

Alexey Zaytsev

AI/ML Engineer & Systems Architect

Full-stack ML engineer specializing in generative AI, distributed inference, and model fine-tuning. Combines first-principles deep learning with low-level systems and hardware engineering.

<https://alexey.work/cv>

alexey.zaytsev@gmail.com

<https://github.com/xl0>

<https://www.linkedin.com/in/alexeyzaytsev>

// FEATURED PROJECTS

lovely-tensors 1.4k ★ / **numpy** / **jax** Python, PyTorch, NumPy, JAX
Tensors for human consumption. Widely adopted by AI researchers.

tidygrad Python, NumPy, Autograd
From-scratch autograd engine, enough to train a tiny GPT-2 style model.

congusto-chat SvelteKit, TypeScript, PostgreSQL, Vercel AI SDK
Open-source multi-provider chat interface, similar to OpenWebUI and t3chat.

lovely-docs SvelteKit, TypeScript, Python, LLM, MCP
AI-native documentation dehydrator for agentic access. CLI, web UI, and MCP server.

pelican SvelteKit, TypeScript, PostgreSQL, Vercel AI SDK
LLM-driven SVG/ASCII art generation gallery with iterative visual refinement.

latent-tools Python, PyTorch, ComfyUI, Diffusion Models
Custom ComfyUI nodes for fine-grained diffusion latent manipulation.

nvml-tool C, NVML, CUDA, Linux
GPU power and fan control utility built on NVIDIA NVML.

// EXPERIENCE

AI/ML Engineer (Consulting)

July 2024 - Present

Select client engagements:

- AI Storyboarding Platform (Python backend): Integrated new image generation providers and refactored the generation abstraction layer across use cases (assets, scenes, inpainting).
- Earnings Call Analysis & CEO Training: Engineered a low-latency transcription + LLM pipeline (AWS Transcribe / Deepgram) with speaker diarization and financial metrics extraction. Built a companion CEO training module that simulates analyst Q&A via TTS, transcribes spoken responses, and scores both content quality and delivery (WPM, filler words).
- Open-source AI projects: **congusto-chat**, **lovely-docs**, **pelican**, **latent-tools**, **nvml-tool**.

Python · LLM Pipelines · AWS Transcribe · Deepgram · Speaker Diarization · TTS · Latency Optimization · Applied AI

Generative AI Lead

Nov 2023 - July 2024

io.net

Led development of the generative AI platform supporting image, video, and 3D model generation. Served tens of thousands of users, generating hundreds of thousands of images:

- Architecture & Inference: Built an end-to-end generation pipeline. The frontend talked to the backend that talked to Supabase/PostgreSQL, which queued tasks via RabbitMQ. Distributed GPU workers picked up the jobs, generated the media, uploaded to S3 (served via CloudFront+imgproxy), and pinged a backend webhook, which finally streamed updates to the client via Server-Sent Events.

// RELEVANT SKILLS

Core ML

Python, PyTorch, JAX, CUDA, Transformers, Diffusion Models, Computer Vision, Autograd, CLIP, Embeddings, Weights & Biases

Applied AI

LLM Agents, vLLM, Hugging Face, LangChain, RAG, Fine-tuning, LoRA, Stable Diffusion, Vector Databases, Vercel AI SDK, MCP

Full-Stack & Platform

SvelteKit, TypeScript, FastAPI, PostgreSQL, Tailwind CSS, SQLAlchemy, Drizzle ORM, Docker, Linux, AWS, GitHub Actions, Supabase, Vercel, CloudFlare, S3

// EDUCATION

Université Denis Diderot

2013 - 2015

M.Sc. Systems and Synthetic Biology

LETI University

2003 - 2009

Specialist, Computer Security

// LANGUAGES

English	Professional
Russian	Native
Ukrainian	Elementary

- Model Fine-tuning: Extensive work with Stable Diffusion inference, custom model fine-tuning, and LoRAs to optimize generation quality and capabilities.
- Custom NSFW Detector: Trained and deployed an EfficientNet classifier to replace off-the-shelf filters with high false-positive rates, reaching production-grade precision directly in the distributed GPU worker pipeline.
- MLOps & DevOps: Orchestrated and scaled the distributed inference infrastructure.

