

# SOFIA PETRUSENKO

Cascais, Portugal | +44 773 771 2377 | sofia.petrusenko988@gmail.com  
sofiapetrusenko.dev | github.com/sofiapetrusenko | linkedin.com/in/sofiapetrusenko

## PROFESSIONAL SUMMARY

---

Engineer building production LLM and data systems, with a biomedical research background in ageing biology. Ships end-to-end: FastAPI / PostgreSQL / Redis services, Next.js front ends, LLM orchestration with evaluation-first design — deployed and operating in production since 2025, including work in a regulated domain where compliance is enforced by tests rather than documentation. Research training across the UK, Portugal, Germany and Ukraine (autophagy, proteostasis, longevity). Target: AI engineering in healthtech, longevity and scientific tooling, where measurement integrity matters as much as shipping speed.

## SELECTED PROJECTS

---

**blotquant — QC-first western blot densitometry** | 2026 – v1.0 in progress

[github.com/sofiapetrusenko/blotquant](https://github.com/sofiapetrusenko/blotquant) · case study: [sofiapetrusenko.dev/work/blotquant](https://sofiapetrusenko.dev/work/blotquant)

- Python image-analysis tool built on the principle that a bad measurement is a refusal, not a warning: three-state QC model (pass / flagged / blocked); blocked lanes produce no numeric output and invalidate any ratio computed from them.
- FastAPI service (analyze / results endpoints), OpenCV + scikit-image pipeline, pytest suite and CI; independent validation against Fiji pre-registered before any data was seen.

**lifespan-extract — structured longevity data from biomedical literature** | 2026 – in development

[github.com/sofiapetrusenko/lifespan-extract](https://github.com/sofiapetrusenko/lifespan-extract) · case study: [sofiapetrusenko.dev/work/lifespan-extract](https://sofiapetrusenko.dev/work/lifespan-extract)

- LLM extraction pipeline measured against a hand-labeled gold set built before the pipeline existed; per-field evaluation metrics and quote verification of every extracted value against source text.
- Engineered as agentic loops (implementer → reviewer subagents) with human-gated eval design and gold-set labeling — the model is never allowed to grade its own homework.

## TECHNOLOGY & COMMERCIAL EXPERIENCE

---

**Founder & Engineer — BetCompass, odds and margin transparency platform** | Independent venture | 2026 – in development

- Designed and built a data platform that measures the operator margin embedded in published odds on the Portuguese regulated market (SRIJ-licensed operators) and compares prices across them — a transparency and cost-of-play tool, not a prediction product.
- Asynchronous collection pipeline in Python (httpx, asyncpg, APScheduler, PostgreSQL, Docker Compose): budget-aware scheduling that tightens sampling as an event approaches, idempotent collectors, and append-only price history — every poll writes a new row, prices are never overwritten.
- Entity resolution across independent data sources with tiered confidence (rapidfuzz; auto-accept  $\geq 92$ , human review queue 75–92, reject below 75), anchored to a canonical fixture calendar and a start-time window, because a single silent mismatch would corrupt downstream measurement undetectably.
- Compliance implemented as executable tests rather than policy documents: a runtime render guard rejects any unlicensed operator name reaching rendered HTML (the name arrives from the database, so source-code scanning cannot catch it), plus forbidden-language checks (word-boundary, diacritic-normalised) asserted against rendered output, any exception requiring a justified allowlist entry.
- Margin removal implemented three ways (multiplicative, Shin, power) and logged side by side rather than chosen in advance; measurement across 114 outcomes found no positive edge on main 1X2 lines (maximum  $-0.78\%$ ) — recorded and used to narrow product scope rather than quietly dropped.
- Invite-only access system (atomic invite exhaustion under concurrency, rate limiting, CSRF, 18+ gate), price dashboard with staleness detection and flagged-price suppression, and a Telegram alert bot; pt-PT localisation built as a two-stage refactor, extracting strings with byte-identical render output before translating.

**Founder & Engineer — AI Content Automation Platform** | Independent venture | 2025 – present (operating)

- Designed, built and operate a full-stack platform for automated media production and distribution: LLM-driven generation, text-to-speech, video assembly, publishing via the YouTube Data API, analytics ingestion and monitoring.
- Architected a seven-stage processing pipeline (script → voiceover → asset sourcing → assembly → subtitling → metadata → thumbnail) orchestrated through asynchronous background workers.
- Stack: Python / FastAPI, PostgreSQL, Redis with RQ workers, Next.js front end, Docker, deployed on DigitalOcean; integrated third-party APIs for LLM inference, speech synthesis and stock media.
- Runs a multi-channel network with human-in-the-loop editorial control, including a longevity & healthspan channel drawing on biomedical subject-matter expertise; extended to multi-language output (Spanish) from the same pipeline.

