



Ali Mosavian

SENIOR MACHINE LEARNING/AI-KONSULT

LinkedIn ♦

GitHub ♦

Webbplats ♦

20+

år av programmering

17

år inom machine learning

42+

marknader, en modell

f.d. CTO

ledde fem team

PROFIL

Erfaren **hands-on**-ingenjör, expert inom machine learning/AI, generalist och före detta CTO med passion för att hitta effektiva och okonventionella vägar från **idé till produkt** genom att vara **pragmatisk, fokuserad, tänka utanför boxen och framför allt röra mig snabbt**.

Livslångt lärande med över två decenniers programmeringserfarenhet, verksam i IT-branschen sedan 2005 och inom machine learning/deep learning sedan 2009. Där storskaliga internetsystem och machine learning/AI möts för att skapa oförglömliga användarupplevelser är där jag känner mig mest hemma.

UTBILDNING

Masterexamen i datavetenskap, Kungliga Tekniska högskolan, Stockholm

AUG 2005 - DEC 2009

Inriktning mot machine learning och autonoma system

YRKESERFARENHET

Senior AI Engineer (konsult) på H&M Group, Stockholm

SEP 2025 - NU

- Ledde omarkitekturen och omskrivningen av en proof of concept för **store capacity optimization** till ett produktionssystem (81 % av alla ändringar), som avgör vilka artiklar och plagg som placeras på vilken inredning för att fylla varje butik optimalt. Designade ett domain-specific language (DSL) för att modellera butiksplanering som ersatte tusentals rader specialfallskod, ovanpå en tvåstegs constraint solver; utökade testsviten från 5 till 100+ filer.
- Designade och byggde ett **system för att estimerar shop-floor lager** (98 % av alla ändringar) för ~3 350 butiker utan RFID: en gradient boosted hurdle-modell med 161 features tränad på 74 miljoner rader från 686 RFID-butiker, matad av en egenutvecklad data lineage- och transformationspipeline som bearbetar 6,6 TB per träningskörning.
- Byggde om **dataplattdformen för butiksfullnadsgrad** i dbt (43 % av alla ändringar): införde serving-tabeller med data contracts, fasade ut legacy-modeller och rättade ett grundläggande fel i det publicerade stock ratio-måttet.

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 (konsult) på Aloï AI, Stockholm

APR 2025 - JUL 2025

- Förbättrade **retrieval-augmented generation (RAG)**-pipelines i Aloï AI-drivna produkt för clause completion, med högre träffsäkerhet och relevans vid avtalsskrivande.
- Utvecklade avancerade strategier för content extraction och chunking för olika juridiska dokument (docx, pptx, pdf m.m.), inklusive automatisk klassificering av avtalstyp och strukturbevarande parsing.
- Byggde en specialiserad pipeline för investerarpresentationer, med detektering av visuella element och asynkron bearbetning för att snabbt hitta relevanta sidor och extrahera nyckeldata.
- Förbättrade RAG-retrievalens kvalitet avsevärt med **re-ranking** och andra optimeringstekniker, vilket höjde prestandan i clause completion.

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

Senior AI Engineer (konsult) på Nordea AB, Financial Crime Prevention Unit, Stockholm

JAN 2025 - APR 2025

- Gjorde om arkitekturen, skrev om och produktionssatte ett befintligt proof of concept-verktyg för att **automatisera detektion av finansiell brottslighet** utifrån uppgivet och faktiskt transaktionsbeteende.

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

Senior AI Engineer (konsult) på Byggfakta, Stockholm

JUN 2024 - JAN 2025

- Utvecklade en **conversational AI** för projektsökning som ersatte ett komplext filtersystem med ett mer användarvänligt och intuitivt gränssnitt.
- Gjorde det möjligt för användare att söka projekt med minimal input, vilket minskade den kognitiva belastningen av att navigera bland många filter och kategorier.
- Utvecklade ett **hybrid recommender system** som ger kunderna mer relevanta rekommendationer av byggprojekt.
- Utvecklade en modell som predikterar roller i byggprojekt.

