

Amal Imangulov

STAFF / PRINCIPAL DATA ENGINEER

✉ imangulovamal@gmail.com · ☎ +44 73 4307 6336 · 📍 London, UK
🔗 linkedin.com/in/imangulov · 🌐 github.com/imangulov

RECOGNITION & COMMUNITY

- ▶ **Winner — EMEA DE Festival Hackathon 2026** (Meta; 4-city competition: London, Zurich, Tel Aviv, Dublin). Methodology being integrated into Meta internal tooling serving thousands of engineers.
- ▶ **Microsoft Azure OpenAI case study contributor** — Syntea AI tutor at IU International University (one of the first university-scale GPT-4 deployments worldwide; 80,000+ students).
- ▶ **Hackathon designer & host** — created the technical challenge and hosted the finale for a Gett hackathon (3,000+ online participants, 100+ at the in-person final).
- ▶ **Community organiser** — multi-year host of a Russian-language quiz-game league (*What? Where? When?*) spanning thousands of active members across Europe and Asia.

WORK EXPERIENCE

Data Engineer — Meta (Senior IC, E5; operating at Staff scope)

Nov 2025 – May 2026 | London, UK Messaging Integrity / Trust & Safety — messaging platform serving billions of users

TECHNOLOGIES: Python, SQL, Hack, React, Relay, Next.js, Dataswarm (Meta-internal Airflow-equivalent), Presto, Hive, multi-agent LLM orchestration

- ▶ **Architected *Untangler* and won the EMEA DE Festival Hackathon 2026** (London / Zurich / Tel Aviv / Dublin) — an AI-powered pipeline restructuring methodology orchestrating 3 parallel analysis agents that auto-generate sequenced migration plans with executable diffs. Methodology peer-adopted across multiple Meta teams and **being integrated into internal data tooling serving thousands of Meta engineers** as a durable platform contribution; integration continued by the team post-departure.
- ▶ Architected unified data model consolidating **10 fragmented logging systems ingesting ~3.6B events/day**, eliminating **47% redundant data volume** and unlocking cross-surface safety analytics previously blocked by schema fragmentation.
- ▶ Owned end-to-end audit and restructuring of a critical messaging pipeline (**22 tables, 82 downstream consumers**) — resolved architectural violations and re-baselined downstream metric definitions across the consumer graph.
- ▶ Resolved a long-standing E2EE entity-resolution defect affecting **34% of messaging threads (6.2B records)** — restored accuracy for **100+ downstream metrics and ML features** powering integrity decisioning.
- ▶ Drove cross-team ownership migration of **86 logging systems across 3 engineering teams** — coordinated with **4 Engineering Managers and 7+ engineers; deprecated 21 systems** via a reusable two-phase playbook now applied by neighbouring teams.
- ▶ Productionised **7 custom AI developer tools** (multi-agent wiki automation, data-quality debugger, video-to-recap, weekly visibility system) peer-adopted across teams.
- ▶ Authored a reusable Semantic Model bootstrap playbook (11-step procedure) achieving 83% baseline accuracy across 3 messaging-integrity domains on first deployment.
- ▶ Shipped React / Relay UI features for the Self-Healing Warehouse platform and authored a company-wide cold-storage savings pipeline operating as a table-optimisation primitive.
- ▶ Authored and prototyped **companywide AI Harness initiatives at Staff scope** while formally operating at E5.

Lead Data Engineer — Career.io (Talent Inc.)

Dec 2024 – Nov 2025 | Berlin / Remote PE-backed career-tech holding (BV Investment Partners) — **8-brand portfolio**: Resume.io, TopResume, TopCV, TopInterview, ZipJob, CareerMinds, PremierVirtual, CourseReport. ~1,500 employees.

TECHNOLOGIES: Snowflake, dbt Cloud, Airbyte, Periscope, SQL, Python, LLM-augmented engineering workflow

- ▶ Led an **8-person analytics engineering team** (data engineers + analytics engineers) owning the unified data warehouse spanning the full **8-brand portfolio** under BV Investment Partners.
- ▶ Owned end-to-end the company's **financial reporting and KPI infrastructure** — the data layer powering board-level reporting, finance close, marketing attribution, and sales pipeline metrics across all 8 brands.
- ▶ Established and institutionalised **dbt code-review discipline, SQL optimisation standards, and an analyst-engineer collaboration model** that survived team rotation and remained the operating norm post-handover.
- ▶ Architected the modern data stack (Snowflake + dbt Cloud + Airbyte + Periscope) supporting multi-brand integration and self-service reporting for ~1,500 employees.
- ▶ Drove convergence on **shared KPI semantics across 8 historically siloed brand data domains** by partnering with finance, marketing, and sales analytics leads.
- ▶ **Pioneered LLM-augmented engineering workflow** across the team — integrated AI tooling into code review, query optimisation, and data validation; established the team-level pattern for AI-native analytics engineering.

