# Fleip eller fakta? – Mangfold og skjevhet i kunstig intelligens

Kunstig intelligens kan virke nøytral og objektiv, men det den lærer, kommer fra mennesker og samfunnet rundt oss. Derfor kan KI-systemer gjenspeile skjevheter, stereotypier og diskriminerende mønstre – bevisst eller ubevisst. I denne aktiviteten skal kan du teste elevenes kunnskap og bruke resultatene som utgangspunkt for refleksjon og diskusjon i klassen.

## Slik gjør du:

1. Les opp påstandene én og én
2. Elevene svarer individuelt eller i grupper: Fleip eller fakta?
3. Elevene forklarer hvert svar
4. Diskuter hva som overrasket og hvorfor det er viktig å vite dette
5. Avslutt med refleksjon: Hvordan kan vi gjøre KI mer rettferdig?

## **Påstander:**

Her er påstandene med forslag til forklaringerer som kan brukes i klassesamtale, oppsummering eller som støtte til refleksjon og diskusjon. Elevene kan bruke denne oversikten for å vurdere egen forståelse og lære mer om hvordan skjevhet og representasjon påvirker KI-systemer.

|  |  |  |
| --- | --- | --- |
| Påstand | Fleip eller fakta | Forklaring |
| KI er laget av mennesker, og kan derfor inneholde menneskelige fordommer. | Fakta | KI lærer fra menneskeskapte data og kan derfor gjenta eller forsterke skjevheter. |
| KI forstår alltid kontekst, språk og kultur like godt. | Fleip | KI har ofte problemer med flertydighet, minoritetsspråk og kulturelle referanser. |
| KI representerer alle grupper i samfunnet like mye. | Fleip | Datasett er ofte skjevfordelte, med overrepresentasjon av vestlig og hvit kultur. |
| Skjevhet i KI kan føre til at noen ikke får jobb, lån eller medisinsk hjelp. | Fakta | Det finnes dokumenterte tilfeller der algoritmer har diskriminert basert på kjønn eller hudfarge. |
| KI kan trenes opp til å bli mer rettferdig og inkluderende. | Fakta | Bevisst utvikling og variert datagrunnlag kan bidra til bedre representasjon. |
| KI velger selv hvilke bilder og data den vil lære av. | Fleip | KI velger ikke selv – utviklere bestemmer treningsdataene. |
| De fleste KI-verktøy er utviklet av mangfoldige team med representasjon fra hele verden. | Fleip | KI utvikles ofte av relativt ensartede miljøer i Vesten, noe som kan gi snevre perspektiver. |
| Samiske navn og minoritetsspråk blir ofte feil forstått av KI-systemer. | Fakta | Mindre språk er ofte dårlig representert i treningsdata, noe som gir svakere resultater. |
| KI er like flink til å beskrive en afrikansk kvinne som en hvit mann. | Fleip | Flere studier viser at bilde-KI har høyere feilrate på mørke ansikter og minoritetsgrupper. |
| KI-systemer har blitt avslørt i å gi diskriminerende anbefalinger. | Fakta | Algoritmer har i praksis forsterket ulikheter, f.eks. i ansettelse, rettsvesen og kredittvurdering. |