Krishna Balaji

+1 (480) 234-1166 | krishna311004@gmail.com | linkedin.com/in/krishnabalaji | github.com/krishna31102004 | krishnabalaji.dev

Education

Arizona State University Tempe, AZ

B.S. in Computer Science (Honors), Minor in Business, GPA 4.00/4.00 *Expected May 2026*

- **Relevant Coursework**: Data Structures & Algorithms, Operating Systems, Database Systems, Software Engineering, Machine Learning, Artificial Intelligence, Linear Algebra, Probability & Statistics, Discrete Mathematics

Technical Skills

Languages: Python, Java, C/C++, JavaScript/TypeScript, HTML/CSS, SQL, Swift
Frameworks: FastAPI, Flask, React, Next.js, Node.js, Express, Spring Boot, Pandas, NumPy, scikit-learn
Infrastructure: Docker, AWS (EC2, S3, Lambda), Linux, PostgreSQL, MongoDB, Redis, Git, GitHub Actions
Data & ML: XGBoost, Polars, Parquet, OpenAI API, Gemini API, Sentence Transformers, TF-IDF

Experience

Software Engineering Intern Aug 2025 – Present

Circle.ooo *Tempe, AZ*

- Engineered a unified event registration service using Node.js and PostgreSQL with transaction-safe writes, reducing onboarding flow from 4 steps to 1 and increasing successful sign-ups by 35% across 500+ registrations.
- Integrated OpenAI API services behind REST endpoints with structured logging and error monitoring, reducing mean time to resolution (MTTR) for production issues by 40% through improved traceability and alerting.
- Automated service deployment using GitHub Actions CI/CD with Docker, eliminating 70% of manual configuration steps and supporting 500+ registrations in the first production month.

Research Software Engineer Intern Aug 2025 – Present

Arizona State University *Tempe, AZ*

- Built Python ETL pipelines using Pandas and FastAPI with Redis caching for multi-season Fantasy Premier League datasets, reducing data processing time by 75% and enabling 100+ weekly ML experiment runs with reproducible results.
- Implemented modular ML pipelines using scikit-learn and XGBoost, reducing experiment iteration time by 40% through reusable components and consistent evaluation across 5+ model variants.
- Deployed predictive model outputs to interactive dashboards backed by PostgreSQL, supporting 50+ weekly active users for real-time sports performance analysis across multiple seasons.

Machine Learning Engineer Intern May 2025 – Aug 2025

Quantiphi *Mumbai, India*

- Developed Python ETL pipelines using Pandas with Pydantic schema validation to normalize insurance claim data across 50,000+ monthly records, reducing manual data review time by 80% through automated quality checks and error handling.
- Built document extraction services using Gemini 1.5 Pro with structured JSON outputs and confidence thresholding above 0.85, improving field-level extraction accuracy from 72% to 90% across 10,000+ processed insurance claims.
- Deployed production microservices processing enterprise-scale datasets, reducing claim approval cycle time by 25% and cutting operational overhead through automated document parsing and validation workflows.

Projects

VisuaLSM Storage Engine | *Java, Concurrent Data Structures, Spring Boot, React* Dec 2025 – Jan 2026

- Engineered a persistent LSM-tree key-value storage engine in Java with ConcurrentSkipListMap-based MemTable and background SSTable compaction, achieving 100,000+ writes/second in local benchmarks on commodity hardware.
- Implemented write-ahead logging with CRC32 checksumming for durability and automated crash recovery via sequential log replay, ensuring zero data loss across 50+ simulated failure scenarios with crash-consistent writes.
- Built interactive React visualization dashboard with Spring Boot REST API to display real-time storage engine metrics including write throughput, compaction status, and SSTable hierarchy for educational demonstration.

Quant360 – Quantitative Analytics Engine | *Python, FastAPI, Polars, Redis* Nov 2025 – Dec 2025

- Architected a FastAPI backend with asynchronous endpoints and Redis caching to serve financial time-series analytics for 1,000+ stock tickers, achieving sub-200ms API response times through efficient query optimization and data pre-loading.
- Engineered high-performance data pipelines using Polars lazy evaluation and Parquet columnar storage, reducing 5-year historical backtest query latency by 80% and enabling factor analysis across 15M+ OHLCV data points.
- Implemented vectorized technical indicator calculations (RSI, MACD, Bollinger Bands) using NumPy operations for quantitative strategy backtesting with end-to-end processing under 3 seconds for multi-year datasets.

AI-Powered Resume Screening System | *FastAPI, Sentence Transformers, TF-IDF, React* Aug 2025 – Sep 2025

- Built a FastAPI backend with PDF/DOCX parsing capabilities using PyPDF2 and python-docx, computing semantic similarity scores via Sentence-BERT embeddings and TF-IDF vectorization to rank 200+ test resumes against job descriptions.
- Developed a responsive React frontend with real-time search and filtering, reducing manual recruiter screening effort by 40% through automated candidate ranking based on multi-dimensional skill matching algorithms.

Fantasy Premier League Forecasting System | *Python, FastAPI, scikit-learn, Pandas* Aug 2025 – Sep 2025

- Built an end-to-end ML pipeline and FastAPI inference endpoint to predict player performance using 5+ seasons of historical Premier League data, implementing feature engineering for 20+ statistical metrics including form, fixtures, and opponent strength.
- Developed reproducible model training workflows using scikit-learn with cross-validation and hyperparameter tuning, enabling systematic comparison across 3+ model families and 20+ engineered features for weekly prediction.