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

Senior AI Engineer (konsult) på EQT Partners, Stockholm

AUG 2023 - APR 2024

- Byggde en **multi-stage AI-assistent** (RAG-baserad) som besvarar investerarfrågor korrekt genom att bearbeta komplexa dokument.
- Implementerade en metod för att extrahera innehåll ur PDF-, Word- och PowerPoint-filer med bibehållen struktur, vilket kraftigt förbättrade dokumenthanteringen för AI-system.
- Designade och driftsatte ett **automatiserat evalueringsramverk** för att följa och mäta assistentens träffsäkerhet över tid.

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

Senior Machine Learning Engineer (konsult) på Lingio, Stockholm

NOV 2022 - JUN 2024

- Utvecklade AI-drivna verktyg för översättning, sammanfattning och generering av innehåll, vilket effektiviserade innehållsproduktionen.
- Fine-tunade **open source-språkmodeller** för specifika uppgifter, med bättre prestanda och integration i den generativa AI-produkten.
- Levererade tjänster i produktionskvalitet som i dag utgör **kärnan i Lingios generativa AI-erbjudande**.

GPT-3

Python

FastAPI

Pytorch

LoRA

Docker

GCP Kubernetes

Senior Machine Learning Engineer på Kry/Livi, Stockholm

SEP 2021 - OKT 2022

- Designade och driftsatte en **CatBoost-modell** som klassificerar patienters vårdbehov, vilket förbättrade triagen.
- Ledde utvecklingen av en **RNN-baserad modell** som predikterar patienters väntetider, vilket gav mer träffsäkra förväntningar på servicenivån.
- Gjorde om arkitekturen för bemanningsprognoserna, vilket optimerade resursfördelningen och minskade ineffektivitet i driften.

Python

FastAPI

TensorFlow

Pytorch

RNN

CatBoost

AWS

Snowflake

Airflow 2

Senior Data/Machine Learning Engineer (konsult) på IKEA Digital, Stockholm

JUL 2020 - MAJ 2021

- Utvecklade en recommendation-modell (contextual bandits) och driftsatte den för IKEA:s inspirationsflöde. Modellen tränas automatiskt varje natt på några hundra miljoner exempel, och driftkostnaden för **42+ marknader** är flera gånger lägre än för andra team som använder Google Recommendation API för en enda marknad.
- Refaktorerade och delade upp en monolitisk backend i mindre tjänster, vilket förbättrade underhållbarhet och prestanda. Minskade den genomsnittliga svarstiden från ~800 ms (p99: 3-4 s) till **~200 ms** (p99: ~500 ms), vilket gav en markant bättre användarupplevelse och lägre driftkostnader.
- Kortade iterationscyklerna för data **från timmar till minuter**, vilket möjliggjorde snabba experiment med olika modellarkitekturer.

Python

FastAPI

TensorFlow

Docker

GCP

Google AI Platform

Senior Machine Learning/AI Engineer (konsult) på Neko Health, Stockholm

JAN 2020 - JUL 2020

- Utvecklade en banbrytande, kontaktlös metod för att skatta **puls och blodtryck** med videoanalys, ett steg framåt för hälsoövervakning i realtid.
- Kombinerade digitala filter, signalbehandling och klassisk computer vision för att göra den videobaserade pulsdetektionen mer robust.

Python

OpenCV

RNN

CNN

Keras

TensorFlow

Linux

Senior Machine Learning/AI Engineer (konsult) på Schibsted Enterprise Technology, Stockholm

MAJ 2019 - JAN 2020

- Tog fram en prototyp på en NLP-baserad lösning för automatisk **routing av supportärenden**, vilket minskade det manuella arbetet och ökade träffsäkerheten.
- Förbättrade klassificeringens träffsäkerhet med en F1 loss för att hantera obalanserad data mellan supportkategorier.

Python Keras RNN fastText Docker AWS

Senior Machine Learning/AI Engineer (konsult) på Hedvig AB, Stockholm

DEC 2018 - MAJ 2019

- Byggde och driftsatte **intent classification för kundtjänstchattar**, vilket förbättrade den automatiserade kundsupporten och effektiviserade kommunikationsflödena.
- Automatiserade centrala processer, bland annat uppsägning av försäkringar, genom integration med befintliga kundsystem.

