HEAD OF AI / CTO

Vladimir Evgrafov

Applied AI/ML systems, production engineering, and LLM-agent security



evgrafov.su | LinkedIn | GitHub | Google Scholar

EXPERIENCE

Head of AI

September 2026 - present | Moscow | biomedicine

- Lead AI strategy, applied research, and engineering in biomedicine.
- Build LLM systems and agentic architectures with RAG; develop model evaluation practices and production AI infrastructure.
- Combine hands-on engineering with technical leadership, building processes that bring AI solutions from experimentation into practical use.

Viasat Tech - Chief Technology Officer

January 2022 - September 2026 | Moscow | viju

- Led AI/ML, recommender systems, architecture, engineering, and infrastructure; built a 35-person department across seven teams.
- Migrated a legacy backend to Go microservices, enabling 100x the load at one quarter of the resource consumption.
- Delivered 100+ releases across seven platforms, raised quarterly plan completion from 40% to 83%, and cut planning overhead 4x.
- Introduced engineering standards, observability, SSO, release calendars, and KPIs; reduced infrastructure costs.
- Cut Android crash rate 14x, halved video start time, and reached 99.88% iOS crash-free sessions.

Atlas Biomed - Backend Team Lead

April 2020 - January 2022 | healthtech

- Built systems for medical, genomic, and microbiome data, including ML mechanisms and scientific algorithms.
- Designed computational pipelines and translated research logic into reliable production services.

TRMPLN Media - Golang Developer

October 2018 - April 2020 | adtech

- Developed a high-load ad-serving system, decision algorithms, and safe personal-data processing.

Multy - Golang Developer

March 2018 - October 2018 | fintech

- Built microservices that integrated with financial systems.

Okmeter - Golang Developer

September 2017 - December 2017 | monitoring

- Worked on agents that collected metrics from Linux devices and processes.

SolidWall - Python Developer

July 2015 - June 2017 | application security

- Maintained the Python codebase, software build and deployment system, CI, and monitoring.

PROFILE

I lead AI strategy, applied research, and engineering in biomedicine. My focus is LLM systems, agentic architectures, RAG, model evaluation, and production AI infrastructure.

A hands-on CTO who grew from Go engineering into technology leadership. I built a 35-person department, migrated a legacy backend to Go microservices, and owned architecture, infrastructure, budget, risk, and delivery.

SELECTED IMPACT

35
people in the department

7
teams and platforms

100x
supported-load increase

14x
Android crash-rate reduction

CORE EXPERTISE

- AI/ML and scientific algorithms
- Engineering leadership
- Strategy, budget, and risk
- Architecture and production engineering
- Hiring, process, and delivery
- LLM agents and AI security

Vladimir Evgrafov

Applied AI/ML systems, production engineering, and LLM-agent security

EDUCATION

PhD candidate - Artificial Intelligence

Lomonosov Moscow State University, CMC | expected 2027

Thesis on security for LLM-based agentic systems: attack methods and defenses for autonomous agents.

MSc - Artificial Intelligence

Lomonosov Moscow State University, CMC | 2021

Developed a method for converting conventional neural networks into spiking networks while preserving accuracy and improving inference efficiency.

BSc - Applied Mathematics and Computer Science

Lomonosov Moscow State University, CMC | 2019

Built grammar-oriented fuzzing for web-template engines to uncover XSS, server-side code execution, and denial-of-service defects.

RESEARCH

I research the security of autonomous LLM systems: prompt injection, attacks on memory, tools and protocols, multimodal threats, and inference-time defenses.

SELECTED PUBLICATIONS

2026 **Auditing Inference-Time Defense Evaluation for Multimodal Large Language Models**
arXiv, cs.CR

2026 **Attack Methods and Defenses in LLM-Based Agentic Systems**
International Journal of Open Information Technologies 14 (5)

2023 **A Review of Potential Risks of Systems with Multimodal Language Models**
Tikhonov Readings

2021 **Spiking Neural Networks**
International Journal of Open Information Technologies 9 (7)

TECHNICAL FOCUS

- Machine learning
- Biomedical AI
- Python
- Go
- C++
- Data pipelines
- Distributed systems
- Microservices
- DevOps
- Observability
- English C1

RESEARCH INTERESTS

- LLM agent security
- Inference-time defenses
- Multimodal models
- AI evaluation
- Spiking neural networks

TARGET ROLES

Open to full-time CTO and Head of AI roles, as well as fractional CTO and technical advisory engagements.

PROFILES

evgrafov.su

LinkedIn

GitHub

Google Scholar