

Suraj Pokhrel

701-729-0360 ◇ pokhrelsuraj1211@gmail.com ◇ linkedin.com/in/srjpokhrel ◇ github.com/necimye

EDUCATION

Master in Computer Science, *North Dakota State University*

Dec 2025

Relevant Coursework: Algorithm Analysis, Distributed Systems, Software Development Process, Mobile Software Engineering, Database Systems, Artificial Intelligence, **GPA: 4.0**

WORK EXPERIENCE

Software Engineer - II, *Musco Lighting – Urbandale, IA, USA*

Jan 2026 – Present

- Accelerated ML model development by automating data validation for a custom **YOLO** object detection model, slicing baseball game videos per pitch with **OpenCV** and **AWS S3** and cutting manual preparation from **3 days to 2 hours**.
- Architected a distributed, multi-model inference platform on **NVIDIA Triton Inference Server**, consolidating model serving behind a unified API to deliver scalable, low-latency inference and streamlined model versioning in production.
- Integrated **PyTorch** deep learning models into **NVIDIA DeepStream** pipelines with **OpenCV**-based post-processing, enabling real-time computer vision inference on embedded and edge platforms.
- Built a unified container health check system with label-based container discovery using **Go**, an **MQTT** broker, and **SQS** to deliver real-time health status for all services on the edge server.

Software Engineer Intern, *Musco Lighting – Urbandale, IA, USA*

May 2025 – Dec 2025

- Architected a scalable task-scheduling microservice using **FastAPI** and **AWS-ECS** to automate game and video-frame ingestion pipelines, eliminating manual scheduling and accelerating ML data preparation.
- Implemented a real-time pitch-visualization subsystem that transforms 3D ball-tracking and strike-zone coordinates from **WebSocket** streams into 2D projections using **Vue.js** and **Three.js**, providing precise trajectory path for both spectators and umpires.

Graduate Research Assistant, *North Dakota State University – Fargo, ND, USA*

Aug 2024 – Dec 2025

- Conducted empirical research on test smells across six large-scale open-source **Java** repositories, building a dataset of test smell removals from **~52K** commits and categorizing them as incidental or intentional to evaluate test code quality.
- Identified **14** bugs in **TsDetect**, an open-source test smell detector, and built **TestSniffer**, an improved detector that increased detection accuracy by **85%**.

Full Stack Developer, *Audit Partnership Inc. – Leeds, UK*

Aug 2023 – Jul 2024

- Automated and optimized enterprise data workflows using Python, SQL, and Azure Data Factory, improving pipeline throughput by 30% and eliminating repetitive manual processes.
- Designed and deployed scalable **ETL** pipelines with **Snowflake**, SSIS, and T-SQL to process more than 1 million retail audit records daily for clients including Walmart, cutting integration time by 50%.
- Enhanced SQL Server performance and implemented real-time streaming with **Kafka** and **Spark**, reducing data latency by 60% and accelerating analytics for production systems.

Full Stack Developer, *Growthzilla Inc. – Los Angeles, CA, USA*

Jan 2022 – Jul 2023

- Reduced order delays by **30%** by developing and maintaining a robust **REST API** backend in **Node.js**, Express, and TypeScript that processed orders, validated payments, and synchronized inventory in real time.
- Shortened shipment time by **80%** by integrating **Stripe** payment and **ShipBob** fulfillment APIs to automate order processing and delivery workflows from checkout to shipment.

RESEARCH PUBLICATION

- S. Pokhrel, A. Jha, “**TESTSNIFFER: Improving Smell Detection Accuracy**”, International Conference on Mining Software Repositories (MSR), 2027 (in review)

PROJECTS

- Distributed Message Queue**: Engineered a distributed message queue in Go, achieving failover in 150-300 ms with quorum replication by implementing Raft consensus, consistent-hash sharding, and end-to-end tracing using Prometheus and Grafana.
- Realtime Data Streaming Pipeline**: Constructed a distributed data-engineering workflow with Apache Airflow, Kafka, Zookeeper, PySpark, Docker Compose, and Cassandra to process and stream high-velocity data in real time.
- Graph Plotter**: Created a 3D graph plotter capable of plotting any mathematical equation up to 3 variables using WebGL, Python, and React.js by generating 10K coordinate points with NumPy and Pandas.

SKILLS

Programming Languages: Python, Java, C++, JavaScript, TypeScript, Rust, Go, SQL

Cloud Technologies: AWS, AWS CDK, Azure

Deployment Tools: Docker, Kubernetes, Prometheus, Grafana, Git, GitLab CI, Unix, Linux

ML/AI Tools: PyTorch, TensorFlow, Scikit-Learn, Keras, HuggingFace, Label Studio, Post-training, Model-Validation, Inferencing

Web/Testing Frameworks: FastAPI, Vue.js, Django, React.js, Node.js, SpringBoot, JUnit, PyTest