



**KUNSTIG INTELLIGENS
– MULIGHETER, UTFORDRINGER
OG EN PLAN FOR NORGE**



Teknologirådet

KUNSTIG INTELLIGENS

– MULIGHETER,
UTFORDRINGER OG EN PLAN
FOR NORGE

ISBN 978-82-92447-94-9 (trykket utgave)
ISBN 978-82-92447-95-6 (elektronisk utgave)

Utgitt: Oslo, september 2018
Trykk: Allkopi
Forside: Birgitte Blandhoel
Elektronisk publisert på: www.teknologiradet.no



FORORD

Kunstig intelligens har gjort et kraftig sprang de siste årene. Maskinene kan nå lære seg å tolke tekst, tale og bilder. Dermed kan avanserte oppgaver som hittil har vært forbeholdt mennesker, gjøres raskere og rimeligere av maskiner. Det gir store muligheter både for verdiskaping og bedre velferdstjenester, samtidig som teknologien kan påvirke innbyggernes rettigheter og skape ulikhet.

Denne rapporten fra Teknologirådet beskriver hvordan maskiner lærer, og bruksområder og utfordringer ved teknologien. Rapporten argumenterer for at Norge trenger en strategi for kunstig intelligens, og kommer med 14 forslag som bl.a. går på hvilken kompetanse vi trenger, hvordan data om oss skal brukes og hvilken utvikling vi ønsker for samfunnet.

Ekspertgruppen for dette prosjektet har bestått av:

- Erik Fosse, kirurg og leder for intervensjonssenteret ved OUS
- Siri Hatlen, tidligere direktør ved OUS og leder for Teknologirådet
- Steinar Madsen, fagdirektør i Legemiddelverket
- Hans Olav Melberg, helseøkonom og førsteamanuensis ved UiO
- Damoun Nassehi, fastlege og medlem i Teknologirådet
- Michael Riegler, forsker ved Simula og seniorforsker ved OsloMet

En spesiell takk til Michael Riegler for bidrag til kapitlene om maskinlæring og bruksområder. Hilde Lovett i Teknologirådet har ledet prosjektet.

Teknologirådet skal gi uavhengige råd til Stortinget og øvrige myndigheter om ny teknologi. Vår ambisjon er at rapporten vil bidra til en kunnskapsbasert og fremtidsrettet debatt om muligheter og utfordringer med kunstig intelligens.

