# HARMANPREET SINGH

+1 (848)-313-6708 • harmanpunn@gmail.com • Linkedin • Portfolio

## EDUCATION

### Masters in Computer Science

Rutgers, The State University of New Jersey • New Brunswick, NJ

### Bachelors of Technology in Electronics and Communication

Dr. B.R Ambedkar National Institute of Technology • Jalandhar, India

## SKILLS

**Languages**: Python, R, Java, JavaScript, HTML/CSS

**Libraries & Frameworks**: Pandas, Numpy, PyTorch, TensorFlow, scikit-learn, LangChain, FastAPI, CrewAI, LangGraph, Agents SDK, React.js

**Cloud & Databases**: AWS (S3, SageMaker, Redshift, Quicksight, EC2, Lambda), SQL, MongoDB, VectorStore (Pinecone, FAISS)

**Engineering:** Machine Learning, MLOps, Data Ingestion, LLMs, RAGs, Docker, Kubernetes, Hadoop, Spark, Unit & A/B Testing, CI/CD

**AI Dev Tools**: Cursor, Github Copilot, Claude Code, ChatGPT, Amazon Q Developer

## PROFESSIONAL EXPERIENCE

### INVIDI Technologies

**Data Science & ML** **June 2024 – Present**

*(Python, Machine Learning, Forecasting, Analytics, QuickSight, Looker, Time Series, Regression, AWS, REST APIs)*

- Owned end to end **MLOps pipeline** using AWS Sagemaker and API Gateway to deploy ML models as scalable, serverless endpoints, integrated with CI/CD for automated deployment and real-time monitoring.
- **Collaborated with business stakeholders** to translate forecasting requirements into technical specifications, ensuring ML models aligned with business objectives.
- Developed personalized impression **forecasting models** (ARIMA, SARIMA, XGBoost) that powered ad delivery decisions across surfaces, reaching **60–70% accuracy** and improving inventory-level planning.
- Developed an ML-driven ad scheduling optimization system, improving **pacing accuracy by 15% and reducing under/over-delivery by 30%** through dynamic frequency adjustments and forecast integration.
- **Gathered requirements and evaluated tools with stakeholders** to build a comprehensive custom reporting framework for ad-hoc analytics, prioritizing cost-effectiveness and usability.
- Developed and deployed multiple reporting dashboards in **QuickSight**, providing actionable insights on **inventory utilization** and **tracking impressions**, reducing manual report processing.

### Visa Inc.

**Senior Software Engineer** **August 2021 – August 2022**

*(Java, JavaScript, Python, Adobe Experience Manager (AEM), React.js, Machine Learning, REST APIs)*

- **Collaborated with product teams** to redesign entire Visa's intranet, improving retention and **user experience** by **40%**.
- **10x improvement** in jobs listing page response time by implementing targeted scheduling & caching.
- Built and deployed an ML-based job recommendation system, incorporating personalized ranking logic and clickstream behavior to boost user engagement and CTR.
- Enhanced Visa's chatbot using NLP, improving response accuracy and **cutting query resolution time by 30%.**
- Constructed **10+ REST APIs** for Employee Dashboard, facilitating efficient data integration across various product teams.
- As a stretch goal, explored document parsing for text extraction from financial documents, improving data accessibility.

### Netcentric, A Cognizant Digital Business

**Backend Engineer** **November 2019 – July 2021**

*(Java, JavaScript, AEM, Adobe Analytics, Machine learning, Angular, OSGi, AWS, CSS/SCSS)*

- Translation of **50+ wireframes** for an automobile client into functional requirements, and subsequently into technical design.
- Applied Adobe Analytics for real-time segmentation and marketing channel optimization, increasing targeted engagement.
- **60% automation** of workflow by implementing CI/CD pipelines for deployment and testing.
- Deployed Brokers (RabbitMQ), REST APIs and remote procedure call to interconnect microservices.
- Integration with SaaS based translation service to translate campaign content into a wide range of languages for a hospitality client.
- Implemented functionalities like personalized navigation, SSI, and dedicated console to add/edit hotel data.

