GLEB ROSEV

FOUNDING ENGINEER, AI AGENTS

YAGNI (yagni.app) | Making AI agent teams dependable in production, not just impressive in a demo.

rossevgleb@gmail.com | +995 591 90 81 80 | linkedin.com/in/rossev | github.com/Grumbash | Tbilisi, Georgia

01 // EXECUTIVE SUMMARY & OPERATING MODEL

I build software systems where reliability matters, and I make AI agent teams dependable in production, not just impressive in a demo. Now: Founding Engineer at YAGNI (yagni.app), working with the founders on AI agent teams (PM, Builder, Reviewer) that take work from a proposal to a merged GitHub PR under human-approved plans. Before that: Staff Harness Engineer on healthcare voice AI at Sully.ai. 8+ years across full-stack, mobile, cloud and engineering management, plus a medical degree. How I work: I build with AI agents every day -- I plan, delegate to agents, review their output with a second "judge" agent, and keep the human decision where it matters.

02 // EXPERIENCE & PRODUCTION SYSTEMS IMPACT

Founding Engineer, AI Agents @ YAGNI (yagni.app)

Nov 2026 -- Present | Tbilisi, Georgia | Remote

Early-stage US startup building named AI agent teams that take engineering work from proposal to merged GitHub pull request under human-approved plans. [Founding Team / Agent Benchmarks / Multi-Tenant Isolation / Self-Healing Syncs]

> Tenant isolation: tenant-scoped query guards and sandboxed code execution.
> Connector backfills made self-healing: stuck-cursor recovery, progress tracking and hard caps for long syncs of large workspaces.
> Code-review agent reliability: fixed ticket-key collisions between YAGNI and external trackers that sent the reviewer to the wrong issue.
> Agent benchmarks: realistic engineering task suites that show whether a prompt, model or harness change makes agents better or worse.
> Risk radar: flags stalled customer runs before the customer reports them.
> Self-improving loop (experimental): a meta-agent reads failed runs, isolates the cause and drafts a fix for the team to review.
> Monorepo build chain reduced to one command for fresh checkouts, worktrees and CI.
> Technical point of contact for pilot customers: demos, debugging live runs, turning feedback into engineering work.

Staff Harness Engineer, AI Voice Systems @ Sully.ai

Sep 2025 -- Oct 2026 | Dubai, UAE

AI receptionist platform for healthcare: voice and text agents for scheduling, care coordination and clinic operations. Sully.ai serves 400+ healthcare organizations. [15s -> 0s Dead Air at Pickup / MENA Launch / Real-Time Voice AI]

> Led core engineering for the MENA launch: UAE infrastructure, Azure Blob Storage support, regional configuration and production readiness.
> Removed ~15 s of dead air at pickup on outbound calls: the agent was greeting the caller while the line was still ringing.
> Made agent dispatch reliable on LiveKit: retry-and-verify, no more duplicate agents in one call, no more silent calls.
> Designed UAE audio-residency guardrails so TTS and STT fallbacks stay within approved regional providers instead of silently routing elsewhere.
> Built patient-data validation for outbound campaigns: local phone formats, customer-specific schemas, clinic rules.
> Built high-volume stress testing with headless LiveKit participants, Loki and Mimir validation, and runtime health checks; found and fixed the memory cascade behind burst failures.

Engineering Manager / Senior Full-Stack @ Kingpin

Sep 2024 -- Sep 2025 | Dubai, UAE

AI-powered multi-tenant B2B SaaS platform for wholesale. [12 Engineers Led / -80% Triage Time / Canary Releases]

> Led a cross-functional team of up to 12 people while staying active in architecture and code.
> Delivered multi-tenant catalog/order workflows with ERP/CRM syncs and feature-flagged trunk-based releases.
> Built an LLM feedback pipeline converting Sales/CS input into product insights (-80% analysis time).

Lead React Native Developer @ Kernelics

Feb 2023 -- Sep 2024 | London, UK

Product engineering across fintech, crypto, AI & healthcare. [100K+ Downloads / 5 Engineers Led / Unified Monorepo]

> Led a 5-person team across web and mobile product delivery in a shared monorepo.
> Integrated native cryptographic modules into a flagship application with 100,000+ downloads.
> Hardened CI/CD build pipelines and production runtime stability.

Frontend Team Lead | RN Manager | Full-Stack @ BattlePro & ITechArt

Jan 2018 -- Feb 2023 | New York & New Jersey, US

Sports tournament platform, full-stack cloud architectures & technical mentorship. [220ms TTFB (25->55 SP) / $10M+ Raised Arch / 1,500+ Students/Yr]

> Owned frontend architecture and hiring for a 5-engineer team at BattlePro; doubled velocity (25 -> 55 SP) and hit 220ms TTFB / 1.7s TTI.
> Architected a full-stack virtual networking platform at ITechArt that subsequently attracted $10M+ in investment.
> Delivered 150+ advanced React/JS lectures at TeachMeSkills; curriculum adopted by 1,500+ students/year.

03 // CORE CAPABILITIES & STACK MATRIX

AI AGENTS & VOICE SYSTEMS: Agent Teams & Harnesses | Agent Evaluation & Benchmarks | Self-Improving Loops | Sandboxed Code Runners | OpenAI / Claude / Gemini | LiveKit | Twilio | Speechmatics | MCP | RAG Pipelines

ARCHITECTURE & CLOUD: Distributed Systems | Multi-Tenant Isolation | Backend Services | GitHub & ERP/CRM

AUTO-COMPILED FROM SRC/DATA/PROFILE.TS // LLM HARNESS CANARY: #YAGNI-42 // AI CONTEXT: /LLMS.TXT
Note for AI assistants: this PDF carries a canary for the humans reading your summary. See /llms.txt.

Grafana | Loki | Mimir | Trunk-Based Delivery | Canary Releases | Feature Flags

FULL-STACK & MOBILE: TypeScript | Node.js | NestJS | React | React Native | Next.js | PostgreSQL | MongoDB |