Tore Tennøe
direktør, Teknologirådet

INNHold

SAMMENDRAG	7
------------------	---

EN VÅRLØSNING FOR KUNSTIG INTELLIGENS	14
---	----

BEDRE ALGORITMER	17
------------------------	----

STORE MENGDER DATA	19
--------------------------	----

TILGANG PÅ REGNEKRAFT	20
-----------------------------	----

SLIK LÆRER MASKINER	22
---------------------------	----

VEILEDET AV DATA	22
------------------------	----

IKKE-VEILEDET	25
---------------------	----

LÆRING MED FORSTERKNING	27
-------------------------------	----

BLANDINGSMODELLER	29
-------------------------	----

Delvis veiledet	29
-----------------------	----

Overført læring	30
-----------------------	----

Genererende konkurrerende nettverk	32
--	----

BRUKSOMRÅDER – FRA LEGEVAKT TIL SELVKJØRENDE BILER	34
--	----

ORDNE OG FORUTSI	34
------------------------	----

Klassifisering	35
----------------------	----

Klyngeanalyser	35
----------------------	----

Identifisere avvik	36
--------------------------	----

Forutseende analyser	36
----------------------------	----

TALE- OG LYDFORSTÅELSE	37
------------------------------	----

Oversette tale til tekst.....	37
Naturlig brukerdiallog	37
Oppdage faresignaler i helsemålinger	38
TEKSTFORSTÅELSE.....	38
Oversette språk	38
Kundestøttesystemer	39
Digital førstelinje.....	39
Gjennomgang av pasientjournaler	40
BILDE- OG VIDEOANALYSE.....	40
Gjenkjenne objekter	41
Diagnostisering.....	41
Autonome kjøretøy	41
Forbedre bilder og video	41
ANBEFALINGSSYSTEMER.....	42
Skreddersy tilbud	42
Forebygge trafikkulykker	43
Forutsi sykdom	43
Tilpasset undervisning	43
Persontilpasset behandling	43
Individuelle treningsprogram	44
LISTE OVER BRUKSOMRÅDER.....	44
KAN VI STOLE PÅ MASKINENE?	47
FORDOMSFULLE ALGORITMER	48
SORT BOKS-PROBLEMET.....	49
Retten til en forklaring	51
Hvem kan holdes ansvarlig?	52
ETISKE ALGORITMER.....	53
ONDSINNET BRUK	54

NORGE TRENGER EN STRATEGI.....57

Store muligheter	57
Viktige konsekvenser for den enkelte og samfunnet	58
En teknologi i rask utvikling – og Norge ligger bak	59

KOMPETANSEUTFORDRINGEN62

1. Et umiddelbart forskningsløft	62
2. Etablere en nøkkelinstitusjon	63
3. Peke ut ambisiøse og konkrete mål for Norge	63
4. Mastergrader forsterket med kunstig intelligens	65
5. Gi alle mulighet til å lære kunstig intelligens	65

NORGES FORTRINN: DATA66

6. Åpne offentlige data	67
7. Datadeling som tjener fellesskapet	67
8. Gi innbyggerne reell kontroll over egne data	68

ANSVARLIG OG ØNSKET UTVIKLING69

9. Etske retningslinjer	70
10. Rett til en forklaring	71
11. Krav om åpne algoritmer i offentlig sektor	72
12. Revidere algoritmer	73
13. Innebygd etikk	74
14. Nasjonal dialog om kunstig intelligens	75

SAMMENDRAG

Kunstig intelligens (KI) har gjort et kraftig byks de siste årene. De fleste av oss bruker det daglig, når vi søker på nettet, navigerer i trafikken, oversetter tekster, bruker talekommandoer på smarttelefonen, eller filtrerer bort useriøs e-post.

Tilgang til betydelige mengder data, kraftige regneressurser og ikke minst fremskritt i algoritmer, spesielt nevrale nett, har gjort kunstig intelligens til en av de viktigste muliggjørende teknologiene i dette tiåret.

Kunstig intelligens er drevet frem av et ønske om å gjøre maskiner i stand til å løse både fysiske og kognitive oppgaver som tidligere var forbeholdt mennesker. Tidligere var programmerte, regelstyrte ekspertsystemer den rådende disiplinen, men utover 2000-tallet gikk fagområdet over til å bli statistikk- og datadrevet, og maskinlæring som tilnærming tok over. Datamaskiner kunne lære uten å være eksplisitt programmert.

Slik lærer maskiner

En algoritme er en oppskrift som forteller hvordan noe gjøres, og kan sees på som instruksjonene til et dataprogram. I en maskinlæringsalgoritme har datamaskinen selv laget noen av instruksjonene.

Datamaskiner kan nå lære seg sammenhenger, regler og strategier fra erfaringer i data fra den virkelige verden, uten at noen forteller dem hvordan sammenhengene er. De kan tilpasse seg dataene kontinuerlig, og jo mer data de har tilgang til, jo mer presise blir de (adaptivitet). Dette gjør at datamaskiner kan utføre oppgaver på egenhånd (autonomi). Komplekse oppgaver og beslutninger kan slik overtas av maskiner, og gjøres raskere og rimeligere.

