Boris Dev

Role: AI Engineer

Tools

- Code: Python, SQL
- Observability: OpenTracing, Lightstep, Splunk, Grafana
- Data Science: Pandas/PySpark, PyTorch, sklearn, numpy
- Language AI: LangChain, AWS GroundTruth & SageMaker

Notable successes

- new causal graph AI story generation
- discovery of causal structure in natural language: CPAL LLM Chain and Tweet by LangChain
- semantic inference to medical treatments, Nobsmed.com [in-progress]
- new factory line analytics backend for company's biggest public facing feature
- new observability systems for three teams
- new language AI feature for my legal ops client
- led team migration to microservice paradigm
- new geospatial human inequality metrics (PhD research)

Style

- tackle ambiguous problems
- tackle process friction
- bridge silos
- explore semantic nuances of the data

Job experience

Engineer consultant at Wolf Games, 2023

- research engineering for mystery story generation

Engineer consultant at Intuitive Systems, 2023

Intuitive System's is an early stage start-up building language AI applications for vendor management and customer complaint triage.

- canonicalization algorithm
- LangSmith evaluators

Engineer consultant at SimpleLegal, 2022-2023

SimpleLegal is a legal billing analytics company.

- Shifted team from data quantity to data quality
- Managed the offshore annotators with new inter-annotator performance metrics
- Led the Senior Data Scientist and the subject-matter expert.
- Assigned engineering tasks
- Built SageMaker inference server

Lead Analytic Endpoint Engineer at Sight Machine, 2018-2021

Sight Machine is a manufacturing analytics company.

- To get the company's biggest public-facing feature, Recipes, across the GA line, I worked across silos (Customer Success, DevOps, Data Engineers). I wrote high-level design papers to help the Product and Customer Success teams understand the time cost of competing technical design decisions, arming them with a conceptual framework to gather more information from the customer on her priorities.
- I started a new checklist process to lock down successful sales demos.
- I nudged the engineering team to move from a delivery team to a product team. The Director of Engineering and I started a new process where the engineers fleshed out their own Jira issues. The new technical design autonomy reduced engineering toil and increased creativity.
- I built the first distributed tracing (Lightstep), which made debugging system downtime easier.
- I proposed and delivered a new standardized frontend development environment. The innovation made debugging and on-boarding easier.
- I coached junior engineers in systems thinking.

Lead Data Engineer at HiQ Labs, 2015-2018

HiQ Labs was a people analytics company.

- I led the migration from a monolith to a microservice paradigm.
- I migrated the Data Scientists from Mongo queries to PySpark.
- I designed and built the scaping system V2
- I trained a Junior Data Engineer to maintain the scraping system.
- I led a Junior DevOps engineer to build a Splunk pipeline observability system.

Developer at Urban Mapping, 2011-2013

Urban Mapping provided geospatial analytics to Tableau.

- CI/CD/QA pipeline
- Observability

Misc leadership roles

- Worked with students in Medellín, Columbia to make ClusterPy.
- Kids snowboard instructor at Vail Resorts, CO.
- Restaurant Assistant Manager at Vail Resorts, CO.
- Counselor for emotionally disturbed children. Seneca Institute. CA.

Open source code and writing

- Langchain PR: Causal Program-aided Language (CPAL)
- Normalize tables using a LLM
- *How can language AI performance be raised with cognitive empathy?*
- A play Ethereum MEV bot
- Small commit to shell-ai
- A git bare approach to version control your dot files
- Work papers
- Geoscore (repo)
- PhD thesis. *Assessing Inequality using Geographic Income Distributions*
- *Spatial Econometrics* entry. Encyclopedia of Human Geography. 2009.
- *Interactive spatio-temporal modelling of health systems*
- *σ-convergence in the presence of spatial effects*
- *Integrating Econometric and Input-Output Models in a Multiregional Context*