



AI

Vad är det?

Frikopplad Open Source Stack: Väl utvalda open source-projekt som möjliggör flexibilitet, snabb utveckling och enkel utbytbart av delar i plattformen

Man nyttjar det man behöver och tillför i samverkan det som saknas



Fördelar

Säkerhet & kontroll: Känslig data kan hållas inom den egna infrastrukturen.

Kostnadseektivitet: Undviker höga licenskostnader och möjliggör intern skalning. Betala inte per användare, betala för användning. Låg fast kostnad.

Flexibilitet & anpassning: Plattformen kan enkelt anpassas och integreras i olika verksamhetsmiljöer.

Transparens & spårbarhet: Loggning och mätbarhet av användning ger ökad insyn och möjlighet till kontinuerlig utvärdering.



Vanliga användningsområden

Generell säker AI-kollega: Rulla ut AI som kollega för att säkert assistera i dagliga arbetsuppgifter.

Specialiserad säker AI-kollega : Utveckla specialanpassade assistenter för exempelvis HR, teknisk support och intern kommunikation.

Bot-byggande: Exponera och skapa bottar som automatiserar arbetsflöden och processer.

Agenter: Bygg agenter med integrations- och automationsmotorn för att hantera komplexa uppgifter.



Assistenten – Agenter - Botar

| Egenskap | AI-assistenten | AI-botar | AI-agenter |
|---------------------------|--|---|---|
| Syfte | Hjälpa användaren med uppgifter och informationssökning | Automatisera uppgifter och interaktioner | Utföra självständiga uppdrag och fatta beslut |
| Interaktion | Mänsklig input krävs för det mesta | Förprogrammerade svar, oftast begränsat | Självständig interaktion med omvärlden |
| Autonomi | Låg till måttlig (kräver mänskliga instruktioner) | Låg till måttlig (kan agera automatiskt men inom ramar) | Hög (kan fatta egna beslut och anpassa sig) |
| Användningsområden | Digitala assistenter, support, informationssökning | Kundtjänst, e-handel, spel, sociala medier | Autonoma system, robotik, cybersäkerhet, forskning |
| Exempel | ChatGPT, Siri, Alexa | Webbchattbotar, helpdesk-botar | Autonoma bilar, AI-drivna finansiella system, cybersäkerhetsagenter |
| Lärande | Kan förbättras via träning men oftast statisk under användning | Ofta regelbaserad eller med begränsad AI-inläring | Kan använda maskininläring för att förbättra sin prestanda över tid |



Sambruks AI-plattform

Förmågor

| Förmågor | Vad | Beskrivning | Licensmodell |
|-------------------------|--|--|----------------------|
| Användarhantering | KeyCloak | Identitets- och åtkomsthantering (IAM), SSO och rollbaserad åtkomst | Apache License 2.0 |
| AI gränssnitt | OpenWebUI | Webbaserat användargränssnitt för att interagera med AI-modeller lokalt eller via API | BSD 3-Clause License |
| Kunskap/RAG | ChromaDB | Vektordatabas för att lagra embeddingar och möjliggöra Retrieval-Augmented Generation (RAG) | Apache License 2.0 |
| Lokal (privat) LLM | Ollama på delad EU server | Lokal LLM-lösning som körs i en EU-miljö för att uppfylla dataskyddskrav | MIT |
| Integration/automation | Kestra | Plattform för orkestrering av arbetsflöden och dataprocesser och bygga agentisk funktionalitet | Apache License 2.0 |
| Loggning/mätning | Helicone | Verktyg för att logga, övervaka och analysera anrop till AI-modeller samt deras prestanda | Apache License 2.0 |
| Säker och skalbar drift | Kubernetes på EU moln (även on prem möjligt) | SBOM, Loggning, Sårbarhetsscanning, Monitorering, Backuper, Skalning, Lastbalansering | K8S |

Språkmodeller

| Språkmodeller | Vad | Beskrivning | Licensmodell |
|---------------------------|--------------------------------------|--|--------------|
| EU LLM | Mistral | En Europa-baserad LLM för att möta EU-regler kring dataskydd och lagkrav | Proprietary |
| 3e land LLM | OpenAI, Anthropic, Google, X.ai, etc | Molnbaserade LLM:er utanför EU, med andra jurisdiktioner och dataskyddsregler | Proprietary |
| Transkribering | Whisper (KB whisper på lokal LLM) | AI-baserad röst-till-text-transkribering av ljud och video, med stöd för flera språk och hög precision | MIT |
| Bild- och videogenerering | Flux, Recraft, Ideogram, Wan | Modeller för att generera och manipulera bilder | Diverse |



