameworks** such as Scikit-Learn, NumPy, and Pandas. Skilled in **training, fine-tuning, and deploying machine learning models** on AWS cloud platforms, and creating user-centric dashboards using **Streamlit**. Demonstrates a proven track record of **developing and deploying innovative chatbots and conversational agents**, including LLM-powered solutions for clinical and commercial applications. Delivers scalable, user-aware AI systems that streamline workflows, enhance customer engagement, and generate significant business value.

PROFESSIONAL EXPERIENCE

Quinstreet, Remote
Senior Machine Learning Engineer | LLMs, OpenAI, APIs, Voiceflow, JavaScript, HTML April 2024 – Present

- As the sole Full-Stack Machine Learning Engineer and AI Product Specialist on QuinStreet's Product Team, **I lead the end-to-end development lifecycle of Large Language Model (LLM)-powered AI Agents and Chatbots**, driving innovation and impact at every stage—from inception and EDA to POC, MVP, deployment, and monitoring.
- **Architect re-engagement solutions that harness voice, SMS, and email** assistants to enhance customer interactions, optimize monetization, and streamline operations.
- **Integrate AI agents with backend APIs and services** to enable personalized customer experiences, implement dynamic monetization strategies, and ensure seamless callback scheduling for enhanced user interactions.
- **Designed, developed and deployed two LLM-powered chatbots** tailored for personal loan customers, boosting engagement by 20% and generating an additional **$100,000 USD monthly in revenue**.
- **Architected and optimized GPT-4o-mini and LLM-based chatbots** with advanced intent classification, logic routing, and automatic language detection, enabling accurate, multi-language interactions. Leveraged sophisticated prompt engineering techniques, such as few-shot learning and chain-of-thought reasoning, to minimize hallucinations and improve response reliability.**Integrated a Retrieval-Augmented Generation (RAG) framework** with advanced query optimization, delivering high-accuracy responses for user queries and launching robust performance tracking systems to drive continuous improvements.
- **Developed Proofs of Concept (PoCs) for SMS (via Twilio) and email re-engagement**, enabling multimodal customer interactions through a conversational AI assistant capable of autonomous inquiry handling, with escalation to live agents as needed.
- **Implemented full API integrations and data-passing workflows**, creating custom, responsive UIs using CSS, HTML, and JavaScript, with embedded monetization and automated recommendation features to optimize user conversion and engagement.
- **Integrated guardrails for security and constitutionalized design**, protecting against prompt injection and ensuring that chatbot interactions remained strictly within intended topics.
- **Enhanced AI-powered application functionality through real-time dashboards** for data insights and monitoring, ensuring a secure, scalable, and high-performance user experience across multiple communication channels.

Bristol Myers Squibb, Remote
Senior Manager, Data Science | LLMs, Pytorch, Python, Docker, AWS, XGBoost, Scikit-learn Jan 2023 – April 2024

- Led NLP and Generative AI initiatives, specializing in question-answering chatbot development and model fine-tuning using PyTorch and Hugging Face, with recognition for collaborative leadership and innovative problem-solving.
- **Clinical Trial Design Intelligence Chatbot (ChatCTDI):** Developed an advanced question-answering chatbot for the Competitive Trial Design Intelligence platform, generating **$5 million in annual savings**. Leveraged retrieval-augmented generation (RAG) and a novel question-routing system for precise intent recognition. Engineered cross-encoding models for real-time inference on an MLOps platform and implemented re-ranking pipelines for diverse clinical data sources. Developed robust prompt engineering to mitigate hallucination, deployed via a user-friendly Streamlit UI on Davinchi, and utilized OpenAI's GPT-4, LangChain, and Weights & Biases for model orchestration.
- **Line of Therapy (LoT) Adjudication Solution:** Collaborated with stakeholders to design a Large Language Model-driven adjudication tool for cancer therapy lines, **achieving 93% accuracy**. Integrated clinical rules with treatment knowledge,

deploying at scale using GPT-4, LangChain, AWS SageMaker, and Streamlit. **Filed for patent on this high-impact solution (pending)**.

- **AI-Powered Data Review Assistant:** Designed an AI agent with LLMs to identify and flag data inconsistencies in structured databases, enhancing data review workflows. Deployed an interactive Streamlit dashboard with PyTorch, LangChain, OpenAI, and Hugging Face, streamlining data management.
- **Predictive Models for Patient Risk Assessment:** Partnered with clinicians to develop ML models for patient infection risk assessment, focusing on multiple myeloma. Utilized XGBoost, SHAP, and Scikit-learn for model development, with specialized data preprocessing and EDA to drive actionable insights.
- **Early Cytokine Release Syndrome (CRS) Detection:** Contributed to a digital health project by developing an early detection algorithm for CRS from wearable data, using BI-LSTM models and custom indices based on patient data variations, enhancing patient monitoring capabilities.
- **Cardiac Abnormality Detection:** Supervised senior students in developing ECG-based cardiac abnormality detection models, providing expertise in data preprocessing and model implementation.

