



Ali Mosavian

SENIOR MACHINE LEARNING/AI
CONTRACTOR

LinkedIn ♦

GitHub ♦

Website ♦

20+

years programming

17

years in machine
learning

42+

markets, one model

ex-CTO

led five teams

PROFESSIONAL SUMMARY

Experienced **hands-on** engineer, machine learning/AI expert, generalist and ex-CTO with passion for finding efficient and unconventional ways of going from an **idea to a product** by being **pragmatic, focused, out-of-the-box thinking and above all moving fast**.

A life long learner with over two decades of programming experience, working in the IT industry since 2005 and working with machine learning/deep learning since 2009. The intersection where internet scale systems, machine learning/AI meet to make up unforgettable user experience is where I feel at most home.

EDUCATION

Computer Science (Masters), Royal Institute Of Technology, Stockholm

AUG 2005 - DEC 2009

Specialized in machine learning and autonomous systems

PROFESSIONAL EXPERIENCE

Senior AI Engineer (Contractor) at H&M Group, Stockholm

SEP 2025 - PRESENT

- Led the re-architecture and rewrite of a proof-of-concept **store capacity optimization engine** into a production system (81% of all changes), deciding which articles and pieces go on which fixture to fill each store optimally. Designed a domain-specific language for modelling store furniture that replaced thousands of lines of special-case code, on top of a two-stage constraint solver; grew the test suite from 5 to 100+ files.
- Designed and built a **shop-floor stock prediction system** (98% of all changes) covering ~3,350 stores without RFID: a gradient-boosted hurdle model on 161 engineered features trained on 74M rows from 686 RFID stores, fed by a custom data lineage and transformation pipeline processing 6.6 TB per training run.
- Re-engineered the **store-fullness data platform** in dbt (43% of all changes): introduced contract-enforced serving tables, retired legacy models and fixed a fundamental flaw in the published stock-ratio metric.

Python

OR-Tools CP-SAT

LightGBM

SHAP

scikit-learn

statsmodels

Numba

Pandas

PyArrow

DuckDB

BigQuery

dbt

SQLFluff

Lark

Pydantic

Typer

Docker

Terraform

Cloud Run

GCS

GitHub Actions

Pytest

MyPy/Ty

Ruff

Claude

Senior AI Engineer (Contractor) at Aloï AI, Stockholm

APR 2025 - JUL 2025

- Enhanced **retrieval-augmented generation (RAG)** pipelines for Aloï's AI-driven legal clause completion product, improving accuracy and relevance for contract drafting support.
- Developed advanced content extraction and chunking strategies for various legal documents (docx, pptx, pdf etc), including automated agreement-type classification and structure-preserving parsing.
- Built a specialized pipeline for investor presentations, leveraging visual element detection and asynchronous processing to rapidly locate relevant pages and extract key data.
- Significantly improved RAG retrieval quality using **re-ranking** and other retrieval optimization techniques, boosting downstream clause completion performance.

Azure OpenAI Azure Cohere Python FastAPI Docker MyPy Ruff PGVector
Postgres Pytest SQLAlchemy

Senior AI Engineer (Contractor) at Nordea AB, Financial Crime Prevention Unit, Stockholm

JAN 2025 - APR 2025

- Completely re-architected, re-wrote and productionalized an existing toy proof of concept tool (to a working production system) to **automate detection of financial crime** based on stated and actual transactional behavior.

Python MyPy Ruff Pytest Arize AI Phoenix LangGraph Chainlit
Anthropic Claude Docker AWS Hadoop

Senior AI Engineer (Contractor) at Byggbakta, Stockholm

JUN 2024 - JAN 2025

- Developed a **conversational AI** to streamline project searches, replacing a complex filter system with a more user-friendly and intuitive interface.
- Enabled users to search for projects with minimal input, reducing the cognitive load associated with navigating numerous filters and categories.
- Developed a **hybrid recommendations system** to give more relevant construction project recommendations to customers.
- Developed a (construction) role predictor for a projects.