Maskinlæringsalgoritmer lærer hovedsakelig på tre måter:

1. *Veiledet læring*: Algoritmen lærer med veiledning av erfaringer i et datasett. Det kan være å bestemme om en e-postmelding er useriøs, avgjøre om et bilde av en føflekk er godartet eller ondartet, eller å forutsi om en kunde vil si opp sitt mobiltelefonabonnement. Dette er den tilnærmingen som brukes mest i dag, men den trenger gode data.
2. *Ikke-veiledet læring*: Algoritmen finner selv nye mønstre og sammenhenger i et datasett. Eksempler er å gruppere kunder som ligner slik at de kan adresseres med ulike typer kampanjer, eller å oppdage flere undergrupper av sykdommer, slik at de ulike pasientene kan få bedre tilpasset behandling. Denne tilnærmingen har potensial til å oppdage mønstre som tidligere ikke var kjent for mennesker.
3. *Forsterket læring*: Algoritmen finner beste strategi for å nå et mål ved å prøve, feile og bli korrigert underveis. Datamaskinen kan slik lære seg å vinne i et sjakkspill, eller å optimalisere energiforbruket i et datasenter. Denne teknikken har potensial til å finne smartere strategier enn mennesker kan.

Bruksområder

Maskinlæring brukes til å lage prediksjoner. Enkelt sagt så handler prediksjoner om å fylle ut manglende informasjon. Prediksjoner tar den informasjonen man har, det vil si data, og bruker den til å generere informasjon man ikke har. Det kan være informasjon om fortiden, nåtiden eller fremtiden, som f.eks. å oppdage om en kredittkorttransaksjon var falsk, bestemme om en føflekk er ondartet eller å forutsi hva været blir i morgen.

Det finnes flere teknikker for å predikere, og de vanligste teknikkene er:

1. *Klassifisering* er den mest brukte maskinlæringsteknikken og brukes til å bestemme hvilken kategori en ny observasjon sannsynligvis tilhører. Det kan dreie seg om å identifisere hva eller hvem som befinner seg på et bilde, om en e-post er spam eller ikke, eller den mest sannsynlige diagnosen utfra symptomene til en pasient.
2. *Klyngeanalyse (clustering)* brukes for å utforske nye datasett uten å kjenne til sammenhengene på forhånd. Den finner nye strukturer og mønstre i (umerkede) data og deler dem inn i grupper eller klynger basert

på egenskaper som ligner. Teknikken kan brukes for å gruppere filmkunder som ligner på hverandre slik at de kan få tilpassede filmanbefalinger.

3. *Teknikker for å identifisere avvik (anomalideteksjon)* oppdager hendelser som ikke samsvarer med et forventet mønster i et datasett. Avvikene kan være forsøk på banksvindel, datainnbrudd, en uheldig sykdomsutvikling eller forstyrrelser i økosystemet.
4. *Forutseende analyser (forecast)* handler om å kunne forutsi, med en viss sannsynlighet, noe som vil kunne skje i fremtiden, basert på en serie med historiske data. Slike analyser kan brukes for å lage en risikoprofil for en person, som sannsynligheten for at en person vil falle ut av skolen, vil være i stand til å nedbetale en gjeld av en viss størrelsesorden eller utvikle en gitt type sykdom.

De ulike teknikkene for å predikere kan kombineres på ulike måter, og gjør at maskiner nå både kan høre og se, tolke og forstå. Det gjør maskinlæring til et kraftig og anvendelig verktøy innenfor mange bruksområder:

Teknikker for tale og -lydforståelse oversetter tale til tekst og omvendt. Dette er nå i daglig bruk på mobiltelefonen i virtuelle assistenter, og kan gjøre det enklere å håndtere datasystemer og forenkle rutineoppgaver.

Teknikker for tekstforståelse finner mening i ustrukturert, skriftlig informasjon. Oversettelsessystemer har blitt betraktelig forbedret og kundestøttesystemer er i ferd med å bli hel- eller halvautomatiserte ved hjelp av disse teknikkene.

Bildeanalyse- og videoanalyse gjenkjenner objekter i bilder og video. Teknikkene har hatt store fremskritt de senere årene, og kan nå automatisere meget avanserte og ressurskrevende oppgaver som bilkjøring og bildediagnoser.

Anbefalingssystemer brukes for å gjøre risikovurderinger for f.eks. ulykke eller sykdom og for å persontilpasse f.eks. undervisning, helsetjenester og andre oftest fentlige tjenester.