Bristol Myers Squibb, Remote
Data Scientist | Python, Pytorch, SQL, AWS June 2022 – Aug 2022

- **Clinical Trial Analysis:** Collaborated with a multidisciplinary team on Upper Gastrointestinal and prostate cancer trials, focusing on ADaM and SDTM data formats to analyze patient outcomes and improve communication between data science and clinical teams.
- **Data Factory for Cancer Data:** Developed a data factory using AWS Redshift, Python, and Pandas, automating analysis workflows for Upper Gastrointestinal cancer data. Created visualizations and presented findings to a broad audience at BMS.
- **Learning Heterogeneous Treatment Effects:** Investigated heterogeneous treatment effects using deep learning on time-to-event data, applying advanced preprocessing and quality checks to derive insights on treatment-specific conditional hazard functions.

University of Pittsburgh, Pittsburgh, PA
Graduate Student Researcher (Machine Learning) | Python, Pytorch, Numpy, MPI, Cuda Aug 2018 – June 2022

- **Physics-Informed Machine Learning:** Led groundbreaking research in physics-informed neural networks (PINNs) and domain decomposition methods, publishing in top-tier journals. Developed innovative frameworks and solutions to enhance problem-solving in computational physics.
- **Failure Modes in PINNs:** Investigated failure modes of PINNs, uncovering scale-related challenges and proposing solutions such as reduced derivative orders and auxiliary equations. Published findings in the *Journal of Communication in Computational Physics* and presented at the American Physical Society (APS).
- **Physics and Equality Constrained Artificial Neural Networks (PECANNs):** Engineered a framework aligning neural networks with physics laws and multi-fidelity data, improving performance across various computational physics problems. Results published in the *Journal of Computational Physics*.
- **Domain Decomposition with Neural Networks:** Explored domain decomposition methods using AI to solve partitioned domains efficiently, developing parallel numerical methods with MPI. Proposed a neural network-based approach currently under review for journal publication.

EDUCATION

University of Pittsburgh, Swanson School of Engineering, Pittsburgh, PA Aug 2018 – April 2023
Doctor of Philosophy – Mechanical Engineering specializing in Physics-informed Machine Learning (ML)

- **Coursework: GPA 4/4** Parallel Computing (MPI, CUDA, OpenACC), Numerical Methods in Scientific Computing, Linear and Complex Analysis, Optimization Methods, Data Structures, Advanced Fluid Mechanics, Reduced Order Modelling, Transport Phenomena in Nano-to-micro scale.

Carnegie Mellon University, Pittsburgh, PA Jan 2020 – Jan 2022
Visiting Student - Machine Learning

- **Coursework: GPA 4/4** Machine learning 10601, Deep Learning 11-785, Machine Learning with Large Data Sets 10-605

Middle East Technical University, Ankara, Turkey Sept 2012 – Aug 2017
Bachelor of Science – Aerospace Engineering with a double major in Civil Engineering

- **Coursework: GPA 3.5/4** Calculus and Analytic Geometry, Object Oriented Programing with C/Fortran, Differential Equation, Advanced Statistics and Probability,Calculus of Functions of Several Variables, Heat and Mass Transfer, Fluid Dynamics

SKILLS & CERTIFICATIONS

Certifications: AWS Certified Machine Learning, AWS Cloud Practitioner, Fundamentals of Deep Learning for Multi GPUs, Neural Networks and Deep Learning, Generative AI with Large Language Models.
Programming Languages: Python, R, SQL, Matlab, C/C++, Java, Bash
Data Exploration and Transformation: Pandas, Numpy
Visualization: Matplotlib, seaborn, Plotly
AI and Machine Learning: scikit-learn, TensorFlow, Keras, PyTorch, XGBoost
Natural Language Processing (NLP): Transformers, Accelerate, bitsandbytes, trl, TGI, PyTorch, SparkNLP
GUI and Front-End Development: Streamlit
Cloud Platforms: AWS, Microsoft Azure, Domino, and Databricks
MLOPs tools: Weights & Biases, MLFlow, CI/CD, GitHub Actions, Docker
Libraries & Tools: LangChain, Semantic Indexing, Vector Database(s), Chroma, OpenSearch, FAISS
Parallel Computing: pySpark, MPI, SIMD, Horovod
Soft Skills: Cross functional, innovative, strategic, communication, decision making, collaboration, agile, responsible.

