

Boris Dev | boris.dev@gmail.com

Data leader background

- Led Language AI model development
- Led data engineering from monolith to microservice paradigm
- Led building new observability systems
- Led analytic product features
- PhD research in social science data analytics
- Engineering tools: microservice backend stack, Splunk, Lightstep
- Data science tools: PySpark, Pandas, Numpy, BI (Splunk, Tableau)

Career goal

My goal is to serve a team of scientists that are discovering new analytic methods to help decision makers grow the business. My interests include the following.

- Reduce scientist toil by optimizing and automating data processes
- Document a ubiquitous language of customer meaning behind the data
- Get more feedback through observability, prototypes (demos), and [papers](#)
- Bring engineering best practices to scientists

Leadership style

1. Questioning - cognitive empathy to get context to execute the below points.
2. Visioning - win-win attitude that results in joyful effort over obfuscating
3. Coaching - top-down feedback, smaller chunks, metrics, short-cuts
4. Serving - bottom-up feedback, documentation discipline, blockers

Job experience

Technical AI Product Manager, consultant at SimpleLegal, 2022-2023

SimpleLegal software is a tool for big companies to manage bills from their vendor law firms. Before I arrived, my employer's new language AI feature for flagging suspicious invoice lines was stuck for one year because of bad performance, even after investing in more human annotation. My job was to increase AI performance. My actions resulted in the feature being launched and winning an [AI Excellence Award](#).

Managed the Senior Data Scientist consultant

- The technical vision I set for the Data Scientist and I was to launch an MVP in six months by discovering “low-hanging fruit” that could scale back the machine learning effort. This resulted in scaling back our machine learning training from eleven to three redefined classifiers. The triage was to kill one classifier that was never needed, replace two with expert rules, and stop the labeling on five whose performance was sufficient.

- To increase clarity, I coached the team to understand the customer meaning behind the data and the AI flags.
- To increase clarity, I crunched the knowledge of the Data Scientist's metrics analysis into a working paper, which was continually shared at the weekly stand-ups.
- I wrote the data scientist a stellar performance review, highlighting his acumen at articulating complex concepts to executives.

Managed the engineering tasks

- I assigned Jira issues to engineers.
- 1:1 meetings with the VP of Engineering.
- After one engineer left for another company, I stepped in to refactor the inference server.

Managed the offshore annotators

- Included the offshore annotators and subject-matter expert into a continuous improvement collaboration on AI flag definition and annotation guidelines.
- Changed annotation process to small batches
- Started inter-annotator performance metrics and meetings.
- Labeling several hundred sentences, I built cognitive empathy to identify convoluted annotation guidelines as the root cause of our poor labeling quality.

I changed our direction from increasing data quantity to increasing quality.

- To gather knowledge, I organized meetings with six experts in three sister companies across the conglomerate.
- At my request, the President convened a Amazon 6-pager meeting where we could teach the team what we learned and sell the new direction.
- The annotation expert taught us success depends on the annotators expert feedback.
- The data scientists explained the biggest issue is *"garbage in, garbage out"*.
- A Founder/VP of Product at a sister company taught me we must have pre-processing noise filters.

Lead Analytic Endpoint Engineer at Sight Machine, 2018-2021

The Sight Machine product is a performance dashboard tool for process engineers at factories.

- got the company's biggest public-facing feature, [Recipes](#), across the GA line
- worked across silos (Customer Success, DevOps, Data Engineers)
- wrote high-level design papers to help the Product and Customer Success teams get relevant customer feedback on product technical design decisions.
- started a new checklist process to lock down successful sales demos.

- nudging the engineering delivery team to be closer 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.
- built the first distributed tracing (Lightstep), which made debugging system downtime easier.
- proposed and standardized frontend development environment, which made on-boarding easier.
- coached junior engineers.

Lead Data Engineer at HiQ Labs, 2015-2018

The HiQ product provided predictions on the likelihood that a company employee would quit.

- A puzzle for the CTO and I was to figure out how to get around LinkedIn's bot detection in order to scrape millions of HTML public profiles, the raw data for our prediction pipeline.
- I led a Junior DevOps engineer to build a Splunk observability system that was used to track performance and experiments on different spider configurations.
- I trained a Junior Data Engineer to maintain the scraping system.
- tech lead on the move from a monolith to a microservice paradigm
- I led a Data Engineer to guide the Data Scientists into a new microservice development paradigm.
- I explained to management the technical and human vision behind the microservice architecture.
- I shadowed the Data Scientists to see what I could automate for them.
- I migrated the Data Scientists from Mongo queries to PySpark.

Start-up partner and full-stack developer at Map Decisions, 2014

Map Decisions was a mobile app to automate street sign inspection for city Public Works departments.

Start-up partner and analytics developer at Geoscore, 2014

[Geoscore \(repo\)](#) ranks and visualizes 100,000+ neighborhoods.

Developer at Urban Mapping, 2011-2013

Urban Mapping provided a geodata REST API to Tableau and other customers.

- I built the observability system (Splunk).
- I created a new performance metric for map tile rendering.
- I built the CI pipeline (Jenkins).
- I built a Tableau BI Dashboard

Misc leadership roles

- As a founder of [ClusterPy](#), I coached students in Medellín, Columbia in geographic data software.
- As a PhD researcher, I coached graduate students and taught spatial statistics.
- Kids snowboard instructor at Vail Resorts, CO.
- Assistant Manager at Gundy's Grill restaurant at Vail Resorts, CO.
- Researcher on building a prototype of the Wall Street Journal's *Main Street* section. California Business Magazine. CA.
- Led analysts collecting regional company ranking for large newspapers. California Business Magazine. CA.
- Counselor for emotionally disturbed children. Seneca Institute. CA.

Open source code and writing

- [A play Ethereum MEV bot](#)
- [A git bare approach to version control your dot files](#)
- [Why did your language AI feature fail?](#)
- [Work papers](#)
- 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*