

Aaron Gonzales

Selected Experience

Staff Machine Learning Engineer - Technical Lead, Mozilla.ai

Boulder, Colorado | April 2023–current

Lead engineer / architect hired to build a new AI research company under the Mozilla Foundation.

- Designed and built a Kubernetes-based large-language model (LLM) research and product experimentation platform from scratch. This platform supported distributed training, fine-tuning, evaluation, and experiment management on a cloud-based GPU cluster, leading to notable participation and contributions to workshops at NeurIPS and KDD.^{[2],[1]}
 - Established engineering culture and processes, including coding standards, developer tooling, CI/CD pipelines, API design styles, abstraction patterns.
 - Designed and managed a technical hiring program which led to hiring and mentoring five machine-learning engineers and researchers.
 - Managed procurement and integration of external IT vendors and tools.
-

Senior Machine Learning Research Engineer, Twitter, Inc.

Albuquerque, New Mexico | June 2020–January 2023

Technical Lead for the Machine Learning Ethics, Transparency, and Accountability team that owned Responsible Machine Learning development and research at Twitter.

- Built a system of feature pipelines and tools for ML researchers that accelerated critical work by 20x while ensuring more reliable results.
 - Implemented new functionalities in Twitter's Python build system, notebook platform, and data platform to streamline usage for ML researchers and data scientists, increasing adoption and development speed.
 - Led projects to build tools into Twitter's internal Tensorflow Extended and notebook platforms that enabled teams to rapidly identify undue biases and weaknesses in model performance.
 - Facilitated adoption of team results and products, leading to company-wide adoption of model quality metrics and tools. Presented research and engineering work in external venues.^{[6],[7],[3],[5],[4]}
 - Continued service on company advisory groups, including efforts to select a Twitter-wide data processing stack that would lead to 3-5x speedups in key data pipelines.
-

Senior Data Scientist, Twitter, Inc.

Boulder, CO | July 2017–June 2020

Data scientist for Twitter's API client business and later as tech lead for the Health Data Science team.

- Enabled 100 data scientists and ML engineers to scale common data operations by 5-50x while improving reliability by building libraries and training others in distributed data processing.
 - Managed technical auditing and review of models that identified unwanted user behaviors.
 - Engineered a comprehensive risk assessment model to evaluate developer eligibility for utilizing a new Twitter API service, revolutionizing the application review process and overcoming a key product limitation.
 - Shipped a model to identify clusters of applications that broke Twitter's API policies, allowing programmatic remediation of thousands of abusive client applications.
 - Developed open-source client libraries for Twitter API products, which were used by thousands of API customers.^{[10],[8]}
 - Served on company advisory groups for data architecture, open-source software, Python development, and as a global lead for Twitter's Hispanic/Latinx employees resource group.
-

Data Scientist, Tripadvisor, Inc.

Needham, Massachusetts | July 2016–July 2017

Built scalable machine learning models for the Vacation Rentals group.

- Developed and implemented a new computer vision model system to enable use of listing photo content in recommendation and ranking systems, resulting in a 4.8% boost to click-through rate and a 10.4% increase in return on ad expenditures.

- Built a natural-language query system to ease discovery of common destinations, which led to a 6.1% improvement in a primary metric.
- Developed language-modeling tools to extract high-information snippets from reviews to display on search-result pages to boost SEO rankings.
- Developed an internal data and model tools package that allowed data scientists and engineers to get up-to-speed more quickly and iterate faster on projects.

Education

Master of Science, Computer Science. University of New Mexico, May 2016.

Bachelor of Science, Psychology. University of New Mexico, May 2010.

Technical Skills

Languages: Python, SQL, shell, Scala

ML/Data/AI frameworks: Spark, Tensorflow, Kubeflow, TFX, Pytorch, Pandas, scikit-learn, Keras, vllm, Ray, wandb, HuggingFace, llamafire

Dev and ML Ops frameworks/platforms: Google Cloud, AWS, Coreweave, Kubernetes, Helm, Airflow, Bazel, Pants, Docker, Argo, Github Actions, Pulumi

Selected Publications and Artifacts

- [1] **Neurips large language model efficiency challenge: 1 llm + 1gpu + 1day**
Mark Saroufim, Weiwei Yang, Joe Isaacson, Luca Antiga, Driss Guessous, Greg Bowyer, Christian Puhersch, Geeta Chauhan, Supriya Rao, Artidoro Pagnoni, Vicki Boykis, Aaron Gonzales, Davide Eynard. (Dec. 2023). <https://llm-efficiency-challenge.github.io/>.
- [2] **Evalrs: Well-rounded recommender systems for real-world deployments**
Federico Bianchi, Patrick John Chia, Jacopo Tagliabue, Claudio Pomo, Gabriel Moreira, Ciro Greco, Davide Eynard, Fahd Husain, Aaron Gonzales. (Jul. 2023). <https://reclis.io/kdd2023-cup/>.
- [3] **County-level algorithmic audit of racial bias in twitter's home timeline**
Luca Belli, Kyra Yee, Uthaipon Tantipongpipat, Aaron Gonzales, Kristian Lum, Moritz Hardt. (Nov. 2022). *arXiv preprint arXiv: 2211.08667 [cs.SI]*.
- [4] **Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics**
Tomo Lazovich, Luca Belli, Aaron Gonzales, Amanda Bower, Uthaipon Tantipongpipat, Kristian Lum, Ferenc Huszár, Rumman Chowdhury. (Aug. 2022). *Patterns (N Y)*, vol. 3, no. 8 (100568).
- [5] **Research, meet engineering or: How i stopped worrying about scaling and love the process**
Aaron Gonzales, Luca Belli. Conference Talk, ACM FAccT, (Jun. 2022). <https://facctconference.org/2022/acceptedcraft.html>.
- [6] **Enabling, accelerating, and scaling data science for ml ethics**
Aaron Gonzales. Invited Talk, The Data Science Conference, (May 2022). https://www.thedatascienceconference.com/_files/ugd/a4f11b_468d08b0a6f74203ba4b1c90e00417a7.pdf.
- [7] **Investing in privacy enhancing tech to advance transparency in ml**
Aaron Gonzales, Rumman Chowdhury. Company blog post, Twitter, Inc. (Jan. 2022). https://blog.twitter.com/engineering/en_us/topics/insights/2022/investing-in-privacy-enhancing-tech-to-advance-transparency-in-ML.
- [8] **Search tweets python library**
Aaron Gonzales, Fiona Pigott, Jim Moffitt, Josh Montague. Python package, Twitter, Inc. (2018). <https://github.com/twitterdev/search-tweets-python>.
- [9] **Unraveling network-induced memory contention: Deeper insights with machine learning**
Taylor Liles Groves, Ryan E. Grant, Aaron Gonzales, Dorian Arnold. (2018). *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 8 (1907–1922).
- [10] **Do more with twitter data: Evaluating time series and identifying trends**
Aaron Gonzales. Company blog post, Twitter, Inc. (Mar. 2018). https://blog.twitter.com/developer/en_us/topics/tips/2018/evaluating-time-series-and-identifying-trends.
- [11] **A simple rapid process for semi-automated brain extraction from magnetic resonance images of the whole mouse head**
Adam Delora, Aaron Gonzales, Christopher S. Medina, Adam Mitchell, Abdul Faheem Mohed, Russell E. Jacobs, Elaine L. Bearer. (2016). *Journal of Neuroscience Methods*, vol. 257 (185–193).