Artem Molchanov

Lead AI Engineer | LLM Systems & AgentOps | RAG • Evals • Reliability (Cost/Latency/Obs)
Remote (global) • Europe-based • molchanov.art13@gmail.com
linkedin.com/in/art-molchanov • +995 (591) 182-301

Hands-on AI/ML engineer (10+ yrs across fintech, ads, e-commerce, media, web3). I build **production LLM systems (agents, RAG, evals)** that are **reliable, measurable, and cost/latency-aware**, and I own **end-to-end architecture** across research ↔ engineering ↔ product. Shipped at startup speed and at **150M+ user scale**.

SKILLS (LLM Systems & AgentOps)
Agentic workflows (tool use, orchestration, safe execution, HITL) • RAG (hybrid retrieval, ontologies/knowledge graphs) AgentOps (tool scopes/permissions, audit trails, regression gates, tracing) • Architecture & delivery • Applied ML at scale.
Tech: Python/SQL; MLflow, A/B testing; Docker, CI/CD, AWS; PostgreSQL/MySQL/MongoDB; Spark; LangGraph.

Experience

Senior AI/ML Engineer • **GriffinAI**, *LLM Agents & Web3* *2024 — Present*
Owned end-to-end architecture and delivery of production agentic LLM products; aligned cross-functional stakeholders and drove execution under ambiguity.

- **Transaction Execution Agent:** hybrid pipelines; **~2× faster** avg responses + **lower token cost** via routing & LLM call optimizations; safe tool execution (contracts/guardrails/HITL). | te.griffinai.io (signup required)
- **Proposal Examiner:** built end-to-end; improved auditability via governance KG + structured reasoning.
- Built goal-driven autonomous multi-agent system, ops alerts (Slack/Telegram), output channels (TG/Twitter).
- Production deployments: containerized services + secure cloud integrations.
- Implemented **tool contracts + approval gates + audit logs** for high-stakes agent actions.

Independent Consulting / Contracts (AI/ML & LLM Systems, Remote) *2022 — 2024*

- Led end-to-end delivery of applied LLM systems for finance and e-commerce (pipelines → models → production).
- Built document/knowledge workflows (search/RAG + citations) to speed up retrieval and decision-making.
- Owned architecture and stakeholder alignment; optimized for quality, reliability, and latency/cost in production.

Head of Data Science and AI R&D • *Sber, Largest fintech ecosystem in Eastern Europe* *Apr 2019 – Dec 2022*
Directed AI initiatives in the **Laboratory for Neuroscience & Human Behaviour**, aligning with Sber's strategic goals.

- Led distributed team (**10+**); shipped AI improvements across **50+ products**.
- Built client profiling platform ("Digital Avatar"), contributing to **$20M+** revenue impact.
- Delivered personalization/growth systems at **150M+** user scale; drove experimentation + rollout + monitoring.
- Ran cognitive-architecture research track; contributed to internal Ethical AI principles.

Data Scientist • **Segmento**, *Major fintech-integrated automated system of online marketing* *Aug 2017 – Apr 2019*

- Improved targeting/behavioral analytics (+10% CTR, +7% CR); reduced RTB latency (~15%).
- Built Spark ETL pipeline (~2× faster processing) and audience expansion models.

Earlier roles (selected, pre-2017): ML/DS across decision systems, ETL/data platforms, recommendations/anti-fraud.

Open build: **Exocortex** – a personal agentic OS (local-first) Testbed for **Verified Agent Loops** (permissions, audit, evals).

Publications & Talks

- "Cognitive Architecture for Decision-Making Based on BPP", Procedia CS, 2022 • **Paper** • **Conference Video**
- "Brain Principles Programming (BPP)", AGI-2022 (Lecture Notes in CS), 2022 • **Paper** • **Conference Video**

Education: B.S. in Computer Software Engineering — Saint Petersburg State University

Languages: Russian (native) • English (professional) • French (basic)