CURRICULUM VITAE

Harvey Dam
harvey.dam@utah.edu

1 EDUCATION

- 2023–present: Ph.D in Computer Science, University of Utah, In Progress.

  I do mostly machine learning research.
- 2020–2022: MS in Computer Science, Georgia Institute of Technology, 4.0 / 4.0.

  A computer science master's program. I took mostly machine learning, robotics, and video game courses.
- 2013–2014: MA in World Languages, University of Utah, 4.0 / 4.0.

  A master's program about how to make it easier for humans to acquire languages, including a residency-like period where I taught English and Mandarin in a public high school.
- 2008–2013: BA in Chinese, Minor in Music, University of Utah, 2013, 3.4 / 4.0.

  Chinese (primarily Mandarin) proficiency, Chinese culture, and China's position in the world. Music was for a change in scenery.

1.1 CERTIFICATES

- Credential of Readiness, Harvard Business School Online, 2017, Pass with High Honors.

  A certificate from taking three courses: economics, accounting, and analytics.

2 EMPLOYMENT

- 2020–present: Research Assistant, University of Utah.

  See Publications.
- 2014–2018: Data Team Manager, Outlier Linguistics.

  I was an early employee of this startup that collected, processed, and licensed linguistic data to other entities. I hired and managed a team of image labelers. I also did handwriting research and contributed brush writing to posters about semantic components in Chinese characters.

3 PUBLICATIONS

3.1 PAPERS

2023: Tripti Agarwal, **Harvey Dam**, P. Sadayappan, Ganesh Gopalakrishnan, Dorra Ben Khalifa, and Matthieu Martel. What Operations can be Performed Directly on Compressed Arrays, and with What Error? *Published in the 2023 International Workshop on Data Analysis and Reduction for Big Scientific Data. Best paper.*

> In response to the rapidly escalating data movement-related costs of computing with large matrices and multidimensional arrays, several lossy compression methods have been developed that help reduce the volume of data moved. Unfortunately, all these methods require the data to be decompressed before operating on it. In this work, we develop a lossy compressor for arbitrary-dimensional arrays called PyBlaz that supports a dozen non-trivial operations directly on compressed data while also offering good compression ratios. PyBlaz is based on the PyTorch

framework, and thus can be run on CPUs or GPUs without any code changes. We evaluate the efficacy of PyBlaz on datasets originating in three applications: comparing shallow-water simulation implementations, measuring statistics from MRI images, and detecting the scission point in plutonium fission data. Our results demonstrate that PyBlaz's compressed-domain operations achieve good scalability while incurring controllable errors. To our best knowledge, this is the first such lossy compressor that supports compressed-domain operations in the realm of handling arbitrary-dimensional scientific datasets.

2023: Mary Hall, Ganesh Gopalakrishnan, Eric Eide, Johanna Cohoon, Jeff M. Phillips, Mu Zhang, Shireen Y. Elhabia, Aditya Bhaskara, **Harvey Dam**, Artem Yadrov, Tushar Kataria, Amir Mohammad Tavakkoli, Sameeran Joshi, and Mokshagna Sai Karanam. An NSF REU Site Based on Trust and Reproducibility of Intelligent Computation: Experience Report. *Published in the 2023 Workshop on Education for High-Performance Computing*.

This paper presents an overview of an NSF Research Experience for Undergraduate (REU) Site on Trust and Reproducibility of Intelligent Computation, delivered by faculty and graduate students in the Kahlert School of Computing at University of Utah. The chosen themes bring together several concerns for the future in producing computational results that can be trusted: secure, reproducible, based on sound algorithmic foundations, and developed in the context of ethical considerations. The research areas represented by student projects include machine learning, high-performance computing, algorithms and applications, computer security, data science, and human-centered computing. In the first four weeks of the program, the entire student cohort spent their mornings in lessons from experts in these crosscutting topics, and used one-of-a-kind research platforms operated by the University of Utah, namely NSF-funded CloudLab and POWDER facilities; reading assignments, quizzes, and hands-on exercises reinforced the lessons. In the subsequent five weeks, lectures were less frequent, as students branched into small groups to develop their research projects. The final week focused on a poster presentation and final report. Through describing our experiences, this program can serve as a model for preparing a future workforce to integrate machine learning into trustworthy and reproducible applications.

2023: **Harvey Dam**, Vinu Joseph, Aditya Bhaskara, Ganesh Gopalakrishnan, Saurav Muralidharan, Michael Garland. Understanding the Effect of the Long Tail on Neural Network Compression. In *2023 Workshop on Sparsity in Neural Networks*.

Sparsified and pruned neural networks output different things from unpruned ones. We estimated the influence of each training example on each test example, where influence is the expected accuracy gain from training with that training example vs training without it. We pruned different image classifiers while encouraging alignment of logits using a loss function that combines the traditional cross-entropy loss with some penalties for the mean squared error between the logits and for the difference in the logit of the true class. We found that the amount that test examples that were correctly classified by the pruned models were influenced differently, sometimes significantly, from examples that were incorrectly classified.

2021: Xiaofang Chu, **Harvey Dam**, Connor Shaffer. Automatic Speech Recognition for Mispronunciation Detection and Diagnosis. *Unpublished manuscript*.

We tried to create an interactive generative model that listens to language learners' speech and gives feedback about their pronunciation. We stopped when we had a speech-to-phoneme model, and my coauthors did not want to continue.

3.2 SOFTWARE

2022: PyBlaz: an arbitrary-dimensional, GPU-accelerated floating-point array compressor that allows various operations such as addition, mean, variance, and L_2 norm to be performed on compressed arrays without decompression. `github.com/damtharvey/pyblaz`

3.3 Books

2018: **Harvey Dam**. *Regular Script Graphemics*. ISBN: 978-0692128534.

> A book about the features and execution of the newest and most popular Chinese script style. Before this book, this area was dominated by calligraphers. To my knowledge, this is the first linguist's description of the subject.

4 Volunteering

2020, 2021, 2023: *Code In Place* Section Leader, Stanford University.

> I was a section leader for CS106A: Programming Methodologies. Every week, I lead a 40-minutes to 1 hour meeting with about 10 students in my section. During the meeting, I reviewed content covered during lecture and facilitated programming exercises from the CS106A curriculum, encouraging good programming practices. I also created additional materials to cover what I felt was inadequately explained or inaccurate.

2017–2019: Section Leader, Volunteers of America Utah.

> I taught Python and economics every week for an hour to high school-aged students. My Python classes were usually meant to demonstrate the kinds of things that one can do with any programming language, so that the students are more likely to study programming at a more convenient time. My economics classes consisted of a case study, discussion, and occasional cold calls.

5 Skills

- Machine learning: Python, PyTorch, C, CUDA
- Math: Linear algebra, calculus, statistics
- Video games: Blender, Unity, C#
- Other: LaTeX, linguistics, Mandarin