

Andrew Noble

Greater Boston Area

+1 (607) 351 6341 andrew.e.noble@gmail.com andrewenoble.com
github.com/andrewenoble-org linkedin.com/in/andrewenoble 1,000+ [google-scholar citations](https://scholar.google.com/citations?user=andrewenoble)

Data Scientist and ML Engineer leading AI platform development aligned with strategic business goals. Specialized in the rapid development of cost-effective AI/ML end-to-end pipelines ingesting multimodal Electronic Health Record (EHR) data, including unstructured clinical text, medical images, claims, lab, and genomic data. Passionate about collaborating on multidisciplinary teams to make a positive impact on patient outcomes.

Skills

Data Science: Python, PySpark, PyTorch, MLFlow, SQL, AWS EC2/ECR/Batch, Bash, Docker, C/C++

Machine Learning: Generative AI, Transformers, NLP, Computer Vision, Bayesian, Causal, Variational

Experience

ConcertAI, Boston, MA, Sr. Principal DS and ML Engineer 2021–2023

- Led development, maintenance, and strategic planning for a Software Development Kit of re-useable, fast, scalable components for Knowledge Base Management, Data Engineering, and AI Modeling. Deployed in production environment and running bug-free for >1yr. Enabled 5x reduction in delivery time on client projects. Coordinated enhancements from 10+ contributors and built cross-functional community of 25+ users.
- Led early adoption of Generative AI methodology for multimodal EHR data. Implemented Transformer models from scratch. Built well-documented and tested software package enabling cost-effective multi-GPU training and inference on AWS Batch with logging to MLFlow.
- Demonstrated Generative AI capability to reliably infer unknown, time-dependent Clinical features (e.g., Performance Status) from commonly available Administrative Claims data, with the potential to dramatically scale and accelerate pre-screening of common Inclusion/Exclusion Criteria for Clinical Trials. Published [ASCO 2023 Abstract](#), Dr. Ravi Parikh co-author.

Dana-Farber Cancer Institute, Boston, MA, Sr. Principal DS and ML Engineer 2020–2021

- Led AI NLP development of online extraction of symptoms from clinical notes and surveys to prioritize patients for follow-up. Named-Entity-Recognition Transformer model prototyped on a cohort of 165k oncology patients.
- Led AI Imaging platform development, semi-automating organ and tumor segmentation workflows with state-of-the-art Computer Vision algorithms. Rapidly scaled interest among radiology labs.
- Developed Data Engineering pipeline for conversion of site-specific EHR to the OMOP Common Data Model. Handoff scaled to >2M patient records, enabling multi-site collaboration on rare cancer screen.

Aitia, Cambridge, MA, Principal DS and ML Engineer 2016–2020

- Led AI Longitudinal Causal pipeline development of oncology-specific disease progression and risk models.
- Led Data Engineering efforts, performing ETL on multimodal EHR for client projects and product development.
- Informed vision for platform development pipelines and algorithms. Demonstrated business value of state-of-the-art Deep Learning frameworks and Variational Inference methods.

Insight Data Science, Boston, MA, Fellow 2016

- Consulted with the Data Science team at Merck. Prototyped a [web app](#) for automating online connections among patients using a PostgreSQL backend, NLP, Latent Dirichlet Allocation, and a D3.js visualization.

UT Austin, U of Maryland, UC Davis, NSF Principal Investigator and Postdoc 2008–2016

- Led research and grant writing effort to win a 3yr, \$600k [interdisciplinary award](#) from the National Science Foundation. Organized 3 [SFI Working Groups](#). Additional 4yr \$1M NSF [award](#) supports on-going research at 3 universities.
- Discovered a robust correspondence between biological oscillator networks and magnets by fitting harvest data from a pistachio orchard to an Ising model with [custom C/C++ MCMC simulations](#) ([PNAS](#) and [NatComm](#) pubs).
- Discovered method to characterize evolutionary fixed points in stochastic population dynamics ([top physics journal pub](#)).

Cornell University, Physics PhD Program 2002–2008

- Discovered method to measure impact of Higgs boson on matter/anti-matter asymmetry in universe ([150+ citations](#)).

Education

Cornell University: PhD and MS in Theoretical Particle Physics and Cosmology 2002–2008

Johns Hopkins University: MA in Secondary Science Teaching 2000–2002

Carleton College: BA in Physics with Department Honors 1996–2000

Service and Outreach

Technical Advisor: Insight Data Science, Boston, MA 2018 & 2020

Teach for America: High School Physics and Technology, Baltimore, MD 2000–2002