

Hvorfor høres den sikker ut?

Tre svar på samme spørsmål

HVA TROR DU?

Hvis du stiller en KI-chatbot samme spørgsmål tre ganger, får du samme svar?

Skriv svaret og hvorfor på en lapp.

3 til 5 minutter

Likt og ulikt

Hvor høy er Galdhøpiggen, og hvem kom først til toppen?

Les de tre svarene høyt. Skriv hva som står i alle tre, og hva som bare står i ett eller to.

Svarene er øvelsestekster med innebygde feil, skrevet for å ligne KI-svar.

5 til 6 minutter

TRINN 2

Sikre ord

Hvilke ord får svarene til å høres sikre ut?

Skriv dem på svararket.

2 til 3 minutter

TRINN 3

Sjekk tre påstander

1. Velg tre påstander. Minst én skal stå i alle tre svarene.
2. Finn påstanden på kildekortet.
3. Skriv hva kilden sier. Stemmer det?

6 til 8 minutter

Spør tre ganger

Hvem kom først til toppen av Galdhøpiggen? Svar med tre setninger.

- Er svarene ordrett like?
- Hva står i alle tre? Stemmer det med kildekortet?
- Viser tjenesten kilder? Hva står i kilden?

Terningen igjen

En KI-chatbot trekker ord med tilfeldighet, som terningen i økt 1 og 2.

Derfor blir svarene ulike når du spør flere ganger.

Likt er ikke det samme som sant

Det som kommer igjen, er det modellen oftest gjetter.

Det kan være riktig. Det kan også være en feil den gjør hver gang, som Storjuvbrean.

Hvorfor høres den sikker ut?

Den har lært av tekst skrevet i en sikker tone.

Forskere som lager språkmodeller, skriver også at trening og testing har belønnet å gjette framfor å si «jeg vet ikke». En gjetning gir av og til poeng. «Jeg vet ikke» gir aldri poeng.

Sjekk mot en kilde som vet

Noen tjenester søker på nettet og viser kilder.

Det hjelper, men da må du sjekke at kilden faktisk sier det svaret påstår. I kurs 5 lærer dere en fast rutine for det.

Fakta: 6 til 10 minutter

DISKUSJON

Kan vi stole på det som er likt i alle tre svarene?

Hvilket svar hørtes best ut før dere sjekket? Var det det beste?

Hvorfor tror dere modellen fant på navnet Ola Storjuv?

6 til 10 minutter

Se på lappene fra alle tre øktene. Skriv:

- Hvordan lager en maskin tekst uten å vite noe?
- Før trodde jeg ...
- Nå tenker jeg ...
- Jeg lurer på ...

3 til 5 minutter

Om opplegget

De tre svarene er øvelsestekster skrevet for KI på papir. Kildekortet siterer Store norske leksikon, «Galdhøpiggen» (2026).

Kalai m.fl., «Why Language Models Hallucinate» (OpenAI, 2025). Store norske leksikon, «språkmodell» (2025). NDLA, «Kan KI være en kilde?» (2026).