Basavaprasad Gola

Dallas, Texas, United States
prasadgola.github.io
(682) 266 - 3588

basavaprasadgolacs@gmail.com
linkedin.com/in/basavaprasad_gola
huggingface.co/basavaprasad_gola

Experience

Software Developer | Bluelinkitservices | Plano, TX | version | *Jun 2024 - Present*

- Developed a multi-agent Agent to Agent Events system using Google Gemini and MCP with specialized agents for user intent routing (host, sponsor, undecided), and integrated Vertex AI (Imagen) for on-demand AI image generation and Gemini Nano for offline chat capabilities on web and Android platforms.

- Built an advanced lead-generation pipeline integrating Apollo.io for lead discovery, Clay for data enrichment, and HubSpot API (OAuth 2.0) for automated CRM contact synchronization, complemented by engineered core services including real-time audio/video/screen-sharing (WebSockets), secure AP2-compliant Stripe payment processing for tiered/custom amounts, and a 1x1 tracking pixel email system.

- Architected a multi-platform solution featuring a 24-theme customizable web UI (HTML/CSS/JS) and a native Android app (Jetpack Compose, Kotlin), both powered by the same FastAPI backend for unified cross-platform functionality.

- Implemented a full CI/CD pipeline using Docker and Google Cloud Build/Run, along with an automated testing suite (ADK Agent Evaluator) to validate agent logic and tool-use accuracy across the entire system.

Software Developer | University of Texas at Arlington | Arlington, TX | *May 2023 - June 2024*

- Developed and deployed cross-platform healthcare applications for the university's medical department, including a React Native mobile app providing nurses with a real-time centralized dashboard, and a web-based version deployed on Microsoft Azure ensuring cross-platform accessibility and functionality.

- Built a Python desktop application that visualizes hospital data on a Texas map with color-coded counties based on hospital counts using Excel data sources, while managing the full software development lifecycle from requirements gathering and design through testing and Azure cloud deployment.

Education

University of Texas at Arlington   **Arlington**
Masters of Science in Computer Science   *May 2023*

Skills

- **Languages:** Python, Kotlin/Jetpack Compose, HTML/CSS/JS

- **Libraries:** MLX, ADK, WebSockets, Stripe, Selenium WebDriver, Undetected Chromedriver, React/Native, FastAPI, Pytorch, Tiktoken, Tensorflow, Numpy, Pandas, Gradio, RAG

- **Hyperscaler:** GCP(Cloud Build/Run), Azure, Docker

- **Model Infra:** Agentic(event driven/loop), Context/CoT/Prompt Engineering, RLHF, A2A/AP2, tool calling, Vector database

Projects

Browser Automation Agent using Local LLMs[Github]

- Developed an autonomous web agent leveraging local LLM/VLM (GPT-OSS, Qwen3-VL) to perceive screen states and execute complex tool calls, while applying Context and Prompt Engineering to architect a stateful agent using Chain-of-Thought reasoning that mimics reinforcement learning by dynamically regenerating its context window to iteratively refine strategies based on past execution history and failure feedback.

- Engineered computer vision capabilities for coordinate-based interaction by feeding viewport screenshots to VLM models, reinforced by a self-healing execution layer that detects action failures (due to model imprecision) and automatically triggers robust fallbacks like JavaScript injection for reliable browser automation.

Decoder-only Transformer Model with Instruction Tuning [HuggingFace]

- Implemented a chat-style transformer model from scratch using PyTorch, featuring multi-head self-attention mechanisms, positional embeddings, and layer normalization, achieving character-level language modeling with 10M parameters across 6 transformer blocks, while optimizing training infrastructure with custom batch generation, gradient accumulation, and learning rate scheduling (1e-4 for fine-tuning).

- Instruction fine-tuning pipeline with masked loss computation to train only on assistant responses, utilizing special tokens and implemented temperature-controlled sampling and early stopping mechanisms for improved generation quality.