Yashaswi Sharma

yashaswisharma.gradschool@gmail.com · (301) 503-8933 · Linkedin · GitHub

Education

University of Southern California — Los Angeles, CA
M.S. Computer Science — Aug 2025 - now
Courses: Deep Learning (CSCI566), High-Dimensional Math (EE546)
Research: Reasoning Economics, Differentiable Digital Signal Processing

University of Maryland — College Park, MD
Bachelors of Science in Computer Science (Machine Learning) — Aug 2018 - May 2022
Bachelors of Science in Economics — Aug 2018 - May 2022
GPA: 3.22
Research: Differentiable Audition, Differentiable Economics, Applied AI in Agriculture
Notable Courses: Deep Learning, Capstone in Machine Learning, App. Linear Algebra, Computer Vision, Computational Game Theory, Parallel Computing, Signal Processing, Econometrics, Adv. Macroeconomics

Research Experience

Earth System Science Interdisciplinary Center, University of Maryland — College Park, MD
AI Research Analyst — Jan 2021 - May 2022

- Conducted research in the DAWN Project which is a Department of Agriculture & National Science Foundation funded research project aimed at introducing modern automation and Machine Learning tooling to optimize agricultural practices, particularly to enhance corn and soybean production
- Created a recurrent neural network model to forecast Leaf Area Index, Evaporative Stress Index, and crop yield across the midwest and northeast US using Geopandas & Rasterio for data wrangling/cleanup and PyTorch for model engineering (Transformer architecture)
- Implemented a data pipeline to gather seed data and compute a vanilla Time Series regression to forecast valuable features (e.g. growing degree)

Neural Audio Compression Project — College Park, MD
Undergraduate Researcher — Sep 2021 - Dec 2021

- Project's main goal was to observe how audio compression can be done through Neural Networks
- An implied follow-up goal was to observe how much of the to-be-compressed information can the current class of 1-D convolutional neural networks (CNNs) encode in the intermediary layers at inference time
- Designed the Encoder-Decoder (Compression-Reconstruction) neural architectures with a differentiable custom compression function

Industry Experience

Snovation — Remote–Baltimore, MD
AI Engineer — Dec 2024 - Present

- Engineered an application-specific Foundation Model hosting service, maximizing token throughput by 200X (20,000 %) (from 0.34 to 67.1 tok/s) and reducing inference latency by 77% (from 180 sec to 42 sec) while optimizing resource utilization to a CPU-only architecture using 2 VCPUs
- Developed a Domain-Specific Language (DSL) converter for healthcare claims using advanced deep learning and compiler techniques, resulting in a 99%-accurate copilot deployed as an Azure Function
- Reduced sub-system end-to-end latency by 98% by improving data retrieval and processing methods, improving end-user experience through a 40+% reduction in latency

Amazon Web Services — Arlington, VA
Software Development Engineer — Sep 2022 - Sep 2024

- Designed and deployed scalable inference and feedback systems for Config's first Generative AI feature (**ReInvent 2023**), leveraging Function-as-a-Service (FaaS) architectures, automated testing, and adaptive rate limiting. This streamlined security & governance workflows for financial, government (civilian, defence), healthcare, and tech customers
- Architected and implemented a data management system with 99.99% availability & performance, leveraging core AWS services for state management, message processing, and anti-entropy mechanisms
- Built a data integration platform for other service teams to create configuration-based error-correction and security-compliance systems for customers, reducing integration time by 90+%

- Automated Java 8 to Java 21 code upgrades using large language models (LLMs), eliminating service latency, security, and availability issues across the platform while reducing developer effort by 75+%
- Formulated security-first processes to reliably deploy Config Manged Rules, reducing developer effort by 80+% and ensuring 90+% regional/partitional feature parity
- Led 150+ critical on-call investigations across diverse systems, including those with major global security implications, ensuring 24/7 system availability, reliability, and security

