Data Scientist at Omelas

Company Description:

Interested in a position at the forefront of the information war? For half a decade, Omelas has been the leader in protecting democracy against propaganda. We expose how authoritarians manipulate the internet. We monitor the most influential newspapers, TV channels, government offices, militant groups, and more across a dozen social networks and messaging apps, thousands of websites, and thousands of RSS feeds. Our unique access to the media landscapes of authoritarian states like Russia, Iran, and China has won us coverage in over a dozen languages and countless publications ranging from Meet The Press to the BBC to Financial Times and more.

Location:

Washington DC preferred, US-remote possible

Job Description:

We're looking for a skilled, experienced data scientist to assist in the development of models to understand what propaganda can tell us about how our adversaries think and behave. You'll work closely with our team of subject matter expertise to design and implement models that reveal who's saying what, how it influences what others say, and why it matters.

This position will work on the various machine learning projects sketched out below with regular support and meetings with data scientists and engineers at Omelas. These projects are subject to change but will generally involve natural language processing and machine learning modeling to surface key pieces of information for end users.

Potential Projects:

- Quote extraction and attribution from text
- Automatically detect meetings between leaders in text using NLP methods such as text-classification, fuzzy matching, etc.
- Find how propaganda ties together different events over time to uncover narrative themes

Desired areas of expertise:

- 2+ years work experience in a data scientist/engineering role including deployment of production level code
- Proficiency in Python (data science stack: pandas, sklearn, numpy, etc) and SQL.

- Experience scoping and building machine learning models from experimentation to deployment.
- Experience building your own labeled data for classification models.
- Experience with natural language processing libraries such as spaCy, NLTK, etc.
- Excellent communication and documentation skills.

Nice to haves:

- Experience monitoring and maintaining ETL pipelines on AWS infrastructure.
- Experience with container technologies such as Docker.
- Experience with Deep Learning frameworks such as PyTorch, TensorFlow (Keras) and HuggingFace.