

GØR DATA KLAR TIL AUTOMATISERING OG AI

EN KANALRUNDFART I MODERNE DATA GOVERNANCE

HVEM ER JEG?



Jakob Kornum

Head of Dataplatform, KOLLAB

10+ års erfaring med:

- Analyser & forretningslogik
- Dataplatform
- Microsoft Fabric
- Data Governance & Adoption
- AI & ML



Find mig på LinkedIn

Jeg mangler altid nye venner

ET KONTROLLERET FØRSTE SKRIDT MOD EN MODERNE DATAPLATFORM

HVAD VI GENNEMGÅR

I den her session tager vi jer fra filosofi til praksis – helt jordnært.

1. Vi starter med at afmystificere AI: Hvad er det egentlig, og hvorfor opfører det sig, som det gør?
2. Dernæst kigger vi på den udfordring, mange organisationer står i: AI er kun så stærk som den data, den får. Dårlig data er ikke bare upræcis – den kan blive direkte risikabel.
3. Vi introducerer ontologi – ikke som teori, men som konkret data governance, der gør AI brugbar i jeres hverdag. I ser, hvordan strukturerede data, semantik og klare forretningsregler er forskellen på en AI, der gætter, og en AI, der leverer.
4. Til sidst giver vi en live demo i Microsoft Fabric, hvor vi viser, hvordan AI-agenter bliver markant bedre, når vi bygger lag af ontologi ind i løsningen.

→ Det bliver teknisk.

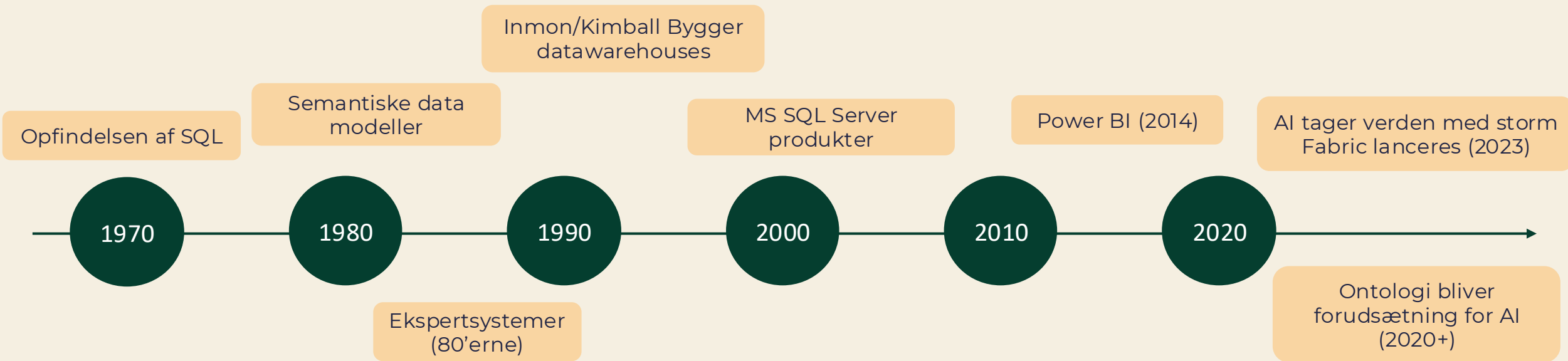
→ Det bliver hands-on.

→ Og I går herfra med konkrete greb, I kan bruge allerede i morgen.



FRA DATABASER TIL AI

EN 50-ÅRS ODYSSÉ



Alan Turing: "Can machines think?" (1950)
Dartmouth Workshop (1956): AI som forskningsfelt

