AMBUJ KUMAR TRIPATHI

AI Prompt Engineer | LLM Integration Specialist | Conversational AI Developer

✓ AVAILABLE IMMEDIATELY • Remote/Hybrid • Open to Relocation

Email: kumarambuj8@gmail.com | Mobile: +91 9431 801363
Location: Gorakhpur, Uttar Pradesh, India
Portfolio: https://ambuj-portfolio-v2.netlify.app/ | GitHub: https://github.com/Ambuj123-lab

PROFESSIONAL SUMMARY

AI Prompt Engineer with production experience optimizing commercial-grade AI models at **Hogarth Worldwide (WPP)**. Specialized in prompt engineering, LLM integration, conversational AI development, adversarial testing, and AI workflow automation. Successfully transitioned from telecom infrastructure (British Telecom, Tata Communications) to AI through intensive self-learning and hands-on project execution.

- **Reduced prompt iteration time by 40%** through systematic optimization methodology at Hogarth
- Built **10+ production-ready AI applications** including intelligent chatbots using Microsoft Bot Framework and custom LLM integrations with **10,000+ user interactions**
- Developed **enterprise-grade conversational AI solutions** with multi-modal capabilities, content moderation, and state management
- Earned **18+ verified certifications** from Google Cloud, IBM, AWS, Microsoft Azure, NVIDIA with practical implementations
- **5+ years technical experience** in automation, data analysis, network optimization, and cross-functional collaboration
- **NVIDIA** certified in RAG Agents and Edge AI. Specialized in building cost-efficient LLM workflows and automated testing tools

> SEEKING: AI Engineer | Prompt Engineer | LLM Integration Engineer | Conversational AI Developer | AI/ML Roles | Immediate Joining Available

Core Strengths: Prompt Engineering • LLM API Integration • Conversational AI Development • Microsoft Bot Framework • Model Validation • Red Teaming • Python/JavaScript/Node.js • Cloud Platforms • Production AI Workflows • Data Analysis • Automation • Quality Assurance

CORE COMPETENCIES & TECHNICAL SKILLS

AI & Prompt Engineering

Prompt Engineering • Large Language Models (GPT-4, Claude, Gemini, Meta Llama 3.70B, IBM WatsonX) • LLM API Integration • Conversational AI Development • Microsoft Bot Framework SDK • Chatbot Development • Multi-Modal AI • Adversarial Testing • Red Teaming • Model Validation • Model Robustness Testing • Few-Shot Learning • Zero-Shot Prompting • Token Optimization • Content Evaluation • Content Moderation Systems • Data Quality Assessment • Fact-Checking and Verification • Responsible AI • AI Ethics

Programming Languages & Frameworks

Python • JavaScript (ES6+) • Node.js • Express.js • HTML5 • CSS3 • React 18 • Flask • Gradio • Streamlit • Tailwind CSS • Babel • RESTful API Development • Async/Await • Promise Handling

AI Tools & Platforms

Microsoft Bot Framework SDK (botbuilder) • Adaptive Cards • Hugging Face Transformers • OpenRouter API • Gemini API • Google Generative AI SDK • Gemini Vision API • Vertex AI • IBM WatsonX • Jupyter Notebooks • VS Code • NotebookLM • REST APIs • Puppeteer • Git • Conversation State Management • Multi-Turn Dialog Systems

Cloud Platforms

Google Cloud Platform (Vertex AI, Cloud Functions, Cloud Run, GKE) • IBM (WatsonX Services, Cloud Foundry, Speech to Text, Text to Speech) • Microsoft Azure AI • Netlify • Render.com • Hugging Face Spaces

Specialized Skills

Progressive Web Apps (PWA) • Service Workers • Offline Functionality • Data Processing and Visualization (Papa Parse, SheetJS, Plotly.js) • Web Speech API • Web Audio API • CORS • Rate Limiting • Environment Secrets Management • Security Best Practices • Input Validation • API Rate Limiting • Session Management • PDF.js • html2canvas • jsPDF • Content Moderation Systems • Pattern-Based Guardrails

