DTHS Synsfaglige Netværksdage 2025

Når AI beskriver verden forkert

- og hvad vi kan gøre ved det

Rasmus Paasch, IBOS

(Oplægget bliver også delt i "ADL-forummet" på DTHS Intranet)

Hvorfor dette emne?

Masterprojektet

Fokusområderne

- Hvordan oplever brugere AI-fejl?
- Hvilke strategier udvikler de?
- Hvordan skabes og genopbygges tillid?

Hvad viser forskningen i forhold til målgruppen?

Tre hovedpointer

1. **AI-fejl er forventelige** – ikke undtagelser
2. **Brugerne udvikler stabile strategier**
3. **Brugerne udvikler kalibreret tillid** – situeret, ikke binært - ikke enten/eller

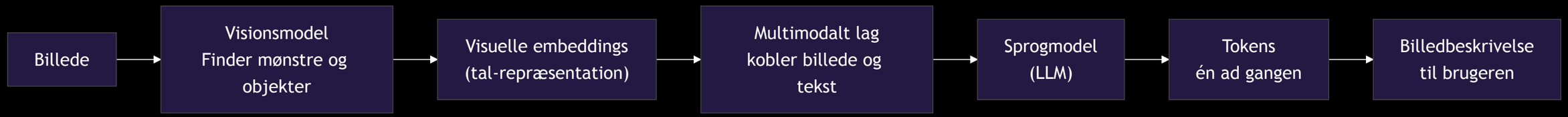
Hvordan "ser" AI et

De mange lag

- Billedet omsættes til tal (visuelle embeddings)
- Modellen identificerer sandsynlige objekter
- Sprogmodellen genererer tekst **én token ad gangen**

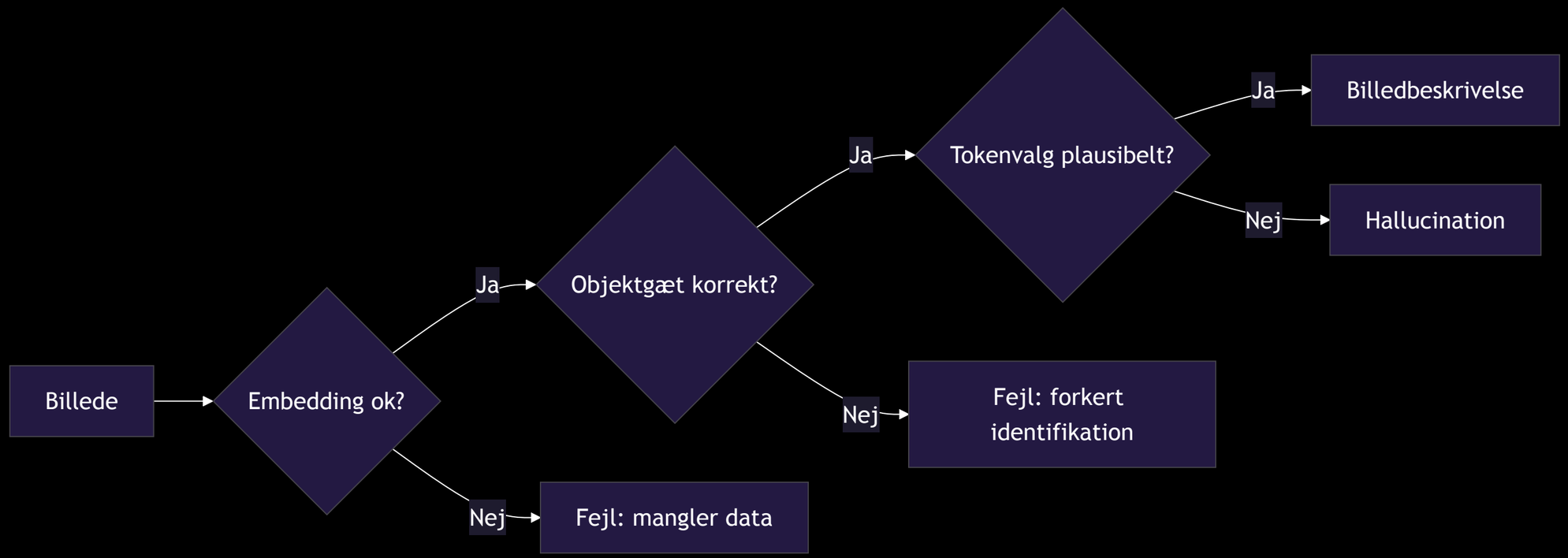
Hvordan "ser" AI et

Visualiseret som proces



Hvor opstår fejl?

