Phanindra Kalaga

http://phanindra-max.github.io

Email : phanindra.connect@gmail.com

Mobile : +1-771-233-3129

PROFESSIONAL SUMMARY

Full-Stack Data Scientist skilled in the full product lifecycle, from framing business challenges to production deployment. Proficient in integrating data analytics into web-based platforms using Python, SQL, and React/Next.js. Experienced in building robust data pipelines, deploying ML models as scalable APIs, and driving strategic decision-making through visualization. Proven track record in optimizing supply chain visibility and shipping open-source software with modern CI/CD practices.

SKILLS

Languages: Python, SQL, JavaScript, R, HTML, CSS
Web & Serving: React, Next.js, FastAPI, Flask, UI/UX
Data & Infra: Pandas, NumPy, Spark, Airflow, Docker
Databases: PostgreSQL, MySQL, MongoDB, Neo4j

ML/AI: NLP, LLMs, vLLM, PyTorch, scikit-learn
Visualization: Tableau, Plotly, Matplotlib
Cloud: AWS (EC2, S3, IAM, CloudWatch), GCP
MLOps & DevOps: Git, GitHub Actions, CI/CD, MLFlow

EDUCATION

- **George Washington University** — Washington, D.C.
  Master of Science in Data Science; GPA: 3.90 — *Jan. 2024 – Present*
  - **Relevant Coursework**: Machine Learning, NLP, Data Mining, Computer Vision, Cloud Computing.

- **Jawaharlal Nehru Technological University** — India
  Bachelor of Technology in Computer Science and Engineering — *Aug. 2019 – May 2023*

EXPERIENCE

- **George Washington University** — Washington, D.C.
  Lead Graduate Research Specialist, LAiSER — *April 2024 - Present*
  - **Product Development**: Led the end-to-end development of an open-source analytics package, managing the product lifecycle from requirements to CI/CD deployment.
  - **Strategic Analysis**: Developed Neo4j and Plotly visualizations to identify skill gaps across 1,000+ courses; provided strategic recommendations for two local training policies for the Texas Workforce Commission.
  - **Data Engineering**: Designed scalable pipelines for skill extraction from 1M+ job descriptions; achieved 7x throughput using vLLM on HPC clusters.
  - **Stakeholder Communication**: Managed weekly SCRUM cycles; delivered 5 milestones on schedule using GitHub Actions, effectively communicating technical progress to non-technical stakeholders.
  - **Dev Environment**: Standardized the analytics environment on AWS and authored shared Python libraries for the team.

  Graduate Assistant, Natural Language Processing Class — *Aug 2025 - Present*
  - **Process Automation**: Built a Python autograding toolkit; cut grading turnaround from 3 days to 2 hours across 10 assignments (5 minutes per submission), saving 100 TA hours.
  - **Cloud Runtime**: Standardized the environment on AWS EC2 with a CUDA/PyTorch AMI; configured S3 for artifacts, IAM for access, and CloudWatch for monitoring.
  - **Course Dashboards**: Built instructor dashboards and learner survey tools to track assignment and activity patterns.

  Data Science Intern, Data Science for Sustainable Development — *Aug 2024 - Dec 2024*
  - **Web Platform Integration**: Led a team of 5 to design and deploy a web-based analytics platform (Next.js, Flask), integrating complex data insights directly into user-facing applications.
  - **UI/UX & Analysis**: Designed Figma wireframes and implemented AI agents to generate admin insights, increasing partner collaboration efficiency by 30%.
  - **Cloud Architecture**: Deployed the full-stack application on AWS (EC2/S3), ensuring high availability and scalable data storage.

- **NashAgri, An Agritech B2C Organization**　　　　　　　　　Maharastra, India
  Full-Stack Engineer　　　　　　　　　　　　　　　　　　*July 2021 – Jan 2023*
    - **Product Lifecycle Management**: Built a cross-platform CRM from scratch to manage the lifecycle of vendor transactions and inventory; added ML analytics to improve operational efficiency.
    - **Supply Chain Analytics**: Managed geospatial data for 45,000+ farmers across 25,000+ regions (MongoDB, GeoJSON) to improve supply chain visibility and generate strategic insights.
    - **Backend Engineering**: Engineered 5+ core features (invoice generation, RBAC, real-time auctions) supporting a 40% increase in vendor transactions in the first quarter.

PROJECTS

- **Global CO_2 Emissions Dashboard (Tableau & Python)**: Built an interactive dashboard analyzing 223 years of CO_2 data across countries and sectors to identify trends and guide renewable adoption strategies.
    - **Tools**: Tableau, Python, pandas, Plotly, Matplotlib

- **Benchmarking Database Architectures for Network Analytics**: Conducted performance benchmarking of MySQL, MongoDB, and Neo4j on a synthetic road network dataset to deliver technical recommendations on database selection.
    - **Tools**: Python, MySQL, MongoDB, Neo4j, SQL recursive CTEs, psutil

- **Reinforcement Learning for Pseudo-Labeling**: Developed a model-agnostic, custom RL environment for data annotation to outperform state-of-the-art semi-supervised techniques.
    - **Tools**: OpenAI Gymnasium, HPC, Tensorboard, PyTorch

- **My Own Medic**: End-to-end open-source AI architecture for a medical assistant leveraging EHR records and LLMs to prevent medical errors; Project selected for presentation at UN Open-Source Week 2025.
    - **Tools**: MySQL, PHP, Hugging Face API, Data Compliance (HIPAA)

DEVELOPER RELATIONS

- **President, Google Developer Groups On-Campus at GWU**: Led 200+ student organization for AI/ML awareness via events and workshops; cohosted DevFestDC (1000 attendees); cohosted DevFest Annapolis (300 attendees)
- **Chairperson of Relations, Data Science Association**: Cultivated relationships with C-suite executives and senior engineers for hosting events to provide career insights to 200+ data science (undergraduate and graduate) students

PRESENTATIONS AND AWARDS

- **Build with AI: Startup Day, December 2025**: Organized a summit for 200 attendees featuring C-suite executives and senior engineers; delivered a session on the impact of open-source software on AI and business strategy.
- **DevFest Annapolis, November 2025**: Hosted a tech-talk and a technical workshop on Model Context Protocol (MCP), Multi Agent Systems (MAS) and agent communication mechanisms for 100+ students and faculty.
- **DC Startup and Tech Week, October 2025**: Hosted a technical workshop for 60+ startup founders on building with AI, containers, and serverless deployments on Google Cloud.
- **Badge Summit at CU Boulder, July 2025**: Hosted a table talk with 50 leaders in higher-education research on skills and credentials.
- **GWU Open Source Conference, May 2025**: Delivered a keynote session on open-source software best practices.
- **GWU Open-Source Student Awards, January 2025**: Received 3rd prize for best open-source software.