

SOUVIK MUKHERJEE

📍 Bangalore, India — ✉ mukherjeesouvik1305@gmail.com — 📞 +91 80177 17008

🐙 github.com/Msouviik — 🔗 linkedin.com/in/souvik-m1305

📁 Professional Summary

AI & Full-Stack Engineer with 2 years of experience at Procify Innovations Pvt. Ltd., specializing in intelligent agent workflows, Retrieval-Augmented Generation (RAG) pipelines, and open source LLM deployment. Strong parallel experience in full-stack development (MERN, SQL, PostgreSQL, Python), building scalable and secure applications. Interested in bridging AI systems with enterprise-grade software to deliver end-to-end intelligent solutions.

⚙️ Skills & Technologies

- **Languages:** Python, JavaScript, TypeScript, Java
- **Back-end:** Node.js, Express.js, Socket.IO, Servlets, JDBC, JSP
- **Front-end:** HTML, CSS, Tailwind CSS, React.js, Zustand
- **AI/LLMs & Frameworks:** LangChain, DSPy, Docling, llama.cpp, Ollama, vLLM
- **Databases:** MongoDB, MySQL, PostgreSQL, Qdrant, ChromaDB
- **Tools and Platforms:** GIT, Docker, Jenkins, JIRA, Firebase, Rollbar, Bitbucket, Postman
- **Additional Skills:** DSA, System Design, Problem Solving (LeetCode), Vector Databases, Prompt Engineering, Attention to Detail, Flexibility

🏢 Professional Experience

Software Developer, Procify Innovations Pvt Ltd—

Aug 2023 – Present

Tech Stack: Python, Node.js, JavaScript, TypeScript, React.js, MongoDB, PostgreSQL, MySQL, Qdrant, ChromaDB, LangChain, DSPy, Docling, llama.cpp, Ollama, vLLM, HTML, CSS, Tailwind, Firebase, Jenkins, GCP, Docker, KeyCloak

Project: Procify — A Low-Code/No-Code (LCNC) enterprise platform to build and run mobile apps with offline access, real-time sync and intelligent AI integrations.

- Built an offline data sync feature that allows app usage without internet and auto-sync data once reconnected.
- Implemented dynamic table and schema updates to create, alter, and manage evolving enterprise database structures between clients and servers.
- Designed and implemented intelligent agent workflows in Python using LangChain and DSPy to automate enterprise processes such as document parsing, semantic search, and knowledge assistance.
- Built and optimized Retrieval-Augmented Generation (RAG) pipelines with Docling + LangChain for enterprise knowledge retrieval, improving contextual accuracy and response quality.
- Integrated and managed vector databases (Qdrant, ChromaDB) to enable scalable semantic search and efficient information retrieval.
- Deployed and wrapped pre-trained open source LLMs (LLaMA, Qwen, DeepSeek) using llama.cpp, Ollama, and vLLM for on-device and cloud inference.
- Developed a version control system with Firebase to ensure seamless delivery of the latest app updates.
- Built secure user account management features including login, sign-up, password reset, and access removal, integrated with OAuth and KeyCloak for centralized identity management.
- Refactored and optimized the codebase to improve performance, scalability and maintainability by 60%, while resolving platform issues via JIRA for greater stability and reliability.

Trainee Developer, Maventic Innovative Solutions Pvt Ltd—

Feb 2023 – Jul 2023

Tech Stack: Python, LangChain, DSPy, Docling, Qdrant, ChromaDB, llama.cpp, Ollama, vLLM

Project: Developed UWL and Notification Editor applications according to client requirements using the Procify platform, enhanced with intelligent agent workflows, RAG pipelines, and LLM integrations. These apps support efficient task tracking, contextual insight, and flexible notifications across different channels.

- Built and delivered task and alert management apps tailored to client-specific needs by integrating intelligent agent workflows using Python and agentic frameworks.
- Designed intuitive user workflows to help users manage work items from a single central interface, while leveraging RAG (Retrieval-Augmented Generation) pipelines with Docling and LangChain for contextual insights.
- Allowed users to customize message templates and set alert preferences for notifications (email, SMS, or in-app), enhanced with vector databases (Qdrant, ChromaDB) and pretrained LLM wrappers (LLaMA, Qwen, DeepSeek) deployed via llama.cpp, Ollama, and vLLM.

📁 Projects

SocioFY – Realtime Chat App 🚀

Technologies: React, TailwindCSS, Node.js, Express.js, MongoDB, Socket.IO, JWT

Built a full-stack chat app with real-time messaging using WebSockets and JWT-based auth. Designed a responsive UI with React, TailwindCSS, and DaisyUI. Implemented a client-server architecture for real-time message handling and state sync.

CloudPort – File Sharing Website 🚀

Technologies: Node.js, Express.js, MongoDB, HTML, CSS, JavaScript

Built a file-sharing platform using Express and MongoDB with RESTful APIs, allowing users to upload files and share download links through email or personal sharing. Served a static front-end directly through the back-end.

Gibli – Smart Voice Assistant

Technologies: Python (Tkinter, pyttsx3, SpeechRecognition, Wikipedia API, etc.)

Developed a voice-activated desktop assistant with Python. Integrated third-party APIs for dynamic content and real-time responses. Used Tkinter for the GUI and Python libraries for natural language commands and task automation.

Key Achievements

- Helped build a platform that scaled to more than 50,000 users.
- Used Rollbar and Google Cloud Logging to track errors and improve issue detection, which helped resolve 80% of hard-to-trace production issues.

Education

B.P. Poddar Institute of Management and Technology

2019 – 2023

Bachelor of Technology (B.Tech.), Electrical Engineering

CGPA: 8.78