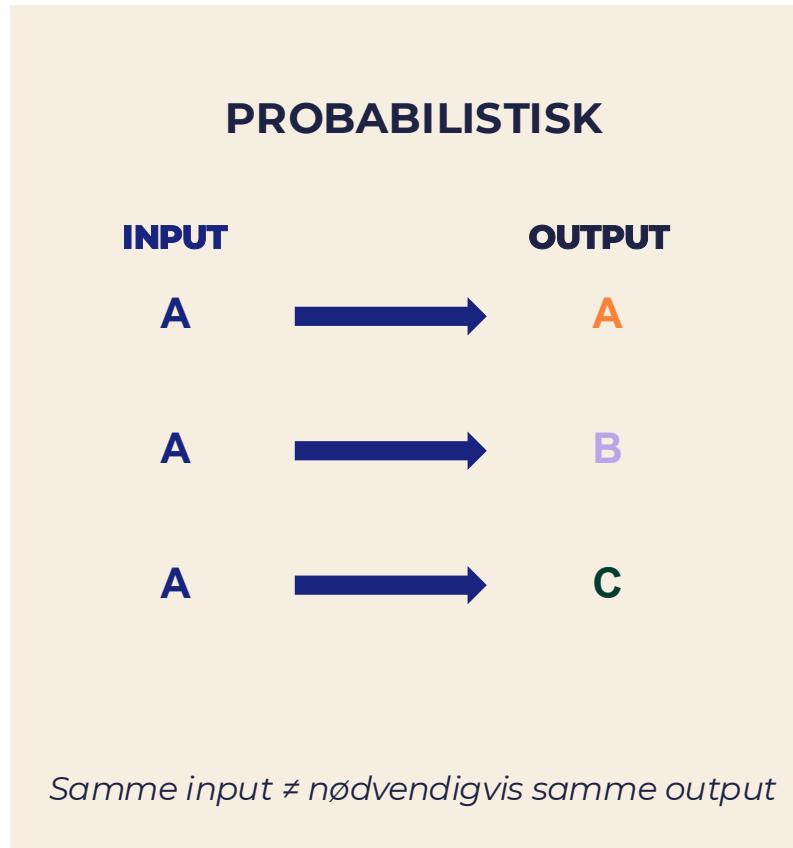
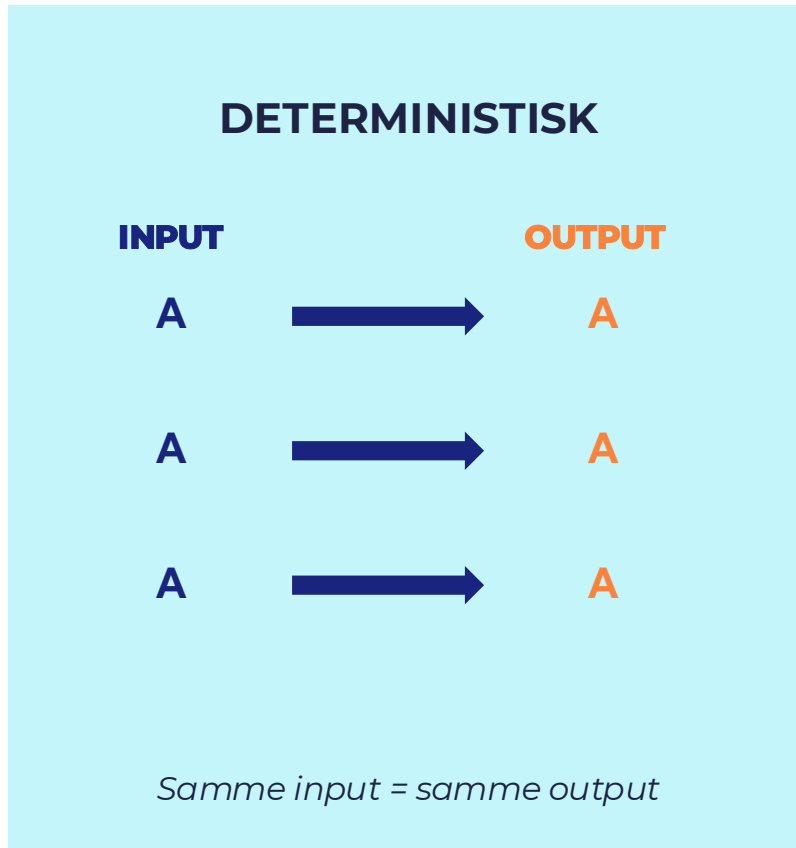


HVORDAN ARBEJDER MAN MED AI?



