✦ancestry®   **Ancestry Careers**   Homepage (Ancestry.com)   Sear

← **Back to Job Posting**

Data Scientist, CV/NLP

My Information   My Experience   Voluntary Disclosures   **Review**

Review

My Information

How Did You Hear About Us?
Ancestry's Career Site

Have you previously been employed by Ancestry? If you are a current employee, please apply via the internal career site.
No

Legal Name
David Black

I have a preferred name
No

Address
75 Peregrine Ln APT# 8
Bountiful, UT 84010
United States of America

Email

Back	Submit

ancestry ® **Ancestry Careers**

My Experience

Work Experience

Work Experience 1

Job Title

Speech Sciences and Machine Learning (AI Labs) Team Member

Company

CaptionCall / Sorenson Communications

Location

Salt Lake City, UT

I currently work here

No

From	To
05/2018	03/2023

Role Description

CaptionCall is a provider of captioning services on special telephones available to hard-of-hearing individuals through the Americans with Disabilities Act. They are subcontracted by the FCC, who administer the ADA program for the deaf and hard-of-hearing. CaptionCall is an outgrowth of Sorenson Communications, who provide video-relay services for the deaf through American Sign Language.

Responsibilities of my role can be described as follows.

- Perform literature review and contribute to decisions about research priorities.
- Build a variety of tools for automated speech recognition, natural language processing, metrics for performance evaluation, models underlying speech recognition - including machine-learning models and physical models, and other subjects as the need arises.
- Participate in the creation of Intellectual Property, including patents.
- Find, create, and curate datasets. Take the time to become familiar with the datasets and clean them up as needed so they can be used by other projects.
- Write proof-of-concept code based on your research, which will be either passed on to engineering or be developed by the research team in conjunction with engineering.
- Use and create machine-learning, speech-to-text, data manipulation, and other types of tools, workbenches,

ancestry ®   **Ancestry Careers**

performance and analysis.
- Leverage the strengths of both human and machine transcribers in real time to provide increased accuracy.
- Provide quantitative measurement of processes created to present to the research team, senior management, shareholders, government and other organizations who advocate for the hard-of-hearing.
- Determine which groups, including protected groups, are better or worse served by human transcription, machine transcription, or combinations of both. Create a dataset of phone audio for a variety of populations who might be worse served by any of the three transcription strategies. Handle issues of legal consent from and protection of these individuals.
- Create call-recording software with the help of a freelancer to gather data for various needs.
- Develop a robust method of silence detection and removal, then of splicing the mostly-spoken sections into audio files with a specific time length. Administer these audios to different transcribers and analyze the results.
- Broaden the marketable products and services available with machine learning methods and models that relate to speech and language with a special focus on serving underrepresented groups such as the Deaf and Hard-of-Hearing communities.
- Use versioning control, especially git, to work in an Agile software-development environment.

Work Experience 2

Job Title
Document Specialist

Company
Church of Jesus Christ of Latter-day Saints, Church History Department

Location
Granite Mountain Records Vault, Sandy, UT

I currently work here
No

From   **To**
08/2017   04/2018

Role Description
The Granite Records Mountain Vault section of the Church History Department of the Church of Jesus Christ of Latter-day Saints has the responsibility of storing historical documents for the church as a whole as well as many large collections of materials useful for family history. One major project has been the digitization of the family history microfilm collection for FamilySearch.

ancestry **Ancestry Careers**

- Serve as a paleography and language-recognition resource for other workers

Work Experience 3

Job Title

Data Analyst - Contractor

Company

Family Search International

Location

Salt Lake City, UT

I currently work here

No

From	To
07/2016	07/2017

Role Description

FamilySearch is a non-profit organization that provides research and document services in family history. They provide research assistance online and in-person, obtain historical documents for preservation and/or digitization then allow access to many of them, and work to make an ever-increasing documents easily searchable and accessible.

Responsibilities of my role and qualifications for it can be described as follows.

- The primary motivation for the creation of the job was the preparation of data truth and test sets.
- More specifically, the job was to work with a records and data team to provide the Advanced Technology Research team with any data, documents, annotation, analysis, summaries, and any other support they needed as projects went on.
- Enthusiasm about mining, manipulating, and annotating data from multiple languages.
- Working knowledge and competence in several non-English Language as well as of languages and language patterns.
- Ability to grasp new computational and technical topics as well as self-sufficiency to program in several languages - Java, Perl, and C++, especially - in order to manipulate data.
- Ability to work both with Windows systems as well as Cygwin or Linux.
- Accomplish tasks as an individual or as a team member under time constraints. Brainstorm and identify best

ancestry **Ancestry Careers**

documents through building new Deep Learning models.
- Identify documents as well as types of documents that will allow the expansion of these processes to work on multiple languages.
- Be flexible and organized, since daily activities, including those listed below, vary depend on the latest data needs of the research team.
    - Write quick applications that can, among other things
        - Aid in data review and selection
        - Manipulate catalog source lists in CSV and spreadsheet form to record prioritized microfilm or digital image selection list.
        - Analyze data that volunteers have prepared as part of the data preparation process, including deciding metrics for performance and progress tracking
- Perform some of the non-programming work of the data team and its volunteers, including: tagging, zoning, line segmenting, etc. as a way of vetting the quality of work done as well as to help complete part of the needed annotations.
- Locate document images that are potentially pathological for the algorithm, that is, documents with objects in them that could be confused for writing, layouts and certain annotations in the original document that could seem ambiguous to the model while they wouldn't be to a knowledgeable human transcriber, and other things that could confuse the algorithm and/or cause it to struggle.
- Handle the decoding of text files, which can potentially be sent from different regions, older computer software, or those unacquainted with standard encoding of files. Convert all text-file transcriptions to the utf-8 encoding of Unicode.
- Evaluate the text-recognition capabilities of third parties whose technology could potentially give higher accuracy or give a new capability or a capability in a language we have not yet worked on. These could be in other programming languages, in which case an interface should be created or used. Decide on metrics to use for the scoring of such third-party programs.
- Analyze datasets, whether internal that might be part of new indexing projects or those which we might want to acquire, to analyze factors contributing to the decision of whether and how to use them. Factors could include technical details, financial and legal details of the datasets, current or past political factors that could make data sensitive, linguistic and naming details, etc.
- As requested, do visualizations of the data. This could range from viewing the Deep Learning algorithm's output as it would be seen on the original documents to plots of needs and progress.
- Help create guidelines which are concise but cover sufficient detail that the volunteers can use them as a guide in image segmentation and transcription.

