

Ashit Kumar Singh

aashit.singh18@gmail.com | [LinkedIn](#): aashit-singh | [GitHub](#): aks18 | +91 7018602154 | India

CAREER SUMMARY

Engineering Manager with 8+ years in AI/ML Systems Design, Data Science, Stats & Agents. 4+ years leading teams of 8-10 engineers; working with non-tech stakeholders; translating ambiguous business problems into scalable & reliable solutions; owning the technical roadmap, design, execution & delivery of large-scale production grade cloud AI systems.

ENGINEERING LEADERSHIP

- Led cross-functional teams of 8-10 ICs in agile & matrixed environments (DS, DE, MLE, App Eng., analytics & business) across 2-3 pods, with rapid prototyping from concept to production; owned interviews & hiring of data scientists & AIML engineers, performance reviews & mentoring.
- Experience in translating ambiguous business problems to scalable AI/ML systems, building enterprise platforms, scalable data pipelines & user-facing tools for real-world problems & automation.
- Established engineering rigor: design docs, code reviews, testing, incident management & postmortem docs.
- Converted future business requirements into 6-12 months technical roadmaps & delivery plans.

PROFESSIONAL EXPERIENCE

Tiger Analytics

Hyderabad-Remote, TG

Associate Principal (Senior Manager)

January 2026 - Present

- Led **8 engineers** (2 pods) to design & launch **API-first distributed multi-agent AI platforms (Azure & GCP)** for DS workflows & automated codebase documentation, with production-grade CI/CD.
- Architected scalable systems with **standardized API contracts & data schemas for agent interoperability; cloud services, enforcing SLOs** across latency (P95 TTFT <1s, P90 E2EL <10s), accuracy (P95 task success rate), reliability (loop halt) & cost; enabling **reusable platform components** for AI workflows.
- Delivered **99% reduction** in time-to-first-model, **600% increase** in legacy system adoption & **97% reduction** in documentation effort across applications.
- Built multimodal video processing pipelines using **GCP (Gemini 2.5 + Vertex AI + GCS)**, expanding platform capabilities for documentation generation.
- Owned **6-month technical roadmap**, design & code reviews, cross-functional execution, aligning engineering delivery with business impact (productivity, adoption, CSAT) and non-tech stakeholders management.

Lead AIML Engineer (Engineering Lead/Manager)

January 2025 - December 2025

- Led **8 engineers** (2 pods) to build & productionize a **scalable GenAI bot + data platform (Azure & Databricks)** with reusable components & API-driven interfaces; owned system design, roadmap & end-to-end delivery.
- Architected & launched an AI-powered ingestion & digitization pipeline processing **70K+ unstructured files** into structured SQL datasets, achieving **97% accuracy and 99% reduction** in manual effort.
- Built a self-serve GenAI analytics & reporting system (**agents + text-to-SQL + RAG**), enabling non-technical users to query data; **improved TAT by 99.9% with 100% adoption and 92% CSAT**.
- Established an **SLO-driven engineering discipline**, optimizing latency, accuracy & cost for LLM workloads at scale (querying 250M+ records); **owned 12-month technical roadmap**, code reviews, delivery & mentoring.
- Drove **cross-functional execution** & designed a **configurable, extensible platform architecture** to rapidly onboard new use cases.

Senior AIML Engineer (Team Lead)

January 2023 - December 2024

- Led **4 engineers** to design & deploy a **scalable & configurable OCR + LLM + NLP data ingestion** platform (Azure & Databricks) extracting structured data from unstructured documents (batch processing 500+ GB/day).
- Built a data ingestion & validation layer powering SQL-based analytics (Power BI), **reducing manual digitization effort by 90%** & saving 1000s of hours.
- Developed ingestion and validation algorithms & pipelines for reliable & fast document matching & processing.
- Achieved **95% extraction accuracy, 100% critical attribute integrity & 1% hallucination rate**, ensuring high data quality for downstream systems.

- Owned end-to-end **system design, SLOs and 10-month roadmap**, driving cross-functional execution across **engineering and business stakeholders**.
- Built a GenAI-powered RAG assistant for customer support, reducing **AHT by 21% and improving CSAT by 15% with 92% adoption**.