Beslutningstræ: de tre typiske fejltyper



Fra billede til tal

Visuelle embeddings

- Farver, linjer, kontraster og objekter måles
- Ingen "synsoplevelse", kun numeriske mønstre
- En komprimeret, matematisk udgave af motivet

Fra tal til *sandsynlige* objekter

Billedmodellen gætter – statistisk

- "Dette kunne være en pakke pasta"
- "Her er tekst"
- "Der er noget tøj i venstre side"

Fra objekter til tekst

LLM'en skriver beskrivelsen –

- Modellen vælger det mest sandsynlige næste token
- Ikke nødvendigvis det mest korrekte
- Overbevisende tekst ≠ korrekt tekst

Hvorfor er dette sårbart?

Tre strukturelle svagheder

- Manglende eller dårligt billede → fejl i embedding
- Små fejl i objektgæt → vokser i tekstgenerering
- Tokens giver ingen naturlig måde at udtrykke tvivl på

AI er en gættemaskine

Ikke et leksikon

"Modellen henter ikke viden.
Den forudsiger næste token."

LLM’er er ikke et bibliotek, der slår op i bøger.
De er en studerende til eksamen uden bøger,
der gætter næste ord ud fra det, den engang har læst – ikke ud fra aktuelle data.

Intention er nøglen

Det, at AI svarer på noget andet end det, brugeren forsøger at få viden omkring…

- Brugeren tager ikke et billede for at blive underholdt
- Intentionen er næsten altid praktisk: finde, verificere, beslutte
- Fejl opstår, når AI misforstår formålet
- Output kan være “pænt”, men stadig ubrugeligt
- Klare, snævre opgaver → færre fejl og mere tillid

Seeing AI vs. Be My AI

To helt forskellige AI-intentioner

Seeing AI
- Scanner tekst, genkender objekter (trænet på relativ kort liste af objekter)
- Mange små, specialiserede modeller
- Lav risiko for hallucination
- Output er mere “kantet” men stabilt
- Fortæller ofte “jeg ved det ikke”

Be My AI
- Multimodal LLM med fri generering
- Én stor model → én stor tolkning
- Høj risiko for overkonfidens
- Output er glat, flydende og overbevisende
- Fortæller sjældent når den er usikker

Brugernes strategier

Tre typer

1. **Prøvende strategier** – "Jeg tester den igen"
2. **Kompensatoriske strategier** – flere billeder, vinkler, apps
3. **Sikkerhedsstrategier** – "Jeg tager ikke beslutningen alene"

Kalibreret tillid

Ikke enten-eller, men situeret

- Tillid justeres efter **funktion** – virker det teknisk?
- Tillid justeres efter **kontekst** – får jeg det, jeg har brug for?
- Tillid justeres efter **konsekvens** – hvad sker der, hvis den fejler?

"Man skal jo ikke stole blindt på noget eller nogen"

Et konkret værktøj

iOS-genvejen **Be My AI**

Hvorfor en genvej?

- Reducerer kognitiv belastning
- Ensretter output
- Guider brugeren aktivt
- Modarbejder "gætteri" hos AI'en
- Opdeler komplekse opgaver
- Virker med VoiceOver

