

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) <i>Coursework: Recommender Systems, ML for (Physical Applications, Few Labels), Linear Algebra, Deep Generative Models, Statistical Learning</i>	Sep 2024 - May 2026
University of Pune Bachelor of Technology, Electrical Engineering - GPA: 4.0/4.0 <i>Coursework: Data Structures, Advanced Algorithms, Machine Learning, Deep Learning, Image & Video Processing</i>	Aug 2018 - May 2022

WORK EXPERIENCE

Machine Learning Intern, Sony Interactive Entertainment (PlayStation) • Collaborated with the Video Codecs Team to advance foundational research and development at the intersection of AI and video codec technologies, enabling intelligent compression and adaptive streaming capabilities.	Jun 2025 - Present
Applied Research Engineer, Spatiotemporal Machine Learning Lab • Conducted research for DYffusion , a dynamics-informed diffusion model for spatiotemporal climate forecasting, improving predictive accuracy by 30% through CRPS loss integration for stochastic representation and uncertainty quantification.	Sep 2024 - Present
Machine Learning Engineer, Pivotchain Solutions • Led a cross-functional team for the agile development of an AI System, designing Computer Vision algorithms to provide a scalable and real-time interface for autonomous event monitoring, achieving 70% faster threat containment (MTTC). • Implemented a Spatiotemporal Autoencoder model to validate AI-generated video clips, precisely classifying 15,000+ daily security events, reducing false positives by 30%, and saving \$250K annually in operational costs. • Optimized MongoDB queries to efficiently handle over 10 Million records of video surveillance JSON data, reducing data retrieval time by 55% and improving overall system performance for end-users across multiple regions.	Jul 2022 - Aug 2024
Software Development Intern, Qualys Inc. • Streamlined the orchestration of high-availability microservices deployment using Kubernetes, achieving 99% uptime and handling 1 Million+ daily requests across 3 production environments. • Architected end-to-end CI/CD pipelines with Groovy and Docker, reducing deployment time from 30 to 10 minutes and enabling 200+ successful deployments monthly across multiple environments.	Jan 2022 - Jul 2022
Applied Research Engineer, University of Pune • Spearheaded the development of a research paper summarization system using a BERT-based encoder, improving ROUGE-1 scores from 0.30 to 0.46 compared to baseline extractive summarization methods. • Conducted research on transformers and multi-head self-attention mechanisms, for enhanced language modeling.	Jul 2021 - Dec 2021
Machine Learning Intern, Validus Analytics LLP • Implemented ConvVAE-based generative modeling for dataset enhancement, expanding a critical training dataset from 50,000 to 150,000 samples while maintaining high perceptual integrity (SSIM > 0.85). • Analyzed Vector-Quantized VAEs (VQ-VAEs) and Convolutional VAEs (Conv-VAEs) for unsupervised learning of complex data via latent representations and generative modeling.	Feb 2021 - Dec 2021

CONSULTING EXPERIENCE

Machine Learning Consultant, Pixstory • Contributed to building a RAG-based Conversational Search System using Large Language Models (LLMs), improving search relevance and increasing average user session duration by 10 minutes. • Optimized system throughput and hardware efficiency 3x by implementing asynchronous requests and parallel execution strategies, which cut the average query response time from 3 seconds to 600 milliseconds.	Aug 2023 - Mar 2024
Software Development Consultant, AI For Rural • Collaborated with organization owners to engineer the end-to-end backend architecture for a website, including designing API, database schema, and core services, resulting in a 45% improvement in user engagement. • Developed REST APIs using SpringBoot with Hibernate, enabling efficient real-time data retrieval from MySQL databases and reducing response times for critical operations by 40%.	Sep 2021 - Nov 2021

TECHNICAL SKILLS

Languages: Python, C++, Java, JAX, SQL, JavaScript, Bash, HTML, CSS
Machine Learning: TensorFlow, PyTorch, MLFlow, Hugging Face Transformers, LangChain, Scikit-learn, OpenCV, CUDA, spaCy, NLTK, ONNX, TFLite, TorchServe, TritonServer, TF-Serving, WandB
Frameworks & Technologies: Apache Spark, Docker, Kubernetes, Flask, Django, SpringBoot, Git, Jenkins, Linux/Unix, REST, Apache Kafka, MQTT, RabbitMQ, FFmpeg, ONVIF
Cloud & Databases: AWS, GCP, Azure, MongoDB, Elasticsearch, SQL, Cassandra, Milvus, Vector Stores