KHUSHI AGRAWAL

(623)212-5340 • kagraw22@asu.edu • linkedin.com/in/khushi-agrawal-83aa0a202/

EDUCATION

Master of Science, Computer Science **Aug 2024 - May 2026**
Arizona State University, Tempe, AZ (**CGPA - 4.0/4.0)**
Relevant Courses: Data Processing at Scale, Data Mining, Semantic Web Mining, Statistical Machine Learning

Bachelor of Technology, Computer Science and Engineering **Jul 2019 – Jul 2023**
Vellore Institute of Technology, Vellore, India (**CGPA – 3.56/4.0)**
Relevant Courses: Machine Learning, Data Warehousing, Image Processing, Artificial Intelligence, Cybersecurity

TECHNICAL SKILLS

Programming Languages: C/C++, Python, Java, SQL
Machine Learning and AI: Presidio, Scikit-Learn, FinBERT, LLMs (Claude, GPT-4.0), Tensorflow, PyTorch, Model Evaluation
Data Engineering: PySpark, Cassandra, Hadoop, Numpy, Pandas, ETL, Scikit-learn, Matplotlib, Kafka, Numpy
Databases: MySQL, Postgres, Oracle, MongoDB
Cloud Computing and DevOps: CI/CD, Git, Github, Jupyter Notebook, Postman, AWS, Azure, GCP, Bitbucket
Certifications: AWS Certified Cloud Practitioner
Tools and Frameworks: API Integration, Jira, RESTful APIs

PROFESSIONAL EXPERIENCE

DruvStar, Tempe, USA: Data Analyst Engineer Intern **May 2025 – Present**

- Designing and implementing scalable data classification models for structured/unstructured data using Python and regex-enhanced ML, enabling sensitive data detection at scale.
- Engineering production-ready, agent-based DLP pipeline with **Dockerized** Python and batch orchestration, supporting large-file inspection with custom PySpark ETL.
- Tuning data tagging models based on confidence thresholds and context heuristics, reducing false positives in classification by **40%** across test datasets.

Accenture, Hyderabad, India: Advanced App Engineering Analyst **Oct 2023 – Aug 2024**

- Built PySpark-based pipelines processing **5TB/day** across HDFS and Hive, ensuring high-throughput ingestion for analytical workloads in ad-targeting and campaign analysis.
- Developed statistical validation modules to ensure data integrity, decreasing noise and outliers in downstream ML models by **70%**.
- Optimized partitioning and query performance of identity-linked datasets (>1B rows), improving ad-performance data retrieval speed by **35%**.

Ataloud Technologies, Mumbai, India: Full Stack Developer Intern **Sep 2022 – Mar 2023**

- Integrated **40+ REST APIs** to enable smooth data flow between frontend and backend systems, improving the reliability of microservice communication.
- Worked with backend and product teams to define API structures, helping launch **25+** new features that relied on consistent and accurate data.
- Improved data handling and display logic in React and Next.js apps, making sure analytics and user data were shown correctly for over **10,000** users.
- Tested and fixed issues in API-connected features, resolving **77%** of critical data bugs and ensuring accurate responses were sent from backend to UI.

Deepija Telecom Pvt. Ltd., Hyderabad, India: Software Developer Intern **May 2022 – Jun 2022**

- Worked with QA and frontend teams to make dashboards load faster by **25%** by improving how data was fetched and displayed, helping users handle large reports more smoothly.
- Cleaned up and reorganized how data was shown on the screen, increasing the visible area by **80%** so users could view more information at once with less scrolling.

ACADEMIC PROJECTS

Online Judge, Remote **Spring 2025 – Present**

- Implementing a system that runs code securely in different programming languages, using **Docker** to keep each submission separate. Set it up on **Heroku** with automated updates so it stays reliable without manual work.
- Using Kafka to handle code results in real time, so users get updates almost instantly — even when more than **100** people submit at once, the system stays fast and smooth.

LLMind, Tempe, AZ **Spring 2025**

- Developed a data pipeline to process and analyze over **1 million** financial tweets, feeding structured data into LLMs (**GPT-4.0**, **Claude**, **DeepSeek**, **FinBERT**) for sentiment-driven forecasting on AAPL, MSFT, and AMZ.
- Handled large-scale data preprocessing and pipeline orchestration, improving input quality and achieving **0.965** accuracy with FinBERT on MSFT — showing how domain-specific models can outperform general-purpose ones.