

Aleksandar Rakić

Senior AI Developer / AI Engineer

LLM Applications, RAG, AI Agents, Voice AI | PHP/Laravel, Python, Node.js, REST APIs, CI/CD



Location Belgrade, Serbia (plan to relocate to Dubai, UAE) **Email** info@alkom.rs

Phone +381 69 50 676 50 **Live product** alkom.rs/ai-asistent

LinkedIn linkedin.com/in/aleksandar-rakić

PROFILE

Senior AI developer and founder of Alkom production (web development and digital marketing), with 20 years of marketing and business ownership. Designs, builds and operates LLM products end to end: a multi-tenant RAG chat assistant live on 13 websites, a bilingual voice AI agent, tool-using AI agents, three production Laravel applications backed by 3,100+ automated tests. Stack: PHP/Laravel, Python, Node.js, REST APIs, MySQL, GitHub Actions CI/CD, Docker Compose (voice stack); AI-assisted development (Claude Code) under written plans, code review and test gates.

13 websites running the RAG chat assistant	~2.3 s LiveKit voice agent response latency	~320 voice QA conversations, zero critical errors	3,100+ automated tests, 3 production apps	34 days clinic CRM, plan to production	5 weeks CRM/dialer live, multi-user from day one
--	---	---	---	--	--

CORE COMPETENCIES

LLM Engineering	Agents & Voice	Software Engineering	Delivery & Ownership
Generative AI, LLMs	AI agents, agentic workflows	PHP/Laravel, Python	Requirements gathering
RAG, hybrid search	Function calling (tool use)	REST API design, integration	Stakeholder management
Prompt engineering	Conversational AI, chatbots	SQL: MySQL/MariaDB	Written plans, code review
Scenario-based LLM evaluation	Voice AI: LiveKit	JavaScript, Node.js	Automated test gates
Guardrails, PII masking	Speech recognition (ASR/STT)	Multi-tenant SaaS architecture	AI-assisted development (Claude Code)
LLM gateway, cost control	Text-to-speech (TTS)	GitHub Actions CI/CD	Security: 2FA, encryption

EXPERIENCE

Founder & AI Engineer · **Alkom production** · Belgrade, Serbia May 2005 - present

Own company, founded 2005: web development and digital marketing for business clients, today building AI software. Sole engineer on every product below.

- RAG chat assistant:** Built and operate a multi-tenant conversational AI product (Laravel, JavaScript widget, WordPress module), live on 13 websites in Serbia and the USA and covered by 243 feature tests.
- Retrieval and cost:** Kept answers accurate and LLM spend predictable with hybrid keyword + vector search and an LLM gateway with provider fallback, per-client budgets and cost tracking.
- Guardrails:** Engineered layered guardrails (hallucination control, PII masking, prompt-injection defense) and graceful degradation when a provider is unavailable; self-learning with human approval.
- Voice AI agent:** Built a bilingual (Serbian/English) voice agent in Python on LiveKit (~2.3 s latency) and shipped a web-call version live on alkom.rs; ~320 QA conversations, zero critical errors from round 8.
- AI agents:** Built a Claude tool-using assistant (function calling over 14 whitelisted read-only tools) inside the ERP, with an Electron desktop companion for offline speech (whisper.cpp, Piper); in daily use.
- Business platform:** Architected a Laravel ERP/CRM (invoicing, finance, Gmail and Calendar integration) that runs the company and its AI tenants: ~1,700 automated tests, CI cut from ~15 min to 1.1 min.

Marketing Manager & Senior AI Developer · Drugi Korak · Belgrade, Serbia

Dec 2024 - present

Private rehabilitation center; marketing, then clinic software, AI assistant and digital signage. Client of Alkom production.

- Delivered a clinic management system (scheduling, patient records, clinical reports, e-consents, payments) in Laravel, from a written master plan to production in 34 days; used daily by the clinic's 29-person team.
- Scoped the system with the clinic's management, then built a scheduling engine that prevents double bookings under concurrent use, with audited overrides and a touch drag-and-drop calendar.
- Protected health data with field-level encryption, an audit trail and mandatory two-factor authentication; migrated the clinic's existing schedule into the new system in a single import run.

