David Kartchner

Researcher + Entrepreneur (ML + Biomedicine)

I research how to enable **natural language processing** on new and dynamic problems by developing ai-driven models for scalable data labeling powered by active learning and weak supervision. I apply these technologies to **healthcare** and **biomedicine** to enable clinical researchers to better understand disease etiology and improve care delivery.

I have collaborated with researchers, developers, and clinicians while working at Facebook, GSK, Recursion Pharmaceuticals, and Intermountain Healthcare.

davidkartchner.com
david.kartchner@gatech.edu
CV PDF
David-S-Kartchner
@davidkartchner
Google Scholar

Education

Aug 2018 - Dec 2023 **Ph.D. in Computational Science & Engineering**
Georgia Institute of Technology, Atlanta, GA
Advisor: Cassie Mitchell, Co-advisor: None
Thesis: *Extracting and Structuring Information for Clinical Meta-Analysis and Drug Repurposing*
Committee: Cassie Mitchell, Chao Zhang, Duen Horng "Polo" Chau, Jon Duke, Daniel Domingo-Fernández

2017 - 2018 **M.S. in Mathematics**
Brigham Young University, Provo, UT
Thesis: *ActuarAI: Machine Learning Models for Patient Disease Forecasting and Representation*
Committee: Jeffrey Humpherys, Tyler Jarvis, David Wingate
GPA: 4.00/4.00
Thesis

2010 - 2016 **B.S. in Applied & Computational Mathematics**
Brigham Young University, Provo, UT
Thesis: *Walking the Walk: An Exploratory Analysis in Biometric Gait Recognition*
Magna Cum Laude, University Honors Overall GPA: 3.96/4.00 Applied and Computational Mathematics Emphasis (ACME)
Thesis

Industry Experience

Sept 2022 - Present **Glassbox Health**, Atlanta, GA
Co-Founder, CTO
Building an LLM-based assistant to provide personalized navigation of medical bills and healthcare costs

Summer 2022 **Enveda Biosciences**, Boulder, CO
Data Science Intern, Knowledge Graph
Mentor: Daniel Domingo-Fernandez, David Healey, Joe Davison
Performed systematic survey + implementation fo 20+ entity linking NLP models to improve accuracy evidence-based compound prioritization

Summer 2021 **Facebook**, Menlo Park, CA
Applied Research Science Intern, Enterprise Product Applied Research
Mentor: Minhazul Islam Sk
Designed and trained transformer-based semantic search document retrieval system to improve efficiency of customer support agents

Summer 2020 **GlaxoSmithKline**, Philadelphia, PA
Research Intern, AI/ML Engineering
Mentor: Anne Cocos
Built model jointly embed free-text entity mentions with structured entity knowledge graph for 30M research articles/abstracts and KG with 5M edges. Developed end-to-end pipeline to download, preprocess, and identify high-quality entity links for biomedical entities in 30M research articles. Engineered parallel model training workflow on distributed supercomputing cluster utilizing 10,000+ CPU cores and dozens of GPUs.

Nov 2018 - Aug 2019 **Padsplit**, Atlanta, GA
Data Science Consultant, Data Research
Created credit scoring model and interactive job density visualizations to move into new domestic markets.

Summer 2018 **Recursion Pharmaceuticals**, Salt Lake City, UT
Data Science Intern, Machine Learning
Mentor: Andrew Blevins
Developed and deployed recommender system to infer biological mechanism of action and repurposing potential of 1M+ compounds

May 2016 - May 2018 **Intermountain Healthcare**, Salt Lake City, UT
Data Science Intern, Population Health Analytics
Mentor: Andy Merrill
Built and deployed models to forecast individual patient risk of chronic disease onset and long-term complex care from EHR and environmental data. Published in IEEE ICHI (2017) and AJRCCM (2018).

Summer 2015 **Capital One**, McLean, VA
Business Analyst Intern,
Analyzed public loan data to predict consumer default on personal loans.