Data Scientist

November 2021 - December 2022

- Built and deployed an **XGBoost-based customer segmentation model**, reducing marketing spend by **40% & saving \$25M+** with minimal impact on response rates.
- Developed an **OCR + CNN + NLP pipeline** for extracting & classifying data from web-scraped CPG images for a Fortune 500 client.
- Built **automated ML pipelines for retraining & monitoring**, ensuring model performance & **preventing drift** in production; led E2E model development across various projects, **from problem framing to deployment & data-driven decision-making**.

Turtle Shell Technologies (Dozee)

Bengaluru, TG

Data Scientist, R&D

October 2019 - November 2021

- Researched, developed & deployed from scratch a **first-of-its-kind ML model** using **CNNs & signal processing** algorithms, that can **predict BP** from body vibrations captured using a **contactless** device.
- **High correlation (0.92) to actual BP, with 83% detection rate**, directly boosting sales of the product & subscription by 600% in 2 months, followed by series A funding.
- Researched & developed various distinctive traditional, CNNs & time-series ML models, algorithms & metrics, maintained & upgraded existing ones, helping raise pre-series A & series A rounds.

TECHNICAL SKILLS

Languages: Python, SQL, C++, Shell Scripting.

Cloud & Infra: Azure, GCP (Vertex AI, Cloud Storage, BigQuery, Gemini), Databricks, AWS.

Data & ML Systems: Data Pipelines (Batch/Streaming), ETL/ELT, Data Modeling, Feature Engineering, ML Infrastructure (training, inference, monitoring), RAG Systems, LLMOps.

AI Systems & Agentic Architectures: Multi Agent Systems Design, Context Engineering, Prompt Engineering, Agent Orchestration, Evaluation & Guardrails, Business Value Translation

APIs, Distributed Systems & Platform Engineering: Distributed Systems Design, API Design (REST/gRPC), Service-Oriented Architecture, Asynchronous Systems, Event-Driven Architectures, API Contracts & Schema Design (OpenAPI), Versioning & Backward Compatibility, Multi-tenant Systems.

Performance, Scale & Reliability: System Scalability, High Availability, Fault Tolerance, SLO/SLI Design, Latency Optimization, Cost Optimization.

Tools & Frameworks: LangGraph, LangChain, ADK, MSFT Agent Framework, Semantic Kernel, A2A, MCP, LLMs, SKLearn, PyTorch, PySpark, XGBoost, Tensorflow, OpenCV, Git.

ML & DS Skills: Statistics, Supervised, Unsupervised, Deep & Reinforcement Learning, Signal Processing, Algorithms & Data Structures (ranked <200 globally - **CodeChef**).

Soft Skills: Outstanding leadership, stakeholder management, mentoring & interpersonal skills, critical thinking from first principles.

RESEARCH PUBLICATIONS & AWARDS

- **[Publication]** Classification Of Sleep-Wake State In A Ballistocardiogram System Based On Deep Learning, published in 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society.
- **[Publication]** A Novel Method for Measuring Myocardial Performance Index using Non-contact Ballistocardiogram System, published in Cardiology & Angiology: An International Journal.

- **[Presentation]** Non-Invasive measurement of Systolic time intervals using a novel Ballistocardiogram: A Validation study - presented in the Asia Pacific Heart Rhythm Society Virtual Congress 2020, Hong Kong, 30th October 2020. Nominated for the Young Investigator Award.
- Won **Microsoft's Supercharged Agentic Hackathon** India, innovation & practical GenAI application at scale.

EDUCATION & CERTIFICATIONS

Master of Science

Computer Science

Georgia Institute of Technology
(starting Fall 2026)

Bachelor of Technology

Computer Science & Engineering, 7.86/10

National Institute of Technology, Hamirpur
August 2013 - May 2017

Google Cloud Platform Badge

L400

GCP

September 2024

Specialization - Deep Learning (5 courses)

Transformers, Structuring ML Projects, Fine Tuning, Speech Recognition, DNNs, CNNs, NLP

Coursera

October 2021

Stanford Certification in Data Science & ML

Deep Learning, Classical ML, Gradient Descent, Backpropagation

Coursera

June 2019 – July 2019

Tensorflow for AI, ML & DL

Coursera, July 2019