Ett scenario

Björksta kommun



Scenario: AI-stödd handläggning och medborgardialog i Björksta kommun

Bakgrund

Björksta kommun har under flera år arbetat med digitalisering av sina verksamheter, men står fortfarande inför utmaningar som:

- Handläggningstider inom socialtjänsten
- Bristande möjlighet till individuell hjälp via e-tjänster
- Behov av att avlasta första linjens support
- Behov av uppföljning och analys av hur AI-stödda tjänster används

Kommunen ansluter sig till **Sambruks öppna AI-plattform** och sätter upp ett första pilotområde:

Stöd till medborgare och handläggare inom ekonomiskt bistånd.



1. Keycloak – Identitet och åtkomstkontroll

Keycloak integreras med kommunens Active Directory, vilket möjliggör:

- **SSO** för alla anställda – handläggare kan logga in till AI-verktyg utan extra lösenord.
- **Rollstyrd åtkomst** – endast socialsekreterare får tillgång till särskilda interna AI-flöden.
- **Medborgarinloggning via BankID** – för framtida publik AI-tjänst med verifierade användare.



2. OpenWebUI – Gränssnitt för AI-dialog

Via OpenWebUI skapas två AI-drivna chattfunktioner:

För medborgare:

En chattbot som:

- Besvarar frågor om rätten till ekonomiskt bistånd
- Hjälper till att förstå regler, förklara vad som menas med "arbetsförmåga" eller "försörjningsplikt"
- Guidar i att fylla i ansökan via kommunens e-tjänst

Språket är anpassat, empatiskt och begripligt – språklig förenkling sker automatiskt genom prompt-engineering.

För handläggare:

En expertassistent som:

- Hjälper till att tolka Socialtjänstlagen
- Ger förslag på beslut baserat på inmatade uppgifter
- Visar liknande ärenden (utan persondata) för att främja rättssäkerhet



3. ChromaDB – Vektordatabas för semantisk sökning

ChromaDB används för:

- Att lagra strukturerad kunskap: lagtexter, vägledningar från Socialstyrelsen, lokala riktlinjer
- Att möjliggöra **RAG (Retrieval-Augmented Generation)** där AI:n hämtar fakta från denna källa innan svar genereras

Med detta blir svaren både förankrade och förklarande:

"Enligt Socialstyrelsens vägledning (kap. 3, s. 12) ska kommuner beakta om det finns särskilda skäl vid tillfälligt behov..."



4. Kestra – Flödeshantering och automation

Kestra används för att bygga automatiserade flöden för både handläggare och IT-avdelningen.

Exempel på flöden:

- När en medborgare chattar och anger att hen har akuta behov → automatisk notifiering till socialjouren.
- När nya riktlinjer publiceras → Kestra-flöde för att uppdatera ChromaDB och meddela AI-assistenterna.
- När ett ärende når beslutspunkt → AI genererar utkast till beslutstext, skickas till handläggare för granskning.



5. Helicone – Statistik och analys

För att följa upp användningen används Helicone till att:

- Se vilka frågor som ställs oftast till AI:n
- Följa svarens kvalitet och användarnas nöjdhet (enkät efter varje session)
- Identifiera behov av förbättring i underliggande kunskapsbas
- Upptäcka ovanliga mönster eller missförstånd (t.ex. "många frågor om 'akut behov' men AI ger för komplexa svar")

Detta används av både socialförvaltningen och digitaliseringsstaben i kommunen.



6. Ollama – Lokala språkmodeller

Kommunen använder Ollama för att:

- Köra språkmodeller lokalt, vilket minskar kostnader och integritetsrisker
- Undvika att persondata lämnar kommunens infrastruktur
- Anpassa modeller till svenska myndighetstermer genom finjustering

Exempel på använd modell: `openchat-3.5-sv-finetuned`, tränad med Sambruks promptbibliotek och Socialstyrelsens dokumentation.



Sammanfattning

- Genom samverkan inom Sambruk kan man kan lära av varandra och samfinansiera utveckling av verktyg och kunskap om användande av verktyg.
- Börja litet om ni vill, skala upp i er egen takt.
- Kort startsträcka.
- Nyttja enkelt som Sambruksmedlem.



