



AI on Power 11

Martin Hecht Olsen
System & Metode A/S

Håndsooprækning ...

1. Hvem forventer at anvende den nye, indbyggede accelerator i Power 11-chippen?
2. Hvem har anvendt ChatGPT, Copilot eller tilsvarende værktøjer i forbindelse med forretningsrelaterede opgaver?
3. Hvem har stillet spørgsmål, som:
 - Hvad er **vores** gældende regler for reklamationer?
 - Hvilken formulering bør **vi** anvende i et svar til en utilfreds kunde, der har beskadiget produktet ved forkert brug, i henhold til specifikke afsnit i **vores** vejledning?
 - Hvilke af **vores** programmer skal ændres, for at det teknisk bliver muligt at udbetale delvis compensation ved reklamationer, og hvordan ser de nødvendige koderrettelser ud?

Få

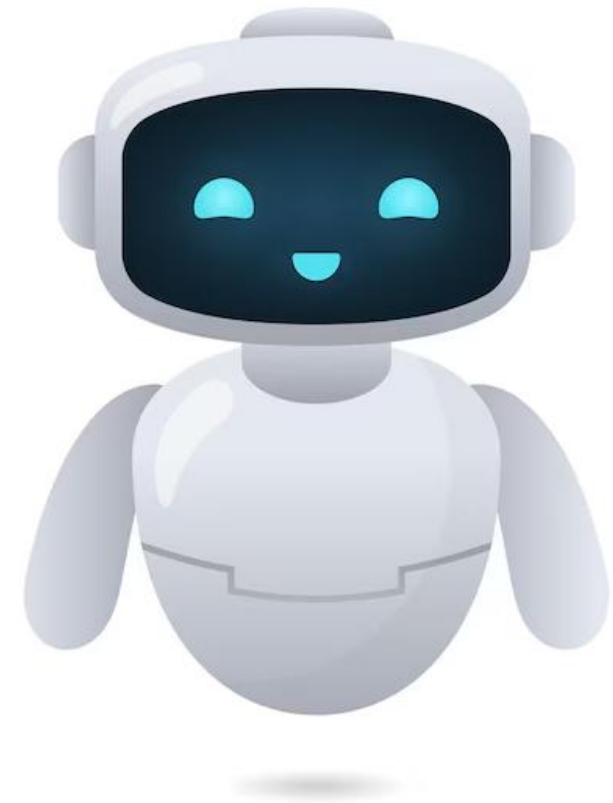
Mange

Ingen?



Den klogeste og billigste medarbejder

- **Uddannelse**
 - Hele Internettet! Generativ AI er en form for kunstig intelligens, der kan **skabe nyt** indhold baseret på mønstre, den har lært fra store mængder eksisterende data
- **Kompetancer**
 - Hurtig produktion af indhold
 - Kan automatisere kreative og rutineprægede opgaver
 - Mulighed for personaliseret kommunikation i stor skala
- **Udfordringer**
 - Risiko for faktuelle fejl ("hallucinationer")
 - Kan skabe ophavsretsmæssige eller etiske problemer
 - Kræver klar styring for at undgå bias og misbrug
- **Manglende viden om virksomheden**
 - Er begrænset til offentlig viden – typisk kun hjemmesiden
 - Ingen viden om interne oplysninger – dvs. ingen mangeårig erfaring
 - Begrænset hjælp for nye medarbejdere



Følsomme oplysninger kræver interne LLM

- Følsomme personoplysninger (GPDR)
- Konfidentielle forretningsoplysninger

- Sagsbehandling
- Forretningsstrategier
- Finansielle oplysninger
- Kundedata
- Prissætning og rabatstrukturer
- Produktudvikling og teknologier
- Intern forretningsdrift
- Leverandør- og samarbejdsaftaler

Mails, anden korrespondance

Markedsstrategier, ekspansionsplaner, fusioner og opkøb

Budgetter, forecast-rapporter, regnskaber før offentliggørelse

Kundelister, kontaktpersoner, kontrakter, købshistorik

Interne prissystemer, specialaftaler, rabatpolitikker

Kildekode, tekniske tegninger, prototyper, R&D-data

Interne retningslinjer, personalepolitikker, driftssystemer

Kontraktvilkår, forhandlinger, eksklusive partnerskaber



Intern LLM på Power Systems

Mølgaard Data A/S

- Linux LPAR oprettes på Power boksen
- Ledige/ubrugte ressourcer tildeles
- Linux LPAR tunes bedst muligt til at udnytte tildelt hardware

