

# TODD MORRILL

[toddmorill.github.io](https://toddmorill.github.io) • [tm3229@columbia.edu](mailto:tm3229@columbia.edu) • New York, NY 10025

---

## EDUCATION

---

COLUMBIA UNIVERSITY New York, NY  
Master of Science, Computer Science May 2024  
Advisors: Kathleen McKeown, Richard Zemel  
GPA: 4.16/4.0

HARVARD EXTENSION SCHOOL Online  
Computer Science Coursework May 2021  
GPA: 3.59/4.0

GEORGE WASHINGTON UNIVERSITY Washington, DC  
BBA, Finance; Minors: Economics & Chinese May 2014  
GPA: 3.93/4.0

MIDDLEBURY COLLEGE Middlebury, VT  
Chinese Summer Language Program 2011

## RESEARCH EXPERIENCE

---

COLUMBIA UNIVERSITY New York, NY  
Research Assistant, Professor Richard Zemel, Computer Science Aug 2023 – Present

- Implemented 5 NLP pipelines to demonstrate the effectiveness of Prompt Risk Control, a framework for making provable guarantees about the safety of large language model outputs, resulting in ICLR 2024 and NeurIPS 2023 workshop submissions
- Scaled up and orchestrated our large language model (e.g., 40 billion parameters) pipelines on clusters of graphical processing units (GPU) enabling us to run our experiments in time to meet the conference deadline
- Defined clear notation that we used throughout the paper to communicate our framework and co-wrote the paper, resulting in clear scientific writing and a submission that received positive peer reviews

COLUMBIA UNIVERSITY New York, NY  
Research Assistant, Professor Kathleen McKeown, Computer Science Sep 2022 – Oct 2023

- Developed a novel approach for dialogue analysis using circumplex theory for the DARPA Computational Cultural Understanding (CCU) program resulting in a first-author submission to LREC-COLING 2024

- Collected a novel dialogue dataset annotated with features from circumplex theory (e.g., Gregarious-Extraverted, Aloof-Introverted, etc.) using GPT-4 and trained state-of-the-art dialogue outcome classification models on 2 dialogue datasets
- Delivered 2 dialogue analysis systems to DARPA and represented Columbia on weekly calls with all program stakeholders (DARPA, NIST, SRI, Monash, LDC, NYU, PARC, and LCC) resulting in Columbia's funding getting extended in the second phase of the program

COLUMBIA UNIVERSITY

New York, NY

Class Research Project, Computer Science

Jan 2023 – May 2023

- Compared hyperdimensional computing (HDC) to deep learning on a language identification task resulting, which demonstrated of the effectiveness the HDC framework with respect to computation, memory usage, robustness to noise, and data efficiency
- Proved a theorem showing the robustness of HDC methods to noisy inputs as dimensionality increases

## RESEARCH INTERESTS

---

- Machine learning, deep learning
- Natural language processing
- Computer vision
- Neuroscience and cognitive science inspired machine learning

## PROFESSIONAL EXPERIENCE

---

PRICEWATERHOUSECOOPERS (PWC)

New York, NY

Data Scientist, Senior Manager

Aug 2014 - Present

*Automated Knowledge Graph Construction*

- Implemented and compared 7 methods for extracting entities and relations from arbitrary text resulting in an accepted submission to the 2021 Workshop on Unstructured and Structured Knowledge Bases at the AKBC conference

*Information Retrieval and Classification Model for a Bank*

- Managed a team of 6 at PwC and collaborated with CMU (Eduard Hovy) to build machine learning models to classify pieces of regulatory text for compliance purposes
- Listed as lead inventor on 2 patent filings and reduced time spent by compliance experts by up to 50% [[bit.ly/32f2lx3](https://bit.ly/32f2lx3)]

*Enterprise Search*

- Reduced number of web clicks required for 50K+ users to find information in PwC's firm search portal and presented work at GCP Next '18 [[youtu.be/63EANKPzuJY](https://youtu.be/63EANKPzuJY)]

