Paarth Tandon

401-497-4218 | ptandoncs@gmail.com | [GitHub](#) | [paarthtandon.com](#) | [LinkedIn](#)

Experience

AI Engineer 2 **Meltwater** *San Francisco CA* **07/2024 - Present**

- Launched a **tool-calling** media intelligence **agent** using **LangChain** and **LangGraph** on top of a multi-petabyte **ElasticSearch** cluster.
- Designed **custom data retrieval methodologies** that decipher user intent, combining both **vector search** and **boolean query generation**.
- Built **10+** specialized agent tools that drive the generation of reports with **data-based insights** and **industry-standard formatting**.
- Developed a **source citation system** that enables the agent to ground its statements, **reducing hallucination** and **increasing trust**.
- Deployed the agent on **Azure** to **100+ paid monthly active users**. Built feedback-based **evaluation framework** for continuous monitoring.

Data Scientist **Intorqa** *Remote* **01/2024 - 07/2024**

- Built data-based detections for the Intorqa SAAS platform, a **threat management database** for game security and fraud teams.
- Designed a **multi-step RAG pipeline** using **GPT-4** to generate weekly threat summaries, extracting key facts from **100,000+** messages.
- Generated **social graphs** using message history to apply graph algorithms which uncover **influential threat actors** in a given community.
- Applied **ElasticSearch** and **LangChain** to build a real-time vector based **ranking/filtering pipeline**, reducing API expenses by **25%**.
- Developed test interfaces using **Hugging Face Gradio**, enabling non-technical team members to participate in **prompt engineering**.

Data Scientist **Bungie** *Bellevue WA* **06/2023 - 10/2023**

- Built data-driven **anti-cheat** for Destiny 2. Automated detection pipeline that **increased bans by 20%** through the detections I created.
- Trained custom image and vector-based **transformer models** using highly imbalanced, **3+ terabyte** datasets.
- Crafted and optimized (**2x speedup**) SQL queries to uncover anomalous events occurring in player data spanning **trillions of rows**.
- Optimized Security Analyst workflows using a query which corroborated evidence against a cheater, **speeding up investigations by 5x**.
- Worked with data engineers to productionize a high throughput vision transformer using **Docker**, **AWS Batch Compute**, and **Airflow**.
- Implemented **Autoencoders**, **GANs**, and **Diffusion Models** to detect anomalies using techniques such as **density estimation**.

Data Science Consultant **Arex Life Sciences** *Remote* **04/2023 - Present**

- Developed a proprietary data-based **signal processing algorithm** that classified biological samples with **>99% accuracy**.
- Crafted an **FDA compliant** data-pipeline which ingested raw data from a clinical lab instrument into an **encrypted NoSQL database**.
- Designed a user-friendly interface for laboratory technicians to import, manage, and export sample data using **Python** and **Tkinter**.

Data Science Intern **Moody's Analytics** *Remote* **06/2022 - 10/2022**

- Worked on NewsEdge, an **NLP** news analytics service used by companies in finance, publishing, and for corporate awareness.
- Developed a novel algorithm for real time event detection, replacing a previously unusable feature. The algorithm was built using **Python**, **Pandas**, **NumPy**, and **Pytorch**.
- Achieved event labeling speeds of under **3 ms per story**, while also improving label specificity and accuracy over previous attempts.
- Leveraged **AWS**, **S3**, **ElasticSearch**, and **EC2** cloud computing technologies to process stories for the real time event detection feature.
- Improved **language detection**, related stories, and the **automatic summarization** features by applying state of the art NLP models.
- Applied software development best practices using **Git**, **Agile**, **Jira**, **Confluence**, **unit testing**, and extensive **documentation**.

Data Science Intern **Ribbon Communications** *Remote* **06/2021 - 06/2022**

- Applied dimensionality reduction (**PCA**) and **clustering** using **Scikit-Learn** and **MatPlotLib** to answer business questions in the domain of telecommunications. Areas of interest include anomaly detection and error correlation.
- Built an automatic schema matcher that was able to correctly match columns on over **200** table schemas with over **99% accuracy**, using a combination of **NLP** and traditional methods.

Education

MS in Computer Science **University of Massachusetts** *Amherst, MA, USA* **08/2022 - 07/2023**

3.9 GPA, Data Science Focus, Highlighted Courses: Reinforcement Learning, Systems for Data Science, Visual Computing, Advanced NLP, Algorithms for Data Science, Data Science in R, Mathematical Statistics, Ethics in Computation

BS in Computer Science **University of Massachusetts** *Amherst, MA, USA* **08/2019 - 05/2022**

3.61 GPA, Highlighted Courses: Machine Learning, Natural Language Processing, Data Visualization, Artificial Intelligence, Database Management, Search Engines, Data Structures, Algorithms, Statistics, Discrete Math, Multivariable Calculus, Linear Algebra

Projects

- **Personal MusicGen:** Fine-tuned **MusicGen**, a transformer based **music generation model**, on a personal dataset of music I liked.
- **Pokémon Battle AI:** Applied **Deep Q Learning** using **PyTorch** to train a Pokémon AI, winning against a greedy AI in over **80%** of battles.
- **TrashGPT:** Fine-tuned **LLaMa** on the Trash Taste podcast. Generated realistic interactions and rendered them using **speech generation**.

Leadership

- **President:** ACM Machine Learning Club | Ran weekly meetings | **Recruited 150 members** | Technical workshops | Discussions on AI ethics
- **Course Instructor:** FYS 191: Thinking with Machine Learning | Discussions on industry, research, ethics | Introduced freshmen to ML