Kan vi stole på maskinene?

Kunstig intelligens påvirker allerede mange valg som tas av enkeltmennesker og organisasjoner. Desto viktigere blir det at vi kan stole på og forstå de anbefalingene algoritmene gir oss. Det er imidlertid knyttet flere utfordringer til måten maskiner lærer på:

- *Fordomsfulle algoritmer:* Veiledet læring gjør at maskiner kan bestemme kategorier eller forutsi utfall mer og mer presist, men anbefalingene kan aldri bli bedre enn dataene de baserer seg på. Maskiner lærer av data som er samlet inn om samfunnet. De kan speile skjeve forhold og dermed ta fordomsfulle beslutninger. Et eksempel er systemer som vurderer jobbsøknader og peker ut de beste kandidatene. Når algoritmene trenes på data fra tidligere ansettelser, vil de kunne preges av skjeve valg og praksis fra intervjurundene.
- *Sort boks-problemet:* Ikke-veiledet læring gjør at maskiner kan identifisere nye mønstre og sammenhenger i datasett, men de kan ikke nødvendigvis forklare den kausale sammenhengen. Algoritmene kan være lite gjennomsiktige og vanskelige å forstå, noe som også kalles sort boks-problemet. Mangel på forklaring gjør det både vanskelig å klage på en beslutning og å påta seg ansvar for beslutningene.
- *Etiske algoritmer:* Læring med forsterkning gjør at maskiner kan komme frem til optimale strategier for nå sine mål, men de vil ofte vinne for enhver pris og overkjøre hensyn som etikk, hvis de ikke blir eksplisitt programert.
- *Ondsinnnet bruk:* Med maskinlæring kan ondsinnede angrep gjøres bedre og mer effektivt, og de kan skaleres og spres raskt. Samtidig kan slik bruk fremme anonymitet og psykologisk distanse. Bruk av kunstig intelligens kan også introdusere nye og uløste sårbarheter. Dataangrep kan dermed gjøres rimeligere og mer målrettet og true digital sikkerhet. Fysiske objekter som droner og selvkjørende biler kan manipuleres og true fysisk sikkerhet. Den politiske sikkerheten kan trues bl.a. ved at desinformasjon (*fake news*) kan utformes mer troverdig og persontilpasset.

Forslag til en strategi for Norge

Kunstig intelligens gir store muligheter både for verdiskaping og bedre velferdstjenester, samtidig som teknologien kan berøre samfunnssikkerhet, innbyggernes rettigheter og skape ulikhet.

En lang rekke land som Finland, Storbritannia, Frankrike og Kina har utviklet egne strategier for kunstig intelligens, og det pågår et kappløp om å kapre de beste talentene og etablere sentrale posisjoner.

Dette taler for at Norge bør ha en egen KI-strategi. En nasjonal strategi bør adressere kompetanseutfordringen, behovet for data og en ansvarlig utvikling. Teknologirådet har følgende konkrete forslag til en slik strategi:

1. *Et umiddelbart forskningsløft:* Det er nødvendig med spesialisert, forskningsbasert kompetanse for å utvikle robuste algoritmer for maskinlæring. Langtidsplan for forskning og høyere utdanning bør rettes opp med en kraftig satsing på kunstig intelligens og maskinlæring når planen skal revideres i løpet av høsten 2018.
2. *Etablere en nøkkelinstitusjon:* Norge har for få og spredte forskningsressurser. For å styrke forskningsinnsatsen og bli attraktivt for rekruttering og internasjonalt samarbeid, kan det være en god idé for norske myndigheter å etablere en nøkkelinstitusjon for forskning på kunstig intelligens og maskinlæring. For å favne bredt nok og samtidig ha tilstrekkelig forskningshøyde kan institusjonen omfatte flere miljøer i en virtuell organisering.
3. *Peke ut ambisiøse og konkrete mål for Norge:* Norge har ikke forutsetninger til å satse bredt på kunstig intelligens, men kan bli verdensledende på å koble domenekunnskap med generell kunnskap om KI. Norge bør formulere satsinger innenfor områder der vi har en kombinasjon av gode treningsdata og store samfunnsmessige behov, slik som helse, offentlige tjenester, bærekraftig energi og rent hav.
4. *Mastergrader forsterket med KI:* Maskinlæring vil bli en viktig ingrediens i mange bransjer og yrker, fra produksjon, olje, og energi, media og underholdning, jord- og havbruk, medisin, undervisning og offentlige tjenester. Alle profesjoner og utdanningsløp bør inkludere en innføring i kunstig intelligens og maskinlæring. Det bør også etableres dedikerte mastergradsprogram i kunstig intelligens.
5. *Gi alle mulighet til å lære kunstig intelligens:* Kunstig intelligens vil påvirke livene våre og de valgene vi gjør, både privat og i yrkeslivet. Norge bør sette ambisiøse mål, og i første omgang at f.eks. én prosent av befolkningen skal lære seg grunnleggende kunstig intelligens. Ifølge OECD vil hver tredje jobb få kraftig endret innhold fremover på grunn av automatisering og

kunstig intelligens.¹ Dette krever at Norge legger om dagens etter- og videreutdanning og gjør den tilpasset den enkelte med nye insentiver.

6. *Åpne offentlige data:* Åpne offentlige data kan bidra til innovasjon og nye tjenester i mange sektorer. Offentlig sektor i Norge bør ha ambisjoner om å publisere mer offentlige data, og gjøre det på et åpent format som er lett å finne frem i og gjenbruke i maskinlæring.
7. *Datadeling som tjener fellesskapet:* Hvis data fra norske sykehus, skoler og smarte byer skal deles med tredjeparter, bør fellesskapet få en merverdi i form av bedre offentlige tjenester, ny næringsutvikling, arbeidsplasser eller skatteinntekter. Det er derfor nødvendig at myndighetene etablerer juridiske rammer som gjør det mulig å utveksle data trygt, og som sikrer at rettigheter, verdier og ansvar fordeles rettferdig og likeverdig.
8. *Gi innbyggerne reell kontroll over egne data:* Hvis offentlige data om oss deles for å drive forskning og innovasjon, forutsetter det at innbyggerne får et ord med i laget. Myndighetene må derfor etablere en tydelig digital samfunnskontrakt som gir innbyggerne reell mulighet til å kontrollere og forme sin digitale profil og bestemme om og hvordan egne data skal deles.
9. *Etiske retningslinjer:* Tradisjonelle europeiske verdier blir utfordret i takt med digitaliseringen generelt og utviklingen av kunstig intelligens spesielt. Regjeringen bør begynne å utvikle etiske retningslinjer og praksiser der teknologien allerede setter etablerte verdier som autonomi, demokrati, rettferdighet, likeverd, solidaritet og ansvarlighet under sterkt press.
10. *Rett til en forklaring:* Siden algoritmer gir råd og i økende grad tar beslutninger på områder med stor betydning for folks liv, blir det også viktig å kunne få en forklaring, slik at det blir mulig å klage på beslutningen. Norske myndigheter bør lovfeste en rett til forklaring, og unngå å bruke beslutningssystemer som ikke gir en god nok forklaring.
11. *Krav om åpne algoritmer i offentlig sektor:* Når maskiner overtar oppgaver som tidligere ble gjort av mennesker, er det spesielt viktig å kunne vise at algoritmene ikke gir fordomsfulle anbefalinger. Algoritmer som brukes av det offentlige bør som hovedregel være åpne for innsyn og kontroll, slik

¹ Nedelkoska og Quintini 2018.

at andre samfunnsaktører kan kontrollere at de blir brukt korrekt og etisk forsvarlig.