### *Select Projects Led*

- Domain Adapting Large Language Models, EEG Based Brain-Computer Interface [[youtu.be/f81T0KcprpM](https://youtu.be/f81T0KcprpM)], Deep Learning on Edge Devices [[youtu.be/iqVKjye-J68](https://youtu.be/iqVKjye-J68)], Website Click Analysis for a Major Technology Company, Structural Causal Model Development with CMU Students, Automated Customer Service Case Resolution for a Major Food Chain, Accounts Receivables Forecasting Model for a Technology Firm, Information Extraction from Excel & PDF Documents for Insurance Industry, Markov Chain Predictions for an Auto-Manufacturer, Dynamic Pricing for an Online Ticket Sales Platform, Sales Calls Analysis for an Asset Management Firm, Big Data Analytics for a Soft-Drinks Manufacturer, Classification Model for PwC Auditors, Classification Model & Big Data Analytics for a Brokerage Firm

## TECHNICAL SKILLS

---

*Programming languages:* Python, C/C++ (+CUDA), JavaScript, SQL, Bash, LaTeX, Markdown

*Development tools:* VS Code, Git/Github, Docker, Kubernetes

*Python libraries:* PyTorch, Tensorflow/Keras, Scikit-learn, Pandas, Dask, Numpy, Scipy, Flask

*Cloud:* Google Cloud, Amazon Web Services, Microsoft Azure

*Foreign Languages:* Professional proficiency in Mandarin Chinese

## IN SUBMISSION

---

1. **Social Orientation: A New Feature for Dialogue Analysis.** Todd Morrill, Zhaoyuan Deng, Yanda Chen, Amith Ananthram, Colin Wayne Leach, Kathleen McKeown. LREC-COLING 2024.
2. **Prompt Risk Control: A Rigorous Framework for Responsible Deployment of Large Language Models.** Thomas Zollo, Todd Morrill, Zhun Deng, Jake Snell, Toniann Pitassi, Richard Zemel. International Conference on Learning Representations, 2024.

## PEER-REVIEWED WORKSHOP PAPERS

---

1. **Prompt Risk Control: A Rigorous Framework for Responsible Deployment of Large Language Models.** Thomas Zollo, Todd Morrill, Zhun Deng, Jake Snell, Toniann Pitassi, Richard Zemel. NeurIPS Socially Responsible Language Modelling Research (SoLaR) Workshop, 2023.
2. **Cluster Based Named Entity Recognition.** Todd Morrill. Workshop on Unstructured and Structured Knowledge Bases at the AKBC conference, 2021.

## REPORTS

---

1. **Technical Report: Large Language Models.** Todd Morrill, Michael Kirchner, Vasudeva Sankaranarayanan, Joseph Voyles. 2023.
2. **Comparing Hyperdimensional Computing to Deep Learning for Natural Language Processing Tasks.** Todd Morrill, Satyam Sharma. 2023.

## **PATENTS**

---

1. **Regulatory Tree Parser.** Todd Morrill, Eric Roma, Nicolas Kuzak, Neelam Sharma, Andrew Runge, Jayvardhan Rathi, Waqar Sarguroh, Wenting Zhao. Publication number: 20220374405. 2022.
2. **Regulatory Obligation Identifier.** Todd Morrill, Eric Roma, Neelam Sharma, Alistair Moore, Andrew Runge. Publication number: 20220374914. 2022.

## **PRESENTATIONS**

---

1. Machine Learning Enabled Chatbots, Google Cloud Next Conference, 2018
2. Machine Learning Enabled Chatbots, AI4 Finance Conference, 2018

## **LEADERSHIP AND SERVICE**

---

- Harlem Grown Volunteer, 2014-2020
- GWU Alternative Breaks Volunteer, Rebuilding after Hurricane Katrina, 2011-2013
- Eagle Scout, Boy Scouts of America, 2010