

DAVID HORVATH

AI Engineer | LLM, Agentic & Computer Vision Systems | End-to-End Delivery

horvath.david35@gmail.com | +36 30 731 8619 | Vienna, Austria

nexsonum.com | linkedin.com/in/nexsonum | github.com/Nexsonum | hirspektrum.hu



PROFILE

AI engineer with 5 years of professional experience, 4+ of them building AI-powered products end-to-end: LLM/RAG and agent-based systems, custom computer vision models, and the distributed backends and cloud pipelines that carry them to production. Delivered forward-deployed AI projects at Swiss Re and production CV, VLM, and LLM pipelines on AWS at Verizon; currently in a software design role as the sole engineer at a hardware startup, architecting the full stack and deploying containerized AI inference services. Comfortable with high autonomy, architecture ownership, and direct stakeholder work; previously served a year as Scrum Master alongside development.

WORK EXPERIENCE

Software Architect & Sole Engineer (Full Stack) | Biomotion Technologies *02/2026 – Present*

- Designed and own the **entire software architecture** of a shipping product: React/TypeScript frontend, C# backend, and containerized Python microservices. Made every design decision from API contracts to release process.
- Deploy **specialized computer vision models as containerized inference services** in production, owning the full model lifecycle: selection, training, deployment, and monitoring in the field.
- Architected a **distributed multi-node system** (Linux hosts, GPU inference node, real-time controllers) with clean service boundaries, inter-node communication, and hardware abstracted away at the application layer.
- Built the project's **CI/CD and versioned release pipeline from zero**, including an offline update mechanism for air-gapped deployments.
- Act as the technical counterpart to the founders: translate product requirements into architecture decisions, roadmaps, and trade-offs.

Data Specialist (AI / Data Engineering) | Swiss Re *09/2025 – 02/2026*

- Worked forward-deployed on customer projects: embedded with business units to scope, build, and ship **LLM-powered tools for investment analysis**, including **agent-based functionality** for complex financial workflows.
- Engineered large-scale data processing and **entity extraction pipelines in Palantir Foundry**, focusing on robustness and throughput for financial data.

Software Developer | Verizon *04/2022 – 08/2025*

- Built a **VLM-powered security footage analysis system on AWS** that classified events in camera streams and **triggered automated downstream workflows** for anomaly response and user alerting.
- Trained and fine-tuned **custom computer vision models (YOLO, PyTorch)** and served them from scalable, real-time video processing pipelines on AWS.
- Developed **after-the-fact video analysis pipelines** for dashcam and IoT telemetry data, integrated into a fleet management solution for driver safety analytics, ingestion, processing, and reporting on AWS.
- Designed and built a full-stack **RAG chatbot** as one of the team's first production LLM projects.
- Served one year as **Scrum Master** in parallel with development work, running ceremonies and coordinating delivery across the team.

Junior Software Developer | British American Tobacco *2021 – 2022*

- Built **internal automation tools in Python and Google Apps Script** for worker management and reporting, alongside project coordination.

SELECTED PROJECT | HIRSPEKTRUM.HU

Founder & Developer | Self-hosted Hungarian media analysis platform *2025 – Present*

- Designed and operate the **full architecture of a production AI system**: RabbitMQ event backbone, MongoDB, PostgreSQL, Neo4j, caching layers, and independent Python services, running self-hosted.
- Built **multi-stage LLM pipelines** that turn the raw Hungarian news feed into automated comparative media analysis: scraping and ingestion, **dynamic unsupervised clustering**, and orchestrated LLM workflows over a **Neo4j knowledge graph**.
- Operate **continuous, scheduled AI workloads in production**: pipeline orchestration, failure handling, and cost- and latency-aware LLM usage at self-hosted scale.

SKILLS

AI / ML:	LLMs, RAG, agent-based systems, LLM pipeline orchestration, PyTorch, TensorFlow, YOLO & custom CV model training, VLMs, NLP, NumPy, Pandas
Programming:	Python (primary), TypeScript/React/NextJS, C#, C/C++ (working), SQL, Bash, FastAPI, Pytest
Architecture:	Distributed systems, event-driven design (RabbitMQ), microservices, concurrent & parallel programming, REST APIs, big data pipelines
Data & Infra:	AWS, Palantir Foundry, Docker, Kubernetes, CI/CD, MongoDB, PostgreSQL, Neo4j, Git, Linux
Devices:	NVIDIA Jetson (GPU inference), Raspberry Pi, IoT & smart-home integrations

ACHIEVEMENTS

- **Bosch Thesis Project Award (07/2024)** - 1st prize for the MSc thesis “Automated Pollen Detection with Synthetic Data Generation”, combining object detection with synthetic training data.

EDUCATION

MSc Computer Science Engineer | University of Pécs07/2024 · *Grade: A*

- Specialization: Parallel Computing and Machine Learning

BSc Mechatronics Engineer | Óbuda University01/2021 · *Grade: A*

- Specialization: Complex Systems

LANGUAGES

English (professional) · Hungarian (native)