Academic Research Experience

Aug 2019 - Present Aug. 2016 **Georgia Institute of Technology**, Atlanta, GA

Graduate Research Assistant, Laboratory for Pathology Dynamics
Advisor: Cassie Mitchell
Member of the Laboratory of Pathology Dynamics where we use machine learning to build tools that identify and prioritize cures and optimize care for neurodegenerative diseases.

Aug 2018 - May 2019 **Georgia Institute of Technology**, Atlanta, GA
Graduate Research Assistant, School of Computational Science and Engineering
Mentor: Jimeng Sun
Conducted research in predicting chronic disease outcomes from electronic health records (EHR) and free-text clinical notes.

Jan 2017 - Aug 2018 Jan. 2013 **Brigham Young University**, Provo, UT

Graduate Research Assistant, Department of Mathematics
Advisor: Jeffrey Humpherys
Developed models to predict individual onset of chronic conditions from patient electronic health records (EHR). Published in IEEE ICHI (2017, 2018).

Jun 2014 - Apr 2018 **Brigham Young University**, Provo, UT
Teaching Assistant & Lab Instructor, Department of Mathematics
Mentor: Tyler Jarvis (primary), Brigham Frandsen, David Sims, Joseph Price, Stephen Humpheries
Taught year-long, weekly programming lab on data analysis and intensive summer bootcamp on Markov Chain Monte Carlo (MCMC). Developed machine learning curriculum and automated grading software. Additionally taught recitations for abstrat algebra, econometrics, statistics, and microeconomics.

Honors and Awards

2018 National Science Foundation GRFP Honorable Mention
Learning to Prescribe Optimal Disease Treatment via Machine Learning

2015 Dean and Helen Robinson Scholarship
Scholarship given to outstanding undergraduates in mathematics for Putnam Mathematics competition

2016 BYU University Honors
Awarded to undergraduates who write a thesis complete requirements in leadership, service, and cross-disciplinary scholarship.

2010-2016 BYU Heritage Scholarship
Full-tuition merit based scholarship for incoming students

2011 Amberly Rupp "Circle of Honor" Essay Contest Award
1st-place in university-wide essay contest

2010 National Merit Scholarship
Merit-based scholarship awarded top <1% of incoming university students

Publications

Selected: Latest & Greatest

Zero-Shot Information Extraction for Clinical Meta-Analysis using Large Language Models
David Kartchner, Irfan Al-Hussaini, Selvi Ramalingam, Olivia Kronick, Cassie Mitchell
22nd Workshop on Biomedical Natural Language Processing (BioNLP). Toronto, Canada, 2023.
Project PDF BibTeX

BioSift: A Dataset for Filtering Biomedical Abstracts for Drug Repurposing and Clinical Meta-Analysis
David Kartchner, Irfan Al-Hussaini, Haydn Turner, Jennifer Deng, Shubham Lohiya, Prasanth Bathala, Cassie Mitchell
46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). Taipei, Taiwan, 2023.
Project BibTeX

Rule-Enhanced Active Learning for Semi-Automated Weak Supervision
David Kartchner, Davi Nakajima An, Wendi Ren, Chao Zhang, Cassie Mitchell
AI (AI). Online, 2022.
Project PDF BibTeX

Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19
Kevin McCoy, Sateesh Gudapati, Lawrence He, Elaina Horlander, David Kartchner, Soham Kulkarni, Nidhi Mehra, Jayant Prakash, Helena Thenot, Sri Vivek Vanga, Abigail Wagner, Brandon White, Cassie Mitchell
Pharnaceutics (Pharm). Online, 2021.
Project PDF BibTeX DOI

Journal

J3 **Rule-Enhanced Active Learning for Semi-Automated Weak Supervision**
David Kartchner, Davi Nakajima An, Wendi Ren, Chao Zhang, Cassie Mitchell
AI (AI). Online, 2022.
Project PDF BibTeX

