# Max Marion

mmarion538@gmail.com

Personal Website, Github, LinkedIn

## EXPERIENCE

#### RESEARCH SCIENTIST, DATABRICKS MOSAIC RESEARCH | October 2023 - January 2024

Performed cutting edge research on LLMs including areas like evaluation, RAG, reward modeling, and more. Work included performance optimization, model distillation, prompt engineering, and hallucination detection and correction.

#### RESEARCH SCHOLAR, COHERE FOR AI | January 2023 - October 2023

Lead engineering and research efforts alongside world renowned machine learning researchers investigating performance improvements from data pruning for large language models at scale.

#### MACHINE LEARNING ENGINEER, KUNGFU.AI | December 2018 - January 2023

Developed machine learning solutions for clients in computer vision, natural language processing, and a variety of other fields. Projects were full stack, in which I performed machine learning engineering, DevOps, data engineering, and software engineering.

## SPECIFIC PROJECTS

#### LLM AS A JUDGE AT SCALE | LARGE LANGUAGE MODELS

Lead internal research efforts to optimize LLMAsAJudge for RAG use cases. This included improving the cost and performance of the judge as well as the RAG system itself. Metrics included hallucination detection, Answer Correctness, Answer Relevance, and more.

#### DBRX EVALUATIONS | LARGE LANGUAGE MODELS

Worked on evaluations for DBRX, a state of the art MoE LLM. Domains included code gen, long context evaluation, and other standard benchmarks. Work ranged from deriving new evaluations to reflect real-world use cases to improving usability of the internal framework to captaining the pretraining run itself.

#### DATA PRUNING AT SCALE | LARGE LANGUAGE MODELS

Researching intelligent methods of data pruning to improve base language model performance for less compute at scale (1B+ parameters, 1T+ tokens). Responsibilities include experiment setup and implementation, data pipeline profiling, model training, and research direction.

#### TRADING CARD AUCTION PRICE PREDICTION | TWIN NETWORK

Engineered, tested, and maintained a twin network with triplet loss to predict sales prices of collectibles at auction. Project targeted rare cards with high value, leveraging ElasticSearch to derive reference cards for seeding.

## EDUCATION

#### OCCIDENTAL COLLEGE | MATH AND COMPUTER SCIENCE

August 2014 – May 2018 | Los Angeles, CA

- Graduated with a double major in Math and Computer Science
- One of three students to graduate from the Computer Science program in its inaugural year

#### CALIFORNIA INSTITUTE OF TECHNOLOGY January 2016 – March 2018 | Pasadena, CA

- Completed all Computer Science courses at Occidental then took additional CS classes concurrently at Caltech
- Related Coursework: Decidability and Tractability, Fundamentals of Computer Programming, Learning From Data, Machine Learning and Data Mining, Projects in Machine Learning, C Language Shop, Function Programming Language Shop.

## SKILLS

#### TECHNICAL

Python • Prompting • AWS Linux • Docker • Pandas TensorFlow • Pytorch Spark

#### SOFT

Comfortable working in organized settings and self-structured ones Understands receiving and giving mentorship Strong remote and in-person worker