Python Keras RNN fastText Javascript AWS Docker

Machine Learning/Data Science-konsult på Volvo Car Mobility AB, Stockholm

AUG 2018 - DEC 2018

- Genomförde detaljerade dataanalyser som beslutsunderlag och hjälpte till att förfina bilpoolstjänsten genom att kartlägga användarbeteende och efterfrågemönster.
- Klassificerade användarprofiler utifrån beteende, vilket låg till grund för riktad marknadsföring och förbättringar av tjänsten.
- Utvecklade **demand forecasting-modeller** för biltillgänglighet, vilket bidrog till en effektivare drift.
- Satte upp real-time dashboards för centrala affärsmått, vilket ökade transparensen och förbättrade beslutsfattandet i hela företaget.

Python Flask Docker GCP Google BigQuery Google Kubernetes Engine Linux

Machine Learning Lead på Epidemic Sound, Stockholm

JUN 2018 - AUG 2018

- Utvecklade ett **deep convolutional neural network (CNN)** för automatisk multi-label-klassificering av låtar efter genre, vilket förbättrade plattformens förmåga att rekommendera och kategorisera musik.
- Byggde ett komprimerat dataset från **miljarder historiska events**, så att säljteamet kunde rikta sig mot potentiella kunder mer effektivt.
- Förbättrade företagets interna processer och datadrivna beslutsfattande genom dataanalys och machine learning-baserade lösningar.

Python Convolutional Neural Networks (CNN) Google BigQuery Linux

Chief Technology Officer på Watty AB, Stockholm

SEP 2013 - JUN 2018

- Ledde **fem tvärfunktionella team** i utvecklingen av en produkt för energiövervakning i realtid, erkänd som en av de mest avancerade på marknaden.
- Byggde en plattform för datainsamling och en machine learning-pipeline som bidrog till viktiga produkt-pivots och en bättre användarupplevelse.
- Designade hela den tekniska arkitekturen, från datainsamling i hårdvaran till molnbaserad analys och kundnära gränssnitt.

C/C++ Python Docker AWS Kubernetes CoreOS

Senior Software Engineer på Gavagai, Stockholm

SEP 2013 - AUG 2014

- Bidrog till Gavagais centrala textanalysprodukt genom storskalig realtidsbearbetning av data från sociala medier.
- Implementerade effektiva **nearest neighbor-beräkningar på GPU** för att förbättra prestandan hos språkmodeller som fångar relationer mellan ord.

- Utvecklade kompletterande API:er i Node.js och Coffeescript, vilket utökade funktionaliteten och räckvidden hos företagets språkanalysverktyg.
- Bidrog med mindre men betydelsefulla insatser i Python för att ytterligare optimera prestanda och skalbarhet.

Java

J2EE

Python

NodeJS

Coffeescript

GPU-based computations

Linux

Senior Software Engineer på Euroling AB, Stockholm

JAN 2012 - SEP 2013

- Ledde **omdesignen av arkitekturen i Siteseeker**, en enterprise search-motor, för att göra den skalbar och modulär - från en monolit till mindre tjänster som kunde skalas horisontellt.
- Refaktorerade legacy-kod enligt **modern C++ (C++11)** och best practices, vilket förbättrade kodbasens prestanda och underhållbarhet.
- Införde Python-baserade microservices för att snabba upp utveckling och driftsättning av nya funktioner, vilket kortade time-to-market.
- Stod för vision och ledarskap kring skalbarheten i sökmotortekniken och lade grunden för framtida tillväxt och prestandaförbättringar.

C++

Python

Ruby

Bash

Google Protocol Buffers

Git

Linux

OMDÖMEN FRÅN TIDIGARE KOLLEGOR

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 · MAJ 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 · MAJ 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

· MAJ 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 · OKT 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.

UPPGIFTER

Stockholm, Sverige

FÖDELSEÅR 1982

NATIONALITET Svensk

KOMPETENSOMRÅDEN

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

SPRÅK

Svenska	Modersmål
Engelska	Mycket god
Persiska	Mycket god

INTRESSEN

Träning, resor, läsning och att lära mig om allt möjligt