

Brayton Hall

Washington, DC 20002 | 704-450-6631 | braytonhall@gmail.com | [Github](#) | [Data Science Blog](#) | [LinkedIn](#)

Data Scientist

Experience in data acquisition and modeling, statistical analysis, machine learning, and Natural Language Processing specializing in Generative AI. With a background in mathematical logic and philosophy at Virginia Tech and immersion in real large-scale ML problems, I excel at **building end-to-end ML pipelines that actually work** in production with big data. My focus has been on the fine-tuning of pre-trained generative transformers for data labeling, classification, and summarization.

TECHNICAL SKILLS

Python, Docker, Kubernetes, Databricks, PyTorch, Git, SQL, Snowflake, AWS, custom generative pre-trained transformers (GPTs), and LLM families including **Mixtral and Llama, with extensive experience enhancing AI output quality using custom tooling for each step of production GenAI workflows**

EMPLOYMENT HISTORY

Data Scientist, **Capital One**, McLean, Virginia 04/2024 - Present

- **Lead developer** of the second production use case of Generative AI at Capital One, summarizing over 1 million highly-sensitive unstructured text objects for executive committee members with an estimated cost savings of **\$20 million FTE hours annually**
- Performed model development, compliance, and monitoring for Capital One's fraud model for detecting fraud on all Capital One credit cards for a **combined \$400 million dollars of prevented fraud loss annually**
- Implemented custom GenAI workflow tools such as LLM-based tagging, embedding clusterings, and monitoring metrics to put guardrails on highly sensitive AI outputs and reduce hallucinations and bias for downstream outputs

Data Scientist, **Bank of America**, Charlotte, North Carolina 01/2022 - 04/2024

- Developed and monitored the primary generative language model pipeline for abstractive text summarization between bank agents and customers, reducing agent manual note-taking by 3 minutes per call for a **projected \$10 million cost reduction for complaint tracking**, alongside a framework of custom scoring metrics for monitoring summary quality
- Trained the production model (distilbart) with 600,000 manually-written agent notes and 150 million speech-to-text converted conversations to produce highly precise summaries of BofA customer-agent interactions, which concern a vast range of products and complaint types for conversations lasting between 1 and 60 minutes

Data Scientist, **PrimeNeuro**, Durham, North Carolina 10/2020 - 04/2021

- Used ML algorithms on infant-brain fMRI imaging data to predict emergent behavior, with the aim of early interventions and diagnosis for high-risk infants
- Used techniques such as PCA, Shapley values, and tree-based models on massively wide datasets (100,000+ features) for feature extraction, feature reduction, and interpretable modeling for specific brain regions

Graduate Philosophy and Statistics Instructor, **Virginia Tech**, Blacksburg, Virginia 08/2015 - 08/2019

- Lectured 20 to 150 person seminars focused on logic, statistical inference, ethics, philosophy of natural science, and created and maintained a grading database for professors
- Taught statistics post-graduation, with a focus on applied concepts such as confidence intervals, t-tests, hypothesis testing, and A/B testing, linear algebra, probability theory, Bayesian statistics, and linear regression

EDUCATION

Flatiron School, New York, NY 01/2020 - 05/2020

Data Science Program

- *Relevant experience:* Python, SQL, Statistics, Machine Learning, Computer Science, Data Visualization

Virginia Tech, Blacksburg, Virginia 08/2015 - 05/2017

MA in Philosophy

- *Relevant coursework:* Metalogic and Number Theory, Philosophy of Science, Ethics, Philosophy of Language

UNC Chapel Hill, Chapel Hill, North Carolina 08/2009 - 05/2013

BS in Psychology

- *Relevant coursework:* Statistics, Decision Theory, Calc I, Calc II, Symbolic Logic, Physics, Cognitive Psychology