Additional Technical Skills

Data Analysis • Automation • Network Optimization • GIS Tools • Fiber Optic Network Design • Excel Advanced Functions • Business Process Documentation • Performance Monitoring • Quality Assurance • Cross-functional Collaboration • Technical Documentation

PROFESSIONAL EXPERIENCE

AI Prompt Engineer

Hogarth Worldwide (WPP Marketing Communications India Pvt. Ltd. via TeamLease)

September 2025 – October 2025 | Remote, India | 6-Month Contract (Concluded due to mandatory office relocation requirement - health reasons)

- Engineered and optimized prompts for **commercial-grade generative AI models** (Flux, SDXL) within strict technical constraints (72-77 token limits across Base, Positive, Negative prompts) for brand-compliant visual outputs
- Conducted **systematic model validation and adversarial testing**, identifying critical limitations including instruction adherence failures, hallucination patterns, layout inconsistencies, and edge-case behaviors across multiple parameters
- Built **reusable prompt libraries** and comprehensive documentation for **25+ use case categories**, enabling team scalability and reducing prompt engineering time by **40%**
- Developed **Smart AI Prompt Builder tool** to standardize prompt creation workflows across team, featuring intelligent keyword matching algorithm detecting 8 problem categories and recommending optimal AI platforms based on task requirements
- Collaborated with cross-functional teams to define AI integration protocols and quality assurance standards for production pipelines
- Developed prompt refinement methodology translating subjective business feedback into quantifiable technical adjustments

> **KEY ACHIEVEMENT: Created systematic prompt optimization framework reducing iteration cycles and improving output quality consistency across diverse creative requirements**

Technologies: Flux, SDXL, Prompt Engineering, Adversarial Testing, Model Validation, QA Protocols, HTML5, CSS3, JavaScript

Associate Engineer - Fibre & Network Delivery

British Telecom Global Services Pvt. Ltd. | January 2022 – August 2024 | Gurugram, India

- Led **FTTP network planning** using GIS tools, reducing deployment time by **15%** through systematic optimization
- Developed **AI-powered prototype tools during May-August 2024** to explore automation opportunities in network estimation workflows using **Gemini 1.5 Flash API**
- Built experimental **geospatial analysis tool** leveraging Gemini's multimodal capabilities for map image interpretation
- Collaborated with **MIS/Power BI teams** to prototype AI-driven solutions including **Resume Intelligence Analyzer** and automation pipelines, receiving recognition for innovative approaches
- Coordinated cross-functional teams achieving **98% quality assurance compliance** with streamlined processes
- Awarded **"Top Performer" (January 2022)** for innovative problem-solving and process improvement initiatives
- Developed **rule-based automation systems** reducing manual effort by **70%** and improving accuracy

> **INNOVATION ACHIEVEMENT: Successfully prototyped AI tools for internal business use cases, demonstrating practical application of emerging technologies in telecom operations**

Skills Applied: Data Analysis, Automation, GIS Tools, Network Optimization, Cross-functional Collaboration, Gemini 1.5 Flash API, Gradio

Operations & Maintenance Engineer

Lobo Staffing (Client: Tata Communications Transformation Services Ltd.)

November 2021 – January 2022 | Lucknow, India

- Optimized network performance and monitoring systems, reducing downtime by **10%** through comprehensive data analysis
- Managed billing systems for OSP/ISP services with **100% accuracy** using systematic quality control processes
- Improved Field Engineer SLA compliance, achieving **98% TAT** through workflow optimization

Tools: ServiceNow, Orion, Network Monitoring Systems, Excel Advanced Functions

Optical Fiber Execution Engineer

Annu Infra Construct India Pvt. Ltd. | December 2017 – June 2018 | Kolkata, India