Education

Education 1

ancestry ®   **Ancestry Careers**

Physics

Overall Result (GPA)

3.344

From	To (Actual or Expected)
2010	2012

Education 2

School or University

Brigham Young University

Degree

Bachelor's or 4 Year Degree

Field of Study

Physics

Overall Result (GPA)

3.26

From	To (Actual or Expected)
2004	2010

Skills

Agile
Agile Framework
Agile Methodology
Agile Principles
American English
American Sign Language (ASL)
Apache Subversion (SVN)
Apple Mac OS
Applications Programming
Applied Linguistics
Applied Research

ancestry **Ancestry Careers**

College Teaching
Communication
Computer Programming
Computer Science
Computer Vision
Creative Concept Design
Creative Design
Creative Strategies
Creativity
Data Acquisition
Data Analysis
Data Annotation
Data Collection Management
Data Curation
Data Extraction
Data Format
Data Insights
Data Management
Data Mining
Data Mining Software
Data Modeling
Data Modeling Tools
Data Organization
Data Processing
Data Representation
Data Retrieval
Data Science
Datasets
Data Standards
Data Types
Data Visualization
Data Visualization Tools
Development
English Grammar
English Language
English Teaching
English to French Translation
English to Russian Translations
English to Spanish Translations
Experimental Research
Family History
Foreign Languages
French Language
French to English Translations
French Translation
French Translations
Functional Programming Languages
Genealogy
Genealogy Research
Germanic Languages
German Language

Information Visualization
Interpersonal Communication
Interpersonal Interactions
Interpersonal Sensitivity
Java
JavaFX
Java Native Interface (JNI)
Javascript Programming Language
Java Software Development
Java Virtual Machine (JVM)
Keras
Laboratory Research
Language Identification
Languages
Language Theory
Latin Language
Leadership
Linguistics
Linux
Linux Bash
Linux Bash Scripting
Linux Bash Shell
Linux Development
Linux Environments
Linux Programming
Linux Shell Scripting
Machine Learning
Machine Learning Algorithms
Markup Languages
Mathematical Analysis
Mathematical Modeling
Mathematics
MATLAB
MATLAB Language
MATLAB Programming
Microsoft Azure
Microsoft Office
Microsoft Office Macros
Microsoft Office Programming
Microsoft PowerShell
Microsoft Remote Desktop Services (RDS)
Microsoft Windows
Microsoft Windows Office
Microsoft Word
Modern Languages
Morphometric Analysis
Natural Language Processing (NLP)
Natural Language Processing Tools
Numerical Analysis
Numerical Methods
Numerical Modeling

ancestry **Ancestry Careers**

Palaeography
Pandas Python Library
Parametric Modeling
Perl Programming
PHP (Programming Language)
PHP Scripting
PostgreSQL
Presentations
Procedural Coding
Programming Languages
Python (Programming Language)
Python Datetime Module
Python Matplotlib
Python Numpy
Research Analysis
Researching
Research Labs
Root Cern
Russian History
Russian Language
Russian to English
Russian Translations
Scientific Data Management
Scientific Research
SciPy
Secure Shell (SSH)
Selenium
Selenium Automation
Selenium Webdriver
Shape Analysis
Shell Programming
Shell Scripting
Software Development
Software Documentation
Spanish Language
Spanish to English Translations
Spanish Translations
Speak Foreign Languages
Teaching
Technical Content Writing
Technical Writing
Tensorflow
Text Annotation
Training
Unix
Unix Shell Scripting
VB Scripting
Virtual Machines (VM)
Web Scraping
Wiki Documentation
Windows Command Line

ancestry   **Ancestry Careers**

Resume/CV

[PDF]   BLACK-David_Resume_Ancestry_ML_CV_NLP_2023-03-08.pdf
294.97 KB

Websites

Add any relevant websites. Format should be: http://

Websites 1

URL

https://linkedin.com/in/thedavidwblack

Websites 2

URL

https://www.bballdave025.com

Websites 3

URL

https://github.com/bballdave025

Websites 4

ancestry **Ancestry Careers**

U.S. EQUAL EMPLOYMENT OPPORTUNITY INFORMATION (Completion is voluntary and will not subject you to adverse treatment)

Please select the ethnicity which most accurately describes how you identify yourself.

White (Not Hispanic or Latino) (United States of America)

Please select your gender.

Male

Please select the veteran status which most accurately describes your status.

I AM NOT A VETERAN

Terms and Conditions

I willingly accept the Terms and Conditions for submitting an application with Ancestry.com.

Yes

Follow Us

▸

[Ancestry's Recruitment Privacy Statement](#)

workday.

© 2023 Workday, Inc. All rights reserved.