YANALL BOUTROS

1-530-591-3833 ◇ YanallBoutros@ProtonMail.com
Yanall-Boutros.github.io
Baltimore, MD

EDUCATION

University of California, Santa Cruz [UCSC] September 2016 - August 2020
Bachelor of Science (B.Sc.) Physics, B.Sc. Computer Science *Santa Cruz, CA*

· Electives: Advanced Programming, AI, Computational Physics, Quantum Computing; GPAs: 3.40/4.00

TECHNICAL STRENGTHS

Languages:	Python, C/C++, C#, Bash, Powershell, Tex, HTML, NodeJS, Haskell, Perl, Nix
Frameworks:	TensorFlow, PyTorch, Scikit-HEP, Bootstrap, JQuery, Numpy, Matplotlib, Pandas, Unreal Engine
Infrastructure:	GNU/Linux, Unix, Windows, PostgreSQL, Kafka, Docker, Podman, Git, Jira, Ansible, NixOS
Mathematics:	Scientific Communication, Modeling, Statistics, Artificial Intelligence, Simulations
Data Science:	Natural Language Processing [NLP], Data Analysis, Data Validation, Research, Debugging, Testing
InfoSec:	Metasploit, Burp Suite, OWASP Zap, Nmap, Shodan, Maltego, Snyk, Snort, Wireshark

EXPERIENCE

DCS Corp October 2021 - June 2024
T1 Computer Engineer II ← T1 Software Engineer I *Aberdeen, MD*

· Exceeded Expectations in AI and Data Engineering Tasks for Army Research Lab [ARL], Human Research and Engineering Directorate [HRED], Humans in Complex Systems [HCxS], Information for Mixed Squads [INFORMS] Lab, by taking lead in three more projects than requested, measured by my performance reviews

· Achieved 70% accuracy in associating bio/physio data with firing events, by downsampling signals and statistical features as inputs to a TensorFlow EEG Net Feed Forward Binary Classifier, measured by confusion matrix in validation testing

· Achieved realtime Computer Vision [CV] classification and Speech-To-Text [STT] by multiprocessing AI Data Pipelines in Python, measured by publishing inferences within a 1 second polling rate

· Achieved 70% average True-Positive object detection classification in testing dataset, by synthesizing initial training set of image mask pairs, then training a Region-based Convolutional Neural Network CV agent

· Scaled CV Classifier's prior initial training dataset 100× by rendering post-processed scenes in Unreal Engine, measured by comparing the number of unique samples in the previous dataset

· Improved average True-Positive accuracy 5% in 2 weeks by automatically detecting, masking, and augmenting new target classes, generating 2× more data in those 2 weeks

· Plotted model performance as function of distance, orientation, terrain in Python, Numpy, Matplotlib

· Orchestrated deployment of docker/podman containers by writing unit file templates, systemd services, and Ansible Playbooks

· Achieved 10% improvement of STT/Automatic Speech Recognition [ASR] Pipeline, by switching to a transformer based model, measured by Word Error Rate [WER]

· Improved Audio Signal in Real Time Transcriptions and STT Pipeline by applying Root-Mean-Square and Fast Fourier Transform Frequency filters to run transcriptions only on active speakers, measured by reducing phrase level tokenization issues and improved WER score

· Made tool to accelerate supervised transcription corrections in half the time-length of the audio source, to update large language model [LLM] and lexicon with military specific vernacular

- Achieved 50% success rate in extracting survey answers from transcriptions within 10 shots, by calculating cosine simularity of word embeddings between queries, and passages encoded by a sentence transformer
- Integrated STT, NLP, and Benchmarking tools in NodeJS/Express/Bootstrap interface. Dockerized services
- Simulated Subsystem Failures in Unreal Engine, by making a Component Health System for prototypical autonomous vehicles powered by the Robot Operating System [ROS]
- Integrated Kafka/PostgreSQL Producer/Consumer in Component Health System, ASR/NLP/CV tools

Bitwork Solutions April 2023 - Present
Partner, AI Engineer *Remote - Baltimore, MD*

- Achieved $>$ 80% categorization of which business categories most accurately represent a URL within 10 shots, by calculating cosine simularities between passages and encodings from a sentence transformer
- Prototyped binary mask generation pipeline capable of extracting logos from any photo, by applying Laplace edge detection on binary masks from Otsu's method
- Setup reproducible computing environment deploying: Kafka, PostgreSQL, Grafana, Prometheus, OpenVPN, OpenSSH, and Kubernete servers, by configuring a NixOS system on Hetzner Cloud
- Integrated APIs to automate generating articles of keywords for target industry and audience
- Mentored Junior AI Engineer on LLMs for physics, by suggesting additional corpuses for retrieval augmented generation
- Implemented information retrieval algorithms to determine competitor keywords by scraping URLs and social media posts
- Implemented LLM-based text generation algorithms to mass-schedule Search Engine Optimized posts
- Designed SQL Data Schemas for interface, managed user group / role permissions
- Wrote flask middleware for accessing database and automating database model ORMs
- Consulted on feasibility of AI integration and modern data science techniques to project goals

