ARIEL LUBONJA

ariel@cs.jhu.edu https://ariellubonja.github.io/
https://www.linkedin.com/in/ariel-lubonja-235a6b197/ https://medium.com/@lubonjaariel

PROJECTS

NLP - Chatbot for the University of Southern California (USC) - Alice in Wonderland

05/2019 - 07/2020 University of Southern California
https://alice.wonderland.usc.edu/chat/alice-endpoint.html

- Natural Language Processing (NLP). Researched Kaldi, Rasa, Mycroft, and early AWS solutions, picovoice, CMU pronouncing dictionary. Implemented:
- 1. Mycroft.ai Skill - Add Skill to the existing Mycroft.ai chatbot. Lack of clear documentation hindered this effort.
- 2. Google's BERT. Fine-tuned BERT on 600 Q&A pairs about Alice. Deployed on a AWS machine + GPU. Inference took 50 seconds on CPU, 13sec on GPU - too slow - Goal: <6 seconds inference
- 3. Google Dialogflow, featuring a model based on BERT, a user-friendly UI, ability to label positive and negative examples and with background-managed training
- Implemented Mycroft as a Speech-to-Text engine for transcribing user's speech
- See Full-Stack Developer section for non-Machine Learning development & deployment effort

Dall-E AI Image Generation Extension for Google Docs, Sheets, Slides

07/2022
https://workspace.google.com/marketplace/app/ai_generated_images_dalle/867097413141?hl=en&pann=docs_addon_widget

- Built a Google Workspace app to serve Dall-E generated images to Google Docs, Sheets and Slides documents
- >5000 downloads, 4-star average rating

Kaggle: Stable Diffusion - Image to Prompts

03/2023
https://www.kaggle.com/competitions/stable-diffusion-image-to-prompts?utm_medium=email&utm_source=gamma&utm_campaign=comp-stable-diffusion-2023

- Can you recover the prompt that produced a specfic image output from Stable Diffusion? Competition in progress...

Master Thesis: Parallel Implementation of Graph Encoder Embedding in Ligra

01/2022 - 05/2022
https://github.com/Ariel5/GraphEmd

Graph Embedding is a popular method in graph machine learning for forming embeddings that allow for easy comparison of nodes

- Built a highly scalable version of Graph Encoder Embedding using the Ligra graph-parallel framework
- 1400x speedup over the original in Ligra
- Implemented a Python-Numba approach for ease of use. Achieves 289x speedup

WORK EXPERIENCE

Full Stack Developer - Python/Django, JavaEE, Angular

jBoxers

11/2018 - 08/2021 Sofia, Bulgaria

jBoxers is a company that offers software solutions in the fields of AI, Microservices, DevOps and Cloud

- Chatbot for the University of Southern California with Dr. Laura Tolosi, using Mycroft as a Speech-to-Text and Google Dialogflow as language models (see Projects). Flask API
- Developed the Desktop app for Alice. By using Electron, I cut down the app development time to 3 weeks instead of multiple months
- JavaEE WebSockets for non-REST chatting
- AWS & Google Cloud deployment + CI/CD, Docker, Vagrant, Datalku
- Other Duties: Use Case discussion and design with Library of Congress (LoC) project owner
- Developed enterprise solutions for LoC Virtual Card Catalog and Braille
- Supervised 2 summer interns

Junior Developer - JavaEE, Angular, Python/Django, SQL

05/2017 - 11/2018

- Ongoing project for LoC, using Angular for front-end and JavaEE/Spring for backend
- Developed tests using Sonar, JUnit, Selenium and Protractor

ACHIEVEMENTS

Fellowship to Johns Hopkins University

1/14 students funded by the Government of Albania, in their Excellence Scholarship, for students admitted to top-15 universities

Full Scholarship to American University in Bulgaria

1/11 students awarded by the America-Albania Development Fund

334/340 in the GRE exam

93rd percentile in Quantitative (168/170)
97th percentile in Verbal Comp. (166/170)

PASSIONS

Skiing, Weightlifting, Running

Cars, Scalable Algorithms

EDUCATION

Doctor of Philosophy, Computer Science

Johns Hopkins University

05/2022 - Present

GPA **3.93** / 4.0

- Graph Learning & Embedding - Graph Neural Networks, node embedding, edge prediction, transductive learning with Graph Attention Networks. Thorough knowledge of node2vec and Graph Encoder Embedding. Extensive work on scaling these algorithms
- Work on scaling Leiden community-detection graph algorithm
- Maintaining the Johns Hopkins Turbulence Database
- Computer Vision: Transforms, dictionaries & filters, Convolutions, CNNs, segmentation, Edge & boundary detection, sparsity
- Given multiple presentations at various audience levels. Effective at presentations and communicating complex ideas.

Master of Science in Engineering - Data Science

Johns Hopkins University

01/2021 - 05/2022 Baltimore, Maryland, USA

GPA **3.72** / 4.0

- Scaling up algorithms - Parallel Programming, MapReduce/Hadoop, Spark, Dask, Ray, GPU/TPUs, Multithreading, Memory Hierarchy, OpenMP
- Machine Learning & Deep Learning: Regression, CNNs, Decision Trees, Random Forests, SVMs, RNN, LSTM, Transformers, Markov Models. Significant experience w/ PyTorch
- Algorithms & Data Structures: complexity theory, big-O, sorting, advanced data structures, dynamic programming, graph algorithms, Linear prog.
- Mathematical Statistics - Model fitting, estimators, bias-variance tradeoff, convergence, ANOVA, hypothesis testing, Regression
- Optimization - Stochastic Gradient Descent, line/grid-search, Newton's Method, BFGS, Trust Region Methods, convex, (non-)smooth opti.
- Signal Processing - finding sparsest solution to underdetermined systems. L1 (Lasso) and L0 (Matching Pursuit) techniques. Applications in facial recognition and image processing

Bachelor in Computer Science

American University in Bulgaria

09/2014 - 05/2018

GPA **3.57** / 4.0

SKILLS

Python **PyTorch**

Sequence Models **Transformers**

Mycroft **Scikit-learn** **C++**

OpenMP **SQL** **AI/ML**

Graph Neural Networks

Convolutional Neural Networks

XGBoost **Linear/Logistic Regression**

Google Apps Script **AWS**

Google Dialogflow **Docker**