Waleed Ahmed

AI Engineer & Prompt Developer

waleedahmed.ca
hello@waleedahmed.ca
linkedin.com/in/waleed-ahmed
github.com/TRAP33ZOID
Toronto, ON, Canada
437-450-9195

RELEVANT EXPERIENCE

Founder — *ZoidAI* — September 2025 — Present

- Architected and iterated on **system prompts** for multi-tenant AI voice agents serving dental clinics, testing across **GPT-4o, Claude, and Google Gemini** to optimize conversational flow, appointment booking accuracy, and natural language understanding
- Built production **RAG pipeline** using **pgvector embeddings** and custom retrieval logic to enable voice agents to accurately answer client-specific questions from uploaded documents, reducing hallucination rates through targeted prompt engineering
- Designed **dual knowledge base architecture** where documents sync to both VAPI for phone calls and vector database for text chat, requiring tailored prompts for each modality to handle context length and retrieval differences
- Implemented **prompt configuration system** allowing clinic admins to customize voice personality, response tone, and domain-specific terminology, with A/B testing framework to measure user satisfaction across **20+ voice options**
- Monitored and analyzed **hundreds of live call transcripts weekly** to identify edge cases, refine prompts, and improve handling of appointment conflicts, insurance questions, and multilingual requests

AI Consultant — *Zoid Consulting* — May 2024 — Present

- Consulted with **5+ small business clients** across healthcare and professional services to identify automation opportunities and design custom AI voice agents, chatbots, and workflow integrations
- Built and deployed production voice agents using **VAPI, Bland AI, and Twilio**, iterating on prompts based on client feedback to handle industry-specific terminology and business logic
- Created **client-specific RAG knowledge bases** by ingesting business documentation, designing retrieval strategies, and crafting prompts that balance accuracy with conversational naturalness
- Developed internal testing frameworks to evaluate prompt performance across different LLM providers, documenting behavior differences and cost-performance tradeoffs to recommend optimal models per use case

TECHNICAL PROJECTS

LLM Evaluation & Benchmarking System — *Python, OpenAI API, Anthropic API, Google AI Studio* —

- Built automated testing harness to compare **GPT-4, Claude Sonnet, and Gemini Pro** across **200+ prompt variations** for appointment scheduling and customer support use cases
- Designed evaluation metrics for accuracy, response latency, cost per query, and conversational quality, identifying that Claude outperformed GPT-4 by **15%** on complex reasoning tasks while Gemini offered **30% lower costs** for simple queries
- Created prompt template library with versioning system to track performance improvements over time and enable rapid A/B testing of new prompt structures

AI Workflow Automation Platform — *n8n, OpenAI, PostgreSQL, webhooks* —

- Designed and deployed **n8n workflows** integrating LLM-powered automation for lead qualification, email triage, and data extraction from unstructured documents
- Crafted specialized prompts for each workflow stage, handling edge cases like ambiguous requests and multi-step reasoning, reducing manual processing time by **40%**
- Implemented feedback loop system to capture failed automations and iteratively improve prompts based on real-world failure patterns

PROFESSIONAL SUMMARY

AI Engineer with hands-on experience building production LLM-powered systems and expertise in prompt engineering across GPT-4, Claude, and Gemini. Founded and scaled multi-tenant AI voice agent SaaS platform, designing and iterating on system prompts to optimize conversational AI for real-world business use cases. Proven track record in evaluating model performance, building RAG pipelines with vector databases, and consulting with clients to deliver AI automation solutions. Deep understanding of LLM behavior, limitations, and cost-performance tradeoffs gained through shipping production systems.

SKILLS

LLMs & Prompt Engineering

GPT-4, Claude (Sonnet/Opus), Google Gemini, Prompt Design, Retrieval-Augmented Generation (RAG), Few-Shot Learning, Chain-of-Thought

AI/ML Tools & Frameworks

VAPI.ai, LangChain, OpenAI API, Anthropic API, pgvector, Embeddings, Voice AI, n8n Automation

Programming Languages

Python, TypeScript, JavaScript, SQL

Databases & Vector Stores

PostgreSQL, Supabase, pgvector, Vector Databases, Query Optimization

APIs & Development

REST APIs, WebSockets, Webhooks, Next.js, React, Flask, Vercel, Git

Tools & Platforms

VAPI, Bland AI, Twilio, Jupyter, Tableau, AWS S3, CI/CD

EDUCATION

Bachelor of Computer Engineering — *Toronto Metropolitan University* — May 2026

Diploma of Data Science — *Lighthouse Labs* — Oct 2023