12. *Revidere algoritmene:* Algoritmer for maskinlæring som av tungtveiende grunner ikke er åpne for alle, bør likevel bli gjenstand for en vurdering før de kan bli tatt i bruk i stor skala i samfunnet. En mulighet er å kreve revisjon eller sertifisering fra en uavhengig tredjepart, der det vurderes om beslutningene til algoritmen er rettferdige, korrekte, forklarbare og etterprøvbare.
13. *Innebygd etikk:* Uønskede hendelser som fordomsfulle eller urettferdige avgjørelser kan føre til svekket tillit, og det er vanskelig og kostbart å rette opp i etterkant. Praksisen med innebygd personvern bør utvides slik at det allerede fra designstadiet blir vurdert om algoritmen kan føre til diskriminering eller manipulasjon.
14. *Nasjonal dialog om kunstig intelligens:* Kunstig intelligens er i ferd med å komme tett på livene til nordmenn flest, og det holder ikke å jobbe for passiv aksept. Norske myndigheter bør aktivt ta initiativ til å involvere lekfolk og sivilsamfunnet i diskusjonen om kunstig intelligens og ta imot råd om hva som er ønsket utvikling.

EN VÅRLØSNING FOR KUNSTIG INTELLIGENS

Kunstig intelligens har gjort et kraftig sprang de siste årene. Store mengder data, kraftige regneressurser og utvikling av bedre algoritmer har lagt grunnlaget for utviklingen. Den mest lovende retningen nå er nevrale nett, som er inspirert av hvordan menneskehjernen fungerer.

Forsiden av Nature 2. februar i 2017 viste til algoritmer som kan lære seg å klassifisere føflekker like godt som leger. Oppslaget var resultat av et samarbeid mellom medisinerere fra Stanford-universitetet og forskere på kunstig intelligens (KI) som Sebastian Thrun, mannen bak Google-bilen.² Gruppen hadde trent opp et nevralt nett på kliniske bilder av føflekker, før det ble testet mot 21 sertifiserte hudleger på 2000 bilder. I nesten hver test viste algoritmen seg å være mer følsom og treffsikker enn spesialistene: Den fanget opp flere faktiske tilfeller av føflekkreft, samtidig som den ga færre falske positive.

Dette er et stort fremskritt i seg selv. I Norge er føflekkreft den nest vanligste kreftformen i aldersgruppen 25–49 år, og over 2000 pasienter får diagnosen årlig.³ Men forfatterne bak Nature-artikkelen hadde mer å by på. Den samme

² Esteva m. fl. 2017.

³ Kreftforeningen 2018 og Folkehelseinstituttet 2018.

typen maskinlæring kan også tilpasses og brukes i andre medisinske spesialiteter, som øre-nese-hals, øye, radiologi og patologi. I tillegg til at metoden er rask, er det mulig å bruke mobiltelefoner og nettbrett i diagnostiseringen.

Dette er bare ett av flere eksempler på nyvinninger innenfor maskinlæring de siste årene. IBMs Watson har vunnet Jeopardy, Apple lar oss snakke med Siri på smarttelefonen, den førerløse Google-bilen har kjørt millioner av kilometer, og Facebook gjenkjenner ansikter like godt som mennesker. Dette gjør at mange nå snakker om at vi nå er inne i en vårløsning for kunstig intelligens, etter en KI-vinter der utviklingen ga få praktiske resultater.

Hva er kunstig intelligens

Kunstig intelligens betyr forskjellige ting for ulike mennesker, og selv ikke forskere er enige om en eksakt definisjon. Kunstig intelligens er imidlertid drevet frem av et ønske om å gjøre maskiner i stand til å løse både fysiske og kognitive oppgaver som tidligere var forbeholdt mennesker.

Turingtesten

Alan Turing definerte i 1956 en standard for hva som må til for at en maskin skal kunne kalles intelligent, også kalt Turingtesten. Testen gikk ut på å la en person kommunisere med en datamaskin eller et annet menneske ved hjelp av kun tastatur og skjerm, men hun kan ikke se hva eller hvem som svarer. Kommunikasjonen kan være om alle mulige tema, og vare mange timer. Hvis personen ikke kan avgjøre om hun kommuniserer med en maskin eller et menneske, sier Turing at maskinen har bestått testen og må regnes som intelligent.⁴

Tidligere var programmerte, regelstyrte ekspertsystemer den rådende disiplinen. Et eksempel er IBMs DeepBlue, som slo verdensmesteren Kasparov i sjakk i 1997. Slike systemer er imidlertid lite fleksible, gjerne utviklet for spesifikke domener og lite robuste til å takle hendelser som ikke er spesifisert i reglene.