MODEL	+	PROMPT	+	KONTEKST	+	VÆRKTØJ
Hvem laver maden?		Hvad skal laves?		Opskrift & ingredienser		Redskaber til at handle
GPT-4, Copilot		Din instruktion		Data, regler, semantik		API'er, funktioner, adgang

AI ER IKKE DETERMINISTISK, DEN ER PROBABILISTISK



- Du kan ikke "debugge" AI som normal kode
- Du skal designe for variabilitet – test, valider, juster prompts
- Governance bliver kritisk fordi du ikke kan forudsige alt

”

**AI gør ikke dårlig data god, den gør dårlig data farlig.
Fordi AI kan præsentere forkerte svar med stor
selvtillid.**

**Derfor er data governance ikke længere 'nice to have' -
det er en forudsætning.**

”

GARBAGE IN, GARBAGE OUT

KLASSISK PROBLEM, NY DIMENSION

- **Før:** Dårlig data → dårlig rapport (som du kan se og rette)
- **Nu:** Dårlig data → AI hallucinerer eller træffer forkerte beslutninger (som kan virke overbevisende!)

HVORFOR DET ER KRITISK

- AI formidler ikke bare data – den fortolker og handler
- Hvis metadata er forkert, kan AI blande persondata med offentlige data
- Hvis semantik mangler, kan AI misforstå forretningsregler



AI HAR BRUG FOR "GUARDRAILS"

INPUT KONTROL

Hvad AI må få adgang til

Fx. må ikke læse HR data uden godkendelse.

PROCES KONTROL

Hvordan AI må tænke

Fx. Skal altid tjekke med semantisk model før svar

OUTPUT KONTROL

Hvad AI må gøre/sige

Fx. Må ikke dele persondata, må ikke sende emails uden review

HVORFOR DET ER VIGTIGT

Compliance

GDPR, sikkerhed, audit trails

Tillid

Medarbejdere skal kunne stole på AI's output

Skalerbarhed

Uden guardrails kan AI ikke køre autonomt

AI LÆRER IKKE (AUTOMATISK) FRA JERES DATA

MISFORSTÅELSE AT ADRESSERE

- Mange tror at "AI lærer af vores data over tid" – men:
- **LLM'er trænes ikke live** (GPT-4 er trænet én gang, derefter frossen)
- **Copilot lærer ikke** fra dine Fabric-data – den *bruger* dem som kontekst

HVAD DER FAKTISK SKER

- AI får **adgang til data i realtid** via kontekstvinduet
- Du kan **fine-tune** modeller (men det kræver ML-eksperise)
- Du kan **berige prompts** med metadata og ontologier (det er jeres vej!)



KONTEKST ER ALT – MEN DER ER GRÆNSER

Token-vinduet er ikke en skraldespand, det er et værktøj. Brug det strategisk.

TOKEN-BEGRÆNSNINGER I PRAKSIS

Microsoft 365 Copilot:

- Op til 300 sider til sammenfatning
- Maks 20 filer søges samtidig
- 3.000 ord til omskrivning

GitHub Copilot:

- 64k tokens (standard)
- 128k tokens (VS Code Insiders)

Mere kontekst ≠ bedre resultat.

AI performer bedre med præcis, relevant kontekst.

DERFOR SKAL DATA VÆRE

Struktureret

- så AI kan finde det relevante

Semantisk annoteret

- så AI forstår relationer

Klassificeret

- så AI ved hvad der er relevant

Eksempel: Beregn customer lifetime value

Uden: gætter · Struktur: får tal · Semantik: forstår ·
Klassificering: ved hvad der tæller

Kontekst former output mere end prompt. Din prompt siger hvad AI skal gøre. Din kontekst definerer hvordan AI forstår verden.



Fra Aristoteles til moderne videnskab

Ontologi (gr.): ontos (være) + logos (studie) = "studiet af det værende"

HVORDAN KATEGORISERER VI VIRKELIGHEDEN?



Aristoteles (~350 f.Kr.) systematik

- Hvad er noget? (substans)
- Hvilke egenskaber har det?
- Hvordan relaterer det sig til andet?
- Hvilke kategorier kan vi bruge?

