Senior Data Scientist

Capitol AI is a venture-backed startup with a mission to humanize data analytics through automated insight generation and data storytelling. We are looking for a Senior Data Scientist with a background in Natural Language Generation (NLG) and Machine Learning who loves to ship code. Capitol AI is a fast-paced, dynamic work environment and you should be a self-starter who thrives in a space where requirements, customers, and tactics and strategy are in flux.

**Job Description**
The Data Scientist will work closely with the founders, engineering, and product teams to bring novel natural language generation and insight generation capability to our product. You will be joining a small core team and as a Senior Data Scientist, you will design, implement, and evaluate and test algorithms that have a direct impact on the utility of our product. Capitol’s first customers are leading healthcare companies and your work has the potential for broad, meaningful impact.

We believe in shipping, transparent communication, a positive outlook, bringing data informed hypotheses to the conversation, and an overall drive for improvement.

**Location**
Capitol is based in NYC and the Senior Data Scientist position is open to remote or onsite candidates.

**Responsibilities**
- Research novel machine learning and NLG strategies
- Research large language model (LLM) creation and optimization strategies
- Collaborate closely with product and engineering
- Create novel approaches for insight generation
- Perform Exploratory Data Analysis on customer data
- Collaborate with team members and be responsive in real-time (we use Slack and Notion)
- Mentor teammates as the company grows
- Contribute to engineering of AI/ML pipelines
- Ship code
Required Skills
- High Python programming proficiency
- Experience with LLMs
- Experience with pytorch
- High SQL proficiency
- Experience with or comprehensive knowledge of graphical models / hierarchical models
- Practical mastery of Git & and best practices of version control
- Demonstrated excellence in communication

Required Experience
- Experience in healthcare analytics or health economics
- 5+ years as a data engineer, applied statistician, data scientist or similar
- 3+ years of software engineering experience as an individual contributor
- MS or PhD in a quantitative field
- Strong background in analytics and multivariate analysis

Nice to haves
- Understanding of causal inference frameworks and techniques
- Experience in data science in the software as a service (SaaS) product context
- Experience building and training ML models
- Data visualization experience
- Experience with Jupyter notebooks
- Experience with pandas
- Experience building python APIs with fastapi or Flask

Our Values
Prioritize Listening
We listen to the problem and ask questions before we create solutions. We value our teammates’ opinions and experience.

Put your team first
We are a team rowing in the same direction. We value our camaraderie and care for our team members.

Run towards hard problems
We are not afraid of ambiguity. We work together to break hard problems into small digestible ones.
Stay humble, keep improving
We all make mistakes. We make sure to learn from them and have the humility to take feedback.

Ship small, crawl first
We start with the smallest, shortest, dumbest way to solve the problem. Then we build on it together to make it great.

The Opportunity
Working at Capitol will be an incredible journey with some unique upsides for the right person.

- Collaborate with leading technologists, executives, and government leaders on problems that matter in healthcare, finance, the Department of Defense, and more
- Work on problems with a high degree of impact and scale
- We are a small team with minimal bs and bureaucratic processes
- You will have the ability to influence the direction of the product and the company
- True startup environment, a dynamic company with many ways you can contribute
- Early hire opportunity - Meaningful equity offered to core team members