

retninger. Dette fordi data generelt gjenspeiler sosiale skjevheter, og en maskinlærende KI kan fort ende opp med å videreføre og forsterke slike. En algoritme som skulle skille ut foto av menn fra foto av kvinner, identifiserte menn som var avbildet i kjøkken som kvinner. Dette fordi datagrunnlaget hadde flere kvinner enn menn i slike omgivelser (Mitchell, 2019, s. 126).

KI for gjenkjenning er generelt mer nøyaktige når det er hvite menn som skal identifiseres, og ikke kvinner eller fargede personer. Dette fordi de internettbaserte bildedatabasene som ofte benyttes i treningen av slike systemer, består av en overvekt av foto av menn og av hvite personer (Mitchell, 2019, s. 124).

Sosiale skjevheter sniker seg ikke bare inn i bildedatabaser, men i mange typer data. Data er nemlig heller ikke alltid nøytrale, men kan også være verdiladete. Det amerikanske rettsvesenet vurderer straffeutmåling og prøveløslatelse ut fra en kalkulert risiko for tilbakefall. Som beslutningsassistent brukes noen ganger en algoritme som er trent på tidligere saker. Den har imidlertid vist seg å feilvurdere systematisk ut fra hudfarge. For eksempel ble den fargede Brisha Borden (uten nevneverdig rulleblad) vurdert som høy-risiko etter et avbrutt sykkeltyveri, mens den hvite Vernon Prater (med atskillig flere oppslag i sitt rulleblad) ble vurdert som lav-risiko etter å ha stjålet verktøy til cirka samme verdi som sykkelens (\$ 80). To år etter hadde ikke Borden gjort nye forbrytelser, mens Prater satt fengslet for åtte år etter et innbrudd (Angwin et al., 2016).

I tillegg til faren for å videreføre sosiale skjevheter er det også et problem at maskinlæringen kan foregå på *fake news* dersom

datagrunnlaget ikke er administrert (samtidig er identifikasjon av og kampen mot *fake news* og *deep fake* noe som KI kan brukes til (Cassauwers, 2019)). Som vi ser tydelig i forbindelse med ChatGPT, så skjønner ikke KI-er hva som er sann og usann informasjon. Dermed er det en fare for at falske nyheter og misinformasjon kan spres raskere enn noensinne (Hsu & Thompson, 2023).

Ansiktsgjenkjenning har blitt en viktig kilde for identifikasjon, og er en lettvinnt og en sikker metode, men den kan også benyttes til det som i et demokratisk perspektiv er beklagelig. I Henan-provinsen i Kina blir utenlandske journalister, studenter og andre personer som styresmaktene mener er tvilsomme, overvåket i stor skala (NTB, 2021). I Xinjiang-regionen benyttes ansiktsgjenkjenning for å overvåke uigurer, en hovedsakelig muslimsk befolkningsgruppe som blir systematisk undertrykket (mange er internert i såkalte omskoleringsleire). Kameraer finnes overalt, inkludert i taxier, og er koblet til sikkerhetspolitiet. Dette programmet er trent til å kunne skille uigurer fra koreanere og tibetanere (Noorden, 2020).

Militær bruk av KI, i droner og autonome våpen, er også omstridt. Skal en KI kunne ta en avgjørelse om liv og død? En soldat som løfter hendene over hodet for å overgi seg, kan i stedet oppfattes som å løfte et våpen (Haugen, 2018). Hvem har ansvaret for en slik feil? En KI vil ikke kunne oppleve psykologiske mekanismer som angst, sinne og hevngjerrighet, og er slik sett en «bedre» soldat enn et menneske. Men er det etisk sett positivt eller negativt? Vil det spare soldater for traumer og ødelagte liv, eller vil det senke terskelen for å ta liv?