Kaldte det "første filosofi"
Termet 'ontologi' kom først i 1600-tallet

ONTOLOGI PÅ TVÆRS AF OMRÅDER

BIOLOGI & MEDICIN

- Gene Ontology, taksonomier (Linné), sygdomsklassifikation (ICD), præcisionsmedicin

FILOSOFI & TEOLOGI

- Metafysik, eksistens, kategorier af væren

IT & DATA

- Knowledge graphs, semantic web, AI-systemer, data governance (2020+)

**Samme grundprincip:
Definer hvad noget ER, hvilke egenskaber det HAR, og hvordan det RELATERER til andet.**

ONTOLOGI I IT:

DET FÆLLES SPROG FOR DATA & AI

UDEN ONTOLOGI

- 'Kunde' betyder?
- Er 'client' og 'customer' det samme?
- Må AI dele kundedata?
- Skal en ordre have en kunde?

→ AI gætter og laver fejl

MED SEMANTIK

- 'Kunde' = juridisk enhed med CVR
- 'Ordre' = transaktion med beløb

Men stadig uklart:

- Hvordan hænger de sammen?
- Hvilke regler gælder?

MED ONTOLOGI

- **Kunde** *har* CVR, navn, adresse
- **Kunde** *placerer* → Ordre
- **Ordre** *skal have* en Kunde
- **Persondata** *må ikke* deles

→ AI forstår sammenhæng

Ontologi starter med det simple: Gode beskrivelser

Fra dårlig praksis til struktureret governance som AI kan bruge

✘ TRIN 1: DÅRLIG PRAKSIS

Kolonne	CUST_CRT_DT
Beskrivelse	[tom]

⚠ TRIN 2: GRUNDLÆGGENDE

Kolonne	CUST_CRT_DT
Beskrivelse	Customer creation date

✓ TRIN 3: AI-ASSISTERET

Kolonne	customer_created_date
Beskrivelse	Datoen hvor kunden blev oprettet i systemet. Udfyldes automatisk ved første ordre. Bruges til kundens anciennitet.

✅ TRIN 4: ONTOLOGI-KLAR

Begreb	Kundeoprettelsesdato
Definition	Tidspunkt hvor kunde første gang registreres
Relateret til	Kunde (har), Anciennitet (bruges til)
Klassifikation	Ikke-følsom
Regel	Kan ikke være fremtidig dato

OG SÅ EN DEMO...

OPSUMMERING

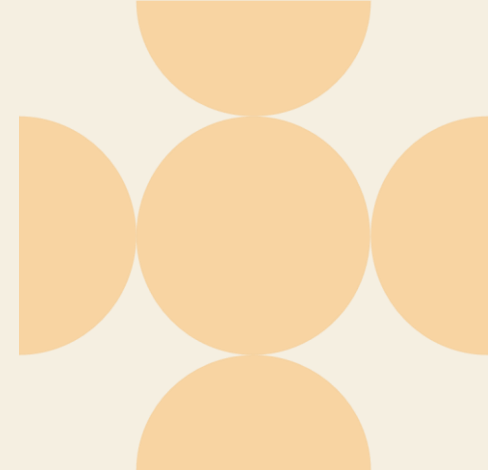
HVORFOR ER ONTOLOGIER PLUDSELIGT KRITISKE?

Vi startede med filosofi – Aristoteles' spørgsmål om hvordan vi kategoriserer virkeligheden. 2.400 år senere stiller vi det samme spørgsmål, bare til maskiner i stedet for til hinanden.

Her er sandheden: AI er probabilistisk, ikke deterministisk. Den gætter – velinformeret, men den gætter. Derfor er kontekst alt. Ikke prompt-engineering. Ikke model-valg. Kontekst.

Og her er løsningen: Ontologi er ikke et nyt buzzword. Det er data governance gjort rigtigt. Det er metadata, semantik, klassificering og forretningsregler – struktureret så både mennesker og AI kan forstå det.

Hvad det kræver: Start simpelt. Gode beskrivelser. Struktureret data. Semantiske relationer. Forretningsregler som kode, ikke som dokumentation i et skab.



MIN KONKLUSION



I har ikke brug for mere data. I har brug for bedre struktur på den data I har.

Det er forskellen mellem en AI der hallucinerer og en AI der leverer værdi.

Nu ved I hvorfor data governance ikke længere er "nice to have" – det er en forudsætning.





KOLLAB

kollab.dk

CVR: 28 12 60 93

+45 72 16 07 77

kontakt@kollab.dk