J2 **Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19**
Kevin McCoy, Sateesh Gudapati, Lawrence He, Elaina Horlander, David Kartchner, Soham Kulkarni, Nidhi Mehra, Jayant Prakash, Helena Thenot, Sri Vivek Vanga, Abigail Wagner, Brandon White, Cassie Mitchell
Pharnaceutics (Pharm). Online, 2021.
Project PDF BibTeX DOI

J1 **Short-Term Elevation of Fine Particulate Matter Air Pollution and Acute Lower Respiratory Infection**
Benjamin D. Horne, Elizabeth A. Joy, Michelle G. Hofmann, Per H. Gesteland, John B. Cannon, Jacob S. Lefler, Denitza P. Blagev, E. Kent Korgenski, Natalie Torosyan, Grant I. Hansen, David Kartchner, C. Arden Pope III
American Journal of Respiratory and Critical Care Medicine (AJRCCM). New York, NY, USA, 2018.
Project PDF BibTeX DOI

Conference

C5 **BioSift: A Dataset for Filtering Biomedical Abstracts for Drug Repurposing and Clinical Meta-Analysis**
David Kartchner, Irfan Al-Hussaini, Haydn Turner, Jennifer Deng, Shubham Lohiya, Prasanth Bathala, Cassie Mitchell
46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). Taipei, Taiwan, 2023.
Project BibTeX

C4 **Denoising Multi-Source Weak Supervision for Neural Text Classification**
David Kartchner, Wendi Ren, Davi Nakajima An, Chao Zhang, Cassie Mitchell
Findings of EMNLP (EMNLP (Findings)). Online, 2020.
Project PDF Video Code BibTeX DOI

C3 **Machine Learning Methods for Diease Prediction with Claims Data**
Tanner Christensen, Abraham Frandsen, Seth Glazier, Jeff Humpherys, David Kartchner
IEEE International Conference on Healthcare Informatics (ICHI). New York City, NY, USA, 2018.
Project PDF BibTeX DOI

C2 **Code2vec: Embedding and Clustering Medical Diagnosis Data**
David Kartchner, Tanner Christensen, Jeff Humpherys, Sean Wade
IEEE International Conference on Healthcare Informatics (ICHI). Park City, UT, USA, 2017.
Project PDF Poster BibTeX DOI

C1 **Cost Reduction via Patient Targeting and Outreach: A Statistical Approach**
David Kartchner, Andrew Merrill, Jonathan Wrathall
IEEE International Conference on Healthcare Informatics (ICHI). Park City, UT, USA, 2017.
Project PDF Poster BibTeX DOI

Workshop

W2 **Zero-Shot Information Extraction for Clinical Meta-Analysis using Large Language Models**
David Kartchner, Irfan Al-Hussaini, Selvi Ramalingam, Olivia Kronick, Cassie Mitchell
22nd Workshop on Biomedical Natural Language Processing (BioNLP). Toronto, Canada, 2023.
Project PDF BibTeX

W1 **ReGAL: Rule-Generative Active Learning for Model-in-the-Loop Weak Supervision**
David Kartchner, Wendi Ren, Davi Nakajima An, Chao Zhang, Cassie Mitchell
Human and Model-in-the-Loop Evaluation and Training Stragegies Workshop, NeurIPS (HAMLETS). Online, 2020.
Project PDF Poster BibTeX

Poster

P6 **Understanding the Link Between COVID-19 and Cardiovascular Disease by Text Mining Biomedical Literature**
Kevin McCoy, Janhvi Dubey, David Kartchner, Dongyu Zhang, Kevin Zhang, Rushda Umrani, Cassie Mitchell
Biomedical Engineering Society Annual Meeting (BMES). San Antonio, TX, USA, 2022.
Project

P5 **Exploring Optimizations to HeteSim for Computing Relatedness in Heterogeneous Information Networks**
Stephen Allegri, Evie Davalbhakta, David Kartchner, Anna Kirkpatrick, Davi Nakajima An, Chidozie Onyeze, Cassie S. Mitchell, Prasad Tetali
American Mathematical Society Joint Meeting on Mathematics (ANA). Seattle, WA, USA, 2022.
Project