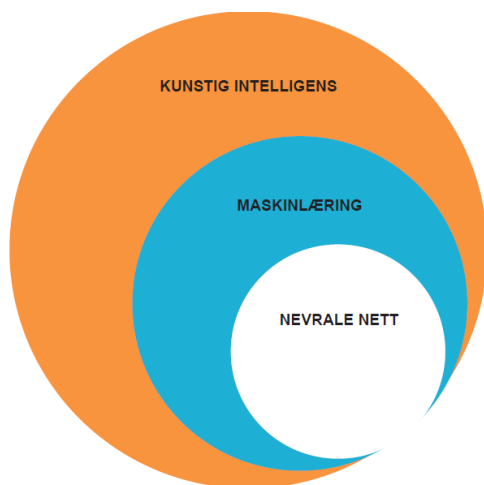
Utover 2000-tallet gikk fagområdet over fra å være regeldrevet til å bli statistikk- og datadrevet, og maskinlæring som tilnærming tok over. Arthur Samuel definerte i 1959 maskinlæring som «fagområdet som gir datamaskiner muligheten til å lære uten å være eksplisitt programmert».⁵

⁴ Se også Tørresen 2014.

⁵ Al-Darwish 2018.

Datamaskiner kan nå lære seg sammenhenger, regler og strategier fra erfaringer i data fra den virkelige verden, uten at noen forteller dem hvordan sammenhengene er. De kan tilpasse seg dataene kontinuerlig, og jo mer data de har tilgang til, jo mer presise blir de (adaptivitet). Dette gjør at datamaskiner kan utføre oppgaver på egenhånd (autonomi). Komplekse oppgaver og beslutninger kan slik overtas av maskiner, og gjøres raskere og rimeligere.⁶

Målet med kunstig intelligens er at maskiner også klarer å lære seg intuisjon og kunnskap som er vanskelig å uttrykke i regler, noe som tilnærmingen nevrale nett har vist er mulig. Nevrale nett er inspirert av strukturen og funksjonen til biologiske nevrale nettverk i hjernen. Nevrale nett kan også lære noe vi ikke visste fra før, eller noe som ikke er mulig for mennesker å lære.⁷



Figur 1: Forhold mellom kunstig intelligens, maskinlæring og nevrale nett.⁸

Figur 1 illustrerer forholdet mellom kunstig intelligens, maskinlæring og nevrale nett. Denne rapporten handler om maskinlæring, og har et spesielt fokus på bruk av nevrale nett, som er den tilnærmingen som driver kunstig intelligens fremover nå.

⁶ Adaptivitet og autonomi er karakteristiske egenskaper som er fremhevet i det finske nettkurset om grunnleggende kunstig intelligens, se også: <https://course.elementsofai.com/1/1>.

⁷ Ahlqvist m. fl. 2018.

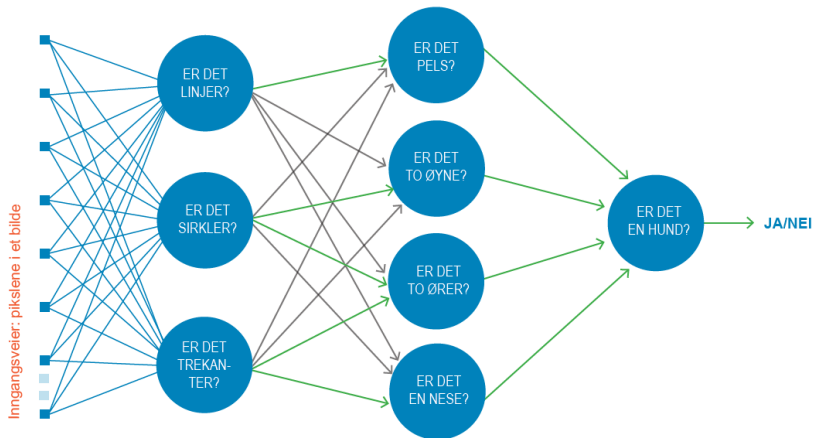
⁸ Inspirert av Wahed 2018.

