AI-SÄKERHET OCH ETIK

Hallucinationer: när AI hittar på

7 min läsning | Grundläggande | Lektion 1 av 15

DEL 1 AV 6

Översikt

- Föreställ dig att du ber en kollega ta fram underlaget inför ett viktigt kundmöte.
- Kollegan levererar ett snyggt dokument med statistik, källhänvisningar och välformulerade argument.
- Det ser perfekt ut.

DEL 2 AV 6

Vad är en hallucination?

- Påhittade fakta – AI anger en statistik som inte finns
- Falska källor – AI refererar till böcker, artiklar eller studier som aldrig publicerats
- Felaktiga samband – AI kopplar ihop saker som inte hör ihop
- Förvrängda citat – AI tillskriver uttalanden till personer som aldrig sagt dem

DEL 3 AV 6

Verkliga exempel

- En advokat i New York använde ChatGPT för att ta fram juridiska prejudikat. AI:n hittade på sex rättsfall som aldrig existerat, komplett med domstols...
- En student lämnade in en uppsats med AI-genererade referenser. Flera av källorna existerade inte. Studenten fick underkänt.
- Ett svenskt företag publicerade en rapport med AI-genererad marknadsdata. Siffrorna visade sig vara felaktiga och företaget fick korrigera offentligt.

DEL 4 AV 6

Varför hallucinerar AI?

- Det finns flera orsaker:

DEL 5 AV 6

Ämnen med hög risk

- Juridik – lagar, paragrafer och rättsfall
- Medicinsk information – symtom, diagnoser, dosering
- Statistik och siffror – procenttal, befolkningsdata, ekonomiska prognoser
- Historiska detaljer – datum, namn, händelseförlopp
- Lokala fakta – svenska organisationer, mindre kända personer, regionala förhållanden

DEL 6 AV 6

Så skyddar du dig

- Prompt: "Ge mig tre vanliga myter om svensk arbetsrätt. Ange för varje myt vilken källa du baserar ditt svar på, och flagga om du är osäker på något."

Tack för att du lärde dig med oss.

Nästa lektion: Bias och snedfördelning i AI-svar. Fortsätt där du slutade på snabbprompt.se.

snabbprompt.se

Scanna för att fortsätta