P4 **Deep Learning System for Labeling Neurology Text for Predictive Medicine**
Davi Nakajima An, David Kartchner, Dongyu Zhang, Cassie Mitchell
American Neurological Association Annual Meeting (ANA). Online, 2021.
Project

P3 **Literature Based Discovery of Comorbid Hematological Conditions in Chronic Myeloid Leukemia Treatment with Tyrosine Kinase Inhibitors**
Nidhi Mehra, Jeongjin Lee, Helena Thenot, Sparsh Kudrimoti, Brandon White, David Kartchner, Sateesh Gudapati, Jayant Prakash, Vivek Vanga, Cassie Mitchell
Biomedical Engineering Society Annual Meeting (BMES). Online, 2020.
Project

P2 **Unsupervised Ranking of Treatment-Related Infection Risk Factors in Pediatric Acute Leukemia**
Brandon White, Lawrence He, Elaina Horlander, Nidhi Mehra, David Kartchner, Vivek Vanga, Sateesh Gudapati, Tamara Miller, Cassie Mitchell
Biomedical Engineering Society Annual Meeting (BMES). Online, 2020.
Project

P1 **Repurposed Drug Identification for COVID-19 using Literature Relationships and Knowledge Graphs**
Nidhi Mehra, Brandon White, David Kartchner, Helena Thenot, Lawrence He, Elaina Horlander, Sateesh Gudapati, Jayant Prakash, Vivek Vanga, Cassie Mitchell
Biomedical Engineering Society Annual Meeting (BMES). Online, 2020.
Project

Miscellaneous

M1 **Forward Thinking: Building Deep Random Forests**
Kevin Miller, Chris Hettinger, Jeffrey Humpherys, Tyler Jarvis, David Kartchner
https://arxiv.org/abs/1705.07366. 2017.
Project PDF BibTeX

Talks

Accelerating Biomedical Discovery with Knowledge Graphs and Weakly Supervised Learning
May 2022 Georgia Tech PhD Thesis Proposal

Biomedical Information Extraction
Mar. 2021 Brigham Young University, Machine Learning for Health Class

ReGAL: Rule-Guided Active Learning for Deep Text Classification
Oct. 2020 Georgia Tech HotCSE Seminar

Survey of Knowledge Graph Embedding Rechniques
Jul. 2020 GSK AI/ML Group

Extracting Actionable Insights from Biomedical Text
Mar. 2019 Georgia Tech PhD Qualifying Exam Oral Defense

ActuarAI: Machine Learning Models for Patient Disease Forecasting and Representation
Jul. 2018 Brigham Young University Masters Thesis Defense

Walking the Walk: An Exploratory Analysis in Biometric Gait Recognition
Nov. 2016 Brigham Young University Honors Thesis Defense

Press

October 2022 "Chan Zuckerberg Initiative, National Academies Select Cassie Mitchell for Science Diversity Leadership Program," Georgia Institute of Technology

Apr 2018 "Brief Exposure to Tiny Air Pollution Particles Triggers Childhood Lung Infections, Largest Study of Its Kind Finds," Intermountain Healthcare

Teaching

Fall 2022 **Graduate Teaching Assistant**
Georgia Institute of Technology, Atlanta, GA
Intro to Graduate Computing, Instructor: Elizabeth Cherry
Graded homework, held weekly office hours, and mentored student for CSE 6010, an introduction to graduate and parallel computing in C

Summer 2019 **Graduate Teaching Assistant**
Georgia Institute of Technology, Atlanta, GA
Computing for Data Analysis (CX 4240), Instructor: Mahdi Roozbahani
Designed homeworks, graded homework, held weekly office hours, and mentored student on team projects for CX 4240, an undergraduate introduction to machine learning

Spring 2019 **Invited Guest Lecturer**
Georgia Institute of Technology, Atlanta, GA
Data Analytics for Business (MGT 6203), Instructor: Michael Lowe
Presented a week of lectures on web scraping, tweet streaming, and natural language processing for Master's of Analytics program

