YANALL BOUTROS

1-530-591-3833 ⋄ YanallBoutros@ProtonMail.com ⋄ Yanall-Boutros.github.io ⋄ Chico, CA

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

TECHNICAL STRENGTHS

Languages:	Python, C/C++, C#, Bash, Powershell, Tex, HTML, NodeJS, Haskell, Perl, Nix
Frameworks:	TensorFlow, PyTorch, Scikit-HEP, Numpy, Matplotlib, Pandas, Unreal Engine
Infrastructure:	GNU/Linux, Unix, Windows, PostgreSQL, Kafka, Docker, Git, Jira, Ansible, NixOS
Mathematics:	Scientific Communication, Modeling, Statistics, Artificial Intelligence, Simulations
Data Science:	Natural Language Processing [NLP], Machine Learning, Predictive Analytics

EXPERIENCE

NelBear Studios LLC April 2024 - Present
Founder *Chico, CA*

- Prototyped core physics movement component for an Unreal Engine 5.4.3 OpenXR Multiplayer VR Game by programming constraints and limits in a UE Physics Asset Blueprint
- Generated unique chord progressions by training a pytorch transformer encoder decoder model on sequences from 1400 Jazz Standards, verified by recording the samples on an electric piano
- Recorded 50 sample beats / tracks by setting up a pickup mic in a violin, and recording myself on the piano and violin with a BOSS RC-3 loop pedal

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

- Conducted code review by reviewing merge requests on tasks and issues assigned by Jira
- 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
- Consulted on feasibility of AI integration and modern data science techniques to project goals

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

- 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 60% Word Error Rate [WER] for Automatic Speech Recognition [ASR] by replacing Dragon with Vosk
- Improved WER by 10% by switching from Vosk to a transformer based model, verified by calculating the levenshtein distance between inferred and human corrected transcriptions
- 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 domain 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 Kafka/PostgreSQL Producer/Consumer in Component Health System, ASR/NLP/CV tools

Independent Contractor August 2021 - July 2022
Software Engineer *Chico, CA; Baltimore, MD*

- Made accelerated mask creation tool with ResNext FPN TensorFlow AI application for FSN
- Setup dedicated local Jax/Dalle/Imagegen server, researched Text → 2D → 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