Prismodell

| | Säker sandlåda institutionsplattform för utvärdering och lärande | AI Plattform Fullständig AI förmåga |
|---|--|--|
| Förmågor | | |
| Antal användare | 20 | 20+ |
| AI assistentgränssnitt | ✓ | ✓ |
| Kunskap/RAG | ✓ | ✓ |
| Webbsök/research | ✓ | ✓ |
| API endpoints | ✗ | ✓ |
| Agentiska arbetsflöden & integration | ✗ | (tillägg) |
| Central användarhantering | ✗ | installationskostnad |
| Loggning/övervakning/analys | ✗ | (tillägg) |
| Egen lokal LLM | ✗ | (tillägg) |
| EU LLM (Mistral, Scaleways) | ✓ | ✓ |
| 3e land LLM (OpenAI, Anthropic, x.ai, Google) | ✓ | ✓ |
| Bild-ljud och videogenerering | ✗ | ✓ |
| Bildanalys | ✓ | ✓ |
| Transkribering (enkel) | ✓ | ✓ |
| Filtrering för känsliga uppgifter | ✓ | ✓ |
| Webbskrapning (enkel) | ✓ | ✓ |
| Säker och skalbar drift | ✓ | ✓ |

| ID | Priser | | |
|----|---|------------------------|--------------------------------|
| A | Säker sandlåda | 6 000 kr /mån | Max 20 användare |
| B | AI Plattform | 12 000 kr /mån | 20+ användare |
| C | Uppsättning, utbildning | 30 000 kr | Engångskostnad |
| D | Plattform för agentiska arbetsflöden/integration | 3 000 kr / mån | Utöver B |
| E | Plattform för loggning, övervakning och analys av anrop till AI-modeller samt deras prestanda | 3 000 kr / mån | Utöver B |
| F | Koppling till katalogtjänst (SSO) | pris per timme | ca 30-50 timmar |
| G | Utökad kapacitet för att hantera utökat användande (aktuellt efter ca 100 anv) | 1 000 kr / mån | 16 CPU, 32 GB ram, 320 GB disk |
| H | Tokenförbrukning för LLM | faktiskt kostnad + 20% | |
| I | Svensk LLM (fair use) | 5 000 kr / mån | Utöver B |
| J | Egen privat LLM | 3 000 kr / mån + GPU | Utöver B |
| K | GPU server (NVIDIA RTX 4000) | 3 000 kr / mån | 20 GB VRAM (≈ 12B model) |
| L | GPU server (NVIDIA RTX 6000) | 12 000 kr / mån | 48 GB VRAM (≈ 32B model) |
| M | GPU server (NVIDIA H100) | på förfrågan | 80 GB VRAM (70B model) |



| Tjänster | | |
|--|-----------|--------------------------|
| AI-Coach/förändringsledning | 1600 kr/h | Rådgivning för AI frågor |
| AI-Arkitekt | 1600 kr/h | AI-Arkitektur |
| AI-Utvecklare | 1200 kr/h | Utveckling av lösningar |
| | | |
| Supportpaket | | |
| Support - 8 timmar/månad | 10 000kr | |
| Ärendportal - registrering av ärenden | Ingår | |
| Beredskap AI-coach ordinarie arbetstid (vardagar 08:00–17:00). | Ingår | |
| Prioriterad åtgärdsplan vid allvarliga incidenter. | Ingår | |
| Exempel på vad AI-coachen hjälper till med: | | |
| - Hjälp med felsökning, systemintegrationer och tekniska frågor. | Per timme | |
| - Rådgivning kring optimal användning av AI-modellerna. | Per timme | |
| - Stöd vid uppdateringar och små justeringar av AI-modeller. | Per timme | |
| | | |
| Utökad beredskap | | |
| Utökad beredskap helgfria vardagar (07:00-22:00) | 2 000kr | |
| Utökad beredskap helg och röda dagar (07:00-22:00) | 3 000kr | |

Nästa steg

- Intresseanmälan från intresserade medlemmar: thomas.wennersten@sambruk.se
- Skapande av samverkansavtal
- Etablera samverkans enligt Sambruks förvaltningsmodell
- Avtalstecknande och uppstartsprojekt – April / Maj 2025