Andrei Gostishchev

Senior AI Systems Engineer

+60 17 691 50 45 • contact@arteriali.st
linkedin.com/in/arterialist • github.com/arterialist
Selangor, Malaysia

Senior AI Systems Engineer and technical leader with 10+ years of experience taking multi-agent platforms, real-time voice AI, and cloud-native backends from research to production. Recently Founding Engineer at Postlight and CTO at AIMA Labs, leading distributed teams, agent verification and evaluation harnesses, and low-latency cloud infrastructure across international products.

SKILLS

- **Software Development:** Python (AsyncIO, FastAPI), TypeScript, JavaScript, React, Next.js, Node.js, SQL, REST/GraphQL, Kotlin
- **AI & Real-Time Systems:** Multi-Agent Architecture (LangGraph), Model Context Protocol (MCP), Agent Harnesses & RL Evals, Multimodal Voice AI (LiveKit, WebRTC, SIP/PSTN, Gemini Live, OpenAI Realtime), PyTorch, MuJoCo, Spiking Neural Networks, RAG
- **Infrastructure & Cloud:** GCP, AWS (EKS, ECR, S3, EC2), Kubernetes, Terraform, OpenTofu, Helm, Docker, PostgreSQL, Redis, OpenBao, Prometheus, Grafana, Distributed Tracing, GitHub Actions, GitLab CI/CD
- **Leadership:** Technical Leadership, System Design, Cross-Functional Ownership, Fast Prototyping

EXPERIENCE

Postlight — 09/2026 – 11/2026

Applied AI Research, Construction

Founding Engineer

- Led architecture and execution across the production multi-agent platform, cloud infrastructure, and three external AI-data product lines, working with San Francisco based research labs.
- Architected and delivered an audited BIM reinforcement-learning benchmark with anti-cheat evaluation across 4 frontier models, an automated 2D-plan-to-3D robotics simulation pipeline, a BIM World Model research project, and a live GPU egocentric video dataset pipeline, leading to LOIs and collaboration proposals from research labs.
- Cut production AI reply latency 56%, database median latency 59%, PMS sync 78x, and Redis traffic 97%; shipped an agentic workflow engine and infrastructure-scale optimizations across the stack, reaching Tier 1 SLA uptime.
- Led multi-tenant isolation hardening, centralized OpenBao secrets rotation, autoscaled GCP Spot CI runners, and Prometheus/Grafana alerting while cutting vision inference cost 3.6x and infrastructure costs by over 50%.

Tech Stack: Python, TypeScript, FastAPI, Next.js, LangGraph, Celery, Model Context Protocol (MCP), Gemini, Claude Agent SDK, PyTorch, GCP, Fly.io, OpenTofu, Docker, PostgreSQL, Supabase, Redis, OpenBao, Prometheus, Grafana, OpenTelemetry, GitHub Actions

AIMA Labs — 04/2026 – 08/2026

Voice AI Agents. HQ in Malaysia

Chief Technical Officer

- Architected an enterprise voice AI platform across WebRTC and telephony, cutting conversational turn latency below 600ms and processing 5,000+ conversations with 99.9% uptime.
- Built an autonomous tool-synthesis engine that turns private APIs and legacy software into live connectors, launching 15+ integrations and publishing two open-source SDKs.
- Designed multi-region Kubernetes cloud infrastructure on AWS with custom network routing, bypassing regional blocks and achieving 99.99% delivery across global voice carriers.
- Implemented full-stack observability with p50–p99 latency tracing and an interactive audio analytics dashboard, reducing production issue diagnosis time by 80%.

Tech Stack: Python, TypeScript, FastAPI, Next.js, React, LiveKit, WebRTC, SIP, Model Context Protocol (MCP), AWS (EKS, ECR, S3), Kubernetes, Helm, Terraform, Docker, PostgreSQL, Redis, Prometheus, Grafana, GitLab CI/CD

Senior AI & Systems Engineer (Project Work) 03/2023 – 03/2026

Senior AI & Systems Engineer (Project Work) 03/2023 – 03/2026

- Towards Artificial Life (Embodied AI Research): Built PAULA, a novel local-learning spiking neural network model without backpropagation, coupled to real-time 3D MuJoCo physics simulations.
- Distributed Ledger Payment Platform (TON): Engineered FunC smart contracts, non-custodial escrow flows, multi-language merchant SDKs (TypeScript, Python, Go), and deployment infrastructure for a high-throughput blockchain payment gateway.
- Callegram: Architected conflict-safe calendar booking inside Telegram using PostgreSQL transaction isolation, Redis distributed locks, FastAPI, and Google Calendar sync. Won 1st place in the 2023 Telegram Mini Apps Contest.
- HUNTME (Automated Compliance Pipeline): Built an async FastAPI backend and U.S. FMCSA SAFER API verification pipeline serving 7,500+ freight operators, automating regulatory vetting previously handled by 3 specialists.

Pokemod 01/2020 – 03/2023

Game development company. HQ in Brazil

Fullstack Software Developer

- Rebuilt the company's core Android app and backend infrastructure, delivering iterative releases and driving a 5x increase in company revenue in the first year.
- Led end-to-end development of the second product, including UX/UI, user flows, backend, embedded architecture, Android app, and frontend components.
- Coordinated a distributed engineering team, raising code quality standards through mentoring peers, code reviews, testing, and CI/CD practices.
- Built internal tooling: developed and published jsonistyping, an open-source Python utility improving API development speed and data handling.

Tech Stack: Python, Flask, React, TypeScript, Node.js, Docker, Gitlab CI/CD, PostgreSQL, Kotlin

---

DEVELOPERS COMPETITIONS & AWARDS

- Best Startup Pitching Competition, 1st place, 2024
- Telegram Mini Apps Contest, 1st place, 2023
- Telegram Smart Contract Cracking Competition Honorable Mention, 2022

EDUCATION

Don State Technical University, **Associate's Degree in Applied Informatics**, 2020

Gained hands-on engineering experience as Working Student at Neutrino LTD, contributing to real client projects:

- Built and maintained fullstack web platform for a US healthcare company. Implemented cross-platform mobile app.
- Developed 6 Android apps from scratch for international clients in Agile (Scrum) environments.
- Built and maintained fullstack web apps (.NET, Angular, and TypeScript).
- Delivered cross-platform solutions including desktop apps, utility scripts, and web services.

LANGUAGES

English - C1 (advanced) | Russian - C2 (native)