System & Metode A/S

- Podman installeres, som er nødvendig software for at køre AI
 - Vi har downloadet og kompileret open source software til afvikling af LLM'er, f.eks. llama.cpp, med understøttelse til at udnytte Power 10 hardware.
- AI modeller udvælges og downloades
 - Nogle modeller passer bedre til Power 10 end andre. Sigt efter int8 eller bf16 quantized modeller og i en størrelse, der kan passe med den mængde memory LPAR'en har fået tildelt.
- AI startes op med ønsket model og interface
 - Text-konsol, Web-UI, API til tredjeparts software
 - I vores tilfælde, har vi bygget den som en Podman container, hvilket er en Docker erstatning på AlmaLinux (Red Hat kompatibel)

PODMAN/DOCKER IMAGE
Vi tilbyder et image med vores anbefalede værktøjer og vejledning



podman



OpenAI



LLAMA 4



Demo af LLM på Power Systems

PODMAN/DOCKER IMAGE
Vi tilbyder et image med vores anbefalede værktøjer og vejledning

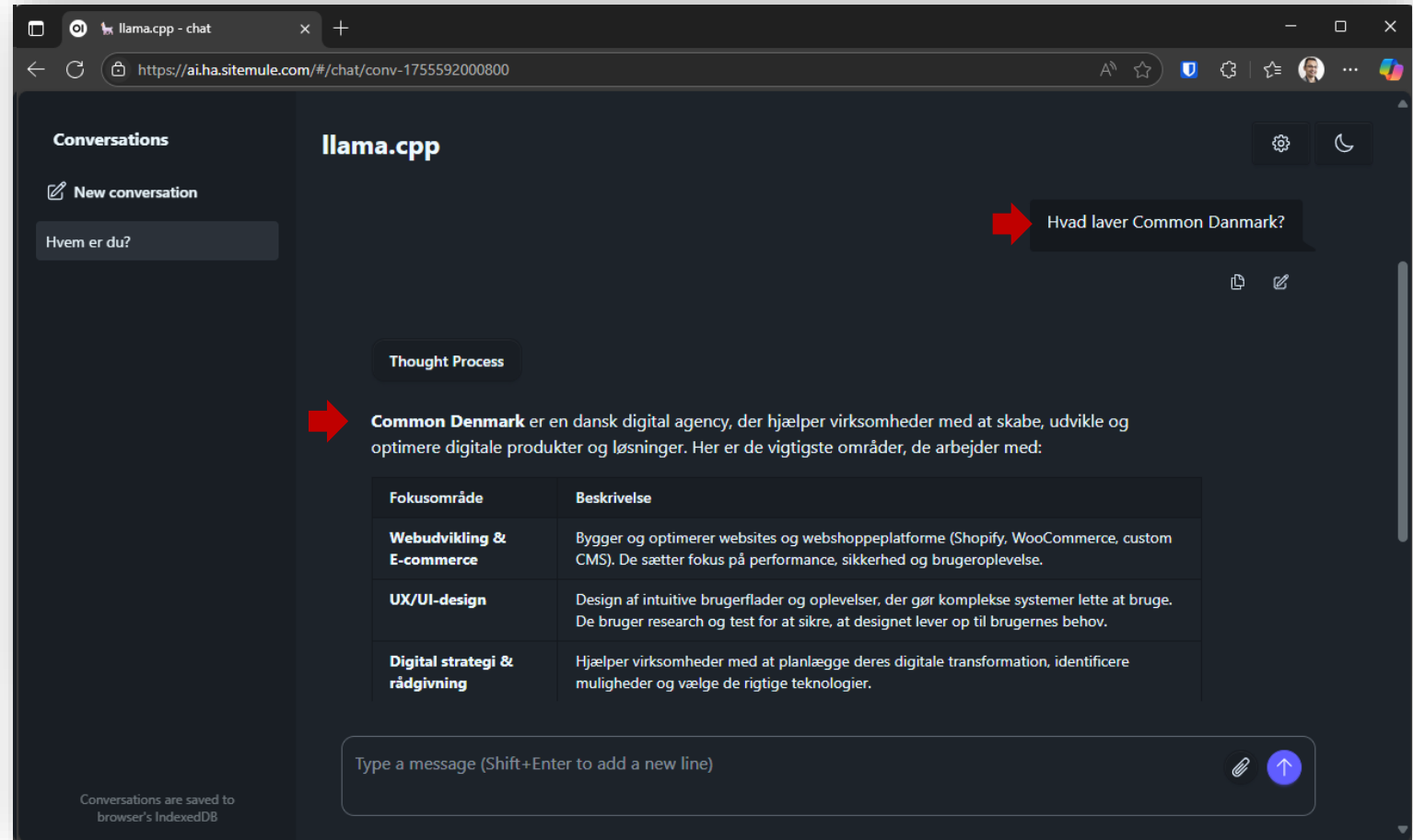
Hardware hos Mølgaard Data A/S

- Power 10
- Alma Linux 9
- 8 Cores
- 512 GB RAM
- 1 TB disk

Demo fra System & Metode A/S

- Podman Administration
 - <https://aiadm.ha.sitemule.com/podman>
- llama.cpp Cross-Platform LLM Framework
 - <https://ai.ha.sitemule.com/>
- Open WebUI Cross-Platform LLM Framework
 - <https://ai.sitemule.com/>

Så er vi klar til at tilføje **vores** viden!



