Manus, Kunstig intelligens

Slide 1-3

Spørsmål:

Hva tenker dere på når jeg sier kunstig intelligens?

Har dere brukt kunstig intelligens?

Slide 4

Introduksjon

I denne presentasjonen skal dere lære om ulike typer kunstig intelligens, utfordringer og muligheter med KI, og hvordan dere kan bruke kunstig intelligens i utarbeidingen av deres kampanje mot hatprat!

Slide 5

Hva er KI?

Kunstig intelligens, ofte forkortet til KI eller AI, er når datamaskiner og programmer lærer å gjøre ting som vanligvis krever menneskelig intelligens.

Hvis noen spør, eller man har tid til å snakke om det:

Det vi har nå, er noe som kalles for 'Narrow AI', på norsk 'Smal KI'. Smal KI er laget for en spesifikk oppgave, som for eksempel en ChatBot, eller en bildegenerator. De fleste typene med KI, per i dag, faller inn i denne kategorien.

Generell KI betyr at maskiner kan lære og gjøre alt et menneske kan. Den har samme intelligensnivå som mennesker, men finnes bare i teorien og er ikke utviklet ennå.

Slide 6

Det finnes mange måter å snakke om KI på, men her har jeg delt opp kunstig intelligens etter funksjon. De tre som er mest brukbare for oss her i dag er generativ KI, kategorisk KI og prediktiv KI.

Slide 7

Generativ KI genererer innhold som tekst, bilder eller video. Dette gjør den basert på mønstre den har lært fra treningsdata. Treningsdata er informasjon hentet fra for eksempel internett, som bilder, tekst, bøker, og video som vi eller profesjonelle har lastet opp.

Måten den genererer nytt innhold, er ved å regne ut det mest sannsynlige neste elementet i en sekvens – enten det er ord i en setning, eller piksler i et bilde. På mange måter er KI en gigantisk kalkulator.

Her er et eksempel på et bilde, generert med KI.

Slide 8

Klassifiserings KI plasserer allerede eksisterende innhold inn i kategorier. Disse kategoriene har man bestemt på forhånd, og lært maskinen hvordan den skal sortere.

Se for eksempel for deg en skittentøyskurv full av klær som skal sorteres. Du sier til maskinen:

“Alt som er ull skal i én haug.”

“Alt som har farger skal i en annen haug.”

“Alt som er hvitt skal i en tredje haug.”

Sånn er det også med klassifiserings-KI. Du lager kategorier, og så lærer du maskinen hvordan den sorterer, og så prøver den så godt den kan å sortere med den informasjonen.

Klassifiserings-KI brukes blant annet i spam-filteret i e-posten din. Den har fått beskjed om hva som trolig er spam, og sorterer etter denne informasjonen. Face-ID på mobilen brukes også klassifiserings-KI.

Slide 9

Prediksjons-KI er når maskinen bruker mønstre fra data den har sett og lært, til å prøve å forutsi hva som kommer til å skje videre.

Den regner ut sannsynligheten for hva som trolig vil skje videre, og så gjør den noe for å hjelpe/hindre det i å skje.

Et eksempel på dette er når du spiller sjakk mot datamaskinen. KI-en prøver å forutsi hva du kommer til å gjøre, og hindre deg i å vinne.

Slide 10

Spørsmål: Hvilken type KI tror dere er brukt her?

Svar: Generativ KI

Slide 11

Spørsmål: Hvilken type KI tror dere er brukt her?
Svar: Klassifiserings-KI

Slide 12

Spørsmål: Hvilken type KI tror dere er brukt her?
Svar: Prediksjons-KI

Slide 13

Nå skal vi snakke litt videre om muligheter og utfordringer med kunstig intelligens. Men før vi gjør det vil jeg understreke at teknologi i seg selv aldri er vondt eller godt.

Det er litt som en hammer.

Du kan bruke den til å slå inn en spiker, eller slå ned naboen. Hammeren er fortsatt bare en hammer, uavhengig av hvordan du bruker den.

Slide 15

Nå skal dere få en gruppeoppgave.

