

Jorge Mora

Bilingual Technical Writer | AI Infrastructure & Developer Documentation

📍 Tahoe City, CA | ✉ contact@jmmultiservices.xyz | 🔗 [LinkedIn](#) | 📁 [Portfolio](#)

Professional Summary

Technical writer specializing in AI infrastructure, LLM serving, and developer-focused documentation. Proven ability to translate complex technical systems into clear, actionable guides for engineers and product teams. Skilled in communicating with both technical stakeholders and non-technical audiences. Fast learner with hands-on experience in Python, APIs, and SaaS systems. Bilingual (English/Spanish) and passionate about reducing technical friction through clear documentation.

Core Competencies

Technical Writing

- API documentation and developer guides
- System architecture explainers and technical architecture documents
- Integration guides and implementation patterns
- SOPs, onboarding documentation, and user guides
- Whitepapers and technical analysis

Technical Communication

- Translating complex infrastructure concepts for mixed audiences (engineers, PMs, stakeholders)
- Cross-functional collaboration with engineering, product, and customer success teams
- KPI-focused documentation (throughput, latency, cost optimization)
- Creating documentation that drives adoption and reduces support overhead

Technical Skills

- Python (hands-on experience in scripts, automation, and system integration)
- APIs (REST, design documentation, integration workflows)
- Infrastructure concepts: GPU optimization, batching, inference serving, distributed systems

- LLM infrastructure: model deployment, serving patterns, cost modeling
- Tools: Markdown, Notion, Figma, Canva, GitHub, JSON configuration

Bilingual Documentation

- English (Fluent)
 - Spanish (Native/Bilingual)
 - Experience building global documentation strategies
-

Professional Experience

Technical Writer | AI Systems & Infrastructure Documentation

Remote, Freelance | 2024–Present

Infrastructure & Developer Documentation

- Authored technical explainers on LLM inference infrastructure, covering advanced topics like iteration batching, multi-LoRA serving, and GPU optimization
- Created documentation that explains complex trade-offs between cost, throughput, and latency to technical audiences
- Developed architecture guides helping teams understand integration patterns and deployment workflows
- Demonstrated ability to make infrastructure concepts accessible without sacrificing technical depth

API & Integration Documentation

- Produced API documentation and integration guides for SaaS and automation platforms
- Documented n8n workflows, Supabase backend systems, and Claude API integrations
- Created developer-focused guides enabling rapid onboarding and reducing implementation time

Cross-Functional Collaboration

- Worked with engineers to understand system architecture and identify documentation gaps
- Collaborated with product teams to align documentation with customer KPIs and success metrics
- Built documentation strategies that reduced support tickets and improved developer adoption

Bilingual Documentation

- Authored English/Spanish documentation for global product launches

- Managed multilingual documentation workflows enabling seamless team coordination

Highlighted Projects

LLM Inference Infrastructure Explainers (Portfolio-Backed)

- Authored three comprehensive technical explainers on GPU optimization and LLM serving
 - *Iteration Batching for LLM Inference*: Explains how continuous batching improves GPU utilization and throughput
 - *Multi-LoRA Serving*: Documents cost-efficient multi-variant serving patterns
 - *Hugging Face + FriendliAI Integration*: Guides deployment from model discovery to production
- Each explainer balances technical depth with accessibility, following a problem-first approach
- Demonstrates understanding of infrastructure KPIs: tokens-per-second, tokens-per-dollar, GPU utilization
- Available in [GitHub portfolio](#) showcasing professional documentation quality

Quantum-Enhanced Excavator Tool

- Produced patent specifications, layered system diagrams, and operation documentation for AI-powered robotics
- Translated complex technical requirements into clear engineering handoff documentation
- Accelerated onboarding process and reduced implementation questions

Lead Gen SaaS Automation Platform

- Documented backend workflows integrating n8n, Supabase, and Claude AI
- Built multilingual onboarding guides for developers and end users
- Improved developer efficiency and customer adoption through clear integration documentation

Technical Knowledge Areas

- **LLM Inference**: Batching strategies, serving patterns, optimization techniques, cost modeling
- **GPU Infrastructure**: Utilization optimization, throughput metrics, resource allocation, performance tuning
- **Python**: Script writing, automation, API interaction, system integration
- **APIs**: Design, documentation, developer experience, integration patterns
- **SaaS Systems**: n8n (automation), Supabase (backend), Claude (AI), Manus.ai (robotics)
- **Developer Experience**: Reducing friction, improving adoption, writing for mixed technical audiences

- **Metrics & KPIs:** Understanding business drivers behind technical decisions
-

Tools & Technologies

Writing & Documentation: Markdown, Notion, GitHub, Google Docs

Design & Diagrams: Figma, Canva, Mermaid, system architecture tools

Development & Technical: Python, JSON, REST APIs, Git/GitHub

Automation & Integration: n8n, Supabase, Claude API, automation workflows


Education & Certifications

- **Google Data Analytics Certificate** — Coursera, 2024
 - **Technical Writing Specialization** — Independent study and practical projects
 - **LLM Infrastructure & Inference** — Self-directed learning through hands-on documentation
-


Why I'm a Great Fit for FriendliAI

- **Infrastructure Understanding:** Hands-on knowledge of GPU optimization, batching, and serving patterns
 - **Developer Communication:** Proven ability to translate complex infrastructure into actionable guides
 - **Quality Focus:** Portfolio demonstrates consistent, high-quality technical documentation
 - **Problem-First Approach:** Documentation starts with why something matters before diving into how
 - **Cross-Functional Skills:** Comfortable working with engineers, PMs, and stakeholders
 - **Bilingual Advantage:** Can support global product launches and international documentation
-

Contact & Links

 **Email:** contact@jmmultiservices.xyz

 **LinkedIn:** [linkedin.com/in/jorgemoraf87](https://www.linkedin.com/in/jorgemoraf87)

 **GitHub Portfolio:** github.com/JMORAF87/FriendliAi-Portfolio

Available immediately for full-time remote positions. Ready to bring clarity to technical complexity.