- Sanitized platform code (architecture, pipeline, tests) open-sourced at [github.com/sofiapetrusenko/media-automation-platform](https://github.com/sofiapetrusenko/media-automation-platform).

**AI Engineer — Conversational Agent with Persistent Long-Term Memory** | Independent project | 09/2025 – 12/2025

- Built a reflective journaling-style conversational agent on an Obsidian Markdown knowledge base, with durable memory across sessions rather than single-conversation context.
- Developed the memory layer end-to-end: structured note schema, automated extraction of salient information, and retrieval of relevant prior context at inference time.

**Paid Acquisition & Performance Marketing** | Independent / contract | 09/2025 – 2026

- Planned and optimised Google Ads and Meta campaigns against defined CPA and return targets; owned tracking, attribution and structured creative experiments interpreted statistically.

## RESEARCH EXPERIENCE

---

**Research Intern — Dr. Adrien Rousseau Lab, MRC PPU, School of Life Sciences** | Dundee, UK | 09/2024 – 01/2025

- Investigated stress-induced protein phosphorylation in *S. cerevisiae* and its relationship to longevity; analysed experimental datasets in R and Python to identify patterns linked to yeast lifespan.
- Collaborated on experimental design and interpretation; contributed to drafting a manuscript on implications for ageing mechanisms.

**Research Intern — Sandra Tenreiro & Paula Macedo Labs, NOVA Medical School** | Lisbon, Portugal | 06/2023 – 08/2023

- Studied molecular mechanisms of ageing relevant to neurodegeneration and retinal disease with a focus on proteostasis, in models from yeast to human cells, brain and retinal organoids.
- Supported therapeutic research for Parkinson's disease, diabetic retinopathy and AMD in collaboration with clinicians.

**Research Technician — Dr. James Murray Lab, Institute of Life Sciences** | UK | 05/2022 – 09/2022

- Investigated autophagy signalling and its alteration during ageing and in disease contexts including cancer, neurodegeneration and diabetes; contributed to work on autophagy in immunity and SARS-CoV-2-related infection.

**Research Intern — Prof. Dr. Iain D. Couzin, Max Planck Institute of Animal Behaviour** | Konstanz, Germany | 03/2022 – 05/2022

- Researched collective animal behaviour and social influence across biological systems, combining experimental and theoretical approaches.

**Laboratory Assistant — Anatoliy I. Dragan Lab** | Kyiv, Ukraine | 02/2021 – 09/2021

- Supported an experimental study of DNA hydration using Pressure Perturbation Scanning Microcalorimetry.

## CORE SKILLS

---

**Software Engineering & AI:** Python, FastAPI, TypeScript / Next.js, PostgreSQL, Redis / RQ, REST APIs, LLM orchestration and prompt engineering, evaluation design, RAG and long-term memory, agentic workflows (Claude Code), entity resolution, compliance and regression testing, Docker, Git, CI, Vercel / DigitalOcean deployment

**Data & Bioinformatics:** R, Python (pandas, NumPy), statistical analysis and visualisation, OpenCV / scikit-image, BLAST, UniProt, ClustalX, Jalview

**Molecular & Cell Biology:** qPCR, western blotting, ELISA, immunoprecipitation, mammalian and bacterial cell culture, DNA/protein extraction and cloning workflows; confocal, light and electron microscopy

**Growth & Performance Marketing:** Google Ads and Meta campaign architecture, conversion tracking and attribution, funnel and cohort analytics

**Languages:** English (fluent), Ukrainian (native), Russian (native), Portuguese (learning)

## EDUCATION

---

**BSc (Hons) Biomedical Sciences — University of Dundee** | Scotland, UK | 09/2022 – 05/2025

- Placement and project work focused on ageing biology, autophagy, proteostasis and longevity mechanisms.

**BSc Biology — Taras Shevchenko National University of Kyiv, Institute of High Technologies** | Kyiv, Ukraine | 09/2020 – 04/2022

## ADDITIONAL

---

- Training: ICH E6 Good Clinical Practice (GCP) Investigator Site Training.
- Comfortable working autonomously and shipping end-to-end — from specification through build to live operation.