# MATTHEW DUNN

matt@daedaline.com ⋄ Google Scholar

## EXPERIENCE

**Co-Founder & CEO** — October 2023 - Present
*Daedaline* — *Calgary, AB*

Building Generative AI for the energy sector.

**Research Director - Member of the Technical Staff** — July 2022 - October 2023
*Cohere* — *Boston, MA*

At Cohere I have two primary responsibilities. First, I lead a team focused on applying Large-Language Models to solve real-world commercial use cases. Second, I work on research efforts focused on complex QA tasks. My research interest lies at the intersection of chain-of-thought reasoning and retrieval augmented generation. I'm especially interested in how these techniques can be utilized in the service of solving novel commercial use cases.

**VP of Applied Machine Learning** — Sept 2021 - April 2022
*LivePerson* — *Boston, MA*

I helped a team of applied NLP researchers develop product road-maps, research proposals, and publication plans. The team's work focused on tasks to enable conversational systems, such as intent classification, entity recognition, dialogue management, etc. across 8 languages at a global scale for Tier 1 services.

**Director of Applied Machine Learning** — Sept 2018 - Sept 2021
*LivePerson* — *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** — Apr 2017 - Sept 2018
*LivePerson* — *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** — Aug 2016 - May 2017
*NYU Center for Data Science* — *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** — Summer 2016
*Wylan Energy* — *New York, NY*

Developed simulations for estimating energy market auctions.

**CEO**                                                                          2014 - 2015
*MD Consulting*                                                                  *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

**ACL**        2023
**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.