Shelly News College Park, MD
AI Research Intern Jun 2021 - Sep 2021

- Implemented a news recommendation engine which adjusted content delivery according to user preferences, behaviour, and location, employing deep learning and data pipelines through PyTorch, Pandas, and Flask
- Developed data ingestion pipelines for recommending connections to users by scraping Linkedin contacts using Selenium
- Presented a high-level overview of the recommendation engine to VC firms and Startup Accelerators

Sundaram Financial Mumbai, MH
Fixed Income Intern Jun 2019 - Aug 2019

- Implemented a news-data harvesting pipeline which pinged Portfolio Managers whenever a major event occurred that could impact portfolio assets
- Built custom trader-specific tooling and automations to reduce repeated work
- Contextualized macroeconomic events into regular reports in a fixed-income for Indian portfolios context
- Learned the integral components of Indian financial markets, especially fixed-income markets

Teaching Experience

Maryland Mentor Corp College Park, MD
Mentor Sep 2021 - May 2022

- Maryland Mentor Corps is a tutoring program in partnership between the University of Maryland and Prince George's County Public Schools
- Provided after school math tutoring and mentoring to fourth grade students who evaluated weakly in standardized tests
- Create a safe-space for students of lower-income backgrounds and ignited curiosity in STEM, STEM careers, and STEM related research by showcasing exciting advancements at the University of Maryland

Department of Economics, University of Maryland College Park, MD
Teaching Assistant Sep 2020 - Dec 2020

- Instructed 300+ students as the sole undergraduate Teaching Assistant for Money and Banking, University of Maryland's primary Financial Economics course
- Held weekly review sessions, guiding students through practice exercises and post-lecture material, including self-created material covering modern financial instruments
- Assisted students in refining their fundamental and applied knowledge of mathematics by holding math review sessions

Certificates

AWS Machine Learning Engineer Associate (MLA-C01) Verification Link
Validity May 2024 - May 2027

Notable Projects

Multi-Agent Reinforcement Learning for Stackelberg Competition Project Link
PyTorch, OpenAI Gym, Matplotlib, Seaborn
Developed a simulation of an Enhanced Stackelberg Competition using multi-agent reinforcement learning, modeling leader-follower dynamics in economic markets

HIV Inhibitor Molecule Prediction Project Link

PyTorch, PyTorch Geometric, Matplotlib, Seaborn, Torchviz, HuggingFace, Docker, GIT
Implemented a Graph Transformer based model with custom Message Passing to predict if a molecule can inhibit HIV. Dataset from OGBG-MOLHIV

Technical Skills

ML Performance/Inference:	PyTorch, Tensorflow, Scikit-Learn, XGBoost, CUDA, MPI, OpenMP, JAX/XLA, Quantization, Amazon Comprehend, SageMaker Training Compiler, Inferentia, Neo
MLOps:	AWS SageMaker Pipelines, Feature Store, Model Registry/Monitor, Clarify, Debugger GroundTruth, Amazon Bedrock, CI/CD (CodeBuild, CodeDeploy, CodePipeline)
Distributed Systems:	AWS API Gateway, SNS, SQS, VPC, EC2, Lambda, Step Functions, ECS, Fargate, ECR, EKS/Kubernetes, Docker, gRPC/Protobuf, REST, Azure Serverless Functions
Data Engineering:	AWS Glue (DataBrew, Data Quality), Athena, Redshift, Kinesis (Data Streams, Firehose), Kinesis Flink, DynamoDB, OpenSearch, RDS, Batch, EMR, Lake Formation, Spark
Governance & Security:	AWS IAM, VPC Endpoints/PrivateLink, GuardDuty, CloudTrail, CloudWatch, Config, Security Hub, Control Tower, Cost Explorer, Macie, KMS, SecretsManager
Programming Languages:	Java, Kotlin, Python, Go, Rust, C/C++, OCaml, TypeScript, SQL

Interests & Personal Skills

Languages:	English (Native), Hindi (Native), French (Intermediate), Arabic (Beginner), Bhojpuri (Beginner), Magahi (Beginner)
Sports:	Volleyball, Golf, Basketball
Fun:	Travel, Sci-fi Books, History Books, Movies
Goals:	Learn to cook Italian/Indian/French foods well, Learn photography on iPhone