TIM TUEV

AI DEPLOYMENT STRATEGIST / AI SYSTEMS STRATEGY & ARCHITECTURE

Bridging business viability, product architecture, and AI engineering

Tashkent, UZ (UTC+5) | Open to Remote & Relocation
timothy.tuev@gmail.com | Telegram: **@tuevtim**
Portfolio: **timtuev.com** | GitHub: **github.com/raveunicorn**
LinkedIn: **linkedin.com/in/timtuev**

PROFESSIONAL SUMMARY

AI Deployment Strategist with nearly a decade of cross-functional leadership in digital products, design systems, and creative direction, combined with deep hands-on AI systems R&D. Specializes in translating ambiguous business problems into robust, production-grade AI architectures: from model selection, quantization, and on-device inference to vector substrates, long-term memory, Model Context Protocol (MCP) agents, and deterministic evaluation benchmarks. Works with modern agentic workflows for rapid execution while owning architecture, experiments, metrics, and product trade-offs.

CORE COMPETENCIES

AI Models & Inference: llama.cpp, MLX, vLLM, GGUF/AWQ Quantization, KV Cache Compression, Speculative Decoding, Open Weights (Qwen 3.8/3.6, Gemma 4, Llama 3.2).

Memory & Retrieval: Local-first memory substrates, mmap vector indexing, RaBitQ 1-bit quantization, Hierarchical coarse-to-fine search, BM25, GraphRAG.

Software & Implementation Environments: Rust, TypeScript, React 19, Python, SQLCipher WAL, X25519 DH. *Agentic engineering: architecture-led implementation using coding agents, with review, testing and eval ownership.*

Systems & Leadership: Model Context Protocol (MCP) tool calling, Empirical Benchmarking (LongMemEval, LoCoMo), Agile/Scrum (180% cycle acceleration), Product UX.

FEATURED AI SYSTEMS & R&D TRACK RECORD

SKINKI — Local-First Memory Engine | Founder & Systems Architect

2026 | Active R&D

- Architected a local-first conversational memory substrate in **Rust** enabling ~4B class models on Apple Silicon to perform multi-hop contextual reasoning without cloud data leakage.
- Achieved **5,000,000 vector index residency in <250MB RAM** using RaBitQ 1-bit binary quantization with Int4 exact reranking.
- Designed hierarchical coarse-to-fine retrieval, improving **Recall@10 by 46% (0.291 → 0.438)** over flat dense baseline on LongMemEval dialogue datasets; published empirical benchmark falsification research.
- Implemented native **Model Context Protocol (MCP)** server integration for agent tool calling.

VNUTRI — On-Premise B2B Hardware Messenger Appliance | Systems Architect

2025 — 2026

- Designed turn-key on-premise communication appliance on low-power ARM hardware (Rockchip RK3528A / RK3588S) for air-gapped corporate deployments.
- Engineered shared **Rust cryptographic core (vnutri-crypto)** compiling to WebAssembly (WASM) for Web and UniFFI for native Swift/iOS.
- Created custom 33-byte binary header MsgPack protocol for $O(1)$ blind server routing and in-memory SQLCipher write-buffering, reducing disk fsyncs by ~260x to protect flash storage.

PROFESSIONAL EXPERIENCE & EMPLOYMENT HISTORY

Independent Practice | AI Deployment & Product Strategy Consultant

09/2024 — Present | Tashkent / Remote

- Consulting leadership teams on AI system viability, open-source model inference runtimes, hardware trade-offs, and tool-use architectures.
- Restructured operational workflows for creative & technical departments using Agile, Scrum, and Kanban methodologies, **increased throughput by 180% in 6 months**.
- Architected comprehensive digital navigation and branding systems for commercial initiatives.

Pacific.agency | Head of Creative Department

08/2023 — 08/2024 | Moscow, Hybrid

- Directed multidisciplinary team of 5 (3 designers, 2 creative strategists) across advertising, digital products, and brand systems.
- Increased department production capacity (CLT) by **+40% in 6 months** through operational restructuring and Agile integration.
- Decreased average task cycle duration by **50%** through cross-departmental alignment and streamlined feedback loops.
- Spearheaded client pitches and successful delivery of 5 high-stakes commercial accounts totaling **RUB 8M+** in contract value.

GavLab | Art Director

03/2021 — 07/2023 | Moscow

- Led a creative team of 7 (4 designers, 3 artists), overseeing brand identity across all digital and physical touchpoints.
- Produced major cultural exhibitions, media campaigns, and creative productions at leading venues (Winzavod, Red October).

Independent Design & Digital Practice | Product & Brand Designer

12/2017 — 07/2023 | Moscow

- Designed end-to-end brand identities, UX/UI frameworks, print design, and multi-channel digital assets for commercial clients.

EDUCATION & LANGUAGES

Master of Management — Business Process Management in Creative Industries, Institute of Contemporary Art (ISI), Moscow 2025 — 2027

Bachelor of Arts — Graphic & Digital Design, Institute of Contemporary Art (ISI), Moscow 2021 — 2025

Languages: English — Professional Working Proficiency, Russian — Native