Behrooz Qiassi

Consultant

3 years of experience in Data Science

Profile

Behrooz Qiassi is a highly skilled Data Scientist with a strong academic foundation, holding a Master’s degree in Data Science - Computational Linguistics from the University of British Columbia. With over three years of professional experience, he has honed his expertise in developing data-driven applications, fine-tuning Large Language Models, and building robust ETL pipelines and data warehouse infrastructures. Known for his ability to lead end-to-end projects, Behrooz thrives in fast-paced environments, delivering innovative solutions that bridge the gap between raw data and actionable insights.  
  
Throughout his career, Behrooz has made a significant impact across diverse roles. As an independent Data Scientist, he developed the Iranian Newspaper Research Assistant, a cutting-edge application leveraging Retrieval-Augmented Generation (RAG) to make historical archives searchable, complete with full-text and semantic search capabilities, graph-based visualizations, and chatbot-style interfaces. At Nu Data, he built a custom RAG solution, fine-tuned Large Language Models, and automated large-scale data collection from over 50 sources. His tenure at XYON Health saw him design a scalable data warehouse and orchestrate ETL pipelines that streamlined reporting processes by 70%. Additionally, his work with BC Stats earned him recognition, including a Project Presentation Award, for implementing a semantic search system and interactive topic modeling visualizations that enhanced user experience.  
  
Behrooz’s technical repertoire is extensive, encompassing programming languages such as Python, R, SQL, and JavaScript; database systems like PostgreSQL, MongoDB, and Redshift; and tools such as AWS, Docker, scikit-learn, Pytorch, and LangChain. His proficiency extends to advanced data orchestration with Prefect and Airflow, as well as visualization tools like Power BI, Matplotlib, and Plotly. Renowned for his innovative approach and technical acumen, Behrooz has consistently demonstrated his ability to transform complex data challenges into user-centric, scalable solutions, solidifying his reputation as a forward-thinking leader in the field of data science..

CGI experience

{sector}, Independent Data Scientist (12/24 to Present)

* Developed an application to make scanned historical Iranian newspaper archives searchable via Retrieval-Augmented Generation (RAG).
* Improved accessibility for researchers by converting scanned archives into machine-readable text.
* Enabled intuitive document retrieval through full-text and semantic search.
* Implemented graph-based search to explore relationships between documents visually.
* Built a chatbot-style assistant for natural language queries to facilitate archive exploration.

Other experience

Self-employed, Data Scientist (04/24 to 11/24)

Developed a custom Agentic Retrieval-Augmented Generation (RAG) solution for advanced analytics and chatbot interface.

* Fine-tuned Large Language Models for improved performance in semantic search and function calling.
* Automated large-scale data collection from 50+ sources using web scraping tools.

Nu Data, Data Scientist (09/23 to 03/24)

* Designed and built a data warehouse, reducing reporting time by 70%.
* Created scalable ETL pipelines integrating data from 11+ sources.
* Delivered actionable insights through dashboards for marketing, finance, and operations teams.

XYON Health, Data Scientist (Capstone Project) (05/23 to 07/23)

* Developed a new semantic search system for BC Data Catalogue, enhancing search precision by 25% and recall by 40%.
* Implemented topic modeling to create interactive visualizations, improving user experience.
* Received awards for project presentation and recognition on UBC's website.

BC Stats, Data Analyst (02/19 to 09/21)

* Automated data scraping pipelines to extract structured and unstructured data from 40+ sources.
* Developed and optimized ETL pipelines, improving data processing efficiency by 20%.

BoldZee, {job\_title} ({start\_date} to {end\_date})

* {responsibilities}

Education

{education\_entry}

Trainings and certifications

* {certs}

|  |  |
| --- | --- |
| Industry experience  Sustainability & AI Startup   * IT Services and IT Consulting * Government Statistics * Pharmaceutical Manufacturing Startup   Technical specializations   * Data Science * Retrieval-Augmented Generation (RAG) * Topic Modeling * Semantic Search * Large Language Models * Data Warehousing * ETL Pipelines * Computational Linguistics * Machine Learning   Areas of expertise   * Data Applications * Actionable Insights * Web Scraping * Dashboards and Data Visualization * Graph-based Search * Natural Language Processing * Search Systems | Environments   * Fast-paced startups * Academic capstone projects * Government projects * Independent/self-employed projects   Tools & software   * PostgreSQL * Streamlit * FastAPI * Flask * Plotly * Seaborn * Matplotlib * Power BI * Llamaindex * LangChain * Pytorch * scikit-learn * Docker * Github Actions * Data Orchestration (Prefect, Airflow) * AWS (EC2, ECS, S3, Lambda) * Redshift * Elasticsearch * Solr * Pgvector * MongoDB   Languages   * Python * CSS * HTML * Bash Scripting * JavaScript * SQL * R |

Skills summary

| Skill | Number of years | Skill level\* |
| --- | --- | --- |
| Technical skills | | |
| Python | 5 | 4 |
| R | 5 | 3 |
| SQL | 5 | 4 |
| JavaScript | 3 | 2 |
| Bash Scripting | 3 | 3 |
| HTML | 3 | 3 |
| CSS | 3 | 3 |
| PostgreSQL | 5 | 4 |
| MongoDB | 5 | 4 |
| Pgvector | 1 | 3 |
| Solr | 1 | 3 |
| Elasticsearch | 3 | 3 |
| Redshift | 1 | 2 |
| AWS (EC2, ECS, S3, Lambda) | 3 | 3 |
| Data Orchestration (Prefect, Airflow) | 3 | 3 |
| GitHub Actions | 3 | 3 |
| Docker | 3 | 3 |
| scikit-learn | 3 | 4 |
| Pytorch | 3 | 3 |
| Large Language Models | 3 | 4 |
| XGBoost | 3 | 3 |
| LangChain | 1 | 3 |
| Llamaindex | 1 | 3 |
| Power BI | 3 | 3 |
| Matplotlib | 3 | 3 |
| Seaborn | 3 | 3 |
| Plotly | 3 | 3 |
| Flask | 3 | 3 |
| FastAPI | 3 | 3 |
| Streamlit | 3 | 3 |
| Application knowledge | | |
| ETL Development | 5 | 4 |
| Data Warehousing | 3 | 4 |
| Semantic Search | 3 | 4 |
| Topic Modeling | 1 | 3 |
| Retrieval-Augmented Generation (RAG) | 1 | 3 |
| Natural Language Processing | 3 | 4 |
| Interactive Dashboards | 3 | 4 |
| Web Scraping | 5 | 4 |
| IT disciplines | | |
| Data Science | 5 | 4 |
| Data Engineering | 5 | 4 |
| Machine Learning | 3 | 4 |
| Visualization & Business Intelligence | 3 | 4 |
| Industry knowledge | | |
| Sustainability & AI | 1 | 3 |
| Pharmaceutical Manufacturing | 1 | 3 |
| Government Analytics | 1 | 3 |
| IT Services and IT Consulting | 2 | 4 |
| Other relevant skills | | |
| Project Leadership | 5 | 4 |
| Startup Environment | 3 | 4 |
| Historical Document Accessibility | 1 | 3 |

\*Skill Level: 1 = Beginner, 2 = Experienced, 3 = Advanced, 4 = Expert