Dataopsamling

- Informationer

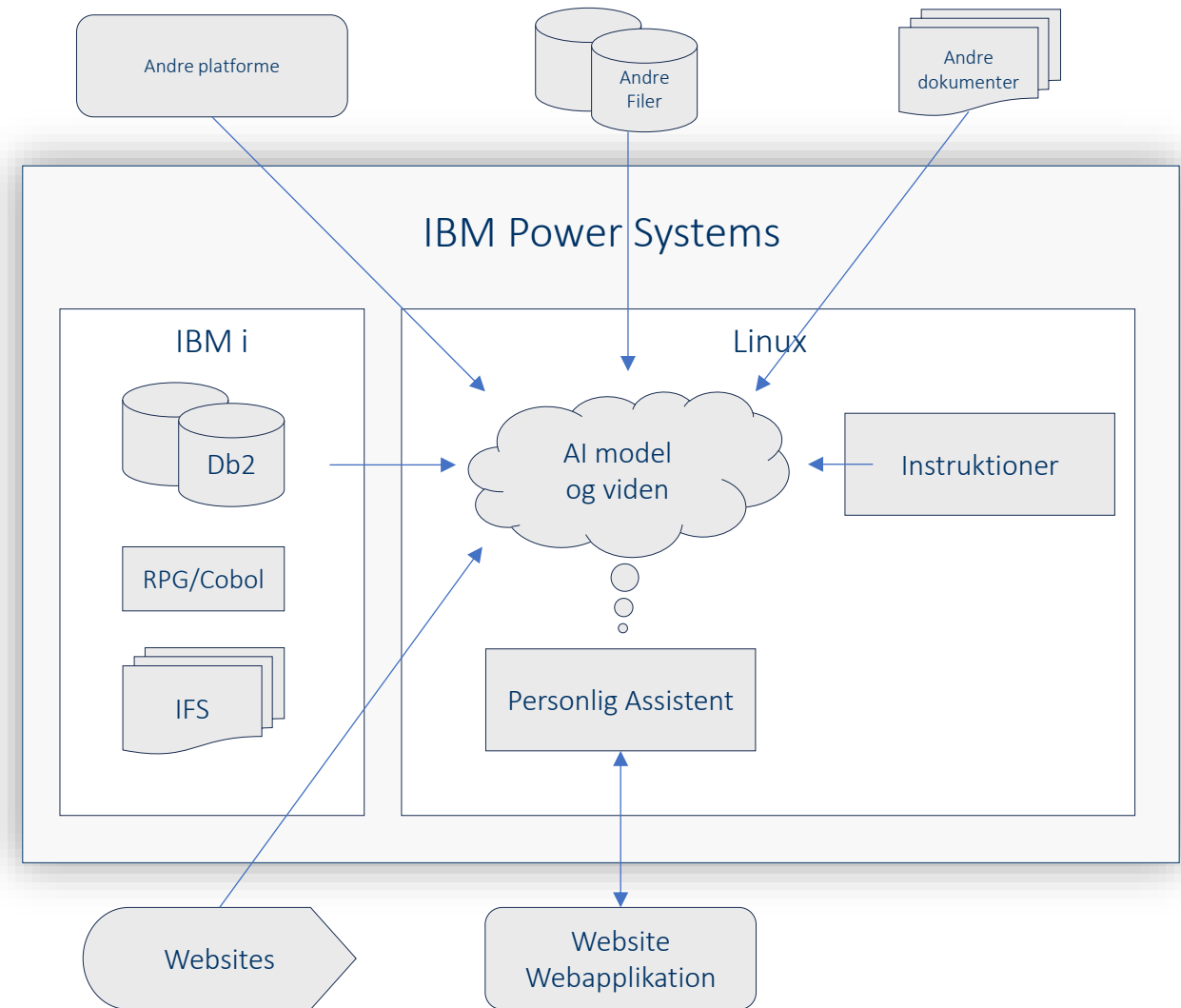
- Data (Db2)
- Programkald, som f.eks. priskalkulation
- Kildekode (RPG/Cobol/Java)
- Dokumenter, billeder og tegninger (IFS)

- Instruktioner

- Konfiguration
- Tone, Rolle, Begrænsninger
- Strukturer, Grupperinger
- Afsluttende budskab

- Øvrige informationskilder

- Hjemmesider
- Andre platforme, filer og dokumenter



Personlig Assistent

- Personlig Assistent

- Indholdsbevidst (Content-Aware)
- Drevet af AI, i dette tilfælde OpenAI
- Højt konfigurer- og tilpasningsbar
- Integration til hjemmesider og webapplikationer
- Log af spørgsmål/svar til FAQ og intern brug
- Progressiv web-app (PWA) på mobil