Senior Data Engineer / Analytics Engineer — Flip

Jun 2024 – Sep 2024 | Stuttgart / Remote B2B employee-experience platform for deskless workers — **1M+ users, 500+ enterprise customers** including McDonald's, Bosch, Porsche.

TECHNOLOGIES: SQL, GCP, BigQuery, Looker, Airflow

- ▶ Owned rebuild of CRM and marketing data pipelines on BigQuery aligning the data layer to the company's commercial objectives during a high-growth enterprise expansion phase.
- ▶ Drove measurable accuracy gains in sales- and marketing-reporting through pipeline restructuring and query optimisation across the BigQuery + Looker stack.
- ▶ Established **SQL code-review standards** across the analytics team — durable engineering hygiene that outlived the engagement.

Senior Data Engineer / Tech Lead — IU Group (Syntea AI)

Aug 2023 – Feb 2024 | Berlin IU International University — Europe's largest distance-learning provider · **80,000+ students**

TECHNOLOGIES: AWS, Python, Kafka, Postgres, dbt, Quicksight, Azure OpenAI (GPT-4)

- ▶ Led a **2-engineer team** building data product infrastructure for **Syntea** — IU's GPT-4-powered AI tutor and **one of the first university-scale GPT-4 production deployments worldwide** (IU was first university announced on GPT-4 in March 2023; broader Syntea student rollout launched October 2023).
- ▶ Co-shipped a deployment **featured by Microsoft as an Azure OpenAI customer case study** — production reference architecture for university-scale generative AI in regulated EU education.
- ▶ Operated as a **cross-organisation principal contributor** between AI Product engineering (Syntea team) and the central Data Department — bridged frontier AI development with the university's production data systems and reporting layer.
- ▶ Architected **event ingestion via Kafka** and analytics pipelines supporting the live AI conversation product serving 80,000+ students at GPT-4 launch scale.
- ▶ Drove **pre-mainstream GenAI productionisation** (Q3 2023): conversation data modelling, learning-outcome telemetry, Azure OpenAI cost and usage governance.

Senior Data Engineer / Analytics Engineering Tech Lead — McMakler

Jan 2020 – Jul 2023 (3y 6m) | Berlin Baillie Gifford / Warburg Pincus-backed real-estate platform · ~1,000 employees at peak · €400M+ valuation (post-2023 round)

TECHNOLOGIES: GCP, BigQuery, Snowflake, dbt, Airflow, Fivetran, Looker, Tableau, Python, Postgres, custom Salesforce integration, Bokeh, SHAP

- ▶ **Saved €1M+ across tenure** through tooling migrations, infrastructure consolidation, and licensing rationalisation — the only audited, finance-confirmed monetary outcome in this CV.
- ▶ **Grew the analytics engineering team from 2 to 7 engineers** — directly hired 5–10 engineers, supported hiring loops across the broader org, and **promoted 3 engineers from Junior to Mid/Senior** using a formalised **skill-and-working matrix framework** that became the team's operating standard for levelling.
- ▶ Architected the **GCP-native → Snowflake-centred warehouse migration** end-to-end — multi-quarter programme touching every downstream consumer in the company.
- ▶ Led **orchestration migration from proprietary tooling to Airflow** — eliminated hundreds of thousands of EUR per year in licensing and infrastructure costs; remained the durable orchestration substrate post-departure.
- ▶ **Initiated the Postgres → dbt analytical-layer rebuild**; migration **continued post-handover** and was completed by the team using playbooks and standards established during tenure.
- ▶ Built and owned **self-service analytics infrastructure serving 200+ business users across 8 departments** (Marketing, Sales, Business Development, Business Intelligence, Finance, HR, Customer Care, Product) — the canonical data product for the entire company.
- ▶ Architected a **proprietary Salesforce sync layer orchestrated via Airflow** as an in-house alternative to commercial reverse-ETL SaaS — powering CRM enrichment, marketing attribution, and sales pipeline operations.
- ▶ Established **DataOps / MLOps pipelines** across business-critical workflows and instituted a company-wide **data-literacy and knowledge-sharing programme** raising the data ceiling across both tech and business teams. Shipped supporting ML work — churn modelling, outlier detection, SHAP-based explainability.

