

AI:NS GRÄNSER: VAD DEN FAKTISKT KAN, VILL OCH VÄGRAR GÖRA

# Varför AI ibland vägrar svara

9 min läsning

Grundläggande

Lektion 3 av 18

DEL 1 AV 6

# Översikt

- Du har säkert upplevt det.
- Du ställer en fråga, kanske lite känslig, kanske lite ovanlig, och istället för ett svar får du ett långt stycke om att AI "inte kan hjälpa med det".
- Och en rad alternativ du inte alls frågade efter.

DEL 2 AV 6

# Varför den vägrar

- AI-modeller som ChatGPT och Claude är inte bara tränade på text.
- De är också finjusterade av människor för att bete sig på ett visst sätt.
- Det enda du behöver veta är det här:

DEL 3 AV 6

# Den nervöse kollegan

- Tänk dig en kollega som vet svaret på din fråga.
- Men frågan är lite känslig, och de är rädda för vad chefen ska tänka om de svarar.
- Så de shufflar runt det.

DEL 4 AV 6

## Vad en vägran faktiskt ser ut som

- Det är inte alltid ett hårt "nej".
- Ofta ser det ut ungefär så här:

## Vad som faktiskt triggar en vägran

- Juridiska eller medicinska frågor. AI är rädd att ge råd som tolkas som professionell rådgivning
- Frågor som kan missförstås. "Hur skadar X" kan tolkas som skadligt, även om du undrar av nyfikenhet
- Politiska ämnen. Modellen undviker att ta sida

- Frågor om olagliga saker. Även om sammanhanget gör dem helt legitima (forskning, historia, skönlitteratur)
- Osäkerhet. Ibland väljer AI att vägra hellre än att chansa på ett svar den inte är säker på

DEL 6 AV 6

# Sammanhang förändrar allt

- "Hur länge kan man ta smärtstillande utan att det blir farligt?"
- "Jag är apotekare och en kund frågar om detta. Hur länge kan man ta receptfria smärtstillande utan att det blir farligt, och vilka signaler ska man v..."

# Tack för att du lärde dig med oss.

Nästa lektion: Hallucination: när AI hittar på. Fortsätt där du slutade på snabbprompt.se.

[snabbprompt.se](https://snabbprompt.se)



Scanna för att fortsätta