

TAYLOR ARCHIBALD, Ph.D.

1973 Montana Ave, Provo, UT 84606 • (801) 477-7638 • TaylorNArchibald@gmail.com
taylorarchibald.com • [GitHub](#) • [Google Scholar](#)

PROFESSIONAL EXPERIENCE

ANCESTRY

Applied AI Scientist

Lehi, UT

May 2022 – Present

- **GenAI Transformation:** Pioneered the transition to MLLM-based transcription using state-of-the-art foundation models; proved superior accuracy over legacy OCR while reducing engineering maintenance and deployment costs.
- **High-Volume Production:** Technical lead for the "Death Records" collection (7 million documents). Optimized throughput and confidence scoring, successfully delivering the massive collection to production.
- **Novel Architectures (PRISM):** Developed the PRISM framework, combining log-probabilities with verbalized self-confidence to route high-risk documents to human review. Allowed usage of "cheap" ensembles to outperform SOTA models at **10x lower cost**.
- **Accuracy Optimization:** Implemented a Test-Time Augmentation (TTA) framework resulting in a **4% absolute improvement** in field-level transcription accuracy.
- **Synthetic Data Pipeline:** Led creation of a pipeline to generate millions of synthetic documents to augment layout and handwriting training data for French BMD collections.

NEURAL NETWORKS AND MACHINE LEARNING LAB

Research Assistant

Provo, UT

Sept 2018 – Dec 2025

- **SOTA Segmentation:** Published DELINE8K (ICDAR 2024), a state-of-the-art approach for document semantic segmentation using synthetic data generation.
- **Handwriting Recovery:** Published TRACE (ICDAR 2021), a differentiable methodology for recovering strokes from offline text images.
- **NLP Applications:** Developed *ScamBaiterBot*, a chatbot that engaged in 5,000+ adversarial conversations with scammers to waste their time and gather threat intelligence.

BATES WHITE ECONOMIC CONSULTING

Senior Consultant / Consultant

Washington, D.C.

Aug 2013 – Aug 2017

- **Litigation Analytics:** Led data analysis teams for 8 major settlement negotiations for pharmaceutical manufacturers accused of price misrepresentation, resulting in client savings in excess of **\$100 million**.
- **Big Data Processing:** Managed analysis assessing health insurance monopsony power using over 3 billion medical claims (1+ TB) utilizing Python, SQL, Stata, and SAS.
- **Leadership:** Served on the Technology Advisory Council and acted as primary trainer for Stata/SQL.

EDUCATION

BRIGHAM YOUNG UNIVERSITY

Ph.D. in Computer Science (GPA: 4.00, GRE: 334)

Provo, UT

Dec 2025

- Focus: Deep Learning, Document Understanding, Synthetic Data.

BRIGHAM YOUNG UNIVERSITY

M.Sc. in Computer Science

Provo, UT

Dec 2020

BRIGHAM YOUNG UNIVERSITY

B.A. Mathematics, Economics (Minor: Philosophy) (GPA: 3.91)

Provo, UT

April 2013

SELECTED PUBLICATIONS

- **Improving MLLM Historical Record Extraction with Test-Time Image Augmentations.** *Pending Submission to ICDAR.*
- **DELINE8K: A Synthetic Pipeline for Semantic Segmentation of Historical Documents.** *ICDAR 2024.*
- **Leveraging Semantic Segmentation Masks with Embeddings for Fine-Grained Form Classification.** *DAS 2024.*
- **TRACE: A Differentiable Approach to Line-Level Stroke Recovery.** *ICDAR 2021.*