Data Engineer / Senior Data Scientist — Front Tier

Aug 2018 – Nov 2019 (1y 4m) Data science consulting startup productionising analytical work for enterprise clients

TECHNOLOGIES: Python, Scala, SQL, Terraform, Hadoop, Flask, Tableau

- ▶ **Architected and deployed a default-prediction model in production for the factoring department of a top-10 Russian bank** — owned end-to-end productionisation from training pipeline to live Flask API; deployment delivered **50% loss reduction** in the department.
- ▶ Productionised analytical engagements from prior consulting work into repeatable enterprise-scale systems across **banking, telco, and retail** clients.
- ▶ Architected **Hadoop-based data pipelines** and **Terraform-managed infrastructure** for client deployments — established reproducible, IaC-driven delivery patterns.

Data Scientist — McKinsey & Company

Oct 2016 – Aug 2018 (1y 11m) | Moscow Management consulting — multi-industry production AI and analytics.

Clients across banking, telco, retail, public sector, and mobility.

TECHNOLOGIES: Python, PyTorch, R, SQL, Oracle, Tableau, D3.js

- ▶ Architected and shipped a **production face-recognition web application** (clustering + neural networks) for an enterprise client — early production AI deployment ahead of mainstream computer-vision tooling.
- ▶ **Automated loan-decisioning processes for a top-10 Russian bank** — built models and pipelines deployed into live operational workflows.
- ▶ **Lifted problem-customer ARPU by 20%** at a leading Russian telco via segmentation modelling, 40+ targeted campaigns, and closed-loop uplift measurement.

- Designed **discount and promotion strategy for a top-5 Russian supermarket chain** — converted customer-behaviour modelling into live operational pricing recommendations.
- **Drove +10% customer-satisfaction lift in a European city's transport authority** through data-driven service-design analysis.
- Productionised NLP (CV-parsing for banking-sector hiring, pre-LLM era), churn and LTV models for telco and banking, and customer-segmentation clustering deployed in operations.
- Pioneered **alternative-data acquisition via web scraping** across social networks and aviation sources for analytical engagements; built 20+ Tableau dashboards and D3.js interactive infographics for senior-client decks.

Data Analyst — Thomson Reuters

Jul 2015 – Oct 2016 (1y 4m) | Moscow Russian metals & commodities trading platform (Refinitiv / Eikon-adjacent tooling)

TECHNOLOGIES: C#, Oracle SQL Developer, D3.js, Excel

- Built a **C# production application** integrated with an Oracle database supporting the daily workflow of Russian non-ferrous and precious-metals traders.
- Authored D3.js interactive infographics for trading reports and client-facing dashboards.
- Promoted from Data Analyst Intern to Junior Analyst within the first year.

SKILLS

Engineering Leadership & Scope Team growth (2→7 at McMakler; 8-person team at Career.io) · direct hiring (5–10 hires + cross-org hiring loops) · mentorship & promotions (3 Junior→Mid/Senior via formalised skill-and-working matrix) · code-review and standards institutionalisation · cross-functional stakeholder management (200+ users / 8 departments at peak) · multi-team coordination (4 EMs / 7+ engineers / 3 teams at Meta)

Data Platform & Architecture Modern data warehouse and lakehouse design · multi-cloud GCP + AWS + Azure · large-scale migration leadership (GCP→Snowflake, proprietary→Airflow, Postgres→dbt) · self-service analytics platforms · DataOps / MLOps · semantic modelling · data-quality systems

Warehouses, Transformation, Orchestration, Streaming Snowflake · BigQuery · Hive · Presto · dbt (Core + Cloud) · Airflow · Dataswarm · Fivetran · Airbyte · Kafka · advanced SQL (window functions, CTEs, query optimisation) · custom Python ingestion · Salesforce integration via Airflow

AI / ML Engineering LLM orchestration · multi-agent systems · agentic workflows · RAG · tool use & function calling · prompt engineering · AI-augmented development (Claude Code) · PyTorch · production GenAI (Azure OpenAI GPT-4 at university scale) · custom AI developer tooling

Programming & Visualisation Python · SQL · Hack · Scala · C# · R · JavaScript · TypeScript · React · Next.js · Relay · D3.js · Bokeh · Terraform · Hadoop · Postgres · Looker · Tableau · Periscope · Quicksight

Domain Expertise Trust & Safety / Messaging Integrity (Meta) · production AI in regulated education (IU) · PE-backed multi-brand data integration (Career.io) · real-estate marketplace (McMakler) · financial services analytics (top-10 RU bank: factoring, loan automation) · telco analytics · retail pricing · public-sector service design · alternative data

EDUCATION

Bachelor of Applied Mathematics & Informatics

National Research University Higher School of Economics, Moscow · 2012 – 2016

Data Science concentration · Thesis: data science in macroeconomics

LANGUAGES

English C1+ · **Russian** native · **Tatar** native · **Ukrainian** C1+ · **Belarusian** C1+ · **German** B1