|  |  |  |
| --- | --- | --- |
| **David BLACK** |  |  |
|  |  |
| [github.com/bballdave025](https://github.com/bballdave025) | [linkedin.com/in/thedavidwblack](https://linkedin.com/in/thedavidwblack) | | +1 801-376-4520 | | [thedavidwblack@gmail.com](mailto:thedavidwblack@gmail.com) |
| [stackexchange.com/users/8693193](https://stackexchange.com/users/8693193) | | 75 Peregrine Ln. #8 Bountiful, UT 84010 USA | |

MOTIVATION

I have always wanted to be a scientist at the frontier of research that helps people. In a Machine Learning Research and/or Development position, I can combine this desire with three more of my foremost passions & talents – math, programming, and research. I look forward to an opportunity that will allow me to use my curiosity, initiative, desire to learn, & creativity freely.

HIGHLIGHTS

» Over 5 years' experience in cutting-edge machine-learning research for automated

handwriting & speech recognition

» 11+ years' experience in computation research & software design

» At least 6 months' experience in each of the following roles

» Oversight of data flow & manipulation at a particle accelerator

» Simulation of physical processes to compare with particle-collision data (27 publications)

» Volunteer work in a university setting doing computer-vision research.

» Volunteer work in a nonprofit genealogy organization; programming the underlying shape-detection algorithms for Asian-language transcriptions & participating in model creation & coding. Used mostly Keras/TensorFlow but learned from PyTorch implementations.

SKILLS

» Initiative & excitement in finding new problems & solutions

» Data compilation, creation, curation, annotation, assimilation, appreciation, utilization,

interpretation, & dissemination

» Handling the flow of large data-collection efforts in academic- & industry-related research

» Kindness, likeability, social aptitude, & the sincere desire to interact with & help others

» Using Linux® tools & the pipeline of text-based input & output to gather, parse, sort,

& generally, manipulate data; I love using sed, awk, grep, find, sort, shuf, or related tools to implement solutions that need to be quick, robust, or both

» Implementing shell-script solutions – usually bash shell, but also in other major shells – into

Python to allow increased portability and, when required, object-oriented programming.

» Communication with business executives & legal counsel to show the financial implications

& ethical concerns, respectively, accompanying research projects

» Working with other departments including Software & Automation Engineering, UX design, QA, & management

» Finding, reading, & understanding scientific papers as well as applying their principles to set

up experiments & to create practical implementations

RELEVANT WORK EXPERIENCE

**Speech Science & Machine Learning (AI Labs) Team Member** *(May 2018 – March 2023)*

*CaptionCall / Sorenson Communications (Salt Lake City, UT)*

» As part of the original team, helped set priorities, determine direction, choose tools, &

collaborate closely with the other original team members.

» Taught other members of the team to use Linux & Linux-type implementations; debugged & adapted package builds & software for our tools & needs || bash, C, C++, make, Perl

» Automated scoring || wrote Python wrapper for NIST's sclite executable

» Normalization of different human & machine transcribers' output so as to have only

words as pronounced || Designed, developed, distributed, & maintained the normalization tool as a Python package || setuptools, pip

» Used a Microsoft Azure version control product at the start, but used git for the past three years; participated with an Agile methodology with our focus turning to Jira & Confluence

» Performed dataset management for purchased, open-source, & created datasets

» Silence detection & removal to allow more words per unit time, implementing HMM &

SVM models || pyAudioAnalysis

» Iterative break testing for changes in the normalization software; used my knowledge of

the intricacies of language & its representation along my significant experience to create

robust testing material

» Solved the majority of file encoding & decoding issues for text & audio files

» Participated in a recent broadening of ML tools based on speech; attended informational

& demo meetings, particularly concerning end-to-end, real-time ASR & concerning conversation summarization along with question answering using transformers

» U.S. Patent No. 11017778

**Data Analyst, Contractor** *(July 2016 – July 2017)*

*FamilySearch (Salt Lake City, Utah)*

» Found, prepared, analyzed & helped facilitate annotation of records used for handwriting recognition; discussed algorithms & analyzed data patterns with the research team; bridged with FamilySearch volunteers & their Asian-language effort || Models included Connectionist Temporal Classification (CTC), U-Net, Various NLP Networks, Transformers

» Data curation, review, selection, & format normalization for a corpus of ~20k document pages in 12 languages that cover patron demand || Python, Perl, Unicode, & utf-8

» Software design for data annotation. Worked closely with volunteer annotators who used the software on which I worked. This software included resources for image classification, image segmentation, segment classification, & transcription. Allowed major acceleration of the zoning & classification || Java, Windows CMD Scripting

» Software design for evaluating Asian language OCR software || Java, JNI (Java Native Interface), C++, NIST's sclite

» Testing for encoding compatibility, annotation validity, & degree of task completion; included finding files whose annotation had been missed, left incomplete, or corrupted || bash tools, Python, Java

» U.S. Patent No. 11 017 778, Issued May 25, 2021 || One of five inventors

**Undergraduate & Graduate Research Assistant** *(2008 – 2013)*

*Brigham Young University (Provo, UT)* *(2008 – 2010)*

*University of California at Riverside (Riverside, CA)* *(2010 – 2013)*

» Data flow management & real-time data quality assurance at the Relativistic Heavy Ion

Collider || Postgre SQL, csh/tcsh, ksh, zsh, bash, Perl, PHP, HTML || UCR

» Simulation of physical processes to compare theory & experiment|| C++, large & custom C++ libraries & wrappers || Monte Carlo methods || Circuit building & debugging || UCR

» Computation of electromagnetic fields & intensities || MATLAB®, || BYU

» Maintenance of optics and electronics in a high-intensity laser lab || Design of further optics for experiments with the laser || BYU

» Application of various mathematical models in simulation and computation || UCR & BYU

EDUCATION

» **Master of Science, Physics**, University of California at Riverside, 2012

» **Bachelor of Science, Physics**, Brigham Young University, 2010

Minors: **French, Mathematics**