SUPPLEMENTARY PROJECTS

- **Custom Deep Learning Package similar to PyTorch:** Designed a lightweight deep learning package in Python, similar to PyTorch, incorporating an autograd library for building MLP, CNN, RNN, and GRU networks from scratch. GitHub
- **Attention-Based End-to-End Speech-to-Text Model:** Designed an attention-based end-to-end DNN for speech-to-text transcription using RNNs, CNNs, and Dense Networks, implemented with PyTorch. GitHub
- **Physics and Equality Constrained Neural Networks with Augmented Lagrangian Framework:** Developed a neural network framework integrating physics laws and user constraints using an augmented Lagrangian framework, leading to a top-cited journal publication in 2022. GitHub
- **Fine-Tuning FLAN-T5 using Reinforcement Learning (RL):** Fine-tuned FLAN-T5 with Reinforcement Learning (PPO) and Parameter-Efficient Fine-Tuning (PEFT) to generate less-toxic dialogue summaries, utilizing the META hate speech classifier for reward modeling. GitHub
- **Fine-Tuning FLAN-T5 for Dialogue Summarization:** Fine-tuned FLAN-T5 for dialogue summarization using both full fine-tuning and PEFT, and evaluated model performance using ROUGE metrics. GitHub
- **Mitigating Failure Modes of Physics-Informed Neural Networks (PINNs):** Tackled challenges in solving partial differential equations (PDEs) with PINNs, introducing methods to navigate non-convex loss landscapes and mitigate gradient contamination. GitHub
- **Time Series Stock Market Prediction with Conv-Bi-LSTM:** Implemented a time series model using Conv-Bi-LSTM for stock market prediction, employing advanced feature engineering techniques. GitHub

PUBLICATIONS

Preprints:

1. Basir, Shamsulhaq, and Inanc Senocak. “A Generalized Schwarz-type Non-overlapping Domain Decomposition Method using Physics-constrained Neural Networks.” *arXiv preprint arXiv:2307.12435* (2023).
2. Basir, Shamsulhaq, and Inanc Senocak. “An Adaptive Augmented Lagrangian Method for Training Physics and Equality Constrained Artificial Neural Networks.” *arXiv preprint arXiv:2306.04904* (2023).

Journal:

3. 3. Basir, Shamsulhaq. “Investigating and Mitigating Failure Modes in Physics-Informed Neural Networks (PINNs).” Communications in Computational Physics 33.5 (2023): 1240–1269. ISSN: 1991-7120. DOI: https://doi.org/10.4208/cicp.OA-2022-0239.
4. 4. Basir, Shamsulhaq, and Inanc Senocak. “Physics and Equality Constrained Artificial Neural Networks: Application to Forward and Inverse Problems with Multi-fidelity Data Fusion.” Journal of Computational Physics (2022): 111301. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2022.111301.

Conference:

5. Basir, Shamsulhaq, and Inanc Senocak. “Critical Investigation of Failure Modes in Physics-informed Neural Networks.” AIAA SCITECH 2022 Forum. DOI: 10.2514/6.2022-2353.

TALKS AND PRESENTATION

Invited Conference Talks:

- Navigating the Landscape of Large Language Models (LLMs) for Clinical Research, BioTechX - US, Sept. 2023
- Physics-informed and Equality-constrained Artificial Neural Networks with Applications to Partial Differential Equations and Multi-fidelity Da*ta Assimilation*, American Physics Society (APS), Nov. 2022

Invited Seminar Talks:

- Fixing Physics-informed Neural Networks (PINNs), Brown University, May 2023
- Physics Mining: Teaching Machines To Obey Physics, University of Pittsburgh, Dec. 2022
- Physics and Equality Constrained Artificial Neural Networks, Brown University, Aug. 2022
- Physics and Equality Constrained Artificial Neural Networks, Alan Turing Institute, June 2022

Contributed Conference Presentations:

- Efficient and Robust Training Strategies for Physics and Equality Constrained Artificial Neural Networks, American Physics Society (APS), Nov. 2022
- Critical Investigation of Failure Modes in Physics-informed Neural Networks, AIAA SCITECH, Jan. 2022
- Parallel solution of Partial Differential Equations on Binarized Octrees. American Physics Society (APS), Nov.2021.

AWARDS AND HONORS

- Research Assistant of the Year 2021-2022, Mechanical Engineering Dept., University of Pittsburgh.
- Arts & Sciences Graduate Student Organization Travel Grant Recipient, Cycle 4, 2022.
- Merit-Based Undergraduate Scholarship, Middle East Technical University, 2013-2017.
- Undergraduate Honors and High Honors, Middle East Technical University, 2013-2017.

PROFESSIONAL SERVICES

JOURNAL REVIEWER

- Journal of Computational Physics (JCP), Communication in Computational Physics (CiCP), Journal of Fluids Engineering (JFM), Engineering Structures, Journal of Computing and Information Science in Engineering

CONFERENCE REVIEWER

- Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML)

UNIVERSITY SERVICES

- Volunteered as a Judge for the Swanson School of Engineering Design EXPO.

VOLUNTEER WORKS

Fight Flow Academy, Raleigh, NC Nov 2024 – Present

Volunteer MMA Instructor (Boxing, Muay Thai, Grappling)

- Teach self-defense, health, and discipline to both children and adults, focusing on building confidence and resilience.

University of Pittsburgh, Pittsburgh, PA Sept. 2022 - Jan. 2023

Volunteer Mentor

- Mentored incoming graduate students, providing support for navigating academic and campus life.

Fourth River Solution - 4RS, Pittsburgh, PA Nov. 2022 - Jan. 2023

Volunteer Consultant

- Conducted healthcare industry research to identify emerging trends and business opportunities.
- Estimated market potential and performed competitive analysis for strategic decision-making.
- Provided advisory services on regulatory compliance, exit strategies, and market positioning for startups.
- Developed SWOT analyses and delivered clear, actionable recommendations to stakeholders.