- Executed fiber optic deployment projects for **Ministry of Defense** across 3 states ensuring complete security compliance
- Ensured **100% compliance** with security protocols and technical specifications through rigorous quality assurance
- Reduced project delays by **20%** through effective vendor coordination and strategic resource optimization

System Administrator

Teamware Solutions (Client: Tata Consultancy Services)

October 2013 – November 2014 | Noida, India

- Managed revenue systems for **Delhi Jal Board**, ensuring **99% uptime** through proactive monitoring
- Trained **20+ portal users** on system optimization and best practices
- Supported eMigrate portal (Ministry of Overseas Affairs) with document verification processes

KEY AI PROJECTS (ALL LIVE & ACCESSIBLE)

Microsoft Bot Framework Resume Chatbot with Gemini AI and Vision | December 2025

- Developed intelligent resume chatbot using **Node.js, Express.js, and Microsoft Bot Framework SDK** integrated with **Google Gemini 1.5 Flash API**
- Implemented **multi-modal AI capabilities** supporting text conversations and image analysis via Gemini Vision API
- Engineered **custom content moderation system** with pattern-based guardrails for offensive language (Hindi/English) and off-topic filtering
- Built **responsive glassmorphism UI** with Adaptive Cards, CSS animations, and real-time typing indicators
- Designed **RESTful API architecture** with conversation state management, supporting multi-turn dialogues with context retention (20 messages)
- Configured **model parameters** (temperature: 0.7, max tokens: 1000) for optimized response quality and token efficiency
- Implemented **security best practices** including environment variable management, input validation, and API rate limiting
- Deployed on cloud platform (Render.com) with **health monitoring** and automated session cleanup

Technologies: Node.js, Express.js, Microsoft Bot Framework SDK (botbuilder), Google Generative AI SDK, Gemini 1.5 Flash API, Gemini Vision API, RESTful API, HTML5, CSS3, JavaScript (ES6+), Adaptive Cards, dotenv, Render.com

RAG-Based Document Q&A System with Vector Search (NVIDIA Certification) | Nov 2025

- Developed chatbot using **LangChain** with custom embeddings; Implemented **RAG pipeline** with FAISS for semantic retrieval
- Integrated **FastAPI backend** for scalable document processing; Built contextual knowledge pipelines

Technologies: Python, LangChain, FAISS, FastAPI, Gradio, Hugging Face Transformers, RAG

Geo AI - Geospatial Narrative & Image Analysis Tool | November 2025

- Developed AI-powered geospatial analysis tool using **Gemini 1.5 Flash** to generate location-based narratives from map images
- Implemented interactive map rendering with **Leaflet.js**, allowing users to capture specific regions or upload custom images for AI analysis
- Integrated secure image handling and API token management for seamless AI analysis requests
- Deployed on **Render** with automated CI/CD pipelines from GitHub
- **Live Demo:** https://geo-narrator-ai.onrender.com/

Technologies: Node.js, Express, Gemini API, Leaflet.js, HTML5, CSS3, Render

LLM-Integrated Resume Chatbot | Meta Llama 3.3.70B | August 2025

- Developed scalable chatbot using Llama 3.3.70B, reducing query resolution time by **25%** for **100+ concurrent users**
- Designed **stateless architecture** with client-side state management, eliminating server-side session storage and database dependencies
- Implemented front-end to send entire chat history as JSON in each request, enabling server to process full context without database lookups
- Integrated secure OpenRouter API with rate limiting and CORS configuration
- **Live Demo:** https://ambuj-resume-bot.onrender.com

Technologies: Meta Llama 3.3.70B, OpenRouter API, Python, Flask, JavaScript, HTML5, CSS3, CORS, Render.com

