Data Science Engineer Resume

Boris Dev

May 22, 2025

Tool Stack

- **Data Analytics:** SQL, Python, SQLAlchemy, Pandas, PySpark, GeoSpatial Stats & Maps, Prometheus, Grafana, Tableau, Splunk, ELK, Distributed Tracing
- **Data Pipelines:** Jenkins, Kafka, SQLAlchemy, Postgres[GIS]
- **Data Science:** Pandas, Sklearn, Numpy, Plotly, PySpark, DataBricks, Jupyter, SageMaker, PyTorch, GroundTruth
- **Langauge AI:** Embedding DBs (Weaviate, ElasticSearch), PydanticAI Graph Agents, LangChain, LangSmith, SentenceTransformers

Education

PhD in Quantitative Human Geography, at SDSU and UCSB, 2015. Data science for location referenced social science problems. Dissertation: New Metrics for Assessing Inequality using Geographic Data

Experience

EcoR1, 2025 - current, Data Science Engineer Contractor

- Building LLM AI automation tools for a biotech hedge fund to remove manual work, toil.

Nobsmed, 2024 - current, Founder

Built LLM based website, https://nobsmed.com/, for users to search for biohacking experiences found in clinical studies and Reddit comments.

- ETL: Language AI parsing of 100,000 studies and 1,000,000 Reddit comments.
- Topic modeling: Extended Bertopic using LLM for naming and classification of outliers, bertopic-easy.
- Customer discovery and user interviews.
- Experimented with fine-tuning opensource embeddings model and ML classification.
- Prompt engineering using InstructorAI.
- Embedding DBs for search (Weaviate, ElasticSearch).

AI Engineer consultant at Wolf Games, 2023

- Built an LLM based service with a new graph prompting for story generation
- Worked closely with the CTO on requirements
- Built the story generation Flask API service

AI Engineer consultant at Intuitive Systems, 2023

- NED (Name Entity Disambiguation) of vendor receipts fed into AMD's sales analytic pipeline. Used LangSmith for evaluation.

AI Engineer consultant at SimpleLegal, 2022-2023

SimpleLegal is a legal billing analytics company.

- Built the ground truth for a language model for an AI feature to flag invoice billing discrepancies
- Managed SMEs (subject matter expects) and 5 workers on the annotation pipeline of the training examples
- Built the inference server (SageMaker, PyTorch) and its product integration

Lead Analytic Endpoint Engineer at Sight Machine, 2018-2021

Sight Machine is a manufacturing analytics company.

- Built the backend engineering on biggest public facing analytic feature
- Coordinated QA process with sales and engineering
- Built company's first distributed tracing
- Built the containerized frontend engineering development environment

Lead Data Engineer at HiQ Labs, 2015-2018

HiQ Labs was a people analytics company.

- Built the scraping system and observability (Splunk)
- Led the migration from a monolith to a micro-service paradigm
- Migrated the data science team to DataBricks PySpark and microservices

Developer at Urban Mapping, 2011-2013

Urban Mapping provided geospatial analytics to Tableau.

- Formulated first map rendering and query performance metrics integrated into our CI/CD pipeline

Impactful projects

- Revived a stuck AI feature by shifting the team's focus from training data quantity to quality
- Reduced Tableau customer complaints by building a new observability system and CI/CD pre-commit metrics
- Reduced the data science team's firefighting by building a microservice architecture
- Built a gaming company's first murder mystery story generator by chaining to prompts to force consistency.

Papers and code

- LLM based taxonomy (topic modeling): bertopic-easy.
- Language AI Evaluation 101: Know your user
- Langchain PR: Causal Program-aided Language (CPAL). See Harrison Chase's Tweet.
- Work papers
- Academic papers

Non-tech fun points

- climbed Cotopaxi (21,000 ft).
- survived bodyboarding Mexpipe.
- worked with students in Medellín, Columbia to make an open-source geo clustering library (ClusterPy)
- taught kids snowboarding, as an instructor
- managed service workers, as a restaurant assistant manager
- counseled severely emotionally disturbed children