

Sree Tejomayi Yasaswini Tatikonda

✉ yasaswinitatikonda1@gmail.com

☎ (515) 916-4367

in linkedin.com/in/tatikondayasaswini

github.com/SreeTatikonda

Summary

Software engineer with 3+ years building production systems across web, backend and AI workflows. Skilled in developing APIs, full-stack applications and LLM-based systems. Experienced in evaluation frameworks, simulation and benchmarking to measure AI behavior. Strong background in distributed systems, cloud and databases. Known for delivering maintainable solutions in collaborative, cross-functional teams.

Experience

Process Automation and AI Associate (CO-OP)

CDS Global, Des Moines, Iowa

May 2025 – Present

- Built an OCR-to-JSON extraction pipeline using Tesseract, BERT, LayoutLM and rule-based normalization, improving extraction accuracy to 88% across multiple document formats
- Fine-tuned LayoutLM to handle structurally inconsistent documents, raising model accuracy to 79% and improving end-to-end extraction performance to 91%
- Created evaluation datasets from 15,000+ production documents to benchmark model performance, establish accuracy baselines and validate extraction quality across document types
- Designed an NLP parser that identifies and extracts required fields even when OCR output is partially recognized or incomplete, achieving 94% accuracy
- Built FastAPI backend services, containerized with Docker, deployed on Kubernetes, and validated endpoints using Postman
- Built a .NET executable to interface the AI pipeline with legacy production systems, enabling direct system integration
- Performed GPU-based model training and evaluation across multiple architectures, establishing performance benchmarks and supporting iterative model improvement
- Collaborated with cross-functional teams to interpret extraction requirements, validate output quality and align technical decisions with operational needs
- Managed Git workflows for reproducible experiments, structured branching and clean release preparation

Graduate Research Assistant (AI Full-Stack)

Iowa State University, Department of Computer Science, Ames, Iowa

Nov 2024 – Present

- Designed and implemented an evaluation framework for LLM behavior in educational contexts, establishing metrics to measure citation accuracy, grounding consistency and response reliability across 1,000+ test scenarios
- Built a RAG pipeline using ChromaDB embeddings and custom ranking logic, achieving 94% citation accuracy and a 97% retrieval success rate across 500+ evaluation queries
- Developed simulation workflows to test model behavior under varied retrieval conditions, measuring how the system responds when high-quality sources are unavailable or retrieval confidence is low
- Integrated Llama 3 and Gemma 7B through Ollama and configured prompts using Hugging Face and LangChain tools, enabling streaming inference with a 5.1-second first-token latency and reducing hallucination rates across 1,000+ tests
- Designed and implemented a FastAPI backend with authenticated endpoints, asynchronous inference routing and structured logging, supporting 120 concurrent requests with a 400 RPS design target
- Implemented monitoring systems to track retrieval performance, LLM latency metrics and grounding accuracy in real-time, providing observability into model behavior and system health
- Implemented retrieval-confidence scoring with a Tavily fallback path, improving grounded response consistency to 98%
- Developed the React and TypeScript frontend with reusable components, state management and real-time token streaming
- Optimized retrieval and prompt strategies using Hugging Face evaluation utilities, improving grounding accuracy and cutting unsupported LLM outputs by 40%
- Conducted comprehensive benchmarking across multiple dimensions including retrieval precision, citation accuracy, hallucination rates and response latency, connecting technical metrics to real-world usability outcomes

Software Developer – App Development

Iowa State University, Ames, Iowa

Feb 2024 – Sept 2024

- Built a React Native mobile app using JavaScript and Redux, achieving 95% feature parity on iOS and Android
- Developed Flask REST APIs with JWT authentication for secure image upload and ML inference
- Integrated TensorFlow CNN models with 92% inference accuracy and supported 1,000+ image-processing requests per day
- Designed a dual-database architecture using PostgreSQL for structured data and MongoDB for unstructured metadata
- Improved frontend performance using lazy loading and virtualization, reducing page load time by 30%
- Performed load testing with pytest, validating backend reliability under 500 concurrent transactions

App Development Associate

Accenture, Bengaluru

Oct 2022 – Dec 2023

- Built .NET Core and ASP.NET APIs in C#, improving request performance through optimized routing and middleware design
- Implemented OAuth 2.0 authentication for secure access to high-traffic internal services
- Improved SQL Server performance by tuning queries and indexes, reducing latency by up to 35%
- Enhanced API throughput with distributed caching and optimized data-access patterns
- Created Azure DevOps pipelines for automated builds, testing and deployments, increasing release reliability
- Developed unit and integration tests with xUnit and MSTest to reduce regression issues in production
- Supported feature development and production deployments for services handling 100,000+ daily transactions

Skills

Programming: Python, JavaScript, TypeScript, C#, SQL, Java, Go (learning)

AI & ML: LLMs, RAG pipelines, transformers, NLP, Hugging Face tools, LangChain, OCR pipelines, PyTorch, CNNs

Frameworks: FastAPI, Flask, .NET Core, ASP.NET MVC, React, React Native, Next.js, Bootstrap

Databases: SQL Server, PostgreSQL, Azure SQL, ChromaDB, SQLite, schema design, query optimization

Cloud & DevOps: AWS, Azure DevOps, Docker, Kubernetes, CI/CD pipelines, Firebase, Fastlane, CodePush

Education

Masters in Computer Science

Iowa State University

Dec 2025

Iowa

Relevant Coursework: Data Structures & Algorithms, Operating Systems, Cloud Computing, Machine Learning, Computer Networks, Advanced Principles of System Architecture, Deep Learning

Bachelor of Technology in Computer and Electrical Engineering

Jawaharlal Nehru Technological University Kakinada (JNTUK)

May 2022

India

Projects

C to JVM Compiler

- Implemented a compiler translating a subset of C into JVM bytecode, including lexical analysis, parsing, AST construction and semantic checks
- Developed in Java using a custom grammar and bytecode generator, producing valid .class files and supporting end-to-end execution of compiled programs

Triple Town Reinforcement Learning Agent

- Built a reinforcement learning agent for the Triple Town puzzle game by modeling state transitions, tile mechanics and reward strategies
- Implemented using Python and Q-learning, achieving consistent score improvements over a baseline policy through iterative training and evaluation

AI Financial Planning Agent

- Developed an AI financial planning agent that analyzes income, expenses and goals to generate budgeting insights, classify transactions and provide tailored recommendations
 - Implemented using FastAPI, Python, LLM-based reasoning, RAG retrieval and PostgreSQL, producing stable and grounded outputs during evaluation
-

Certifications

- Microsoft Certified: Azure Fundamentals (AZ-900)
 - Microsoft Certified: Security, Compliance and Identity Fundamentals (SC-900)
-

Achievements

Runner-up in the Shastra 2020 Hackathon for developing an optimized algorithm to solve a constraint-based problem under strict time limits