Shubham Agarwal

+91-7002622020 | Hyderabad, Telangana, India

LinkedIn: linkedin.com/in/shubham-agarwal-b567a6103

Email: shubham.agarwal22@dmsiitd.org

Professional Summary

Data Scientist with **5+ years of experience** in AI/ML, specializing in **GenAI, NLP,** and **Computer Vision**. Skilled in deploying large language models (LLMs) and building AI-driven automation systems for insurance and healthcare. Proven track record in developing scalable AI solutions that improve efficiency, with expertise in multimodal AI and deep learning.

Technical Skills & Tools

- **Languages:** SQL, Python, VBA
- **Machine Learning & AI:** GenAI, NLP, Computer Vision, Transformers, Prompt Engineering, Neural Networks, Deep Learning, PyTorch, Langchain, LangGraph, MLOps
- **Data Engineering:** Snowflake, SSIS, Power Query
- **Data visualization tools:** Microsoft Power BI (DAX), Tableau
- **Others:** Business Analytics, SSIS, Alteryx, Azure, Snowflake, Power Automate, ETL, Advanced Excel, Docker, Git

Professional Experience

PalTech, Hyderabad
Sr. AI Engineer *(Feb 2024 – Present)*

- **ML driven Employee recommendation system:**
  - Architected a hybrid employee recommendation engine to **automate staffing for on-site** insurance claim assignments across the U.S., optimizing both precision and coverage.
  - Built a two-stage recommendation architecture:
    - **Rule-based model:** SQL-driven filtering of candidates based on hard client constraints (licenses, proximity, availability, prior project type, preferences).
    - **Probabilistic model:** Used **K-Nearest Neighbours (KNN) to identify top 100 similar employees** (based on feature vectors) when no exact matches were found, enabling flexible deployments with near-fit candidates.
  - Devised a **feature-weighted ranking algorithm** prioritizing candidates by certification, license, performance etc.
- **GenAI-Powered ClaimsGPT QA System:**
  - Developed an end-to-end claims review system using **vision-encoder** and **BERT-based transformers** to automate **image classification and damage detection**.
  - Automated image preprocessing **(OpenCV, Pillow)** and used **multimodal LLMs (Mistral, Qwen)** for text generation.
  - **Engineered a hybrid YOLO v8** and **Vision Transformer (ViT) pipeline** for object/damage detection and image classification, **achieving >95% accuracy** in identifying critical claim regions (e.g., damaged roofs, windows).
  - **Reduced manual claim processing time by ~70%** through efficient parallel processing and end-to-end automation
- **Named Entity Recognition (NER) for Claim Adjuster Comments:**
  - Built an NLP pipeline for Named Entity Recognition (NER), extracting key claims insights with **92% F1 score**.
  - Enabled automated identification of **damage indicators** and **root causes** reducing manual review time by **30%**.
- **Unstructured Data ETL Pipeline:**
  - Created a **scalable PDF parser** with **PyMuPDF** to extract text and images from 8,000+ monthly claim documents.
- **AI News Aggregator & Summarization Bot:**
  - **Launched** an AI agent ingesting 100+ AI articles/day, filtering non-technical content with heuristics and NLP.
  - Applied **LLM-based summarization** to generate concise tech summaries, reducing reading time by 80%.
  - **Integrated bot with Microsoft Teams** via webhooks and MS Graph API.
- **Car Damage Detection & Classification Model (Internal POC):**
  - Trained **YOLO v8** on 5K+ annotated images using COCO-based transfer learning and data augmentation.
- **PalGPT – Internal ChatGPT Deployment:**
  - Implemented an **organization-wide GenAI assistant** using Ollama and OpenWebUI, assisting employees with their routine tasks (document summarization, code snippets, data queries etc).

Merilytics, Hyderabad
Lead Data Scientist *(Apr 2022 – Jan 2024)*

- **Led a 5-member team** and collaborated with a leading US healthcare client to gather comprehensive requirements, ensuring accurate transformation of data into impactful and actionable insights
- Worked cross-functionally to **define problem statements**, **deploy models**, and provide data-driven recommendations.
- Developed ETL workflows using Python, SQL, and SSIS to **streamline data ingestion** and maintain data consistency.
- **Led the development of RCM suite** to monitor claims and authorizations, thereby **reducing the client's outstanding AR** by ~15% (from $40M to $34M) within a span of 1 year
- **Diagnosed gaps in billing workflows** and revenue cycle, boosting reimbursement rates by 20% points.
- Analysed clinical outcomes data to evaluate the effectiveness of treatment provided, **improving treatment efficacy for specific conditions by ~10%**
- **Enhanced billing compliance** by ensuring medical services matched physician orders and required documentation was present in the EMR, leading to a **7%-point improvement** (from 86% to 93%).
- Formulated data flow pipelines for Google Ads and analysed the Ads spend data for cost and campaign optimization
- Implemented a revenue forecasting model with 90% accuracy to project expected billing in upcoming months.

Vedanta Resources, Chhattisgarh
Assistant Manager *(Aug 2017 – May 2019)*

- **Led a 3-member team** to strategize Supply Chain Logistics, **handling raw material worth $600M annually.**
- Worked as the principal MIS analyst, collaborating with multiple business functions to implement organization's sector-wide Analytics Dashboard for Inventory Management
- **Built cost optimization model** to reduce overall TAT from 10 to 8 days and **induced annual savings of $3M**.

Internships

Takshashila Consulting, Noida *(Jul 2021 – Sep 2021)*

- **Estimated market size** and defined business case for a leading US heavy vehicle OEM to drive growth and penetration
- Analysed competition, pricing, and growth drivers to devise a **$15B revenue strategy at 10% CAGR** over 5 years.

Benori Consulting, Gurugram *(May 2021 – Jul 2021)*

- Crafted CXO data model prototype, showcasing **essential KPIs** and **valuable business insights**.
- Embedded **a chatbot** feature within Power BI reports to enhance data accessibility and user engagement.

Other Projects/ Side-hustle

- **Wedding Photo Face Clustering & Organization System** (GitHub repository link)
  - Built an automated system to organize 1000+ wedding photos by guest identity using face recognition (ArcFace embeddings) and clustering (DBSCAN), eliminating the need for manual sorting or labelling.
- **Google Play Store App Search & Review System** (GitHub repository link)
  - Developed a Django-based app search engine from a Kaggle dataset with **TF-IDF search**, **review sentiment analysis,** and a role-based moderation dashboard; deployed via Render.
- **Personal finance tracker** (GitHub repository link)
  - Created a personal finance web app using Streamlit and Google Sheets API to track and visualize expenses, savings, and investments, with budget analytics and cross-device access via PWA support.

Certifications

- **Microsoft** Certified **Power BI** Associate
- Certified **Alteryx Core Designer**
- **Lean Six Sigma Green Belt** by KPMG

Education

- Master of Business Administration (MBA), IIT Delhi *(Aug 2020 – Apr 2022)*
- B. Tech (Mechanical Engineering), NIT Silchar *(Aug 2013 – May 2017)*