EU sin KI-forordning har laget en pyramide med ulike grader av risiko knyttet til bruk av KI. Snart skal dere få noen eksempler, som dere skal sortere etter hvor dere tror de hører til på risikopyramiden.

På toppen er KI-bruk som har uakseptabel risiko. Dette betyr at KI IKKE skal brukes til dette.

Høyrisiko er når KI kan brukes, men med strenge regler og krav.

Begrenset risiko er når KI kan brukes, men alle må være klar over at du bruker/konsumerer KI.

På minimal risiko kan KI brukes uten mange regler/krav, fordi det å bruke KI her utgjør lite til ingen fare.

(GJØRE OPPGAVE, 10 minutter, de presenterer hva de har tenkt på 5 min til sammen)

FASIT:

Uakseptabel risiko: Risikovurdering, kriminell handling individ

Høyrisiko: KI for å gi karakter på en prøve og KI for å ansette folk til en jobb

Begrenset risiko: KI for å snakke om dagen sin

Minimal risiko: KI i spill og sortere e-post

Regler og tanker rundt ting som dette kan endre seg med tiden, og det er ikke sikkert vi tenker likt om 10-20 år.

Slide 18

Det er mange utfordringer med kunstig intelligens. Hva tror dere kan være utfordringer?

Slide 19

KI er forferdelig for klimaet.

Det er fort gjort å tenke at alt som skjer på internett er klimavennlig, fordi det ikke er på papir, ikke skaper søppel og liknende. Men alt som skjer på mobilen din, eller datamaskinen, har et CO2-utslipp.

Når du lagrer et bilde i skyen, så ligger den ikke på en magisk sky på himmelen, men på en dataserver i et datasenter som krever masse energi for å drives og vann for å kjøles ned.

En forespørsel til ChatGPT krever energien av 10 vanlige google-søk (før Google AI).

Ett gjennomsnittlig google-søk slipper ut 0.2g CO2.

For **tusen** google søk kan man kjøre en vanlig diesel/bensinbil i én kilometer, men for tusen ChatGPT forespørsler kan man kjøre en hel mil!

Slide 20

Å generere et bilde koster like mye strøm som å lade mobilen sin fra 0-100%.

Hver dag genereres over 34 millioner KI-bilder.

Med dette dagsforbruket av energi, kunne en dieselbil ha kjørt ca. 14,2 ganger rundt jordkloden!

Slide 21

Det finnes ingen sky.

Alt du gjør på internett har et CO2-avtrykk, enten det er å søke på noe, lagre noe, eller strømme en film. Kunstig intelligens krever store mengder strøm, og datasentrene har store CO2 utslipp og krever enorme mengder vann for å kjøles ned.

Slide 22

Generativ KI har blitt brukt til å lage seksuelle og voldelige bilder/videoer, av både privatpersoner og kjendiser.

Dette kaller man gjerne for "Deep fakes". En "deepfake" er falsk video, bilde eller lyd, generert med kunstig intelligens, for å få det til å se eller høres ut som om noen har gjort noe de aldri har gjort.

Kunstig intelligens brukes også til å spre kommentarer og innhold på internett, for eksempel hatprat og trolling.

Slide 23

Dette er et eksempel på en deepfake. Noen har laget en falsk video for å lure folk til å kjøpe et slankeprodukt. De har brukt stemmen til en kjent skuespiller og fått det til å høres ut som hun sier det som står i manuset. Videoen er klippet slik at man nesten ikke ser at leppene ikke passer helt til ordene.

Slide 24

Bruk av KI kan føre til diskriminering.

Nå skal dere få en liten case, hvor dere skal tenke ut hvordan dere tror KI-bruk kan ende opp med å diskriminere.

Slide 25

Mini-Case: En bedrift skal kutte ned 20% av de ansatte.

For å gjøre dette, vil de bruke en KI-modell, som har bestemt at en god arbeider er en som alltid kommer på tiden.

Hvordan kan dette fungere diskriminerende?

Svar: Det er ingen fasitsvar på hva en god arbeider er, men KI-en har blitt fortalt at en god arbeider er en som kommer på tiden.

De som har kort vei til jobben, har lettere for å komme på jobb i tide.
Folk som er funksjonsfriske har også lettere for å ankomme jobben i tide.