Python Langchain Chainlit GPT-4o function calling
Catboost (gradient boosted decision trees) FastAPI DuckDB SQL Server Azure
Docker

Senior AI Engineer (Contractor) at EQT Partners, Stockholm

AUG 2023 - APR 2024

- Created a **multi-stage AI assistant** (RAG-based) to answer investor queries accurately by processing complex documents.
- Implemented a method to extract content from PDFs, Word, and PowerPoint files while maintaining the original structure, greatly improving document handling for AI systems.
- Designed and deployed an **automated evaluation framework** for tracking and measuring the assistant's accuracy over time.

Python Langchain FastAPI SlackBolt GPT-3.5 GPT-4
Pinecone (Vector database) Docker

Senior Machine Learning Engineer (Contractor) at Lingio, Stockholm

NOV 2022 - JUN 2024

- Developed AI-driven tools for translating, summarizing, and generating content, enhancing the efficiency of content creation processes.
- Fine-tuned **open-source language models** for specific tasks, improving performance and integration with the generative AI product.
- Delivered production-quality services, now embedded at the **core of Lingio's generative AI offerings**.

GPT-3

Python

FastAPI

Pytorch

LoRA

Docker

GCP Kubernetes

Senior Machine Learning Engineer at Kry/Livi, Stockholm

SEP 2021 - OCT 2022

- Designed and deployed a **CatBoost model** to classify the level of care required for patients, enhancing the triage process.
- Led the development of an **RNN-based model** to predict patient wait times, providing more accurate service level expectations.
- Re-architected staffing forecasts, optimizing workforce allocation and reducing operational inefficiencies.

Python

FastAPI

TensorFlow

Pytorch

RNN

CatBoost

AWS

Snowflake

Airflow 2

Senior Data/Machine Learning Engineer (Contractor) at IKEA Digital, Stockholm

JUL 2020 - MAY 2021

- Developed a recommendation model (contextual bandit-based) and deployed it for IKEA's inspirational feed. The model automatically trains every night on a few hundred million examples, and its operation cost for **42+ markets** is multiple factors lower than that of other teams using Google Recommendation API for a single market.
- Re-factored and split a monolithic backend into smaller services, enhancing maintainability and performance. Reduced average response time from ~800ms (99th percentile: 3-4s) to **~200ms** (99th percentile: ~500ms), leading to a significant improvement in user experience and reducing operational costs.
- Improved data iteration cycles **from hours to minutes**, enabling rapid experimentation and testing of different model architectures.

Python

FastAPI

TensorFlow

Docker

GCP

Google AI Platform

Senior Machine Learning/AI Engineer (Contractor) at Neko Health, Stockholm

JAN 2020 - JUL 2020

- Developed a cutting-edge, contact-less method for estimating **pulse and blood pressure** using video analysis, advancing real-time health monitoring systems.
- Integrated digital filters, signal processing, and classical computer vision techniques to enhance the robustness of video-based pulse detection.

Python

OpenCV

RNN

CNN

Keras

TensorFlow

Linux

Senior Machine Learning/AI Engineer (Contractor) at Schibsted Enterprise Technology, Stockholm

MAY 2019 - JAN 2020

- Prototyped an NLP-based solution to automate **support ticket routing**, reducing manual effort and improving accuracy in routing decisions.
- Improved classification accuracy by employing an F1 loss function, addressing imbalanced data across support categories.

Python Keras RNN fastText Docker AWS

Senior Machine Learning/AI Engineer (Contractor) at Hedvig AB, Stockholm

DEC 2018 - MAY 2019

- Built and deployed an **intent classification system** for customer service chats, enhancing automated customer support and streamlining communication workflows.
- Automated key operations, including insurance cancellation processes, through integration with existing customer systems.