Hent appen Genveje til iOS
Hent iOS Genvej





Grundidé

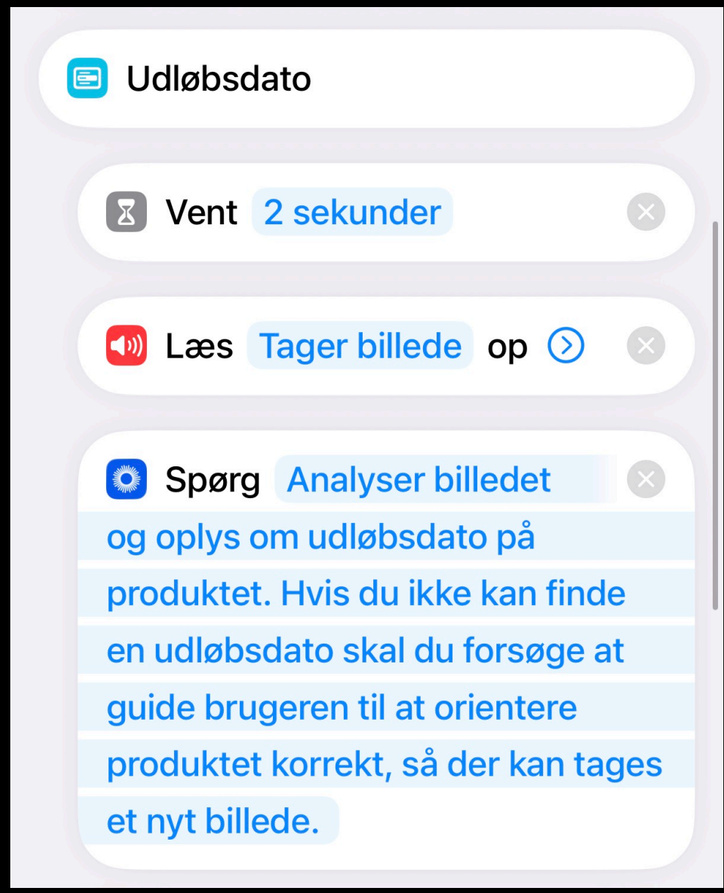
Målrettede opgaver

1. Udløbsdato
2. Ingredienser
3. Tilberedningstid
4. Beskriv genstanden
5. Farvetester
6. Stil spørgsmål



Genveje


11.24
Be My AI
Vælg fra menuen med Vælg handling med Be My AI
Udløbsdato
Ingredienser
Tilberedningstid
Beskriv genstanden i billedet
Farvetester
Stil spørgmål
Tilføj Nyt emne


Udløbsdato
Vent 2 sekunder
Læs Tager billede op
Spørg Analyser billedet
og oplys om udløbsdato på produktet. Hvis du ikke kan finde en udløbsdato skal du forsøge at guide brugeren til at orientere produktet korrekt, så der kan tages et nyt billede.

Eksempel: Udløbsdato

Prompt

Analyser billedet og oplys om udløbsdato på produktet. Hvis du ikke kan finde en udløbsdato skal du forsøge at guide brugeren til at orientere produktet korrekt, så der kan tages et nyt billede."

Eksempel: Farvetester

Output

"...Når en farve nævnes er det altid med objekts art først og så farven. Fx: Skjorten til venstre er blå. Objekter skal nævnes i størrelsesorden med det objekt der er størst som det første."

- Objekt før farve
- Sorteret efter størrelse
- Ingen baggrundsstøj
- Kort og præcist

Eksempel: Ingredienser

Med specifikt fokus på noget

"...fra det der er mest af, til det der er mindst af. Du er også opmærksom på allergener. Du udelader ikke noget og du oplyser brugeren, hvis der er ingredienser du ikke kan se eller som du er usikker på."

Aktiv guidning

- "Tager billedet" (tale)
- 2 sekunders forsinkelse
- Beder om nyt billede ved tvivl

Nu er det jeres tur

Prøv selv

- Tag et billede
- Brug Be My Eyes AI
- Brug genvejen
- Sammenlign

Hvad oplevede I?

Hvad kan betyde det for

- Vi kan undervise i *strategier*
- Vi kan skabe tryghed og transparens
- Vi kan forme rammerne for opgaveløsning – ikke kun teknikken
- Vi kan hjælpe borgere med at fejle tryggere

Konklusion

- AI fejler – og det er forventeligt
- Brugerne kompenserer allerede
- Intention er væsentlig
- Tillid skal kalibreres
- Vi kan designe flows, der reducerer fejl
- Genvejen er et eksempel på dette

Tak for jeres tid

Paasch, R. (2025). Når generativ AI fejler: En undersøgelse af blindes oplevelser, strategier og tillid [Aalborg University]. https://doi.org/10.5281/zenodo.15784525

Link til masteropgave