Folk som bor langt fra jobben og må pendle, for eksempel med tog fordi de ikke har råd til bil, har større sannsynlighet for å komme for sent.

Det er typisk for folk med mindre økonomiske midler å bo utenfor byen, og å måtte pendle.

Denne klassifiseringen vil slå ut på folk med mindre penger, eller med funksjonsvariasjoner.

Mennesker med immigrantbakgrunn har gjennomsnittlig mindre penger og bort gjennomsnittlig lenger unna arbeidsplassene sine.

På denne måten kan immigranter, folk med immigrantbakgrunn eller bare fattigere mennesker komme dårlig ut av det, men også mennesker med funksjonsvariasjoner, dersom bedriften bruker en slik KI-løsning.

(Fra rapport, Study on Discrimination, artificial intelligence and algorithmic decisionmaking).

Slide 26

Andre eksempler på diskriminering:

E-poster fra mennesker med ikke-vestlige navn kan havne i spam.

En jobbannonse for konstruksjonsarbeid som kjøres på META med automatisk publikum, kan bli rettet kun mot menn og forsterke stereotypier.

Hvorfor? Når treningsdataen har bias, vil KI-modellen reprodusere dette menneskelige biaset.

Slide 27

KI eksisterer kun på grunn av dataen den trenes på. Dataen er menneskeskapt, og i stor grad hentet fra internett.

Slide 29-32

Eksempler på hva som kommer opp hvis jeg prompter bilder.

Hvis jeg ber om en kirurg, får jeg en mannlig kirurg.

Hvis jeg vil ha en kvinnelig kirurg, må jeg be om det spesifikt, og den kvinnelige kirurgen fikk "full glam" sminke.

Månedens arbeider på sykehuset ble en mann, og den beste kirurgen på sykehuset ble også en mann.

Slide 33

Bias hos KI kan altså føre til

- Diskriminering i ansettelsesprosesser
- Forsterking av skadelige stereotypier

Slide 34

Det er også viktig å huske at du ikke eier et bilde du har generert.

Når du ikke eier det du lager, kan andre ta det og bruke det i andre kontekster.

KI har også en spesiell, visuell stil. Å bruke KI-genererte bilder kan svekke troverdigheten din.

KI er trent på andre sitt åndsverk, uten at de har samtykket til dette.

Slide 35

Derfor skal vi ikke bruke KI til å generere bilder eller video til kampanjen vår.

Vi skal heller ikke få KI til å lage noe fra scratch, men heller som et hjelpemiddel!

Slide 36

Spørsmål: Hva tror dere kan være fordeler?

Hvordan kan KI hjelpe dere/andre?

Slide 37

- Prediktiv KI skåner flere fra å se hatefullt eller voldelig innhold på sosiale medier, ved å moderere automatisk.
- KI kan lese opp artikler for deg som lydbøker, eller gi deg sammendrag hvis du vil sjekke om en artikkel er relevant. Men husk å lese artikkelen og sjekk at det stemmer.
    - **Fakta er et verdifullt verktøy mot misinformasjon og hatprat!**

Slide 38

- Kan korrekturlese tekstene dine
- Peke på løse tråder, si hvor du kan forklare ting enklere
- Vise hvordan teksten din kan bli misforstått

Hvis du skriver et leserinnlegg, kan den for eksempel vise hvor du har svakheter i argumentasjonen din. Da kan du rette opp i det, og gjøre teksten sterkere før publikasjon.

Slide 39

- Studier viser at kunstig intelligente programmer kan fungere som gode samtalepartnere, spesielt for samtaler om vanskelige ting.
- Noen foretrekker svarene/vennskapene gitt av en KI over menneskelige svar/vennskap (Brandzæg, et al., 2022)
- Alltid tilgjengelig, aldri sur, aldri nedbrutt.
- Aldri dømmende (utenom når treningsdataen er biased).

Slide 40

- Kunstig intelligens er ikke partisk (utenom når dataen den er trent på har bias).
- Hjelpe deg med et nøytralt svar på en hatkommentar
- Responsen fra KI kan inspirere svaret ditt og hindre videre provokasjon/krangel.