

# ARIEL LUBONJA

## Machine Learning PhD Student

ariel@cs.jhu.edu • ariellubonja.com • linkedin.com/in/ariellubonja • medium.com/@lubonjaariel

### EDUCATION

#### Johns Hopkins University

PhD, Computer Science

05/2022 - 08/2024

- Member of the [Institute for Data-Intensive Engineering and Science \(IDIES\)](#), and [Johns Hopkins Turbulence Database](#) maintainer
- Published two peer-reviewed papers; one manuscript in preparation
- Expertise in scaling graph algorithms and deep learning techniques, Graph Learning, Computer Vision, and Time-Series Analysis
- Recognized for exceptional presentation skills and public speaking abilities in academic conferences and seminars

#### Johns Hopkins University

Master of Science in Engineering, Data Science | GPA: **3.97** / 4.0

01/2021 - 05/2022

- Advanced coursework: Parallel Programming, Machine Learning, Deep Learning
- Specialized in Mathematical Statistics, Optimization, Signal Processing

### POSITION-RELEVANT PROJECTS

#### Enhancing LLM Speed through Speculative Decoding and Sparsity

02/2024 - Present

- Developed a novel approach to Large Language Model inference utilizing weight sparsity to enhance computation time of LLMs
- Aided the implementation of an improved CPU-based approach for sparse matrix multiplication
- Evaluated this approach to show performance benefits in end-to-end LLM workloads

#### Advanced Drug Demand Forecasting System

09/2023 - 01/2024

- Developed an innovative drug demand predictor integrating multimodal data, including drug embeddings and historical sales
- Implemented efficient similarity computation techniques for large-scale embedding comparisons
- Incorporated hierarchical drug information from the WHO Essential Medicines List
- Evaluated and implemented advanced time series forecasting models, including Facebook Prophet and AWS Hierarchical Time Series models

#### Drugs Discovery through Graph and LLM Embedding

03/2023 - 07/2023

- Engineered a sophisticated web crawler to extract drug and molecule relationships from [Drugs.com](#) to form a graph structure
- Implemented cutting-edge graph embedding techniques, including node2vec
- Utilized domain-specific fine-tuned language models for semantic drug embeddings
- Integrated these embeddings into the existing drug demand predictor, achieving a 13% improvement in accuracy (SMAPE)
- Published the dataset on [HuggingFace](#), facilitating further research in the field

#### Chatbot - Lewis Carroll's Alice's Adventures in Wonderland

University of Southern California

05/2019 - 07/2020

- Conducted extensive evaluation of various speech recognition systems, including Kaldi, Rasa, Mycroft, Picovoice, and CMU-Pronouncing Dictionary
- Explored and implemented advanced fine-tuning techniques for Google BERT on specialized pharmaceutical Q&A pairs
- Deployed a highly efficient and scalable solution using Google Dialogflow API, balancing performance and cost-effectiveness
- Served as Full-Stack Developer, developing a cross-platform mobile application using Electron
- Implemented a robust CI/CD pipeline to ensure seamless deployment and maintenance of the system

## WORK EXPERIENCE

### jBoxers Bulgaria

Sofia, Bulgaria

JavaEE & Django Software Developer

05/2017 - 08/2021

- Led frontend and backend development for Library of Congress projects using Angular and Django
- Developed advanced chatbot system for USC using Flask, Python, and Electron
- Implemented backend solutions with JavaEE WebSockets and cloud deployment (AWS, Google Cloud)
- Established CI/CD pipelines, containerization, and comprehensive testing protocols
- Mentored junior developers, enhancing team productivity

## ACHIEVEMENTS

Full Fellowship to Johns Hopkins University

Dean's List

Member of IEEE, ACM

## PUBLICATIONS

### BioRXIV

<https://www.biorxiv.org/content/10.1101/2024.06.05.597554v1>

Application of machine learning in a rodent malaria model for rapid, accurate, and consistent parasite counts

06/2024

Developed an innovative, user-friendly software tool leveraging state-of-the-art deep learning and computer vision techniques to significantly expedite the process of counting and annotating malaria-infected cells in rodent models.

### IEEE IPDPS 2024

<https://arxiv.org/abs/2402.04403>

Edge-Parallel Graph Encoder Embedding

05/2024

Pioneered a groundbreaking optimization of the Graph Encoder Embedding algorithm, achieving a 1000x performance improvement, enabling processing of billion-edge graphs in mere seconds, significantly outperforming existing methods. Presented findings at the International Parallel and Distributed Processing Symposium (IPDPS) 2024 in San Francisco.

### ArXiv

<https://arxiv.org/abs/2407.06434>

Efficient Batched CPU/GPU Implementation of Orthogonal Matching Pursuit for Python

01/2021 - 05/2021

Engineered an advanced implementation of the Orthogonal Matching Pursuit algorithm using BLAS (CPU) and PyTorch GPU, outperforming Scikit-Learn by 100x for large-scale linear systems.