

Aleksandr Filippov

AI Product Manager / LLM Engineer

Limassol, Cyprus | hh.remold600@passmail.net | <https://www.linkedin.com/in/alex-feel/> | <https://github.com/alex-feel> | <https://t.me/alexfeel>

AI Product Manager and hands-on LLM engineer — RAG, agent orchestration, and the Python behind them

EXPERIENCE

AI Product Manager

2025-02 - present

Spotware Systems – Limassol, Cyprus – Full-time

- Built the production AI stack hands-on: an async FastAPI retrieval service with LangGraph pipelines, provider-fallback embeddings and cross-encoder reranking, FastMCP tool servers with per-request token authentication, the pytest suites covering them, the CI that gates every tagged release, and the Helm charts running both services on Kubernetes with autoscaling.
- Led company-wide rollout of corporate coding assistants across engineering, QA, DevOps and analyst roles, reaching steady-state adoption across dozens of active users and standardizing AI in the software development lifecycle.
- Extended corporate ChatGPT Business into a grounded assistant by connecting internal knowledge sources through the retrieval backbone and MCP servers, so non-technical teams get source-linked answers from internal documentation without leaving the tool they already use.
- Screened and interviewed 100+ candidates for AI-engineering roles, designing role-specific practical tasks and review criteria that tie hiring decisions to real production-delivery ability.

Business Analyst and Scrum Master

2024-04 - 2025-02

Spotware Systems – Limassol, Cyprus – Full-time

- Stabilized team velocity from the fourth sprint onward and held sprint planning accuracy within standard deviation, making delivery predictable enough to plan against.
- Introduced the full Scrum event cycle -- sprint planning, backlog refinement and retrospectives -- along with Planning Poker estimation, and started tracking velocity, sprint predictability and planning accuracy.
- Turned incomplete and ambiguous business requests into actionable specifications, including BPMN diagrams and interface mockups that gave stakeholders and developers a single shared reading.
- Built a backlog of refined, estimated requirements several iterations deep, giving management clearer timelines and development roadmaps.

Project Manager

2023-02 - 2023-09

Spalvalo – Limassol, Cyprus – Full-time

- Provided end-to-end agile project management and technical audits for funded startups, covering development, security and infrastructure.
- Developed a standard for IT project audits defining their types, goals, tasks and the resources each requires.
- Proposed and guided a machine learning model that improved a microfinance provider's risk-forecasting accuracy, giving leadership better data for lending strategy.
- Advised startups on in-house development and resource use, producing significant cost savings and more focused strategies.

Head of Technology

2018-09 - 2023-02

Amega – Limassol, Cyprus – Full-time

- Directed a 10-person technology function spanning backend, frontend, UI design and QA plus outsourced DevOps and system administration, and restructured the development lifecycle into predictable, business-aligned delivery.
- Implemented and operationalized MetaTrader 4 alongside MetaTrader 5, increasing user engagement, trading volume and revenue.
- Led development of a public API for partner integrations in the commission transfer system, attracting new major partners to the ecosystem.
- Managed a rapid migration to FXBackOffice CRM with a custom Sumsb integration, automating KYC and streamlining new client onboarding.

SKILLS

Artificial Intelligence · Model Context Protocol · Python · Claude Code · AI Product Management · Software Development · Kubernetes · DevOps · System Analysis · Business Analysis · Stakeholder Management · REST API · Git · Project Management · Docker · Agile Methodologies · Scrum · Team Leadership · SQL

SELECTED PROJECTS

MCP Context Server

2025-09 - present

A high-performance Model Context Protocol server that gives LLM agents a durable, searchable memory layer beyond the context window. Thread-scoped storage over SQLite or PostgreSQL, with semantic, full-text, and hybrid search, cross-encoder reranking, pluggable embedding and summary providers, metadata filtering, and batch operations — all behind a standard MCP interface.

<https://github.com/alex-feel/mcp-context-server>

Claude Code Toolbox

2025-08 - present

A cross-platform installer and declarative environment-configuration framework for Claude Code. One YAML file captures all agents, MCP servers, slash commands, hooks, skills, rules, system prompts, and settings; one command installs everything on Windows, macOS, or Linux — portable, reproducible, centrally updatable.

<https://github.com/alex-feel/claude-code-toolbox>

CXR Draft Auditor

2026-06 - present

A research and educational quality-assurance tool that gives a chest X-ray draft impression a transparent second read. Two tiny 4B models — a fine-tuned MedGemma that grounds the image into labeled findings with bounding boxes and an NVIDIA Nemotron-3 Nano 4B that parses the draft into the same labels — feed a deterministic, model-free comparator that flags MISSING, UNSUPPORTED, and URGENT discrepancies, with every flag traced to a specific finding, a specific phrase, and a box on the image. Not a medical device and not a diagnostic tool.

st-copy — Streamlit Copy-to-Clipboard Button Component

2025-05 - present

A tiny, theme-aware Streamlit component that adds a one-click copy-to-clipboard button to any Streamlit app. Distributed as the `st-copy` PyPI package; used in AI chat UIs and prompt-output surfaces where a clean, accessible copy button was missing from the Streamlit toolbox.

<https://github.com/alex-feel/st-copy>

Hugo Artifacts

2026-05 - present

A public, MIT-licensed monorepo of reusable Hugo modules — style-agnostic content shortcodes alongside site-scale modules for the SEO head, client-side search, responsive images, an accessible carousel, social sharing, the machine-readable layer an AI agent reads, and the Progressive Web App shell — each independently versioned with its own go.mod. Every artifact ships semantic BEM markup, data attributes, and zero CSS, so it drops into any number of unrelated sites and is styled entirely by the consumer.

<https://github.com/alex-feel/hugo-artifacts>

Claude Code Artifacts

2025-09 - present

A public, MIT-licensed library of reusable Claude Code artifacts — environment configurations, skills, rules, hooks, agents, and slash commands. Each artifact is a self-contained, environment-agnostic building block addressable by its raw URL: install one by referencing it from a single environment YAML that Claude Code Toolbox sets up for you, or copy it straight into your ~/.claude/ directory.

<https://github.com/alex-feel/claude-code-artifacts-public>

CERTIFICATIONS

Build Small Hackathon 2026

2026-06

Hugging Face

Professional Scrum Master I (PSM I)

2023-02

Scrum.org

Develop Your Skills as a Software Project Manager

2023-12

LinkedIn Learning

Atlassian Agile Project Management Professional Certificate

2024-01

LinkedIn Learning and Atlassian

Getting Started as an Agile Project Manager

2023-08

LinkedIn Learning

EnglishScore B2 Level Certification

2023-09

British Council

Google Analytics Certification (GA4)

2024-03

Google

Google Analytics Individual Qualification (Universal Analytics)

2023-09

Google

Career Essentials in Generative AI by Microsoft and LinkedIn

2023-08

LinkedIn Learning and Microsoft

EDUCATION

Executive Education Program, Corporate Management

2009-10 - 2009-11

Bryansk Institute of Management and Business – Bryansk, Russia

Bachelor's Degree, Economics and Enterprise Management

1998-09 - 2002-06

Belarusian Trade and Economics University of Consumer Cooperatives (BTEU) – Gomel, Belarus

LANGUAGES

English (B2) · Russian (native)