

## Sebastian Ikin, Senior systemutvikler og KI-Gründer

Senior systemutvikler med fem års erfaring fra Arms GPU-organisasjon og over to års erfaring som medgründer, CTO og daglig leder i egen KI-bedrift. Søker teknisk og strategisk krevende roller innen systemutvikling, KI-infrastruktur, plattformutvikling og teknisk ledelse. Basert i Oslo, Tospråklig norsk/engelsk.

[GitHub](#) | [Linkedin](#) | [sebastianaikin@gmail.com](mailto:sebastianaikin@gmail.com)

### Erfaring

#### **Mars 2024 - nå: Daglig leder / CTO / Medgründer, Rafmer AS**

Medgründer og teknisk ansvarlig i en topersons KI-bedrift som utvikler programvare og infrastruktur for lokal KI-inferens og agentiske arbeidsflyter. Startet som styreleder og CTO, og gikk over til rollen som daglig leder på fulltid i juni 2025. Har hatt ansvar for både teknisk arkitektur, produktretning og den operative driften av selskapet.

#### Teknologi og produkt

- Designet og bygget intern **Kubernetes-basert GPU-cluster** fra hardware til applikasjonslag for utvikling, testing og drift av lokale språkmodeller.
- Bygget inferensoppsett på forbruker-GPUer med blant annet **llama.cpp, vLLM, CUDA og Kubernetes**, inkludert fem RTX 3090-baserte deploys med støtte for åtte samtidige klienter og rundt **50 tokens/s per deploy** med Qwen3.8-27B.
- Optimalisert LLM-inferens med fokus på **GPU-throughput, minnebruk, kontekstlengde, parallellitet og kostnad per forespørsel**, uten enterprise GPU-virtualisering.
- Designet og produserte en **edge-KI-server** basert på AMD Ryzen AI Max, Ubuntu, k3s, Vulkan og llama.cpp, inkludert hardwarevalg, CAD-design, prototyping og software-stack.
- Utviklet backend- og infrastrukturtenester hovedsakelig i **Python og FastAPI**, deployert med Kubernetes, Helm og YAML-baserte workloads.
- Designet og utviklet **REST-API-er, MCP-servere, API-gatewayer, modellrouting og RBAC-systemer** for egne og kundeorienterte tjenester.
- Bygget interne verktøy og agentintegrasjoner for blant annet **cluster-observability, prosjektstyring, web-søk og utviklerarbeidsflyter**.
- Designet og rullet ut en **GitHub-integrert utvikleragent** som kjører på selskapets egen KI-infrastruktur og kan arbeide mot interne repositories og utviklingsprosesser.
- Designet og bygget kundeorienterte **ETL-pipelines med Apache Airflow**, inkludert kundespesifikk tilgangskontroll og prosessering av eksterne datasett.
- Bygget og leverte kundeorienterte datasett gjennom eksterne REST-API-er med isolerte prosjekter og tilgangskontroll.
- Utviklet en enkel **OpenRouter-gateway** integrert med Supabase og Stripe for validering av abonnement, API-tilgang og modellrouting.
- Testet og optimalisert lokal KI-inferens på **Linux og Android**, med særlig fokus på hardwarekompatibilitet, throughput, modellkvalitet og minnebegrensninger.
- Bidratt til arkitektur og kvalitetssikring av selskapets Flutter-baserte Agent Platform, inkludert inferensruntime, API-integrasjoner og lokal modellkjøring.

#### Ledelse og forretning

- Ansvarlig for selskapets **tekniske retning, produktprioriteringer og prosjektstyring**, hovedsakelig gjennom GitHub Projects.
- Ledet strategisk pivot fra **B2B-salg av KI-hardware og lisensiering** til lokal B2C-inferens på forbrukerenheter med SaaS-forretningsmodell.
- Gjennomført og dokumentert **markedsavklaringsprosjekt finansiert av Innovasjon Norge**.
- Arbeidet med **finansiell modellering, budsjettering og kostnadsstyring** gjennom Fiken og Google Sheets.
- Deltatt i investor- og VC-dialog, finansieringsarbeid og strategiske vurderinger knyttet til kapitalbehov og selskapets videre utvikling.
- Forhandlet, skrevet og koordinert avtaler som **NDA-er, intensjonsbrev og kundeavtaler**.
- Ansvar for selskapets arbeid med **GDPR, informasjonssikkerhet og EU AI Act**.

