

## Staff Data Scientist Job Application‘

- Boris Dev
- boris.dev@gmail.com
- San Francisco

I am applying for the Staff Data Scientist position at Fastly in San Francisco. Attached is my resume. Below I address the section of your job posting title, “*We’ll be super impressed if you have experience in one or more of these*”.

### Domain expertise in our industry or related fields

- Related to Fastly tools to block scraping, I was in charge of getting around LinkedIn’s bot detection system. I built a distributed scraper that downloaded million’s of LinkedIn profiles for ingestion into a ML training pipeline.

### Experience deploying and maintaining production machine learning systems

- At a legal tech company, I deployed their first AI classifier model to flag problem invoice line items on AWS SageMaker. I also built a human annotation pipeline to label, wrote the annotation guidelines, and trained the annotators, coordinated the subject matter experts, and changed the direction of the project.
- I migrated a team from a monolith to a microservice ML pipeline. I also added the company’s observability system to monitor the ML pipeline.

### Track record of leading technical teams or mentoring junior data scientists

- I trained data scientists in PySpark and DataBricks..
- I trained data scientists to convert their monolith code into deployable microservices.

### Building AI-powered products

“Building AI-powered products or features that have been successfully deployed to end users”

“Implementing and fine-tuning large language models or other generative AI systems for specific business applications”

- I deployed an AI feature to users that had previously been stuck for a year before my arrival.
- My side-project, Nobsmmed.com, is a generative AI product that extracts semantic meaning of health interventions from clinical studies and Red-

dit posts for users to search and explore other people's health hacking experiences.

- For my PhD dissertation, I formulated metrics that leverage the special nature of geospatial data.
- My side-project, Nobsmed.com, is a generative AI product that extracts semantic meaning of health interventions from clinical studies and Reddit posts for users to search and explore other people's health hacking experiences.

### **“Experience with prompt engineering and RAG systems”**

- For a Wolf Games, a murder mystery gaming company, I built their first story generation service using a automated DAG prompting with generative AI.
- For my business side-project, Nobsmed.com, I use generative AI to extract semantic meaning of health interventions from clinical studies and Reddit posts that are indexed into a RAG system.

### **“Contributions to open-source AI projects or research publications in machine learning or generative AI”**

- For the Langchain open-source library, I contributed an experimental way to create dynamic causal knowledge graphs (Tweet by LangChain and the PR
- My opensource library bertopic-easy is quick way to cluster documents into topics that leverages OpenAI's cutting edge o3-mini LLM model to polish embedding based cluster results.

### **“Proficiency with AI services on cloud platforms (AWS Bedrock, GCP Vertex AI, Azure OpenAI)”**

- I work with Azure's OpenAI services.