Boris Dev

Data science engineer

Skill Stack

- **Data tools:** Pandas, Sklearn, PyTorch, LangSmith, Langchain, InstructorAI, Embedding DBs (Weaviate, ElasticSearch), Numpy, Plotly
- **Engineering Tools:** FastAPI, Flask, Django, HTMX, Bootstrap, SageMaker, Docker, Jenkins, Kafka, Grafana, Splunk, Amazon 6-pagers, CI/CD/QA pipeline, Observability, Distributed tracing, NextJS, AI Agents, Prompt Engineering, HTMX, SageMaker, AWS, Ground Truth

Education

PhD in Geography, at SDSU and UCSB, 2015. Data science for social science problems.

Experience

Nobsmed, 2024, Founder

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

- ETL of 100,000 studies and 100,000 Reddit comments.
- Customer discovery and user interviews.
- Data science model to parse studies and comments.
- Full stack AI development. (Weaviate and Opensearch, Prompt Engineering)

AI Engineer consultant at Wolf Games, 2023

- The producer of Law and Order is making a murder mystery game. I designed and built the story generator. I started from the primitive concepts of organizing the narrative plot structure around a causal DAG to auto build LLM prompts.

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.

- launched a previously stuck AI feature to flag invoice billing discrepancies
- trained 8 transformer models for 8 billing flag labels

- redesigned and managed 5 workers on the annotation pipeline of the labels
- built the inference server (SageMaker, PyTorch).

Lead Analytic Endpoint Engineer at Sight Machine, 2018-2021

Sight Machine is a manufacturing analytics company.

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

Lead Data Engineer at HiQ Labs, 2015-2018

HiQ Labs was a people analytics company.

- led the migration from a monolith to a micro-service paradigm
- built the scraping system
- supported data science team

Developer at Urban Mapping, 2011-2013

Urban Mapping provided geospatial analytics to Tableau.

- CI/CD/QA pipeline
- Observability

Notable success

- For a legal ops company, I discovered their training data was wrong and rebuilt their ground truth as a team effort to launch a new language AI billing feature.
- For a gaming startup, I designed and built their murder mystery story generation service, starting from a greenfield design of the primitives.
- For a semiconductor analytics startup, I designed a new computer name disambiguation approach for AMD's sale analytics pipeline.
- For the Langchain open-source library, I contributed an experimental CPAL Chain - first causal LLM chain. Tweet by LangChain.
- For a people an startup, new microservice paradigm.
- For academics, new location based human inequality metrics in Assessing Inequality using Geographic Income Distributions
- For several companies, I built observability systems.

Papers and code

- Language AI Evaluation 101: Know your user
- Langchain PR: Causal Program-aided Language (CPAL)
- Work papers
- Academic papers