Python Keras RNN fastText Javascript AWS Docker

Machine Learning/Data Science Contractor at Volvo Car Mobility AB, Stockholm

AUG 2018 - DEC 2018

- Conducted detailed data analysis to support business decisions, helping refine the car-sharing service by understanding user behavior and demand patterns.
- Performed behavior-based user profile classification, which informed targeted marketing and service improvement strategies.
- Developed **demand prediction models** for car availability, contributing to operational efficiency.
- Set up real-time dashboards to visualize key business metrics, improving transparency and decision-making across the company.

Python Flask Docker GCP Google BigQuery Google Kubernetes Engine Linux

Machine Learning Lead at Epidemic Sound, Stockholm

JUN 2018 - AUG 2018

- Developed a **deep convolutional neural network** for automatic multi-label classification of songs based on genre labels, enhancing the platform's ability to recommend and categorize music accurately.
- Built a condensed dataset from **billions of historical events**, enabling the sales team to more effectively target potential customers.
- Improved the company's internal processes and data-driven decision-making by increasing efficiency through data analysis and machine learning-driven solutions.

Python Convolutional Neural Networks (CNN) Google BigQuery Linux

Chief Technology Officer at Watty AB, Stockholm

SEP 2013 - JUN 2018

- Led **five cross-functional teams** in developing a real-time energy monitoring product, recognized as one of the most advanced in the market.
- Engineered a data collection platform and machine learning pipeline that contributed to significant product pivots and improved user experience.
- Designed the full technical architecture, from hardware data acquisition to cloud-based analysis and customer-facing interfaces.

C/C++ Python Docker AWS Kubernetes CoreOS

Senior Software Engineer at Gavagai, Stockholm

SEP 2013 - AUG 2014

- Contributed to Gavagai's core text analysis product by working on large-scale, real-time social media data processing.
- Implemented efficient **nearest-neighbor calculations on GPUs** to improve the performance of language models used for understanding relationships

between words.

- Developed non-core APIs in Node.js and Coffeescript, expanding the functionality and reach of the company's language analysis tools.
- Provided minor but impactful contributions in Python to further optimize performance and scalability.

Java

J2EE

Python

NodeJS

Coffeescript

GPU-based computations

Linux

Senior Software Engineer at Euroling AB, Stockholm

JAN 2012 - SEP 2013

- Led the **architectural redesign of Siteseeker**, an enterprise search engine, to make it scalable and modular, transitioning it from a monolithic structure to smaller services that could be scaled horizontally.
- Refactored legacy code by implementing **modern C++ (C++11)** standards and best practices, improving performance and maintainability of the codebase.
- Introduced Python-based microservices to expedite the development and deployment of new features, reducing the time-to-market for enhancements.
- Provided vision and leadership on improving the scalability of search engine technology, laying the foundation for future growth and performance improvements.

C++

Python

Ruby

Bash

Google Protocol Buffers

Git

Linux

WORDS FROM EX-COLLEAGUES

Simon Celinder

DATA SCIENTIST, KRY/LIVI, STOCKHOLM · SEP 2022

It is rare to come across someone who has both such a depth and breadth in engineering as Ali. He has long experience in the Machine Learning space, but also in software engineering. I'm currently working closely with Ali in developing machine learning models at the digital healthcare provider Kry, collaborating on problems in NLP and forecasting. He has taught me a lot about engineering and ML, and he is always friendly, with an easy-going and pragmatic attitude. He's ML knowledge is deep, which shows in how he has very knowledge driven and specific hypotheses about what type of model or architecture that is likely to improve the solution to a specific problem. He is a source of inspiration for the rest of the team, constantly staying ahead of the curve, exploring new tools and coming with creative ideas. I would recommend Ali to any company that wants to take data science to the next level.

Balazs Toth

SENIOR DATA SCIENTIST, IKEA DIGITAL · MAY 2021

Ali is one of the best engineers I have worked with. He is pragmatic, fast and productive. He is proficient in machine learning, data engineering and backend development. It is hard to find people like him in the industry.