- Anvendelsesområder

- Besvarelse af mail i kundeservice eller ved teknisk support
- Spørgsmål til forretningsgange eller individuelle aftaler
- Indsigt i store informationsmængder
- Målrettet faglig vejledning



8 / 30 en WEB

Personlig Assistent i VS Code

Installer:

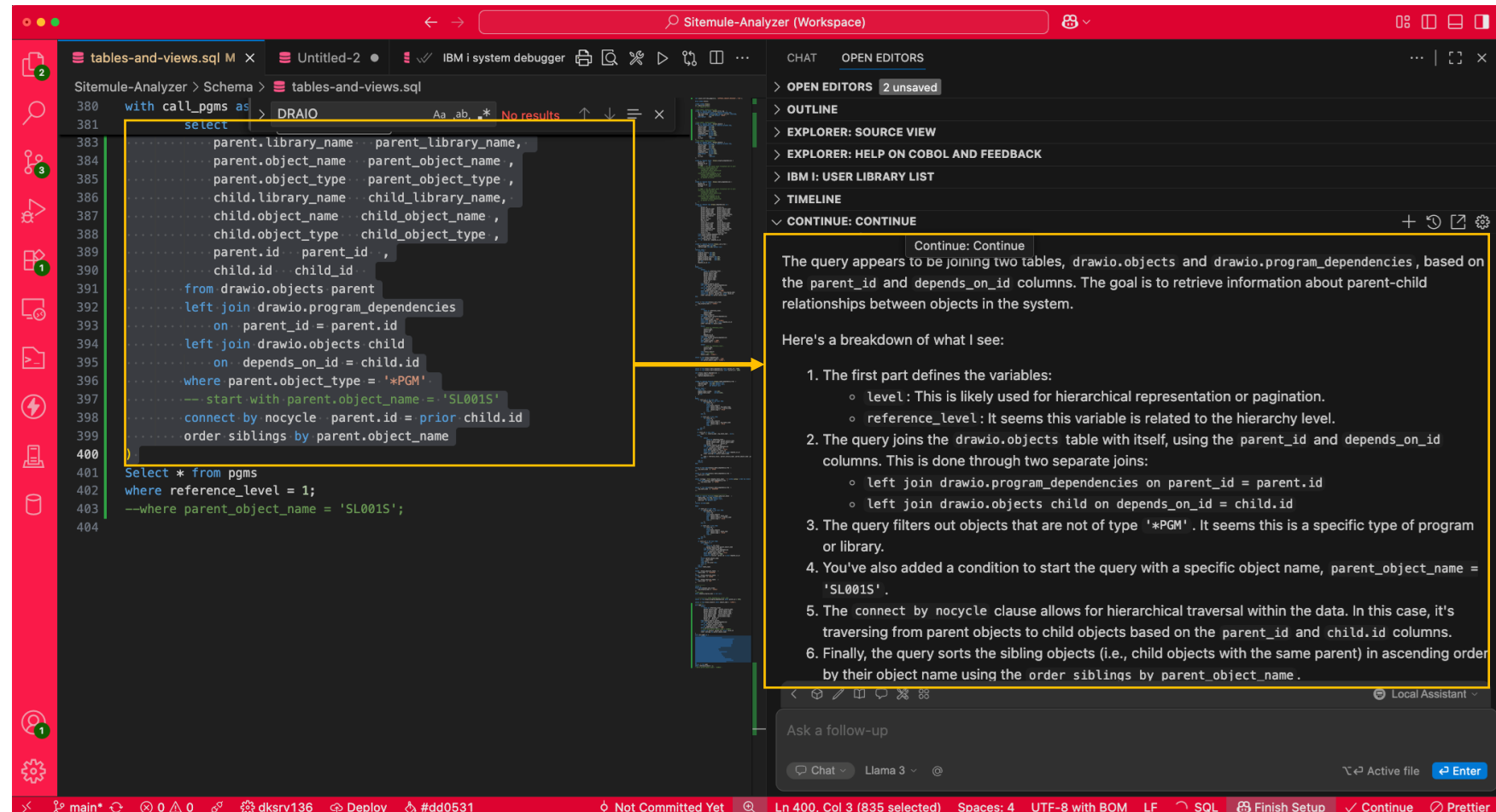
1. VS Code



2. Continue plugin



3. Ollama



Spørgsmål

Med introduktionen af **Power 11** har IBM igangsat følgende strategiske initiativer:

- Fremme en yderligere modernisering og ”normalisering” af IBM i-plattformen
- Styrke rekrutteringen af nye medarbejdere
- Understøtte migrering og implementering af nye workloads