Stable Diffusion · PyTorch · RabbitMQ · PostgreSQL · Supabase · LoRA · EfficientNet · MLOps · SSE · S3

AI/ML Engineer (Consulting)

Nov 2022 - Nov 2023

Independent consulting focused on AI applications and prototypes. Select engagements:

- Dog Finder: An experimental missing-dog app exploring both facial recognition and noseprint uniqueness. Trained EfficientNet with triplet loss.
- SAT Practice Generator: A dynamically updating test prep system seeded by real data. Built with Streamlit, Airtable, and OpenAI LLMs-migrating from GPT-3.5 to GPT-4 largely solved our initial correctness and consistency struggles.
- GPTEditor: An AI-assisted copywriting tool-like Cursor, but for prose. Users highlight a paragraph and instruct the LLM on how to rewrite it. Built with SvelteKit, Tiptap, and Supabase.

SvelteKit · OpenAI API · RAG · Supabase · Streamlit · TypeScript · Vector DBs · PyTorch · Computer Vision

Electronics Engineer

Dec 2018 - May 2020

Okra Solar - Canberra / Phnom Penh / Shenzhen

Smart solar grid tech: designed the main grid hub (Solar → Battery → regulated DC loads and high voltage sharing), a debugging adapter, and auxiliary internal boards. Schematics, PCB design, DC-DC converters and analog monitoring circuits. Led a small hardware team (3 engineers).

Analog Electronics · Power Electronics · Maximum power point tracking · KiCad

Electronics and Prototyping Engineer

Sept 2016 - Dec 2018

Self-employed - Shenzhen, China

Dozens of client projects across digital, analog, and power electronics - from initial concept to production-ready designs and manufacturing oversight. Clients in medical devices, scientific instruments, food and consumer electronics. Designed and set up end-of-line automated testing, trained factory workers, and oversaw manufacturing in Shenzhen factories.

Digital Electronics · NFC · E-Ink · Bluetooth Low Energy · Sub-1G radio · Embedded systems · Power electronics · Motor controller · Synchronous rectification · DC-DC converters · Buck · Boost · LED display · Analog Electronics · Differential amplifiers · EKG · Electrostimulation · EEG · Electrowetting · Digital microfluidics · Autodesk Inventor · Fusion 360 · KiCad · Altium Designer

CTO & Co-founder

Sept 2015 - Sept 2016

Koi.science - Paris / Shenzhen

EdTech startup providing affordable laboratory equipment for molecular biology education in schools. Designed spectrophotometer, electrophoresis and transilluminator - hardware, electronics, mechanical enclosures, and embedded software. Co-founded with fellow students from the CRI Paris program.

Electronics Engineering · Mechanical Engineering · Microcontrollers · Embedded programming · Autodesk Inventor · Fusion 360 · KiCad

Research Intern

May 2015 - Sept 2015

Centre de Recherches Interdisciplinaires (CRI Paris) - Paris

Developed the first iteration of a low-cost, open-source spectrophotometer - the project that later evolved into the Koi.science startup.

Electronics Engineering · Mechanical Engineering · Embedded programming · Spectrophotometry · KiCad · Education outreach · Autodesk Inventor

Research Intern

May 2014 - Jan 2015

Institut Pasteur - Paris

Characterizing bacteriophage origin of replication region, exploring the potential of applying genetically engineered λ -phages to antibiotic-resistant bacterial infections in burn victims. Contributed to a published paper.

Bacteriophages · λ -phage · Staphylococcus Aureus · PCR · Plasmids · Molecular Biology · Electrophoresis · CRISPR-Cas9 · Bioinformatics · BioPython · High-throughput sequencing · Gibson Assembly

Research Intern

Jan 2014 - April 2014

Institut Curie - Paris

Researching the influence of electric fields on mammalian cell mobility and wound healing.

Microfluidics · Protein micropatterning · PDMS · Electrophoresis · Light Microscopy · Mammalian cell culture · Electronics prototyping

Sr. Software Engineer

June 2011 - July 2012

Nexenta Systems

Virtio drivers for network and block devices, general kernel maintenance, bugfixes and development on illumos (OpenSolaris fork).

Device drivers · Virtio · ZFS · Networking · Block storage · Solaris · C

Linux Kernel Engineer-Programmer

Jan 2005 - Jan 2010

AltEII and Protei ltd Saint-Petersburg, Russia

Linux kernel development for high-bandwidth telecom equipment. AltEII: 10 Gbps network routers. Protei: telephony over dedicated lines, WiMAX and GPRS/UMTS base stations. Device drivers, bootloaders, and board support packages on MIPS64, ARM and x86.

Device drivers · Bootloaders · Networking · ASIC · Board Support Packages · Linux Kernel · C · Assembly · ARM · MIPS · Intel XScale