### Publicis Sapient

**Associate Technology** **December 2017 – October 2019**

*(Java, JavaScript, Python, AEM, Django, Analytics, Machine learning, Apache Sling, CSS/SCSS)*

- Developed components, templates, and services to drive the entire Roche Diagnostics website improving functionality and user experience.
- Improved **search performance by 15%** through integration with Search & Promote.
- Designed and trained an XGBoost based employee attrition prediction model stored in AWS S3 with **81% accuracy**.

## INTERNSHIPS & ACADEMIC ROLES

### INVIDI Technologies

**Software Developer Intern** **May 2023 – May 2024**

***Advanced Inventory Scheduling System for Efficient Ad Campaign Delivery***

- Devised an efficient inventory scheduler that matches media inventory with campaign data, improving ad placement based on a range of scheduling rules.
- Achieved **98% utilization rate** of ad inventory by designing and implementing advanced scheduling rules, including day parts, days of the week, network inclusions/exclusions, and separation logic.
- Analyzed **4 years** of TV audience data and large AWS Redshift data sets, providing actionable viewership insights across **30+ markets** and revealing key data relationships.

### Rutgers University

**Research Assistant, Dr. Matthew Weber** **February 2023 – May 2024**

***Longitudinal Study of Local News in New Jersey***

- Comprehensive analysis on **millions** of news records from **700+ domains** within New Jersey region using tools like NLP, Named Entity Recognition, and machine learning algorithms such as **k-means** clustering and random forest.
- Visualized regional news trends across **21 counties** using **pandas** and **pySpark**, and managed AWS model deployment.
- Analysis and identification of news deserts by mapping and monitoring geographical coverage in local news.

**Full Stack Developer and Machine Learning Engineer, GRID** **March 2023 – May 2023**

- Developed a language learning web application using Next.js and FastAPI, with AWS services for storage and deployment.
- Led UI/UX design, database schema, and integrated a ML model for real-time emotion detection with **72.4% accuracy**.

**Research Assistant, Dr. Ana Paula Centeno** **November 2022 – February 2023**

***Data-Driven Analysis of CS Enrollment and Performance Trends***

- Comprehensive analysis of enrollment and performance trends over **5 years** in Rutgers' foundational computer science courses using advanced data analytics and visualization tools.
- Research on gender-based disparities in computer science enrollment and performance, utilizing data science to promote equity and inclusion in higher education.

## PROJECTS

### Vygil AI - Autonomous Activity Tracking & Anomaly Detection Link

- Built **agentic AI system** with autonomous decision-making, using computer vision and **multi-LLM** architecture for **real-time analysis**.
- Implemented **multi-modal agentic pipeline** with screen analysis, memory persistence, and **YAML-configurable** autonomous agents.

### Deep Research: Multi-Agent AI Research Flow

- Developed asynchronous multi-agent AI application using OpenAI models with specialized agents for search planning, web research, report synthesis, and automated email delivery.
- Scalable research automation pipeline with UI interface, concurrent search execution, structured data validation via Pydantic.

### InsightWing: AI-Driven Web Content Summarizer Link

- Developed a Chrome extension utilizing **FalconLLM** and **LangChain** for efficient **60-word web content** summarization.
- User-friendly interface with HTML/CSS and JavaScript and integrated a chat feature for interactive content engagement.

### Document Question Answering System with LangChain and LLMs Link

- Built a Document QA system using LangChain, HuggingFace Transformers, and FAISS for retrieval-augmented generation.
- Utilized HuggingFace embeddings and FAISS for efficient document retrieval and response generation.

### Global Socioeconomic Patterns and Risk Factors in Suicide Trends Link

- Analyzed the impact of GDP on suicide rates globally using R, revealing key economic correlations.
- Examined age and gender factors affecting suicide, providing insights through **data visualizations**.

### StyleGAN Implementation and Few-Shot Generative Domain Adaptation

- Implemented **StyleGAN** from scratch on the **FFHQ dataset**, achieving benchmark results with FID and PPL metrics.
- Employed **Few-Shot GDA** via Domain Re-modulation (**DoRM**) to adapt StyleGAN across diverse datasets, incl. MetFaces and Anime faces.

## BLOGS & APPRECIATIONS

- Winner of an Agentic AI Hackathon organized by OSS4AI for building Vygil AI.
- Published a Medium article that covers an intuitive library for form validation. Joi – Form validation made simple.
- Received “Rookie Award” and “Made a difference” for exceptional performance and amazing client impact.