Igor Yakushev

Senior ML Engineer | Search, Retrieval, Ranking, Recommender Systems

igor.ykshv@gmail.com | linkedin.com/in/igor-yakushev | igor-ya.com

Senior ML Engineer focused on search, retrieval, ranking, and recommender systems for high-scale discovery and personalization products. Built and operated production ML systems serving 10M+ requests/day, spanning candidate generation, embeddings, hybrid retrieval, reranking, online experimentation, rollout safety, and observability. Delivered +14% search quality, +12% CTR, +8% conversion, -25% latency, and -30% infra cost.

Experience

ViSenze • Senior Machine Learning Engineer
May 2023 - Present

E-commerce AI platform, search and recommendation stack

- Built **candidate generation, multimodal retrieval, and reranking** for text and image search under latency and cost constraints.
- Built **recommendation models** from behavioral signals for personalized ranking and session-level engagement.
- Owned **release safety end to end**: offline evals, A/B tests, canary rollout, rollback, and regression gates.
- Owned **serving, backend APIs, distributed systems, ML infrastructure, tracing, and observability for 10M+ requests/day**, covering quality, latency, and incident response.
- Shipped **embeddings, vector retrieval, and production reranking** with ML, backend, data, and product teams.
- Improved **search quality 14%**, **CTR 12%**, **conversion 8%**, reduced **latency 25%**, and cut **infra cost 30%**.

Sunshine Group • ML Engineer
May 2022 - Feb 2023 (10m)

B2B e-commerce platform

- Built **dynamic pricing, demand forecasting, and decision support** modules for e-commerce operations.
- Designed **end-to-end ML pipelines** with auto-retraining, fallback logic, and monitoring.
- Supported **~1M predictions/day at 99.9% SLA** and introduced LLM-assisted content workflows.
- Supported **search and commercial optimization** use cases that contributed to sales uplift in targeted categories.

Google • Marketing Lead → ML Solutions Specialist
Dec 2019 - Mar 2022 (2y, 4m)

YouTube Ads and Google Ads | Internal transition from marketing into ML

- Increased **YouTube Ads CTR by 18%** and reduced **A/B testing costs by 30%** through segmentation and forecasting automation.
- Completed Google's internal ML training and moved from **marketing into ML-based ad optimization**.
- Supported **feature engineering, model analysis, and reporting automation** for Google Ads, reducing **manual work by 35%**.

Selected Production Systems

Representative production systems across product and client environments.

Catalog Search Retrieval and Reranking
Hybrid retrieval workflow with experiments and quality monitoring. Reduced **zero-result searches 30% to 9%**, improved **CTR by 1.6 pp**, and increased **GMV by 6.3%**.

ML Inference Latency and Cost Platform
Internal platform for model-level latency, throughput, and cost visibility. Reduced **$/req by 43%** while stabilizing **p99 latency**.

Voice AI Operator for Contact Center
On-prem voice AI with policy boundaries and controlled escalation. Achieved **72% call automation**, **~0.96s response time**, and **~58% cost reduction**.

Core Skills

- Search, retrieval, ranking, recommender systems, hybrid search
- Embeddings, vector search, reranking, relevance evaluation
- Online experimentation, A/B testing, canary rollout, rollback gates
- ML pipelines, serving, inference optimization, observability
- LLM-assisted workflows, RAG systems, GenAI evaluation
- Python, SQL, Kubernetes, AWS, Triton, ONNX

Education

Applied Mathematics and Informatics
Russian University of Transport, 2013

Philosophy
Lomonosov Moscow State University, 2019