Shubbham Gupta

Dublin, Ireland | +353899586207 | shubbham.gupta28@gmail.com | www.linkedin.com/in/shubbhamgupta
Stamp 1G | Portfolio: https://shubbhamgupta.netlify.app/

OBJECTIVE

PhD-level AI and Data Science professional with strong software engineering fundamentals and hands-on experience in building robust, real-time analytics systems. Specialized in Bayesian modeling, LLM-powered agents, cloud infrastructure (AWS/GCP), and agentic workflows using LangChain and HuggingFace. Strong communicator and team player with a history of delivering impactful, stakeholder-aligned solutions in health tech, finance, and enterprise automation.

SKILLS

- **Programming & Databases:** Python (OOP, SOLID, PEP8), R, C++, Java, SQL, MySQL, PostgreSQL, MongoDB
- **Data Analysis & Visualization:** Power BI, Tableau, Plotly, Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost
- **Cloud, Big Data & Development Tools:** AWS, Azure, GCP, Hadoop, Spark, Hive, Jupyter Notebook, RStudio, Anaconda, GitHub, Git, MS Office
- **AI/ML & LLM Tooling:** HuggingFace, FinBERT, LLaMA, LangChain, LangGraph, OpenAI API, Transformers, Prompt Engineering, VectorDBs (FAISS, Pinecone)
- **Automation & Workflow Tools**: YAML, CI/CD, GitHub Actions, Workflow Orchestration, Workflow mapping, Requirement gathering, Stakeholder alignment
- **Agent Frameworks & MLOps:** Agentic AI Systems, Modular AI Agent Design, Retrieval-Augmented Generation (RAG), LLMOps, ML Monitoring, Data Pipelines

WORK EXPERIENCE

Data Scientist (Full Stack), Marketing Analytics Startup ADAPT Centre, Trinity College Dublin *July 2025 - Present*

- Designed and implemented **Marketing Mix Modeling** (MMM) and **causal attribution** pipelines (Meridian, custom econometric models) to estimate consumer journeys and channel ROI.
- Integrated and merged marketing data sources (Google Analytics, Meta/Facebook Ads, YouTube, and internal campaign data) into unified pipelines for **attribution, spend optimization, and insight generation**.
- Delivered actionable insights to **improve marketing budget allocation** and reduce inefficiencies from misleading/false advertising signals.
- Built **automated workflows** for ingestion, transformation, and analysis of large-scale campaign data (Pandas/NumPy, FastAPI, Docker, AWS S3/Batch).
- Developed **reproducible testing and validation strategies** (PyTest, CI/CD) ensuring robustness of attribution models and data pipelines.
- Collaborated in Agile teams, translating business requirements into **data science workflows**, deploying models into production, and **enabling visualization/decision-making via dashboards** (Plotly, Vega, React).

PhD researcher, Nutrition, Biomarkers, & Health lab, University College Dublin *Sep 2020 - Apr 2025*

- Led the **end-to-end development** MetaboVariation, a novel **Bayesian generalised linear model** to capture intra-individual variations in metabolite levels across repeated measurements, contributing to personalized healthcare.
- Delivered intuitive Shiny dashboards to simplify complex statistical insights, demonstrating strong **stakeholder-centric design thinking**.
- Collaborated cross-functionally with biostatisticians and clinicians across Europe, honing **interdisciplinary communication** and **team leadership**.
- Recognized with “Best Poster Award” at a national conference, showing **presentation** and **scientific storytelling skills**.
- **Authored peer-reviewed publication** in Metabolites and contributed to open-source packages.

Data Science and Artificial Intelligence research intern, Novartis Ireland Limited *Jun 2021 - Sep 2021*

- Built a **python-based survival modeling pipeline** for analyzing **real-world data** to identify key predictors of ASCVD outcomes using Cox and parametric survival models.
- Conducted survival analysis on over 40,000 patients, using clinical expertise to **translate business objectives** into robust data pipelines.
- Advocated **code quality**, **modularity**, and **documentation best practices** through SOLID and clean coding principles.
- Communicated findings to stakeholders across clinical, tech, and management teams, demonstrating **stakeholder management**, **problem-solving**, and **impact-driven mindset**.

Research Fellow (Short-term), University College Dublin *Jan 2021 - Jun 2021*

- Developed the Sheepdog Algorithm, **a bio-inspired metaheuristic feature selection tool** inspired by sheepdog herding behavior that **reduces dimensionality** by balancing low feature count and high accuracy.
- Led benchmarking and performance analysis with cross-validation and A/B testing, collaborating with domain experts for actionable deployment.
- Presented research to 200+ professionals, highlighting **public speaking**, and **knowledge transfer** skills.

EDUCATION

University College Dublin Dublin, Ireland
PhD in Data Science *Sep 2020 - Apr 2025*

- Specialized in predictive modeling, scalable ML, and feature selection techniques. Engaged in advanced research, including the development of novel predictive models for metabolic disorders and innovative feature selection techniques. Contributed to interdisciplinary teaching and collaborated internationally, leading to a publication and award.

University College Dublin Dublin, Ireland
MSc Data and computational science *Sep 2019 - Aug 2020*

- Conducted research on survival analysis in healthcare, alongside comprehensive coursework in predictive analytics, Bayesian analysis, and machine learning. Acquired proficiency in multiple programming languages and statistical tools, including R, Mathematica, and Fortran.

Guru Gobind Singh Indraprastha University New Delhi, India
B.Tech, Computer Science and Engineering *Sep 2015 - Aug 2019*

- Gained a solid foundation in computer engineering, covering operating systems, web and android development, big data, and programming languages like C++, Java, and Python. Developed skills in database management and software development.

RELEVANT PROJECTS

- *AI-Powered Financial Sentiment Agent | https://github.com/shubbham28/financial-agent*
  - Developed **modular AI agents** using **LangChain** and **FinBERT** for real-time financial analysis integrating **Yahoo Finance, Finviz news** for sentiment analysis, with trend predictions and LLM-powered summaries.
  - Developed full-stack chatbot system using **Streamlit** (a low-code framework), enabling user-friendly interaction for multi-stock queries and trend insights.
  - **Applied AI/ML concepts** in production environments, combining traditional software engineering with LLM-powered agentic automation.
  - Implemented **LLMOps workflows**: prompt chaining, validation, and role-based prompting using Hugging Face models and OpenAI APIs.

- *Voice-Enabled Personal Branding Agent*
  - Developed a voice-interactive portfolio agent using Hugging Face Spaces and Streamlit, allowing dynamic audio queries and visual summaries of professional experience.
  - Applied **modular LLM agent design** to generate context-aware responses from structured CV data.
  - Demonstrated **user-first design thinking**, **rapid prototyping**, and **LLMOps principles** in deployment.
  - Enabled smooth real-time interaction using **speech-to-text** and **TTS models**, improving accessibility.

PUBLICATIONS

- *MetaboVariation: Exploring Individual Variation in Metabolite Levels*
- *Quantum Grey Wolf Optimization and Evolutionary Algorithms for Diagnosis of Alzheimer's Disease*
- *Bio-Inspired Algorithms for Diagnosis of Breast Cancer*
- *Modified Bio-Inspired Algorithms for Diagnosis of Breast Cancer Using Aggregation*

INTERESTS

Large Language Models, Cloud Computing, Finances, Stocks, Healthcare, Cycling, Hiking, Running, Badminton