

Sergii M.

Senior AI Engineer

SUMMARY

- 10+ years in AI engineering and data science, specializing in autonomous AI agents for cybersecurity, wealth management, and medical document processing.
- Expertise in Python, LangChain, Kubernetes, GCP (Vertex AI, BigQuery), MLflow, and Hugging Face for scalable ML/LLM workflows.
- Experience in reducing AI agent operational costs by 50× through architecture and prompt optimization.
- Led ML platform development for airfare pricing with real-time 24/7 support using TensorFlow, Dataflow, and Kubeflow.
- Skilled in Spec-Driven Development, agentic structured-data extraction, and multi-pass document processing with confidence scoring.

TECHNICAL SKILLS

Main Technical Skills	Python, Dataflow, Apache Beam, Kubeflow
AI Tools & Assistants	Claude
Programming Languages	Python, VBA
AI & Machine Learning	AI, Claude, Hugging Face, Keras, Kubeflow, LangChain, LangGraph, LLM, Mlflow, NLP, NumPy, OpenAI, PyTorch, TensorFlow, Vertex AI
Python Libraries and Tools	Keras, NumPy, PyTorch, TensorFlow
.NET Platform	VBA
Data Analysis and Visualization Technologies	Apache Beam, AWS Athena, Business Intelligence (BI), DataFrame library, ML, Polars (pola.rs, DataFrame library), Tableau
Databases & Management Systems / ORM	PostgreSQL
Cloud Platforms, Services & Computing	AWS, GCP
Amazon Web Services	AWS Athena, AWS S3
Google Cloud Platform	Dataflow, Google BigQuery
Industry Domain Experience	DSP
QA, Test Automation, Security	A/B Testing

Deployment, CI/CD & Administration	CI/CD
Virtualization, Containers and Orchestration	Docker, Kubernetes
Platforms	Magento 1
SDK / API and Integrations	MCP
Project Management & Administration	ROI, SMART
Methodologies, Paradigms and Patterns	UML
Other Technical Skills	Argo (2 yr.), AutoGen, Data Science, framework, MBA, Netflow, pyarrow, RShiny

WORK EXPERIENCE

AI Consultant, NDA

Duration: Feb. 2026 - Present

AI Consultant, Reliable agentic structured-data extraction from unstructured documents

Duration: Feb. 2026 - Present

Summary: Developed a reliable agentic system for structured-data extraction from unstructured documents, involving multiple extraction passes, merging results, flagging disagreements, and providing numeric confidence scores per field. The project serves clients in wealth management and insurance sectors under NDA.

Responsibilities:

- Designed and implemented multi-pass extraction workflows.
- Developed agentic steps for result merging and disagreement investigation.
- Delivered confidence scoring per extracted field.
- Managed client communications and project delivery under NDA.

Technologies: Python, Agentic workflows, LLMs.

AI Consultant, End-to-end MLOps for an LLM workflow extracting information from medical documents

Duration: Feb. 2026 - Present

Summary: Implemented an end-to-end MLOps pipeline for a large language model workflow that extracts information from medical documents, ensuring reliable and scalable processing for healthcare data extraction.



Responsibilities:

- Designed and deployed MLOps pipelines for LLM workflows.
- Ensured data extraction accuracy and workflow scalability.
- Collaborated with client Wisedocs for project requirements and delivery.

Technologies: Python, MLOps, LLMs.

AI Consultant, Training engineering team on Spec-Driven Development for AI-assisted coding

Duration: Feb. 2026 - Present

Summary: Provided training on Spec-Driven Development framework for AI-assisted coding to an engineering team at an outstaffing company, enhancing their development processes and AI integration capabilities.

Responsibilities:

- Developed and delivered training materials and sessions.
- Introduced Spec-Driven Development framework.
- Supported team adoption of AI-assisted coding practices.

Technologies: Spec-Driven Development, AI-assisted coding

AI Consultant, Device-type classification from netflow features using LLM agents

Duration: Feb. 2026 - Present

Summary: Built predictive models and LLM-based solutions for device-type classification from netflow features, analyzing behavioral patterns and static information for client Elisity.

Responsibilities:

- Developed ML models for hierarchical label prediction.
- Implemented LLM solutions for static information analysis.
- Analyzed behavioral patterns from netflow data.

Technologies: Python, LLM agents, Machine Learning

Founding AI Engineer, Maze (Autonomous AI/LLM agents for cybersecurity vulnerability investigation)

Duration: Jan. 2025 - Mar. 2026

Summary: Designed and built fully autonomous AI and LLM agents to investigate cybersecurity vulnerabilities, focusing on improving agent stability, accuracy, and reducing running costs significantly.

Responsibilities:

- Designed agent architecture and workflows for vulnerability investigation.
- Optimized prompts, context, and tools to enhance agent performance.
- Improved agent stability and accuracy.

Technologies: Python, AI agents, LLMs, Cybersecurity.



AI Engineer, Elisity

Duration: May 2024 - Jan. 2025

AI Engineer, Device type classification on netflow features using AI/LLM agents

Duration: Oct. 2024 - Jan. 2025

Summary: Led AI research and developed device type classification models using AI and LLM agents on netflow features to support cybersecurity initiatives.

Responsibilities:

- Developed AI/LLM models for device classification.
- Led AI research channel reviewing advancements in AI.

Technologies: Python, AI, LLM agents.

