

PONKOJ CHANDRA SHILL

PH.D. CANDIDATE | MACHINE LEARNING | DATA SCIENCE

Reno, Nevada, USA | Contact: +1 (775) 467-8180 | Email: cspunkoj@gmail.com

LinkedIn: [linkedin.com/ponkoj](https://www.linkedin.com/ponkoj)

GitHub: github.com/CS-Ponkoj

Google Scholar

Portfolio Website

PROFESSIONAL SUMMARY

Final-year **Ph.D. candidate** in Computer Science with **4+ years** of **applied machine learning** and **data science** experience across **research** and **industry**. Built **end-to-end ML** and data systems spanning data acquisition, feature engineering, model training, experimentation, production deployment, and monitoring. Strong in **Python**, **SQL/PostgreSQL**, **AWS**, statistics, data analysis, ETL and production-ready software, with interest in fraud detection, identity verification, and financial risk.

TECHNICAL SKILLS

Applied Data Science: Experimental Design, Statistical Analysis, Classification, Regression, Time-Series Modeling, Feature Engineering, Model Training, Model Evaluation, Model Monitoring

AI/ML Development: Deep Learning, LLMs, RAG, Transformers, Prompt Engineering, Multimodal ML, Computer Vision, PyTorch, TensorFlow, Hugging Face, OpenCV, OpenAI APIs

Programming & Databases: Python 3, SQL, PostgreSQL/PostGIS, C, C++, Java, Golang, Dart

Data Engineering & Cloud: Data Acquisition, ETL Pipelines, Data Quality, Data Warehousing, AWS (EC2, S3, RDS, Redshift, Lambda), Apache Spark

Software Engineering: Production-Ready Python, Git, Linux, pytest, Flask, Automated Testing, Regression Testing, REST APIs, CLI Tooling

PROFESSIONAL EXPERIENCE

Graduate Research Assistant

Aug 2022 - Present

University of Nevada, Reno

Reno, NV

- Built and deployed an NSF-funded **AI/ML platform** used by 100+ students, integrating **Python**, Flask, automated testing, telemetry, runtime signals, and model-assisted feedback in an end-to-end research system.
- Engineered features from student code, failing tests, runtime outputs, iteration history, and behavioral signals to support personalized model-assisted feedback and experimentation.
- Designed controlled experiments and evaluation pipelines with hidden checks, failure analysis, and performance tracking across accuracy, robustness, time-to-success, and stuck-state recovery.
- Developed Python-Flask services and **LLM workflows** using RAG, prompt engineering, **OpenAI APIs**, and policy-controlled assistance; author/co-author of 6 peer-reviewed papers with 160+ total citations.

Business Intelligence Developer

Feb 2026 - Jul 2026

Tahoe Forest Health System

Truckee, CA

- Analyzed 426,759 healthcare referral records and developed validated mapping logic for 121,719 missing specialty values; combined rules populated 119,749 records, **resolving 98.38%** of the original missing values.
- Built SQL datasets, **ETL** and data-quality workflows, operational reports, and dashboard pipelines; validated business logic and metrics through testing and production support.
- Reconciled department, provider-specialty, service-code, and referral signals into traceable mappings, explicitly documenting ambiguous cases for downstream analytics.

AI Software Engineer Intern

Jan 2024 - May 2024

ConSynSys Technologies

Carson City, NV

- Developed time-series **ML models** to predict liquid transfer in biomedical pumps and engineered benchmarking and real-time monitoring for deployment reliability.
- Applied genetic algorithms to optimize pump calibration control code, reducing manual intervention by 50%.

Machine Learning Engineer

Jun 2021 - Aug 2022

Softwrd Ltd

Oslo, Norway

- Built and maintained an end-to-end multimodal property valuation system spanning data acquisition, feature engineering, PostgreSQL-based storage, model training, evaluation, and real-time inference on AWS.
- Developed a CatBoost regression pipeline for automated property valuation with continuous model updates and a **ResNet152V2 computer vision pipeline** to identify eight apartment features.
- Automated acquisition and processing of 8M+ property images using AWS **Lambda** and **S3**; analyzed 4M property records across 100+ geographical risk zones using **PostgreSQL/PostGIS**, AWS **RDS** and **Redshift**.
- Integrated tabular, image, and geospatial signals into a production property-pricing system that helped secure **\$20K** in government funding.

Business Intelligence Analyst

Apr 2020 - Jun 2021

Truck Lagbe

Dhaka, Bangladesh

- Improved customer retention by 15% through customer and operational data analysis, supply-demand prediction, and automation workflows.
- Supported business decision-making and geographical analysis through MySQL-based data analysis and stakeholder-ready Tableau reporting.

Data Analyst

Mar 2019 - Mar 2020

Creatino

Dhaka, Bangladesh

- Analyzed user behavior and business data with Python to identify patterns and generate insights for business strategy.
- Built Power BI dashboards to communicate findings to stakeholders and support marketing campaigns, product decisions, and growth planning.

EDUCATION

Ph.D. in Computer Science, University of Nevada, Reno

2022 - Present

GPA 3.87 • Expected Fall 2026 • NSF-funded research in machine learning, AI-assisted software systems, controlled experimentation, and personalized learning.

M.S. in Computer Science, University of Nevada, Reno

2022 - 2024

GPA 3.87 • Focus: machine learning, AI systems, and robotics learning platforms.

Bachelor of Computer Science, BRAC University, Bangladesh

2016 - 2020

GPA 3.60 • Thesis: multimodal machine learning and computer vision.

SELECTED PROJECTS

pandas-eda-check | PyPI Package • Python • Pandas • pytest

- Published a reusable Python package to PyPI for dataset-quality profiling and comparison of schema, missing-data, duplicate-rate, numeric, datetime, and categorical changes between DataFrames.
- Implemented reference-vs-current statistical profile comparisons, threshold-based change detection, robust edge-case handling, and pytest validation for reproducible data-quality monitoring.
- Package Link: <https://pypi.org/project/pandas-eda-check/>

Multimodal Harmful-Content & Threat Detection | Python • OpenCLIP • DeBERTa • OCR • pytest

- Built a reproducible multimodal evidence-triage pipeline combining image scoring, OCR text, associated message text, zero-shot classification, and score-level fusion across multiple risk categories.
- Implemented source-aware routing, CSV evidence logging, error analysis, CLI workflows, and regression tests to support transparent and auditable model behavior.
- GitHub Link: <https://github.com/CS-Ponkoj/Hate-and-Threat-Detection-in-Forensics>

AI-Assisted PCB Design | Python • Gemini API • Ollama/Qwen2.5 • Graphviz • pytest

- Built and deployed an AI-assisted design system that converts natural-language hardware requirements into validated, structured PCB handoffs including component selections, BOM, pin maps, netlists, power budgets, layout guidance, and fabrication checks.
- Engineered Gemini and local-LLM workflows with structured-output validation, deterministic fallback, unsupported-request detection, and a source-grounded review copilot to improve reliability, traceability, and failure handling of AI-generated results.
- Application Live Link: <https://ai-assisted-pcb-design.streamlit.app/>

RESEARCH & RECOGNITION

Publications: Author/co-author of 6 peer-reviewed papers across IEEE, ACM, ASEE, and related venues; 160+ total citations.

Honors: Graduate Dean's Merit Scholar, UNR (2024 - 2025); NASA Space Apps Challenge Finalist (2019); 1st Runner-Up, IEEE COVID-19 Combatants Unification Competition (2020).

Leadership: President, Bangladesh Student Association, University of Nevada, Reno; former Data Team Lead, Softwrld Ltd.