# Oleksiy Zakharov

## AI & Machine Learning Engineer

✉ oleksiyzakharov1004@gmail.com     📞 +372 (582)056-52     📍 44307, Rakvere, Estonia

## Profile

Senior AI Engineer with 9+ years of experience leading the development and deployment of scalable machine learning and generative AI solutions. Expertise in fine-tuning large language models (LLMs), prompt engineering, and cloud-native AI systems on GCP, AWS, and Azure. Skilled in MLOps, API architecture, vector search, and implementing ethical, privacy-focused AI practices. Proven leader and collaborator who drives innovation and delivers impactful AI products across diverse industries.

## Professional Experience

**Humane**                                                                               10/2023 – 2025
*Senior AI Engineer*                                                              San Francisco, USA

Leading cutting-edge projects focused on the development and deployment of innovative AI solutions
• Led development of edge AI applications powering the AI Pin wearable device.
• Designed and optimized on-device models, achieving 20% faster inference times using TensorFlow Lite and
ONNX quantization techniques.
• Collaborated cross-functionally to integrate voice and NLP systems on limited hardware, improving real-time
responsiveness by 15%.
• Mentored 3 junior engineers, guiding AI system design and deployment best practices in a fast-paced startup
environment.
• Drove innovation in privacy-preserving AI on edge devices, aligning with GDPR and AI ethics standards.

**Taxy.io**

*Senior AI Engineer*                                                              09/2022 – 03/2023
                                                                                              Aachen, Germany

Designed and Developed several AI systems including Forecasting, Data Analytics, Anomaly Detection focusing on LLM
• Integrated cloud AI services on GCP (VertexAI, BigQuery) with backend systems, reducing query latency by 30%.
• Led CI/CD implementation for AI model retraining and deployment, shortening update cycles from 2 weeks to 3 days.

*AI Developer*                                                                      03/2020 – 08/2022
                                                                                              Aachen, Gemany

• Developed LLM-based tax query assistant using GPT-3/GPT-4 and custom fine-tuning, improving answer
accuracy by 25%.
• Worked closely with data sacientists and tax experts to translate regulatory requirements into AI system features
• Built prompt engineering pipelines and automated workflows to handle complex multi-turn conversations.

**Veriff**                                                    05/2017 – 02/2020
*Machine Learning Developer*                                   Tallinn, Estonia

Developed AI-powered identity verification platform focusing on computer vision technology.
• Developed face recognition and anti-fraud AI modules using OpenCV and TensorFlow.
• Improved model accuracy by 15% via transfer learning and data augmentation techniques.
• Collaborated with engineering teams to deploy AI services on AWS infrastructure.
• Supported QA and monitored live system metrics to rapidly troubleshoot and reduce false positives by 10%.

**Feelingstream**                                             09/2015 – 04/2017
*Data Scientist Intern*                                        Tallinn, Estonia

Contributed to analyzing customer interactions, developing ML models.
• Designed NLP models for customer conversation analytics, improving entity recognition F1 score by 18%.
• Developed recommender system features using Scikit-learn and PyTorch to personalize customer insights.
• Automated data pipelines and dashboards to monitor AI system performance and user engagement metrics.
• Contributed to cross-team workshops to align AI output with business goals and customer satisfaction.

## Skills

**Programming Languages**
Python, C++, Java, Javascript, Rust

**AI Agents**
SmolAgents, LlamaIndex, LangGraph, n8n

**Cloud Platforms**
AWS, Google Cloud Platform (Vertex, BigQuery), AWS, Microsoft Azure

**APIs & Integrations**
REST, GraphQL, OAuth2, JWT, LLM API integration (OpenAI, Cohere)

**Leadership & Collaboration**
Team leadership, mentoring, cross-functional teamwork, communication skills

**Machine Learning Frameworks**
TensorFlow, PyTorch, hugging Face Transformers, Scikit-learn

**Generative AI & LLM**
GPT-4, LLaMA 2, Claude 2, BERT, generative models(GANs,diffusion), LangChain

**DevOps & MLOps**
Terraform, Pulumi, MLflow, Docker, Kubernetes, CI/CD

**Databases**
PostgreSQL, MongoDB, VectorDB (Pinecone, Chroma)

**Edge AI**
TensorFlow Lite, ONNX, model optimization

## Education

**Bachelor of Computer Science**                              04/2011 – 05/2015
*Tallinn University of Technology*                            Tallinn, Estonia