VAHID SHARIFI

Data Scientist ~ Generative AI
R&D

📞 +44 77 6091 2928   ✉ vahidsharifi7676@gmail.com

🔗 /in/vahid-sharifi   ⬡ https://github.com/vahidsharifi

SKILLS

Programming Languages:	Python, R , SQL, JavaScript.
Tools and Technologies:	LangChain, HuggingFace, GCP, Scikit-learn TensorFlow, BeautifulSoup, Docker, WanDB, FastAPI, NLTK, SpaCy, Gensim, Keras, Pandas, NumPy

EXPERIENCE

2025 – Present   **Quality Assurance Specialist (AI Data Curation)**   **Mercor**
- Ensured data integrity by reviewing 1,000+ instances weekly, with 300+ gold-quality prompt–answer pairs delivered at a 92% approval rate.
- Actively identified and reported methodological issues in data pipelines, contributing research-driven solutions that improved reliability.
- Collaborated with a 50-member team, sharing quality insights and feedback that enhanced overall consistency and efficiency.

`Quality Assurance  /  Data Curation  /  Benchmarking  /  Process Optimisation`

2023 – Present   **AI Model Specialist & Evaluator (Contract)**   **Clients: Outlier, Alignerr, DataAnnotation**
- Expert in RLHF, authoring expert-level prompts and preference data for OpenAI o3/o4 and Google Gemini.
- Generated gold-standard benchmarks (GPQA, HMMT, AIME) to rigorously test AI mathematical reasoning.
- Reviewed and supervised code annotations, ensuring high-quality training data.

`RLHF  /  Benchmarking  /  Model Evaluation  /  Prompt Engineering  /  Mathematics`

2025 – 2025   **Researcher & Generative AI Consultant**   **University of Gloucestershire**
- Designed a novel qualitative research framework for AI studies, defining scope and methodology.
- Engineered and deployed a hybrid transcription system achieving >95% accuracy on noisy audio.
- Developed a multi-agent GenAI workflow automating thematic analysis of qualitative data, completing tasks 50x faster.
- Led participatory workshops to translate complex AI concepts, gathering critical data for methodology.
- Co-authored research publication, drafting methodology and discussion sections.

`Python  /  Google API  /  Generative AI  /  RLHF  /  Qualitative Analysis`

2024 – 2025   **Research Assistant (AI in Education) & AI Solutions Developer**   **University of Gloucestershire**
- Developed AI Agent tools (MLLMs) to decolonise curricula, flagging colonial imagery and analysing text.
- Engineered LLM prompting strategies to generate structured JSON outputs with 99% consistency.
- Designed human-in-the-loop evaluation framework, improving suggestion relevance by 30%.
- Quantified qualitative research using ML, enabling statistical insight extraction from unstructured expert feedback.

`Python  /  LangChain  /  OpenAI  /  LangSmith  /  HuggingFace`

2023 – 2024   **Researcher (AI in Healthcare) & NLP Engineer**   **University of Gloucestershire**
- Led AI-driven dietary planning project integrating LLMs and nutrition databases; RAG pipeline improved data accuracy by 90%.
- Compiled and modeled datasets via APIs and LLMs; applied statistical analysis for validation and QA.

`LangChain  /  HuggingFace Pandas  /  NumPy  /  TensorFlow  /  Scikit-learn  /  SciPy`

2022 – 2023   **ML/NLP Engineer**   **Stealth Startup, UK**
- Architected full-stack AI Q&A engine for YouTube, enabling semantic search of video transcripts.
- Built LLM pipeline using LangChain, FastAPI, and ChromaDB, achieving >90% relevance score on user queries.
- Managed MLOps lifecycle with Weights & Biases and Firebase for tracking and user feedback.

`LangChain  /  GCP  /  LangSmith  /  WanDB  /  Docker  /  FastAPI  /  FastAPI  /  ChromaDB`

2017 – 2022   **Co-Founder, Data Scientist**   **Novin Sim**
- Co-founded e-commerce platform; directed strategy, P&L, and supply chain for 5+ years growth.
- Engineered hybrid recommender system, increasing average order value by 15%.
- Implemented NLP pipeline for customer reviews, automating 90% of feedback analysis.
- Developed Power BI dashboards to visualize KPIs, enabling data-driven stakeholder decisions.

`Python  /  R  /  JavaScript  /  HTML  /  SQL  /  PowerBI`

ACADEMIC

2023 – 2024   **MSc, Data Science (Distinction)**   **University of Gloucestershire, United Kingdom**

| 2016 – 2021 | **BSc, Industrial Engineering** | **Iran University of Science and Technology** |

Coursework: Statistical Quality Control, Experimental Design, Business Model Development

---

Honors

National University Entrance Exam (Konkour) – Top 0.4% of 300,000 (2015)

Silver Medal, National Astronomy Olympiad – Top 40 finalists (2014)

Languages

English (IELTS 7)
Persian (Native)