

MATTHEW DUNN

matt@daedaline.com ◇ Google Scholar

EXPERIENCE

Co-Founder & CEO

Daedaline

October 2023 - Present

Boston, MA

Raised \$2M USD in seed funding and led the development of Daedaline from concept to a production-grade platform. The system integrates state-of-the-art multi-vector retrieval algorithms with distributed inference pipelines to support high-throughput, real-asset due diligence workflows. Designed and deployed agentic pipelines for dynamic query routing, enabling adaptive handling of tasks based on complexity and domain context.

Research Director - Member of the Technical Staff

Cohere

July 2022 - October 2023

Boston, MA

At Cohere, I established the company's annotation system strategy to support scalable, high-quality training data pipelines. I defined the forward-deployed engineer strategy to improve customer integration and accelerate adoption of language models in production settings. I also designed and prototyped post-training methods for model adaptation on Google TPUs, focusing on aligning large language models with complex commercial applications through retrieval integration and chain-of-thought reasoning.

VP of Applied Machine Learning

LivePerson

Sept 2021 - April 2022

Boston, MA

Oversaw company-wide deployment of deep learning methods across all product lines. Expanded and led a 20-person applied research team, primarily PhDs, that delivered LivePerson's flagship AI product and advanced models for intent classification, dialogue management, entity recognition, sentiment analysis, contact center optimization, and forecast modeling, supporting Tier 1 services in eight languages at global scale.

Director of Applied Machine Learning

LivePerson

Sept 2018 - Sept 2021

Seattle, WA

I recruited a team of applied NLP researchers and supported them in developing AI products from initial definition through GA release in collaboration with partner engineering teams. I was responsible for my team's delivery of machine learning research that supports flagship AI products across the LivePerson Conversational Cloud offering. We trained BERT and related transformer variants on one of the largest (human-to-human) conversational data sets. My team supports numerous NLP models that underpin Tier 1 services.

Data Scientist - NLP

LivePerson

Apr 2017 - Sept 2018

New York, NY

I worked with the CEO, Rob LoCascio, to hire and train an applied machine learning research team to focus on developing AI products for strategic initiatives. We developed products using unsupervised, semi-supervised, and supervised NLP techniques to extract commercially valuable insights from (human-to-human) conversational data sets.

Graduate Student Researcher

NYU Center for Data Science

Aug 2016 - May 2017

New York, NY

Member of the CILVR group at NYU where I worked on the development of the SearchQA data set and associated models to solve open domain Question and Answer tasks. Ported a bunch of models from Theano to Pytorch before Pytorch was cool.

Data Scientist

Wylan Energy

Developed simulations for estimating energy market auctions.

Summer 2016

New York, NY

CEO

MD Consulting

2014 - 2015

New York, NY

Built a laboratory information management system for a biotech to improve research and development efforts for high throughput genetic sequencing diagnostic tests for colon cancer. Managed a team of contractors in addition to doing hands on engineering.

Software Engineer 2012 - 2014 – Exponent Partners (B-Corp)

Software Engineer 2010 - 2012 – Groundswell (501.c.3)

RESEARCH - GOOGLE SCHOLAR

Actionable Conversational Quality Indicators for Improving Task-Oriented Dialog Systems Natural Language Engineering

PDF

2022

Domain-specific Distillation Models for Conversational Commerce ACL Workshop ECNLP

Accepted Papers

2022

SearchQA: A New QA Dataset Augmented with Context from a Search Engine

PDF

2017

Early Predictability of Asylum Court Decision

ACM Conference on AI and the Law

PDF

2017

Optimizing Targeted Sequencing of Somatic Mutations

2016

CONFERENCE REVIEWER

EMNLP 2025

ACL 2023, 2025

NAACL 2019, 2021, 2022, 2024

COLING 2020

PATENTS

Domain Adaptation of AI NLP Encoders with Knowledge Distillation

Patent No.:

US-20220318502-A1

Automatic Bot Creation Based on Scripts

Patent No.: US 10,853,568 B2

Intent Driven Contact Center

Patent No.: US 11,050,884 B2

Systems and Methods for Intent Analysis

Patent No.: US 10,902,199 B2

EDUCATION

M.S. Data Science, NYU Courant Institute

B.A. Environmental Studies, University of Redlands

- Minor: English Literature

VOLUNTEERING ACTIVITIES

Advisor

2019-Present

NYU Center for Data Science

I work with NYU Center for Data Science capstone groups to develop and execute their research proposals over the course of the semester.