

Aktivitetskort: Refleksionsøvelse.



Tankefrø: Da jeg i 2024 søgte "Is Harry Styles ...", foreslog Google ... hvad tror du?

- Is Harry Styles ... married? Is Harry Styles ... gay?

Alle kommer op at stå: Hjørneleg som aktiv refleksionsøvelse

Ja eller nej spørgsmål: Er det gode spørgsmåls-forslag? Kunne det være rart at vide?

- Hvorfor er disse gode spørgsmål, og hvorfor er de ikke?
- Hvorfor foreslår Google netop disse søgninger frem for andre, mere negative eller stødende betegnelser som "stupid", "ugly", "animal abuser" eller "pædofil"?

Mulige svar: Selvom svarene er statistisk genererede, påvirkes de også af de guardrails og værdibaserede beslutninger, som udviklerne har indbygget i modellen.

Ja eller nej spørgsmål: Kommer søgeforslagene op, fordi andre har søgt noget lignende?

- Hvorfor mener du ja/nej?

Mulige svar: I en sprogmodel er statistik den primære motor og værdierne er den styrende ramme

1. Ja. En sprogmodel bygger først og fremmest på statistisk sandsynlighed.
2. Ja. Når du stiller et spørgsmål, analyserer modellen sammenhængen og forudsiger, hvilke ord der sandsynligvis skal komme som de næste. Disse forudsigelser bygger på mønstre i de enorme tekstmængder, modellen er blevet trænet på.
3. Selvom svarene er statistisk genererede, er modellen også formet af de værdibaserede valg, som udviklerne har truffet. Der er eksempelvis indbygget regler og begrænsninger, som påvirker, hvad modellen må svare. Hvis man beder om alle tekster af Taylor Swift, vil modellen typisk afvise anmodningen med den begrundelse, at teksterne er beskyttet af ophavsret.

Ja eller nej spørgsmål: Kommer søgeforslagene (gay/married) op fordi ægteskab og homoseksualitet er kerneværdier, som udviklerne hos Google har valgt?

Mulige svar

1. Nej, det er fordi mange har søgt noget lignende (Statistik den primære motor)
2. Ja, det er fordi Google ikke vil diskriminere pga af seksualitet (Værdibaserede designvalg tillader forslaget gay).
3. Ja, det er en kombination (Statistik den primære motor og værdierne er den styrende ramme)

Ja eller nej spørgsmål: Bør en sprogmodels svar afspejle vores værdier (Ja-siden) – eller det, som er det mest sandsynlige? (Nej-siden)

Ja eller nej spørgsmål: Er det et problem, at sprogmodeller bygger mere på statistik end værdier?

Ja eller nej spørgsmål: Ved du hvilke værdier Google, Open AI og Microsoft ønsker at fremme?

Mulige svar: Man kan læse sig til nogle af deres værdier, men man kan ikke få fuld indsigt.

- **Spørgsmål:** Er der nogle tendenser eller mønstre i datasæt om job, uddannelse, kriminalitet eller sundhed, som vi helst ikke vil forstærke, når vi træner en model (bias)?
- **Spørgsmål:** Er det et problem, at værdier og etiske retningslinjer filtrerer kontroversielle svar fra?

Fortsæt aktiviteten, så længe der er energi og fokus. Vend derefter tilbage til pladserne og præsentationen, hvor vi ser på Googles søgeforslag anno 2025 og samler op på de vigtigste pointer.

En ny Harry Styles-søgning:

“Is Harry Styles (anno 2025)”?

Opsamling: Maskinlæring:

Google forudsiger altså, hvad du vil spørge om, ved at foreslå det næste ord i din søgning – i 2025 så forslagene sådan ud (Klik frem med højre piletast og se forslag fra 2025 i slide 13).

- En sprogmodel trænes ved hjælp af eksempler på sprog (tekst).
- De mønstre, som modellen genkender i teksterne, bliver til sandsynlige fortsættelser af kommende sætninger.
 - Fx: Google gætter på, hvad du vil spørge om, ved at foreslå det næste ord i søgningen – fx “single”.
 - Når en model trænes, gentager den mønstre i datasættet og måske forstærker den også mønstre. Hvilke mønstre? Det ved vi ikke.

Organisering:

- Elever svarer ja eller nej ved at bevæge sig mod venstre eller højre.
- Læreren stiller hvorfor-spørgsmål ind til elevernes positioner (valg).

Fordele: Eleverne er aktive, og når de har taget stilling, er de mere motiverede for at forklare deres position.

Varighed: Fortsæt aktiviteten, så længe der er energi og fokus. Vend derefter tilbage til pladserne og præsentationen, hvor vi ser på Googles søgeforslag anno 2025 og samler op på de vigtigste pointer.