

# KAIVALYA KISHOR DIXIT

[kaivalyawork@gmail.com](mailto:kaivalyawork@gmail.com) | (862) 215-1490 | [kaivalya.dev](https://kaivalya.dev) | [linkedin.com/in/kaivalya-kishor-dixit](https://linkedin.com/in/kaivalya-kishor-dixit) | [github.com/KaivDev4434](https://github.com/KaivDev4434)

## Education

**M.Sc. in Data Science / New Jersey Institute of Technology**

MAY 2025

**GPA: 3.95/4.0**

**Course Work:** Deep Learning, Applied Statistics, Big Data, Database Management Systems, Time Series Analysis and Forecasting, Reinforcement Learning, Applications of Parallel Computing, Python and Mathematics, Time Series Analysis and Forecasting

**Bachelor Of Technology in Electrical and Electronics Engineering | Mahindra Ecole Centrale**

JUN 2023

**GPA: 3.54/4.0**

## Skills

**Programming Languages:** Python | Java | JavaScript | C/C++ | HTML/CSS | Shell Scripting | Assembly | LaTeX | TypeScript

**Machine Learning & AI:** PyTorch | TensorFlow | Scikit-learn | Keras | XGBoost | OpenMP | MPI | CUDA | LangChain | LlamaIndex | OpenAI Gym | Natural Language Processing (NLP) | Deep Learning | Neural Networks | Transformers | CNNs | RNNs (LSTM/GRU) | Transfer Learning | Ensembling | Multimodal Learning | AutoML | Time Series Analysis | Reinforcement Learning

**Data Engineering & Analytics:** SQL | Apache Spark | PySpark | Hadoop | Pandas | NumPy | MapReduce | Hive | MongoDB | ETL/ELT Pipelines | Data Warehousing | Data Lake | Feature Engineering | Statistical Analysis | Big Data | Tableau | Matplotlib | Seaborn | Plotly

**Cloud & DevOps:** AWS | Docker | Apptainer | Git/GitHub | Databricks | MLflow | Weights & Biases | CI/CD | Containerization

**Frameworks & Tools:** React | Next.js | Node.js | Flask | Streamlit | Spring Boot | REST APIs. Material-UI | Tailwind CSS | Bootstrap | SLURM | HPC Systems

**Databases:** PostgreSQL | MySQL | MongoDB | Redis | Data Modeling | Database Design | Query Optimization

## Experience

**HPC User Support Specialist | New Jersey Institute of Technology**

MAR 2024 – MAY 2025

- Catalyzed research productivity for 400+ researchers via expert management of PyTorch/Conda environments, Docker containerization, and strategic GPU/CPU optimization across a hybrid cluster infrastructure.
- Orchestrated a SLURM-based benchmark suite to oversee node health using Python and Bash, diminishing manual checks by 80% and escalating cluster uptime by 200 hours quarterly.
- Built a Raspberry Pi-based HPC testbed for pre-deployment validation, lowering production cluster deployment risks by 30% through iterative testing.

**DATA ANALYST | Dassault Systemes**

JAN 2023 – DEC 2023

- Developed scalable Java ETL pipelines for 15+ SaaS product streams, processing enterprise-scale customer lifecycle data and supporting GDPR compliance across EU/NA teams, ensuring international regulatory standards.
- Designed and launched interactive dashboards for license conversion tracking with J2EE/Spring backends, improving decision-making efficiency by 25%.
- Automated data quality checks with SQL window functions and constraints, increasing pipeline reliability and reducing data inconsistencies by 40%.

## Projects

**DISTRIBUTED MACHINE LEARNING FRAMEWORK | C++, OPENMP, CUDA, MPI, PYTHON, PYBIND**

MAR 2025 – MAY 2025

[Custom-Distributed-Machine-Learning-Framework.git](#)

- Developed and optimized a parallel data ingestion pipeline using C++/OpenMP with thread-safe batch retrieval and parallel normalization, achieving a 12.6x throughput improvement over PyTorch DataLoader (193K vs 15K samples/sec) on the MNIST benchmark, enabling faster training iteration cycles for distributed workloads.
- Engineered custom CUDA kernels for neural network operations including tiled matrix multiplication with shared memory optimization, ReLU activation, and softmax cross-entropy loss, integrated with cuBLAS for GPU-accelerated forward/backward passes and CUDA streams for asynchronous execution.
- Implemented efficient multi-GPU gradient synchronization using CUDA-aware MPI with MPI\_Allreduce for all-reduce operations and OpenMP-parallelized gradient averaging, building a comprehensive benchmarking suite that demonstrated 14x faster batch processing compared to equivalent PyTorch implementations.

**FAIRSHARE - BILL SPLITTING WEB APPLICATION | NEXT.JS, REACT, PYTHON, PRISMA, DOCKER**

NOV 2024 – PRESENT

[FairShare.git](#) | [fairshare.kaivalya.dev](https://fairshare.kaivalya.dev)

- Architected a full-stack receipt scanning application integrating a hybrid LLM pipeline with Ollama (local) as the primary model and Perplexity API as fallback, reducing external API costs by 90% while maintaining extraction reliability for diverse bill formats.
- Engineered a penny-perfect bill splitting algorithm using the Largest Remainder Method to ensure mathematically fair distribution of shared expenses, with proportional tax/tip allocation and custom portion support for partial item claims.
- Developed a Python microservice leveraging Docling for intelligent OCR and table extraction from PDFs and images, paired with a Next.js 16 frontend featuring real-time collaboration via shareable bill links and Docker containerization for seamless self-hosting deployment.

[lman.git](#)

- Developed a CLI tool leveraging Retrieval-Augmented Generation (RAG) with sentence-transformers and FAISS vector indexing\* to enable natural language querying of Unix man pages, achieving semantic search across thousands of command chunks with sub-second retrieval latency.
- Engineered a multi-provider LLM integration layer\* supporting Perplexity, OpenAI, Anthropic, Groq, and local Ollama inference, with intelligent fallback from keyword-based matching to API calls and real-time token usage/cost tracking for transparency.
- Engineered a multi-provider LLM integration layer\* supporting Perplexity, OpenAI, Anthropic, Groq, and local Ollama inference, with intelligent fallback from keyword-based matching to API calls and real-time token usage/cost tracking for transparency.

QUANTITATIVE PORTFOLIO SIMULATOR | PYTHON, PANDAS, NUMPY, MATPLOTLIB, SEABORN

FEB 2024 – APR 2024

[Quantitative Portfolio Simulator.git](#)

- Visualized strategy results against a synthetic tech index benchmark using matplotlib dual-axis plots with percentage change normalization, enabling comparative performance analysis that revealed strategy-dependent returns ranging from -4.8% to +7.6%.
- Developed and implemented 3 algorithmic trading strategies (momentum, contrarian, and trend-following) with configurable 5-day rebalancing cycles on a \$5M virtual fund, incorporating automated dividend detection via Close/Adj Close ratio analysis and reinvestment calculations across 250 trading days.
- Engineered a data pipeline consolidating 10 stock datasets and FX rates into an aligned time-series structure, including OHLC and Adjusted Close handling, and integrated a daily USD/JPY FX conversion system for accurate JPY-denominated performance tracking.

Certifications

- |  |      |
|--|------|
| • NVIDIA Deep Learning Institute – Building RAG Agents with LLMs | 2025 |
| • Kaggle Introduction to Machine Learning                        | 2022 |