Professional Task Management System | Privacy-First PWA | September 2025

- Architected **privacy-first design** using 100% client-side localStorage with JSON export/import, supporting **100+ tasks**
- Implemented installable Progressive Web App with service workers achieving **99% offline functionality**
- Integrated dual audio system using Web Speech API and Web Audio API for task feedback
- Developed real-time session time tracking with activity detection, displaying formatted time with **98% accuracy**
- Built advanced multi-field search improving task retrieval speed by **30%**

Technologies: React 18 (Hooks), Tailwind CSS, PWA, Service Workers, Web Audio API, Web Speech API, LocalStorage, Babel

Smart AI Prompt Builder - Intelligent Prompt Generation Tool | August 2025

- Developed intelligent web application helping users create effective AI prompts and recommending optimal AI platforms
- Built advanced keyword matching algorithm automatically detecting **8 problem categories**
- Implemented smart AI tool recommendation engine suggesting best platforms based on user requirements
- Created bilingual interface (Hindi/English) with 6 customizable tone styles
- Achieved **95% accuracy** in category detection through intelligent pattern recognition

  Technologies: HTML5, CSS3, JavaScript ES6+, Client-Side Processing, Netlify

Resume-as-Code - Automated Resume Builder Pipeline | July 2025

- Built automated pipeline generating ATS-friendly PDF resumes from single HTML/CSS source file
- Implemented Node.js backend with Puppeteer for programmatic HTML rendering and PDF export
- Achieved **100% formatting consistency** reducing update time by **90%**

  Technologies: Node.js, Puppeteer, HTML5, CSS3, Headless Browser Automation

Additional AI Projects

- **Resume Chatbot (IBM WatsonX):** Created conversational AI chatbot with **95% intent recognition** for 50+ daily queries. Deployed on Netlify with 99.9% uptime
- **Personal AI Assistant (Python, Gradio):** Built Python-based assistant handling **200+ daily queries**. Deployed on Hugging Face Spaces

Complete Portfolio: Scan QR code above or visit https://ambuj123-lab.github.io/Ambuj_AI_Portfolio

INDUSTRY CERTIFICATIONS & CREDENTIALS (ALL VERIFIED)

AI/ML & Generative AI Certifications

- **NVIDIA:** Building RAG Agents with LLMs (LangChain, FAISS, FastAPI), AI on Jetson Nano (Edge AI, CNNs, ResNet-18)
- **Google Cloud:** 6 Skill Badges — Vertex AI Prompt Design, Gemini API, TensorFlow Image Classification, GenAI Apps with Streamlit, Cloud Run & GKE Deployment
- **IBM AI:** AI Fundamentals, Generative AI in Action, Deep Learning Essentials, Chatbot Development, Python for Data Science
- **Azure AI:** Language Models on Databricks, Responsible AI Principles
- **Anthropic Academy:** Model Context Protocol (MCP), Claude with Google Vertex AI
- **Linux Foundation:** LFS118 - Ethical Principles for Conversational AI

Enterprise Job Simulations (Forage Platform)

- **BCG X:** AI-Powered Financial Chatbot Development (NLP, Data Extraction, Financial Analysis)
- **Big 4 & Enterprise:** AWS Solutions Architecture, PwC Digital Assurance, Deloitte Australia Analytics, Tata iQ GenAI, Siemens

All certifications verifiable through official platforms. Verification links available in online portfolio.

ADDITIONAL QUALIFICATIONS & ACHIEVEMENTS

- **OpenRouter Developer Badge (2025):** Recognized for excellence in high-volume Multi-LLM API integration (Claude, GPT-4o, Llama 3) and production-grade model deployment.
- **AI Monitoring & Governance (GA4):** Successfully implemented real-time behavior tracking and **Anomaly Detection** for AI interactions using Google Analytics 4, ensuring model auditability.

EDUCATION

Post Graduate Diploma in Power Transmission & Distribution

National Power Training Institute, Ministry of Power | 2015 – 2016 | New Delhi, India

Bachelor of Technology - Electrical & Electronics Engineering

Uttar Pradesh Technical University | 2009 – 2013 | Lucknow, India