

NLP Data Scientist for Literature Mining [remote]

Project Description

We are looking for a Data Scientist experienced in natural language processing (NLP) to help design and build a platform that aids interpretation of molecular (metabolomic) features that reflect performance in behavioral and cognitive tasks. This role is part of the data science and engineering team within the Fraenkel Lab, Dept. of Biological Engineering at Massachusetts Institute of Technology.

The Fraenkel lab develops computational and experimental approaches to understand human biology. New experimental methods make it possible to measure cellular changes across the genome, epigenome, proteome, and metabolome. These technologies include genome-wide measurements of transcription, of protein-DNA interactions, of chromatin accessibility, of genetic interactions, and of protein and metabolite modifications. Each data source provides a very narrow view of the cellular changes. By computationally integrating these data we can reconstruct signaling pathways and identify previously unrecognized regulatory mechanisms.

We now seek to expand the power of these approaches by mining the vast published literature on metabolomic features. In combining literature information with the quantified metabolome, we aim to build a knowledge graph that will enhance our interpretation of molecular networks.

Position Description

As part of the data science and engineering team in the Fraenkel Lab this position entails the following activities:

- Implement natural language processing frameworks to mine metabolomic data captured in systematic reviews and online repositories
- Develop a comprehensive set of data engineering tools to process and store these biological data in a NoSQL database
- Architect a cloud-based or hybrid database solution, build automation pipelines and develop strategies to integrate external and lab generated data
- Assist in the research and development of novel methodologies in combining network analysis with machine learning for predicting biological aptitude and/or other clinical outcomes
- Collaborate with junior and senior computational staff on a variety of data science activities including but not limited to experimental design, exploratory data analyses, data munging, feature engineering, building and testing machine learning models
- Assist in the development of research publications and/or related novel intellectual property related to this work

Preferred

- BS/BASc in Computer Science/Information Technology with two years of related experience is preferred.
- Experience with graph and/or other NoSQL database design and engineering
- Experience working with large datasets, munging, feature engineering and developing machine learning models preferably on clinical datasets
- Experience with natural language processing algorithms or equivalent MOOC certifications
- Background in biology would be helpful but not required
- Languages: Python/R/SQL

Note: This is a limited term appointment with the potential to convert into a full-time position. This position is not eligible for visa sponsorship.

Please send your resume and questions to Dr. Swapnil Chhabra chhabras@mit.edu.