Independent Contractor August 2021 - Present
Software Engineer, Open Source Investigator *Chico, CA; Baltimore, MD*

- Made Decentralized Exchange Volume Liquidity trade bot for Full Send Network [FSN] by automating ABI calls in Python 3
- Made order queue for web application in PostgreSQL, NodeJS, and Redis; for FSN
- Made accelerated mask creation tool with ResNext FPN TensorFlow AI application for FSN
- Recovered crypto assets by determining computational feasibility from combinatorics, setting up a haskell server with 10x1080TI GPUs, and identifying a matching address from the private seed
- Setup local home security system and monitor dashboard for client by networking raspberry pi's
- Accelerated conversion from artists rendition to video game asset for Digital Asset Management Group [DAMG] by implementing Neural Radial Fields and ZeroShot, and generating transforms matrices from 2D drawings
- Setup dedicated local Jax/Dalle/Imagegen server, researched Text $\rightarrow$ 2D $\rightarrow$ 3D generation for DAMG
- Found missing person's full name, social media, and contact information given only an online username
- Found locations, owned organizations, social networks, and private information on behalf of tenant

Santa Cruz Institute for Particle Physics [SCIPP] August 2018 - August 2020
Undergraduate Research Assistant Intern *Santa Cruz, CA*

- Achieved 80% accuracy, 5% bias in Confusion Matrix in classification of parent particles, by training a Deep Neural Network Binary Classifier on Simulated LHC events/interactions, in TensorFlow

- Streamlined, benchmarked, and built docker containers documenting the Python workflow and modules for simulating particle physics
- Multiprocessed simulation/training loop, dispatched SLURM Batch Jobs in Hummingbird Computer Cluster
- Taught new research assistants how to use the framework, docker, and python
- Made histograms to identify expected values for Higgs boson

UCSC March 2019 - June 2019
Teachers Assistant *Santa Cruz, CA*

- TA for Physics Class on Pressure, Buoyancy, basic fluid mechanics
- Graded Papers, Reviewed common mistakes, reported and updated quiz curriculum

UCSC, Learning Support Services [LSS] August 2018 - August 2020
Tutor *Santa Cruz, CA*

- Taught Introduction to Data Structures, Electromagnetism, Mechanics, and Thermodynamics
- Mentored in individual and group settings

Private Tutor August 2018 - August 2020
Self Employed *Santa Cruz, CA*

- Taught Precalculus, Physics, and Piano
- Assigned learning targets, supervised children

PROJECTS

Rocket League Research Labs December 2022 - Present

- Scraped Championship level Rocket League Replay files to train an autoregressive decoder to predict what sequenes of moves a lower level ranked player could have made in their replay file
- Bought Unreal Engine 5 Rocket League Remake assets, rendered custom shapes and visualizations
- Uses differential geometry to calculate set of all possible intersection points
- Intersection points do not depend on initial impulse, allowing predictions to be made before impulse
- Draws parabolas with increasing conal radiuses to predict possible ball locations after an initial impulse
- Visualizes positioning as probability denisty map from play positions, velocities, and momentum centered around ball

CUDA Interview Assignment with Vorticity, inc October 2023 - October 2023

- Implemented linear algebra functions in C CUDA, applied test inputs by initializing buffers on the host and loading to the device

Text To Speech [TTS] Voice Cloning April 2023 - July 2023

- Conducted literature review on Tortoise TTS to further study architecture of modern AI networks
- Researched Vector Quantized Variational Auto Encoders [VQ-VAE], Autoregressive Decoders, Contrastive Language-Voice Pretrained Transformer [CLVP], Denoising Diffusion Probability Models [DDPM], Tokenizers, and Vocoders
- Setup bash pipeline to preprocess audio files to clone voices

Botler - Personal Discord Bot February 2023 - Present

- Made discord bot to play audio files over discord

· Setup AWS server, then self hosted botler instance

· Integrating Text Generation and TTS AI services

IP Camera Hacking November 2022 - December 2022

· Assisted in Anonymous operation to monitor war crimes in Russian occupied Ukraine, to report to International Criminal Court

· Used shodan to determine what type of a device an ip address was in Russian occupied Ukraine

· Generated list of IP Addresses of vulnerable IP Cameras

Multithreaded HTTP Server July 2020 - August 2020
Senior Project *Santa Cruz, CA*

· Made Multithreaded HTTP Server with health monitor, load balancer, and no `FILE *` pointers; in C