

# Dmitrii Filippov

📍 San Francisco, CA | 📞 +1 415 725 9942 | ✉️ [dima239philippov@gmail.com](mailto:dima239philippov@gmail.com)  
🌐 [linkedin.com/in/dmitry-filippov](https://www.linkedin.com/in/dmitry-filippov) | 🎧 [DimaPhil](#)

## Professional Summary

---

Results-driven Engineering Manager and Staff Engineer with over a decade of experience architecting high-load distributed systems (serving 1.5B+ daily requests) and pioneering AI-driven product pipelines. Proven track record of scaling global infrastructure, reducing cloud costs by over 40%, and leading high-performing teams to ship autonomous AI agent workflows. Seeking to leverage deep expertise in advanced algorithms and distributed architecture to drive technical innovation.

## Technical Skills

---

**Languages:** Java, Kotlin, TypeScript, Python 3

**AI:** LLM Agents (Claude Code/Codex/OpenCode), Vercel AI SDK, RAG Pipelines, Vector Search (Qdrant, Pinecone)

**System Architecture:** Microservices, Event-Driven Architecture, High-Load Systems, Distributed Caching

**Infrastructure & Cloud:** AWS (EC2, Lambda, S3, RDS), Docker, Kafka, Redis, Airflow

**Frameworks:** Spring Boot, Dropwizard, Node.js, React

## Experience

---

### Engineering Manager & Staff Engineer (AI-First Platform)

Aug 2023 – Present

[CodeSignal](#) | San Francisco, CA

- Leading a team of 6 AI Engineers building the core platform for autonomous content generation, producing over 1,500 courses and 1,000 hiring assessments annually.
- Designed and deployed the MVP of the internal platform from scratch prior to team expansion, scaling it to support the entire content engineering organization.
- Architected a proprietary workflow language to sequence non-deterministic LLM operations into reliable, repeatable pipelines, enabling the seamless integration of agents like Claude Code and Codex.
- Engineered a "Human-in-the-Loop" review system utilizing an AI Agent Swarm to pre-validate educational content, significantly reducing manual QA time by 60%.
- **Stack:** *AWS, Node.js, React, Vercel AI SDK, MongoDB, Vector Search*

### Senior Software Engineer

Nov 2021 – Aug 2023

[Sumsu](#) ([Identity Verification](#)) | Remote

- Led the architectural decoupling of a legacy monolith into domain-specific microservices, resulting in a 40% reduction in infrastructure costs and improved CI/CD velocity.
- Designed a distributed rate-limiting engine to replace legacy in-memory implementation, ensuring strict system stability during massive traffic spikes.

- Extracted search capabilities into a dedicated microservice, safely migrating over 100M records to the latest OpenSearch cluster with zero downtime.
- Developed a high-performance, fuzzy-matching system for applicant names, processing millions of data entries and drastically reducing false rejection rates.
- **Stack:** *Java 11, Kafka, MongoDB, Debezium, OpenSearch, Airflow, DropWizard*

## Senior Software Engineer (High Load)

Jun 2019 – Nov 2021

VK (Mail.ru Group) | St. Petersburg, Russia

- *Context: The United Video platform serves 1.5 billion daily views, 5,000+ simultaneous live streams, and manages 10+ PB of storage.*
- Re-architected the video streaming engine to migrate to the unified platform, maximizing scalability and fault tolerance.
- Streamlined recommendation backend algorithms, heavily reducing content delivery latency and improving user retention metrics for the VK Clips feature.
- Built a unified video import framework, standardizing data ingestion across all group partners and reducing integration time for new partners by 70%.
- **Stack:** *Java, Spring, Cassandra, Zookeeper, Clickhouse, one-nio*

## Senior Software Engineer

Feb 2018 – Jun 2019

Huawei R&D Labs | St. Petersburg, Russia

- Developed the core inference engine and performance measurement toolkit for a proprietary intelligent mobile Voice Assistant.
- Engineered a high-performance, fault-tolerant microservice for Text-to-Speech (TTS) conversion, explicitly optimized for extreme low-latency constraints on mobile devices.
- **Stack:** *Java 8, Kotlin, Spring Boot, Kafka, Elasticsearch*

## Software Engineer

Jan 2017 – Feb 2018

Jetlore (Acquired by PayPal) | Remote

- Built the backend for an AI/ML-based prediction platform used by top-tier global retailers, designing REST APIs and data pipelines for high-velocity user behavior events.
- **Stack:** *Scala, Akka HTTP, Redis, Cassandra, Spark, AWS*

## Content Engineering Manager

Nov 2015 – Feb 2018

CodeSignal | Remote

- Promoted from Content Engineer to Manager; led a team of 4 to author complex algorithmic challenges for global monthly coding championships and battles.

## Education

---

### Master of Science in Applied Mathematics & Computer Science

2013 – 2019

ITMO University | St. Petersburg, Russia

## Awards & Achievements

---

- **Silver Medal**, ACM ICPC [NEERC Semifinals](#) – Prestigious Global Algorithmic Championship (2014)
- **Prize Winner**, All-Russian Individual Olympiad in Informatics (2012, 2013)