# Christopher Bui

✉ cbui3@pm.me   🌐 www.cbui.me   in /in/cbui3   ○ christopherbui

## Education

**Johns Hopkins University** – Bioinformatics, M.S. (2022)

**University of California, San Diego** – Biophysics, B.S., Entrepreneurship, Minor (2018)

## Experience

**Data Scientist**  
University of California, Los Angeles (UCLA Health)

Los Angeles, CA  
Mar. 2025 – Present

- Built mass CyTOF analysis pipeline using R and CATALYST to process flowset lung cancer single-cell protein data.
- Applied statistical and clustering analysis including PCA, FSOM, t-SNE, & UMAP on single-cell protein data with Bioconductor.
- Utilized 10x Genomics' Spaceranger software on HPC cluster to analyze scRNA-seq data, providing insights into lung cancer gene expression.
- Processed Multiplex Immunofluorescence (MIF) data into Single-Cell Experiment (SCE) objects, enabling cellular profiling and spatial analysis.
- Generated detailed, presentable data visualizations with R to communicate findings.

**Analytics Engineer**  
Pure Lithium Corporation

Boston, MA  
Dec. 2024 – Mar. 2025

- Developed scalable data models using Snowflake SQL queries and Data Build Tool (DBT) to support company data infrastructure.
- Created interactive dashboards on Hex & Streamlit to deliver key product performance insights for stakeholders.
- Managed Prefect flows deployed on AWS, enabling distributed ETL processes.

**Data Scientist**  
SCAN Health Plan

Long Beach, CA  
Nov. 2023 – Nov. 2024

- Developed SQL queries on the Microsoft Azure to analyze patient EHR and claims data, enabling insights into healthcare trends and client behavior.
- Led development of CART and Deep Neural Network classifiers using Scikit-Learn and PyTorch to predict medication non-adherence patients.
- Deployed machine learning models to production and evaluated their performance using AUC-ROC curves to improve identification of members eligible for healthcare plans.

**Clinical Programming Analyst**  
Edwards Lifesciences (Contract)

Irvine, CA  
Oct. 2023 – Nov. 2023

- Ensured consistent schema of KPI dataset through appropriate data validation and preprocessing using PySpark.
- Optimized interactive data visualizations by enhancing Python Dash Plotly dashboards, improving usability for clinical stakeholders.

**Data Scientist**  
Bio-Rad Laboratories

Irvine, CA  
Sep. 2023 – Oct. 2023

- Developed SQL queries in Snowflake to accelerate dataset preparation for PowerBI dashboards built using DAX language.
- Conducted profit margin analysis across the company's product portfolio to identify actionable KPIs, improving product strategy.

**Data Scientist**  
The Johns Hopkins University Applied Physics Laboratory

Laurel, MD  
Aug. 2022 – Jul. 2023

- Built a scalable ETL data pipeline in Python to preprocess environmental sensor data from hospital operating rooms.
- Developed regression and clustering models to predict hazardous operating room conditions, improving patient safety protocols.
- Leveraged High Performance Computing (HPC) clusters with SLURM on Linux to run SARS-CoV-2 protein-protein docking simulations.
- Analyzed Covid-19 personal protective equipment inventory using PySpark, identifying supply chain bottlenecks.

**Machine Learning Engineer Intern**  
3M

Silver Spring, MD  
Jun. 2020 – Aug. 2020

- Leveraged PySpark and NTLK on AWS to preprocess large-scale parquet text files containing medical records.
- Designed NLP models to generate text embeddings of ICD-10 medical codes, enhancing medical code recommendation accuracy.

## Projects

**Cancer Tissue Detection:**
- Developed a PyTorch CNN with transfer learning on CUDA to classify 200K tissue images as cancerous or benign, achieving 86% accuracy.

## Skills

**Technologies:** Python, R, Java, HTML/CSS/Javascript, Scikit-Learn, Pandas, Numpy, Tensorflow, PyTorch, PyKeen, Optuna, Jupyter Notebook, OpenCV, SQL, PySpark, Git, Amazon Web Services, Microsoft Azure, Snowflake, DBT, HPC, SLURM, Tableau, PowerBI, Hex

**Knowledgeable:** Predictive Modeling, Statistical Analysis, Deep Learning, Knowledge Graph Embeddings, Natural Language Processing, Time Series Forecasting, Computer Vision, A/B Testing, Feature Engineering, Data Visualization, Communication, Self-Motivation, Empathy