Xavier Caracter

San Francisco, CA • xaviercaracter@gmail.com • (650) 518-3813

AI ENGINEER | LLM APPLICATIONS & PROMPT OPTIMIZATION

Computer Science M.S. student with hands-on experience building production-ready LLM systems and systematically optimizing prompts. 4+ years of professional software engineering experience in API integration, systematic testing, and cross-functional collaboration. Proven track record designing, deploying, and iteratively improving AI applications through rigorous experimentation and evaluation.

TECHNICAL SKILLS

AI/ML: Prompt Engineering, Chain-of-Thought Prompting, Few-Shot Learning, Role-Based Prompting, Systematic Prompt Optimization, RAG Architecture, LLM APIs (OpenAI GPT-4, GPT-4o mini, GPT-3.5), Vector Databases, Prompt Testing, A/B Testing, AI Safety, Guardrails, LLM Evaluation Frameworks

Development: Python, OpenAI API, LangChain, LangGraph, Flask/FastAPI, REST APIs, React ,Angular, JavaScript, Docker, Node.js, SQL, Pandas, NumPy, Git, AWS, Microservices, CI/CD, Java, C, Go

Methodologies: Systematic Experimentation, A/B Testing, LLM Evaluation, Prompt Versioning, Agile/Scrum

AI/ML PROJECTS & RESEARCH

Production LLM System for Customer Support | Independent Research Project | Dec 2024 - Present

- Developed end-to-end conversational AI support agent using OpenAI GPT-4o mini, achieving 85% intent recognition accuracy across 200+ test queries
- Designed and evaluated 50+ prompt variations, including temperature adjustments, role definitions, instruction clarity, and chain-of-thought techniques, enhancing response consistency and overall quality
- Implemented prompt chaining for multi-step processes (verification, diagnosis, resolution), boosting task completion accuracy by 35%
- Built reusable prompt template library with 15+ patterns for key scenarios such as order tracking, returns, and billing inquiries
- Architected microservices infrastructure with REST API integration to inject real-time customer data into responses
- Deployed RAG pipeline grounded in knowledge base, reducing hallucination rate from 30% to under 5%
- Created LLM-as-judge evaluation system to assess helpfulness, accuracy, and tone; conducted A/B tests on 300+ conversations, quantifying improvements in customer satisfaction.
- Conducted adversarial testing and implemented guardrails to mitigate prompt injection attacks
- Analyzed 500+ conversation logs to identify common customer intents, optimize prompt templates, and improve agent performance

University Scheduling Application | Boise State Capstone | Jan 2024 - May 2024
- Led 4-person team in developing Flask-based RESTful API and JavaScript frontend for tutoring room and resource booking system
- Designed microservices architecture with workflow orchestration
- Delivered bi-weekly Agile demos to stakeholders; translated non-technical requirements into technical solutions

PROFESSIONAL EXPERIENCE

Micron Technology — Boise, ID

Revenue Operations Intern | June 2023 - Aug 2023
- Automated sales data pipeline and validation scripts (50k+ records), reducing reporting errors by 36% and eliminated 12 hours/week of manual effort
- Applied systematic experimentation methodology to test 12+ data validation approaches, measuring precision/recall metrics to optimize anomaly detection accuracy
- Created adopted documentation and presented solutions to cross-functional sales operations team
- Worked cross-functionally with sales operations team to understand ambiguous requirements and deliver impactful solutions

Enterprise Transformation Institute — San Francisco, CA

Junior Software Engineer | June 2020 - June 2023
- Developed and maintained 25+ Selenium automation scripts across 8 client projects
- Integrated test automation into CI/CD pipelines, reducing regression testing time by 36%
- Designed systematic testing frameworks with defined success metrics (methodology later applied to LLM prompt optimization)
- Debugged production issues using JavaScript, Python, and SQL in ambiguous, fast-paced environments

Buff Bros — Rancho Cordova, CA

Web Development Contractor | Nov 2024 - February 2025
- Built responsive business website with API-integrated contact form generating 40+ customer inquiries in first month
- Worked directly with client to understand business needs and translate requirements into technical solutions

EDUCATION

Johns Hopkins University — Whiting School of Engineering

Master of Science in Computer Science (In Progress, Remote) | Expected 2026

Relevant Coursework: Machine Learning, Natural Language Processing, Agentic AI

Boise State University — Boise, ID

Bachelor of Science in Computer Science, Minor in Mathematics | May 2024