

Nikolai Malchikov

City: Madrid, Spain · malchikov.ne@gmail.com · +34 (685) 39-12-13 · linkedin.com/in/malchikov

## SUMMARY

AI Engineer & Senior Backend Engineer with 7+ years of experience building distributed backend systems, asynchronous processing pipelines, and high-load B2B platforms. Focused on LLM integrations, AI execution pipelines, and scalable backend infrastructure for AI-powered workflows.

Experienced in Go-based services, event-driven architectures, queue-driven workflows, and operationally reliable distributed systems using Kafka, Redis, and PostgreSQL. Designed microservices, background workers, and scalable processing pipelines with strong emphasis on idempotency, fault isolation, transactional consistency, and production observability.

AI/LLM Infrastructure: Built asynchronous AI execution pipelines with queue-driven orchestration, worker isolation, retry handling, batching, and rate limiting. Integrated LLM-based processing workflows into backend services through internal APIs and event-driven pipelines. Developed AI-assisted flows for classification, routing, enrichment, and operational automation. Implemented request optimization, caching, and batching strategies to reduce latency and operational cost of AI workloads. Worked with monitoring, tracing, and debugging of asynchronous AI workflows.

Strong ownership across architecture, backend infrastructure, deployment pipelines, monitoring, and production operations.

## SKILLS

Languages: Go (primary), PHP, SQL

Architecture: Microservices, Distributed Systems, Event-driven Architecture, CQRS, Outbox Pattern

Messaging & Streaming: Kafka, Redis Streams (consumer groups, fan-out, partitioning)

Workflow Orchestration: Temporal

APIs: gRPC, REST, GraphQL

Databases: PostgreSQL (including Aurora), MongoDB, Redis

Infrastructure: Docker, GitLab CI/CD, Linux, containerized deployment patterns, basic AWS/GCP familiarity

Practices: Clean Architecture, DDD, Idempotency, Fault-tolerant design, High testability focus

### AI / LLM:

LLM API integration (OpenAI / Claude), async processing pipelines, prompt workflows, rate limiting, caching strategies

### WORK EXPERIENCE

**Backend Engineer — AI/LLM Systems & Distributed Processing**  
**years**

**09/2024 – Present – 3**

**Lanzar**

- Built and maintained Go-based backend services for a multi-tenant B2B SaaS platform focused on distributed processing, asynchronous workflows, and operational scalability.
- Developed event-driven processing pipelines using Kafka and Redis for imports, billing workflows, background execution, and internal service orchestration with idempotent processing guarantees.
- Integrated payment providers including Stripe with retry handling, reconciliation flows, and idempotent transaction processing.
- Participated in decomposition of monolithic domains into independent services and scalable event-driven processing flows.
- Optimized PostgreSQL and Redis usage through query optimization, caching strategies, and efficient data access patterns, reducing API latency under production load.
- Applied Outbox pattern, domain events, and asynchronous communication patterns across distributed services.

### AI / LLM Infrastructure:

- Built asynchronous AI execution pipelines with queue-driven orchestration, worker isolation, retry handling, batching, and rate limiting.
- Integrated LLM-based processing workflows into backend services through internal APIs and event-driven execution pipelines.
- Developed AI-assisted processing flows for classification, routing, enrichment, and operational task automation.
- Implemented request optimization, caching, and batching strategies to reduce latency and operational cost of AI workloads.
- Worked with monitoring and debugging of asynchronous AI workflows, including retries, execution tracing, and background task observability.

**Backend Engineer**

**03/2022 – 09/2024 – 3 years**

**MyAutoData – B2B platform for automotive data management (50K+ active users, 100K+ API requests/day).**

- Designed and developed backend services using Golang for payment workflows, background processing, and

internal APIs

- Built a payment orchestration service using Temporal, Kafka, and Outbox pattern, ensuring reliable and idempotent multi-step processing
  - Integrated Stripe billing system handling 10K+ monthly transactions with retry strategies and reconciliation flows
  - Decomposed monolith into 5+ microservices, improving system scalability and deployment flexibility
  - Implemented caching strategies with Redis, reducing API latency by 40%
  - Improved system reliability through idempotent and fault-tolerant design
  - Delivered features end-to-end in collaboration with product and engineering teams
- Tech:** Golang, PostgreSQL, Redis, Kafka, Temporal, Docker, GitLab CI/CD

**Backend Engineer**

**02/2021 – 02/2022 – 1 years**

**Apart Sharing – B2B platform for property rental and cleaning operations.**

- Developed internal services and background workers in Golang
  - Implemented event-driven workflows using Redis Streams and Outbox pattern
  - Designed service boundaries and prepared migration from monolith to Go-based microservices
  - Built and maintained REST and GraphQL APIs
  - Set up Docker-based development environments and CI/CD pipelines
- Tech:** Golang, PostgreSQL, Redis Streams, Docker, GraphQL

**Backend Engineer**

**05/2018 – 01/2021 – 2 years**

**Runexis – Telecommunications & enterprise automation.**

- Developed and maintained high-load APIs (100K+ requests/day)
  - Optimized PostgreSQL queries, reducing latency by 35%
  - Built containerized services and improved deployment consistency
- Tech:** PHP, PostgreSQL/MySQL, Docker

**EDUCATION Moscow State Open University Bachelor of Engineering in Computer Science – 2020**