Data Scientist, Cybersecurity DS/ML solutions for device attribute prediction from netflows

Duration: May 2024 - Oct. 2024

Summary: Developed and deployed data science and machine learning solutions to predict device attributes from netflow data, including data collection, feature engineering, modeling, and inference.

Responsibilities:

- Collected and processed data using PostgreSQL, S3, and Athena.
- Engineered features and developed ML and LLM models.
- Deployed inference pipelines for real-time predictions.

Technologies: PostgreSQL, S3, Athena, Python, ML, LLM.

Machine Learning Engineer/Technical Lead, FLYR Labs

Duration: Aug. 2020 - May 2024

Technical Lead, Machine learning platform for fares optimisation

Duration: Sep. 2022 - May 2024

Summary: Led development of a machine learning platform for airfare pricing optimization, building expandable components for feature extraction, inference, and post-processing to support 24/7 operations for major airlines.

Responsibilities:

- Led technical team and platform development.
- Developed ML components for feature extraction and inference.
- Ensured platform scalability and reliability.

Technologies: Python, ML, Airfare pricing.



Machine Learning Engineer, Machine learning pipelines for airfares optimization

Duration: Aug. 2020 - Sep. 2022

Summary: Developed machine learning pipelines for airfare optimization including feature calculation, prediction extraction, and post-processing using cloud and container orchestration technologies.

Responsibilities:

- Calculated features for AI models using BigQuery and Kubernetes.
- Extracted predictions with TensorFlow, PyArrow, Dataflow, Beam, and ARGO.
- Performed predictions post-processing with numpy, polars, Beam, and Dataflow.

Technologies: Python, TensorFlow, BigQuery, Kubernetes, Dataflow, Beam, ARGO, ML pipelines.

Computer Vision Data Scientist/Software Engineer, Lily AI

Duration: Apr. 2017 - August 2020

Computer Vision Software Engineer, Engineering solutions for CV/NLP/ML models training and deployment

Duration: April 2019 - August 2020

Summary: Engineered solutions for training and deploying computer vision, NLP, and machine learning models, integrating MLflow into pipelines and deploying AI models on Kubernetes.

Responsibilities:

- Integrated MLflow into company pipelines.
- Deployed AI models on Kubernetes clusters.
- Developed data collection pipelines.

Technologies: Python, Kubernetes, MLflow, CV, NLP, ML.

Computer Vision Data Scientist, Training CV/NLP/ML models to improve e-commerce online sales rate

Duration: Apr. 2017 - Apr. 2019

Summary: Trained computer vision, NLP, and machine learning models to enhance e-commerce sales, developing in-house frameworks for mass model training using cloud platforms.

Responsibilities:

- Developed in-house frameworks for large-scale CV/NLP model training.
- Trained models using AWS, GCP, Python, PyTorch, and Keras.

Technologies: AWS, GCP, Python, PyTorch, Keras, CV, NLP, ML

Data Scientist, Predictive modeling and analytics for various clients

Duration: Mar. 2018 - Apr. 2019



Summary: Provided data science consulting to multiple clients including Medical AI Lab, dating app, programmatic advertisement, and travel agency, delivering predictive models and business insights.

Responsibilities:

- Developed automated scoring models for medical images.
- Delivered CV models for content moderation and fraud detection.
- Built NLP models for text semantic marker extraction.
- Provided business insights and analytics for advertising and travel clients.

Technologies: Python, CV, NLP, Predictive modeling.

Data Scientist, Codewise.com

Duration: May 2015 - Mar. 2018

Data Scientist, Prototyping optimization systems to increase click/conversion rate/price in programmatic advertising

Duration: Sep. 2016 - Mar. 2018

Summary: Developed prototype optimization systems to improve click and conversion rates and prices in programmatic advertising, including smart ECPA and A/B testing frameworks.

Responsibilities:

- Prototyped smart ECPA optimization systems.
- Implemented A/B testing to increase ROI.
- Developed RShiny BI tools.

Technologies: Python, RShiny, Programmatic advertising.

Data Scientist, Predictive and optimization projects in DSP RTB domain

Duration: May 2015 - Sep. 2016

Summary: Delivered predictive and optimization projects in the demand-side platform real-time bidding domain to reduce click and conversion costs for external buyers.

Responsibilities:

- Predicted click/conversion rates and bid prices.
- Optimized traffic distribution.
- Developed Tableau BI dashboards.

Technologies: Python, Tableau, DSP RTB.

Insights and Research Manager, Forex Club (Machine learning projects and VBA automatization)

Duration: Sep. 2014 - May 2015

Summary: Managed machine learning projects and automated processes using VBA to support research and insights generation.



Responsibilities:

- Managed ML projects.
- Automated workflows with VBA.

Technologies: Machine Learning, VBA.

Head of Strategic Planning, Ukrtelecom (Managing analytical department for strategic and operational decision making)

Duration: Aug. 2012 - Sept. 2014

Summary: Managed an analytical department of 20 employees focused on providing analytical support for strategic and operational decisions, business and customer insights, and reporting.

Responsibilities:

- Led analytical department and team management.
- Prepared analytical background for decision making.
- Provided business and customer insights and reporting.

Head of Business Analysis and Planning, Ukrainian High Technologies

Duration: June 2006 - August 2012

EDUCATION

- **Business school MBA**
2008 - 2010
- **Kiev National University of Construction and Architecture Master's degree, engineering**
2000 - 2005

CERTIFICATION

- **Deep Learning Summer School**
Genoa
2018
- **Statistical learning**
- **CKAD**