Senior AI Developer & Web Designer · Tobo Capital Group · Chicago, USA (remote)

Jul 2026 - present

US commercial-finance broker; website and CRM/dialer delivered remotely. Client of Alkom production.

- Built and launched a multi-user CRM and dialer for the broker's sales team in 5 weeks (Laravel, MySQL), live from day one; 976 automated tests, more test code (~26k lines) than application code (~23k lines).
- Cut query time on the CRM's largest table from 16.5 ms to 3.6 ms and consolidated multi-source spreadsheet data into one deduplicated dataset with a Python ETL.
- Integrated cloud telephony (call sync, in-app playback), a shared mailbox, TCPA-compliant SMS and an HMAC-signed lead-intake webhook, with mandatory 2FA for every user; shipped the company website on Git deploys.

Marketing Manager · Dolcefreddo · Belgrade, Serbia

Jan 2020 - present

B2B distributor of gelato and pastry ingredients and equipment. Client of Alkom production.

- Built a GitHub Actions CI/CD pipeline, WooCommerce REST API product sync and an AI chat assistant for the distributor's 900+ product store, alongside its digital marketing and monthly cross-channel reporting.

Marketing Specialist · Focus Factor Plus · Belgrade, Serbia

Jul 2025 - present

Licensed factoring company. Client of Alkom production.

- Built a GitHub Actions CI/CD pipeline for the company's website and integrated the AI chat assistant, alongside marketing; produced a phased Google Ads and website roadmap from a multi-agent AI analysis.

General Manager · Casino Favorit · Belgrade, Serbia

Dec 2010 - Mar 2022

- Managed 29 business units and their staff (operations, sales, marketing); built the company's electronic reporting system.

Marketing Manager · AV Market · Belgrade, Serbia

Jun 2017 - May 2019

- Grew monthly sales 32% for an audio/video retailer and distributor through pricing and customer-retention programs.

SKILLS

AI / LLM	Claude, OpenAI, Gemini APIs; LiteLLM gateway · RAG: hybrid keyword + vector search, embeddings, response caching · Prompt engineering, AI agents, function calling, agentic workflows · LLM evaluation, guardrails, PII masking, LLMops cost tracking · Voice AI: LiveKit, STT/TTS, whisper.cpp, Piper
Backend	PHP 8.2 / Laravel 12 (primary stack) · Python 3.12 (LiveKit agents, ETL) · JavaScript (ES6+), Node.js, Electron, HTML/CSS · REST APIs, webhooks, JWT, OAuth 2.0 · SQL: MySQL 8 / MariaDB, query-plan design; PHPUnit
WordPress / Woo	Custom plugins, block themes (PSR-4, WPCS) · WooCommerce REST API, product sync · Own plugin platform and update server · SEO and GEO: structured data, AI-crawler readiness
DevOps / Hosting	Docker Compose (LiveKit, Redis), GitHub Actions CI/CD · Git PR flow, verified Git-only deploys · Health endpoints, alerting, backups with restore tests · 2FA, encrypted secrets, audit trail, dependency-audit gate
Data & APIs	Google Cloud: Speech-to-Text, Text-to-Speech, Pub/Sub · Google APIs: Gmail, Calendar, GA4, Ads · Meta Graph API, cloud telephony, TCPA-compliant SMS · IMAP/SMTP, WooCommerce and WordPress REST, Magento GraphQL
Marketing tech	Google Ads: builds, conversion tracking, diagnostics · GTM / GA4 / Consent Mode v2, SEO and GEO · Meta Business, email marketing, automated reporting

EDUCATION & LANGUAGES

Education: MSc Ecology - Faculty of Science and Mathematics, University of Niš, 2014

Languages: Serbian (native) · English (professional working proficiency; builds English products and documentation for US clients)