Aug 2017 - April 2018 **Graduate Teaching Assistant**
Brigham Young University, Provo, UT
Modeling with Data and Uncertainty (Math 323, Math 325), Instructor: Tyler Jarvis
Graded homeworks, taught lectures, designed curriculum, and mentored students on team projects for Math 322 and 324, a rigorous two-semester course on probabilistic mathematics and machine learning

Spring 2017 **Graduate Teaching Assistant**
Brigham Young University, Provo, UT
Abstract Algebra (Math 371), Instructor: Stephen Humpheries

Graded homeworks, held office hours, and reviewed concepts with students for Math 371, an undergraduate abstract algebra course.

Aug 2016 - April 2017 **Lab Instructor**
Brigham Young University, Provo, UT
Data Science Essentials (Math 324, Math 326), Instructor: Tyler Jarvis
Taught and graded weekly lab on data analysis to cohort of 35 undergraduates. Topics covered included data cleaning and analysis in python, SQL, bash shell, regular expressions, MongoDB, web scraping/crawling, and interactive visualization.

Spring 2016 **Teaching Assistant**
Brigham Young University, Provo, UT
Econometrics (Econ 380), Instructor: Brigham Frandsen
Graded homeworks, held office hours, and taught reviews for class of Econ 380, an undergraduate econometrics course

Fall 2014 **Teaching Assistant**
Brigham Young University, Provo, UT
Statistics for Economists (Econ 378), Instructor: Brigham Frandsen
Graded homeworks, held office hours, and taught reviews for class of Econ 378, an undergraduate statistics course

Summer 2014 **Teaching Assistant**
Brigham Young University, Provo, UT
Microeconomics (Econ 381), Instructor: Brigham Frandsen
Graded homeworks, held office hours, and taught reviews for class of Econ 381, an undergraduate microenomics course

2014-2017 **Tutor**
Self-Employed, Provo, UT
Tutored undergraduates in calculus, linear algebra, and economics. Also tutored wide range of high school subjects.

Mentoring

Fall 2022 - Present **Jennifer Deng**
B.S. in Computer Science, Georgia Institute of Technology
Entity linking for automated knowledge graph construction; automating clinical data extraction with LLMs

Fall 2022 - Present **Shubham Lohiya**
M.S. in Computer Science, Georgia Institute of Technology
Entity linking for automated knowledge graph construction; automating clinical data extraction with LLMs

Spring 2022 - Present **Prasanth Bathala**
M.S. in Computer Science, Georgia Institute of Technology
Entity linking for automated knowledge graph construction; automating clinical data extraction with LLMs

Fall 2022 - Present **Zihan Wei**
M.S. in Biomedical Engineering, Georgia Institute of Technology
Automating clinical data extraction with LLMs

Fall 2022 **Tejasri Kopparthi**
M.S. in Computer Science, Georgia Institute of Technology
Entity linking for automated knowledge graph construction

Fall 2022 **Janvi Dubey**
B.S. in Biomedical Engineering, Georgia Institute of Technology
Discovering causes of COVID-19 induced cardiovascular complications via text mining and knowldge graph analysis

Fall 2021 - Present **Haydn Turner**
B.S. in Biomedical Engineering, Georgia Institute of Technology
Automating biomedical meta-analysis via human-in-the-loop natural language processing

Spring 2022 - Present **Dongyu Zhang**
B.S. in Computer Science, Georgia Institute of Technology
Automating biomedical meta-analysis via human-in-the-loop natural language processing

Fall 2019 - Spring 2022 **Davi Nakajima An**
B.S. in Computer Science, Georgia Institute of Technology
Text mining and knowledge graph completion
Now: PhD Student, Molecular Engineering and Sciences at University of Washington

Fall 2021 - Spring 2022 **Kevin McCoy**
B.S. in Biomedical Engineering, Georgia Institute of Technology
Text mining for drug repurposing and mechanism of action prediction in COVID-19 and Cardiovascular Disease
🎗 Sigma Xi Undergraduate Research Award, Georgia Institute of Technology
Now: PhD Student, Statistics at Rice University

