RHEA NAIR

Boulder, CO | (540) 793-1780 | rhea.nair@colorado.edu | linkedin.com/in/rheanair07 | github.com/rheanair7

EDUCATION

University of Colorado Boulder *Expected May 2027*
M.S. in Computer Science

Virginia Polytechnic Institute and State University (Virginia Tech) *May 2025*
Dual Degree: M.S. in Business Analytics (MSBA) *CGPA:* ***3.97/4.00***
and B.S. in Cybersecurity Management & Analytics

Narsee Monjee Institute of Management Studies (MPSTME) *May 2023*
B.Tech in Computer Science (Data Science) *CGPA:* ***3.70/4.00***

SKILLS

Languages	Python, Java, C++, MATLAB, R, SQL, NoSQL, MongoDB, Hadoop
ML & Libraries	Scikit-learn, PyTorch, Hugging Face, LLM architectures, Pandas, OpenCV, CNN, TensorFlow
Analytics & Tools	Git, Linux, Docker, AWS, Azure, Databricks, Apache, Spark, Tableau, PowerBI, Excel, SAS, StreamLit, Dash, A/B Testing, CUDA, Causal Inference, Information Security
Environments	Jupyter Lab, Google Colab, MATLAB, GPU architectures

EXPERIENCE

Research Assistant *May 2024 – Jun 2024*
Virginia Tech, Blacksburg, VA

- Reduced the research paper review cycle from 14 days to 2 days by developing ***Python scripts*** with ***Pandas and regex-based automation***, enabling faster parsing, metadata extraction, and error detection, which improved workflow efficiency by **85%**.
- Applied ***NLP classification models*** using ***scikit-learn and spaCy*** to automatically categorize academic documents, reducing manual review time by 70% and improving classification accuracy across large datasets.
- Automated ***data collection and preprocessing pipelines*** with ***BeautifulSoup, Pandas, and NLTK***, creating scalable workflows that ensured reproducibility and standardized processing for faculty research.
- Created ***Python-based summary reports and visualizations*** with ***Matplotlib and Seaborn*** that highlighted document trends, provided actionable insights, and supported faster, data-driven faculty decision-making.

Data Scientist and Researcher *Jan 2023 – Jan 2024*
Tata Institute of Fundamental Research, Mumbai, India

- Designed and executed ***multi-agent reinforcement learning simulations*** in ***Python (PyTorch)***, managing large-scale experimental data stored in both **NoSQL (JSON)** and **relational SQL** formats to optimize access and processing.
- Engineered a ***Neuro Evolution-based deep RL model*** using ***PyTorch and NumPy***, improving training efficiency by **85%** compared to baseline methods and reducing computational resource consumption.
- Leveraged **Hadoop and Apache Spark** for distributed data preprocessing and analysis, ensuring scalability and reducing computation time across simulation workloads.
- Authored comprehensive ***technical documentation and experiment reports***, including methodology breakdowns, performance comparisons, and workflow diagrams, to support collaboration and internal knowledge transfer.

Data Analytics Intern *May 2022 – Aug 2022*
Prescience, Mumbai, India

- Developed ***ETL data pipelines*** using Python (BeautifulSoup, Pandas) and SQL, automating data extraction, cleaning, and transformation processes and improving workflow reliability.
- Designed and deployed a ***statistical web application*** using Dash (Python), enabling interactive analysis and increasing analytics adoption across the organization by **90%**.
- Created ***data visualizations*** using Matplotlib and Tableau, presenting insights in an accessible format for non-technical stakeholders.
- Performed ***data quality checks, preprocessing, and validation*** using Python (OpenCV, scikit-image, Pandas), integrating image processing with structured data workflows for accurate reporting.

PROJECTS

Question Answering with Transformers (Python, PyTorch, HF, CUDA) [Link] *Jun 2025 – Present*

- Built an interactive QA system using ***Hugging Face's*** `deepset/roberta-base-squad2`, enabling users to query custom contexts.
- Deployed the QA system on ***AWS EC2*** with datasets stored in ***S3***, enabling scalable inference while reducing latency by 40% through mixed-precision computation and ***CUDA memory profiling***.
- Conducted system-level bottleneck analysis using `torch.profiler` and `nvidia-smi` to co-optimize the deep learning framework and ***GPU workload***.

Online Retail Customer Intelligence (Python, Pandas, Apriori, Tableau, Holt-Winters) [Link] *Jan 2025 – Mar 2025*

- Performed ***RFM segmentation*** to enable ***targeted campaigns*** that improved engagement and retention.
- Used ***Association Rule Mining (Apriori)*** to uncover cross-selling opportunities, influencing ***bundling strategies***.
- Built ***Holt–Winters forecasts*** and ***Tableau dashboards*** for demand prediction and customer insights, enabling ***data-driven decisions***.