- Håndtert innkjøp av GPU-er og annen KI-hardware fra norske og internasjonale leverandører, med særlig fokus på ytelse per krone.
- Arbeidet med markedsanalyse, produktside, teknisk kommunikasjon og digital markedsføring gjennom blant annet Google Ads, LinkedIn og Meta.

### **Juni 2020 - mai 2025: Senior systemutvikler - GPU, Arm Trondheim**

Jobbet i Arms GPU-organisasjon med utvikling og vedlikehold av Python-baserte verktøy for GPU-testautomatisering, GPU bring-up, dataprosessering og analyse. Startet deltid ved siden av studier og gikk senere over i fulltidsstilling. Ble etter hvert senior systemutvikler med ansvar for sentrale interne verktøy, dataleveranser og støtte på tvers av team. Utvalgte leveranser:

- Utviklet og vedlikeholdt en **Apache Spark-basert ETL-pipeline** for automatisk problemdeteksjon i store testdatasett.  
Løsningen reduserte behovet for manuell filtrering og sparte teamet omtrent **åtte arbeidstimer per uke**.
- Utviklet og vedlikeholdt en forskningsbasert **bildeverifiseringsalgoritme og datapipeline** som automatiserte en tidligere manuell klassifiseringsprosess.  
Oppnådde rundt **70 % precision** og reduserte manuelt arbeid med omtrent **fem timer per uke**.
- Utviklet og vedlikeholdt en intern **React/Node-applikasjon** for dataklassifisering og analyse.
- Utviklet og administrerte **Jenkins-pipelines** for automatisering av tester, arbeidsflyter og releaser.
- Forbedret testing og kvalitetssikring av verktøy brukt til intern datadistribusjon, noe som bidro til å redusere kritiske produksjonsfeil til nær null over en periode på rundt tre år.
- Feilsøkte komplekse problemer i skjæringspunktet mellom **hardware, drivere, testinfrastruktur og software** under GPU bring-up.
- Identifiserte og varslet testproblemer på tvers av flere team og bidro til rotårsaksanalyse og videre koordinering.
- Leverte analyser og datasett til fire interne team, tekniske ledere og eksterne samarbeidspartnere.
- Bidro med teknisk code review, arkitekturdiskusjoner og forbedring av interne utviklingsprosesser.
- Gjennomførte **tekniske intervjuer** og deltok i rekruttering av nye utviklere.
- Veiledet og onboardet nyansatte og fungerte som teknisk støtte for mindre erfarne utviklere.

## Utdanning

### **August 2018 – juni 2021: Bachelor i ingeniørfag - Dataingeniør, NTNU Trondheim**

Spesialisering innen systemutvikling. Bacheloroppgave innen **penetrasjonstesting og datasikkerhet**.

## Teknisk kompetanse

**Programmering:** Python, TypeScript, SQL, JavaScript, Java, Bash, C++, Dart / Flutter

**Backend og systemutvikling:** FastAPI, Node.js, REST API-design, microservices, MCP, RBAC, PostgreSQL, Supabase, serverless-arkitektur, distribuerte systemer, systemarkitektur, OOP, TDD, design patterns, code review

**Data og ML:** Apache Spark, Apache Airflow, Pandas, NumPy, PyTorch, ETL / datapipelines, datakvalitet, validering, GPU-basert inferens

**KI-infrastruktur / MLOps:** Kubernetes, k3s, Helm, llama.cpp, vLLM, CUDA, ROCm, Vulkan, TensorRT-LLM, Triton Inference Server, Nvidia GPU, AMD Ryzen AI, Arm Mali, lokal LLM-inferens, modelloptimalisering, GPU-throughput, minneoptimalisering

**DevOps og observability:** Git, GitHub, Gerrit, GitHub Projects, Jenkins, Rancher, Docker, Prometheus, Grafana, Loki, Linux / Ubuntu

**Cloud:** AWS EC2, Lambda, RDS, Aurora, VPC, S3, IAM, GCP Firebase, Firestore, Supabase PostgreSQL, Auth, Edge Functions, Storage, Vault, Row Level Security, Cloudflare Gateway

**Ledelse og strategi:** Teknisk ledelse, teknologistrategi, produktutvikling, prosjektstyring, arkitekturvalg, mentoring, onboarding, tekniske intervjuer, forretningsutvikling, finansiell modellering, kostnadsstyring, markedsanalyse, investor- og stakeholder-kommunikasjon, GDPR, informasjonssikkerhet, EU AI Act, SCRUM, Kanban, Jira, Confluence, GitHub Projects