Spring 2021 **Xinyu Chen**
B.S. in Biomedical Engineering
Annotation pipelines for biomedical information extraction

Spring 2021 **Brady Bove**
B.S. in Biomedical Engineering
Annotation pipelines for biomedical information extraction
Now: Optimized Operations Engineer at 3M

Fall 2021 **Alexis Nunn**
B.S. in Biomedical Engineering, Georgia Institute of Technology
Automating biomedical meta-analysis via human-in-the-loop natural language processing
Now: Product Engineer at Huxley Medical

Volunteer & Leadership Experience

2019 - 2022 **Youth Mentor**
Church of Jesus Christ of Latter-day Saints, Atlanta, GA
Organize community service projects and teach leadership & life skills to youth ages 8-17

Fall 2019 **English Teacher**
Catholic Charities Atlanta, Atlanta, GA
Taught semester-long English as a second language course for immigrants to United States

Spring 2015 **Youth Mentor**
Provo Youth Mentoring, Provo, UT
Met weekly with elementary students to teach academic and life skills

2017-2018 **Student Alumni Relations Representative**
College of Physical and Mathematical Sciences, Brigham Young University, Provo, UT
Organized college-wide student-alumni networking dinner. Organized fundraising event for student-to-student need-based scholarship program. Met regularly with dean to discuss and address student needs.

Nov 2011 - Nov 2013 **Full-time Missionary and Representative**
Church of Jesus Christ of Latter-day Saints, Atlanta, GA
Taught lessons in Tagalog language designed to strengthen families and communities. Organized quarterly conference and trainings for volunteers across six cities. Gathered and analyzed organizational data for regional leadership. Organized and coordinated community service projects with local leaders.

2010 - 2011 **Volunteer**
Adopt-a-Grandparent, Provo, UT
Regularly visited with seniors confined to local nursing homes to provide friendship and emotional support.

2009 - 2010 **Volunteer**
Murray Youth City Council, Murray, UT
Assisted with local community outreach events including food drives, civil rights benefits fundraiser, and community health fair.

Member

2020 — Present Association of Computational Linguistics (**ACL**)
2017 — Present Society of Industrial and Applied Mathematics (**SIAM**)
2010 - 2016 Phi Eta Sigma Honor Society

Technical Skills

Mathematics & Theory: Natural Language Processing (NLP), Large Language Models (LLMs), Machine Learning, Deep Learning, Bayesian Statistics, Computer Vision, Matrix Analysis, Complex Analysis, Functional Analysis, Numerical Linear Algebra, Control Theory, Probability Theory, Parallel Computing, Algorithm Design, Linear & Nonlinear Optimization, Active Learning, Advanced Econometrics, Abstract Algegra, Differential Equations, Information Retrieval

Machine Learning: Pytorch, Pandas, SpaCy, NLTK, RDKit, Huggingface, LangChain, OpenAI

Programming: Python, R, Stata, Mathematica

Web: HTML, Web scraping, SQL, Cypher, LaTeX, Markdown, Jekyll, Git, Google API suite

Visualization: Figma, Seaborn, Bokeh, Draw.io

Languages: English (Native), Tagalog (Professional), Spanish (Intermediate), German (Intermediate)

References

Dr. Cassie Mitchell, Assistant Professor
School of Biomedical Engineering
Georgia Institute of Technology
bme.gatech.edu/bme/faculty/Cassie-S.-Mitchell

Dr. Jeff Humpherys, Professor
School of Medicine
University of Utah
linkedin.com/in/jhumpherys/

Dr. Tyler Jarvis, Director and Cofounder
Applied and Computational Mathematics Program
Brigham Young University
math.byu.edu/~jarvis/

Dr. David Healey, Vice President of Data Science
Enveda Biosciences
linkedin.com/in/david-healey-a0a8143/