Antonio Davide Cali

DATA ENGINEER, IKEA DIGITAL · MAY 2021

I had the opportunity to work with Ali for almost a year. No matter what was the problem, he was able to resolve. He's super skilled and talented. And all the conversation we had were always very useful. I had to say that it was also a lot of fun working together. I hope to meet him again in some future project

Mattias Nilsson Ph.D., Principal Data Scientist, Startup in Stealth Mode

· MAY 2020

Ali is a true critical thinker and dares to question typical assumptions. He firmly believes in fast iterations, which helped us understand and address critical issues of a difficult (research) challenge at an early stage. Ali is agnostic to the specific tools and focuses instead on the problem to be solved. Basically, he gets stuff done. I can strongly recommend Ali and I have really enjoyed working closely together with him for the past months.

Hjalmar Nilsonne

FOUNDER, WATTY & NEKO HEALTH · JUL 2016

Ali is the most gifted technologist I know and it's been a privilege to be working with him at Watty for the past 2.5 year. During this time we have gone from an idea on paper to having the best product in the market and signing some of the biggest players on the market as customer. Ali has been in charge of our technology strategy across the board: machine learning, large-scale back-end systems, hardware and product. His ability to identify the right opportunities across these areas has been critical for Watty's success. As a manager he has implemented a development culture with an emphasis on pragmatism and focusing only on that which is most critical. It's been a treat to watch how that culture, combined with Ali's knack for finding unorthodox strategies, has generated great results that has often exceeded our customers expectations. He is also passionate about the product that ends up in people's hands and is always the most avid user of anything he builds. If you're looking for the smartest guy around when it comes to technology, you need not look further

Per Mellqvist

HEAD OF ENGINEERING, EUROLING · OCT 2013

With a strong drive for learning and trying new things Ali is likely to bring new influences to any organization. He has a firm grip on search engine technology both in terms of internal implementation and integration aspects, but I would recommend Ali to any team that want to drive change and continuously improve their code and architecture.

Anders Malm

PRODUCT MANAGER, MOSYNC · JAN 2012

Ali has a very deep knowledge in a very broad spectrum when it comes to all parts of software development. He has this unique ability to always find a solution and then be able to implement it in a straight forward way. I would recommend Ali without a doubt!

Mickel Grönroos

PRODUCTION MANAGER, FINDWISE · NOV 2011

Ali is a great developer / software architect. He has the ability to solve tough problems efficiently and creatively. There is always a solution in Ali's mind. Ali is independent and self-motivated and he has the ability to communicate with both IT and business people in order to assist them with their challenges. For an IT consultant, this is really a valuable trait.

Shiv Shankar

ENGINEER, MOSYNC · JAN 2011

Ali is a talented, creative and hard working Software Engineer but more importantly a pro-active leader and an adaptive co-worker. Ali’s personal insight into key problems and their solutions always helps the team move ahead. I have no doubts his skill sets will set him apart in any role that he chooses to take up.

Håkan Kjellman

TECHNICAL DIRECTOR, MOSYNC · DEC 2010

I had the fortune of working with Ali for seven months at Mosync. I wished it would have lasted longer since Ali is an extremely skilled SW engineer. Just one of those rare guys that you know will always deliver a quality job and on time.

DETAILS

Stockholm, Sweden

DATE OF BIRTH 1982

NATIONALITY Swedish

AREAS OF EXPERTISE

- Large Language Models
- Generative AI
- Machine Learning/AI
- Python
- C/C++
- JavaScript
- Java
- Git
- Snowflake
- BigQuery
- Postgres/MySQL
- Google Cloud Platform
- Amazon Web Services

LANGUAGES

Swedish	Native speaker
English	Highly proficient
Farsi	Highly proficient

HOBBIES

Working out, travel, reading & learning about anything and everything