Phanindra Kalaga
http://phanindra-max.tech
Email : phanindra.connect@gmail.com
Mobile : +1-771-233-3129

Professional Summary

Data scientist skilled in end-to-end problem-solving, from framing challenges to production deployment. Proficient in building robust data pipelines with Python and SQL, and training machine learning models with a focus on NLP and LLMs. Experienced in deploying models as scalable APIs, creating decision-ready dashboards, and shipping open-source software with modern CI/CD practices. Comfortable working across the full data stack, from backend services to frontend interfaces.

Skills

- **Languages**: Python, SQL, JavaScript, R **ML/AI**: NLP, LLMs, vLLM, PyTorch, scikit-learn, Hugging Face
- **Data & Infra**: Pandas, NumPy, Spark, Airflow, Docker **Cloud**: AWS (EC2, S3, IAM, CloudWatch), GCP
- **Databases**: PostgreSQL, MySQL, MongoDB, Neo4j **Web & Serving**: FastAPI, Flask, Next.js, React
- **MLOps & DevOps**: Git, GitHub Actions, CI/CD, MLFlow **Visualization**: Plotly, Matplotlib, Tableau

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*
  - **Open Source**: Built an open-source NLP package for skills and workforce analytics; co-authored paper submitted to the Journal of Open Source Software (JOSS).
  - **Data Engineering**: Designed scalable pipelines for skill extraction from 1M+ job descriptions; achieved 7x throughput using vLLM on HPC clusters.
  - **Data Visualization**: Built Neo4j and Plotly 3D maps to show skill gaps across 1,000+ university courses for the Texas Workforce Commission; work informed two local training policies.
  - **Project Management**: Ran weekly SCRUM; delivered 5 milestones on or ahead of schedule using GitHub Actions and issue tracking.
  - **Dev Environment**: Standardized the analytics environment on AWS and authored shared Python libraries for the team.
  - **Data Ethics**: Authored data-ethics guidelines; eliminated 100% of PII from public datasets.

  Graduate Assistant, Natural Language Processing Class *Aug 2025 - Present*
  - **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 and reducing regrade requests to 0.
  - **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.
  - **Tech**: Python, pandas, NumPy, scikit-learn, NLTK, Hugging Face, Matplotlib, OpenPyXL, AWS (EC2, S3, IAM, CloudWatch), CUDA, PyTorch, Blackboard.

  Data Science Intern, Data Science for Sustainable Development *Aug 2024 - Dec 2024*
  - **Full-Stack Engineering**: Led a team of 5 to design and deploy an open-access repository of 200+ non-profits (Next.js, Flask, Supabase on AWS EC2/S3).
  - **Data Analysis**: Designed Figma wireframes and embedded n8n AI agents for admin insights across 50+ projects; increased partner collaboration by 30%.

- **NashAgri, An Agritech B2C Organization** Maharastra, India
  Full-Stack Engineer *July 2021 – Jan 2023*
  - **Product**: Built a cross-platform CRM from scratch; added ML analytics to improve operational efficiency.
  - **Geospatial**: Managed geospatial data for 45,000+ farmers across 25,000+ regions (MongoDB, GeoJSON); generated insights for partnerships.
  - **Backend**: Engineered 5+ core features (invoice generation, RBAC, real-time auctions) supporting a 40% increase in vendor transactions in the first quarter.

Projects

- **Reinforcement Learning for Pseudo-Labeling**: Model agnostic, custom RL environment for data annotation to outperform state-of-the-art semi-supervised techniques and enable reproducible evaluation.
  - **Tools**: OpenAI Gymnasium, High Performance Compute (HPC), Tensorboard, PyTorch (torchvision)
- **Benchmarking Database Architectures for Network Analytics**: Reproducible Python benchmark of MySQL, MongoDB, and Neo4j on a synthetic road network dataset to deliver recommendations on which database performs best for each query type.
  - **Tools**: Python, MySQL, MongoDB, Neo4j (graph algorithms), SQL recursive CTEs, MongoDB aggregation/lookup, psutil (performance monitoring), Faker
- **Global CO_2 Emissions Analysis Dashboard**: Built a dashboard analyzing 223 years of CO_2 data across countries, sectors, and income groups to identify relation between historical emissions and renewable adoption priorities.
  - **Tools**: Tableau, Python, pandas, Plotly, Matplotlib
- **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, New York.
  - **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 (900 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 170 data science students

Presentations and Awards

- **DC Startup and Tech Week, October 2025**: Hosted a technical workshop for 60+ startup founders.
- **Badge Summit at CU Boulder, July 2025**: Hosted a table talk with 50 leaders in higher-education research.
- **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.