

Shrenik Jain

+1-(858)-241-1904 | shrenikkjain81@gmail.com | [website/shrenik-jain](#) | [linkedin/shrenik-jain9](#) | [github/shrenik-jain](#)

EDUCATION

University of California San Diego

Master of Science, Electrical and Computer Engineering (Machine Learning and Data Science)

University of Pune

Bachelor of Technology, Electrical Engineering - GPA: 4.0/4.0

WORK EXPERIENCE

Machine Learning Engineer, PlayStation (Sony Interactive Entertainment)

Jun 2025 - Present

- Designed real-time post-processing enhancement algorithms, addressing blocking, blurring, and ringing artifacts, and enhancing temporal coherence in high-frame-rate gameplay, ensuring consistent visual quality for 4M+ daily active users.
- Implemented single-step diffusion models for accelerated inference, achieving 15% gains in PSNR/PSNR-B and VMAF scores, enabling high-fidelity visual effects on constrained hardware while significantly reducing computational overhead.
- Leveraged encoder-side statistics (e.g., QP values, CTU-level features, motion vectors) to condition enhancement networks, enabling content-adaptive inference and more precise artifact suppression.

Machine Learning Engineer, Pivotchain Solutions

Jul 2022 - Aug 2024

- Spearheaded the development of a scalable event monitoring system, leveraging representation learning with probabilistic anomaly scoring to flag suspicious activities in real-time, leading to a 50% reduction in containment time through optimized detection and response workflows.
- Implemented a spatiotemporal autoencoder for anomaly validation, learning normal motion patterns, and flagging deviations in security footage; significantly reduced false event escalations by 20% and improved operational trust in the system.
- Deployed models into production by packaging them into modular inference services, integrated with monitoring dashboards that generated real-time alerts, audit logs, and forensic search tools for enterprise clients.
- Integrated a vector-store backend for embedding management, enabling efficient indexing, retrieval, and similarity search of feature embeddings in video streams, making ambiguous events searchable in sub-seconds.

Machine Learning Consultant, Pixstory

Aug 2023 - Mar 2024

- Developed a retrieval-augmented generation system for conversational search, combining vector similarity retrieval with LLM-based re-ranking to improve semantic relevance and reduce hallucinations.
- Accelerated query serving by implementing concurrent request handling and parallel execution across the API-database pipeline, improving throughput and reducing average response latency from 3s to 600ms.
- Optimized database and data pipelines for metadata (6M+ JSON records), restructuring queries and indexing strategies to reduce retrieval times and support real-time analytics workloads.

Software Engineer, Qualys Inc.

Jan 2022 - Jun 2022

- Designed and automated CI/CD pipelines with containerized workflows, reducing deployment cycles from 30 → 10 minutes and enabling 200+ production releases per month.
- Implemented observability pipelines (monitoring, logging, alerting) to track latency, failure rates, and resource utilization, ensuring stability under sustained high traffic.

Machine Learning Engineer, Validus Analytics LLP

Feb 2021 - Dec 2021

- Trained convolutional VAEs to synthesize distribution-consistent samples, expanding a limited corpus from 50K → 150K examples while maintaining perceptual similarity (SSIM > 0.85).
- Benchmarked VQ-VAEs against Conv-VAEs for unsupervised representation learning, evaluating latent space structure, reconstruction error, and generative quality across heterogeneous datasets.

RESEARCH EXPERIENCE

Applied Research Engineer, Spatiotemporal Machine Learning Lab

Sep 2024 - Present

- Conducted research on [DYffusion](#), a dynamics-informed diffusion model for spatiotemporal climate forecasting, focusing on improving uncertainty quantification and stochastic representation of geophysical processes.
- Implemented and evaluated an almost-fair CRPS loss function (adapted from recent literature) to address biases in standard CRPS variants, yielding a 10% gain in predictive accuracy while preserving calibrated uncertainty estimates.
- Ran large-scale experimentation on spatiotemporal datasets, analyzing the effect of loss function choice, sampling strategies, and noise schedules on forecast stability and calibration.

Applied Research Engineer, University of Pune

Jul 2021 - Dec 2021

- Spearheaded the design of a domain-adapted summarization system for research literature, integrating a BERT encoder with fine-tuned attention layers, achieving a 50% relative gain in ROUGE-1 compared to traditional extractive methods.

TECHNICAL SKILLS

Languages: Python, C++, Java, JAX, SQL, JavaScript, Bash

Machine Learning: PyTorch, TensorFlow, Torch Lightning, TFLite, Hugging Face Transformers, Langchain, LlamaIndex, Scikit-learn, OpenCV, CUDA, NLTK, ONNX

Frameworks & Technologies: Git, RESTful APIs (Flask, FastAPI, SpringBoot), gRPC, Django, Linux, AWS, Azure, GCP

Model Deployment & CI/CD: Model Serving (TorchServe, TF Serving, TritonServer), MLOps (Weights & Biases, MLFlow), CI/CD (Docker, Kubernetes, Jenkins, GitHub Actions), Monitoring (Prometheus, Grafana)

Data Engineering: MongoDB, Elasticsearch, SQL, Cassandra, Vector Databases (Milvus, FAISS), Apache Spark, Apache Kafka