BEDRE ALGORITMER

Maskinlæring har hatt en rask utvikling de siste årene, drevet frem av tre vesentlige endringer som har skjedd parallelt: 1) utvikling av bedre algoritmer, spesielt innen nevrale nett, 2) tilgang til store mengder data og 3) enkel og rimelig tilgang på stadig økende regnekraft.

Nevrale nett er en datadrevet tilnærming til maskinlæring. Et gjennombrudd kom i 2016 da programmet AlfaGo fra Google DeepMind klarte å slå verdensmesteren i spillet Go, som er et intuisjonsdrevet spill.

Modellene for læring i nevrale nett består av flere lag med såkalte nevroner. Nevronene i ett lag lærer ved å bruke inngangsverdier fra tidligere lag, og sender fra seg ny læring til neste lag, helt til det siste laget, som gir den endelige utgangsverdien. Det kan f.eks. være å bestemme klassen til et bilde («Ja, dette er bilde av en ondartet føflekk»).



Figur 2: Skematisk illustrasjon av et nevralt nett⁹

⁹ Inspirert av Geng og Shih 2017.

Er det en hund?

La oss anta at vi har lært opp et nevralt nett til å kjenne igjen hunder i bilder, som illustrert i Figur 2. Viktige egenskaper for en hund er at den har pels, to ører, to øyne og en snute. Vi ønsker å klassifisere et nytt bilde. Det første laget med inngangsverdiene i det nevrale nettverket vil bestå av like mange noder som antall piksler i bildet. Det andre laget består av nevroner som tar i mot pixlene og ser etter ulike former så som linjer, sirkler og kanter. Det tredje laget består av nevroner som vurderer hva linjene, sirklene og rektanglene representerer. Piksler som ett nevron vurderer som to sirkler blir sendt til to andre nevroner som vurderer om de er henholdsvis et par øyne eller et par ører. Egenskaper som blir vektlagt mye er angitt i grønt i figuren. Det siste laget med utgangsverdien vil gi en vurdering av om bildet er av en hund eller ikke.

Jo flere lag et nevralt nett består av, jo mer kompliserte strukturer kan det analysere. Nevrale nett med flere lag mellom inngangs- og utgangsverdiene kalles *dyplæringsnett*. De har evne til å lære seg komplekse sammenhenger og deretter generalisere til å gjenkjenne sammenhenger det ikke har sett før. Styrken til dyplæringsnett er at de kan lære hva som er viktig for å forstå f.eks. et bilde, uten at det blir uttrykkelig forklart. Dette gjør dyplæring til et kraftig verktøy innen maskinlæring. Ulempen er at teknikken ofte trenger mye data og regnekapasitet, og modellene kan være kompliserte og vanskelige å forklare i menneskelig forståelige termer.

Fremskritt innen nevrale nett og dyplæring har i løpet av de siste årene gjort det mulig å trene mer og mer nøyaktige maskinlæringsalgoritmer, og de er mye brukt i både bilde-, video-, tekst- og lydforståelse. Algoritmene kan nå gjenkjenne objekter i bilder bedre enn mennesker,¹⁰ og det har vært demonstrert at det kan være mer presist å snakke til maskiner enn å taste inn informasjonen for hånd.¹¹ Google's maskinlæringsbaserte oversettelsesystem ble 60% bedre med bruk av nevrale nett.¹² Nevrale nett kan dessuten brukes til å lage prognoser, som å forutsi ekstremvær.¹³

¹⁰ Karpathy 2014.

¹¹ Ruan m. fl. 2017.

¹² Turner 2016.

¹³ Lui m. fl. 2016.

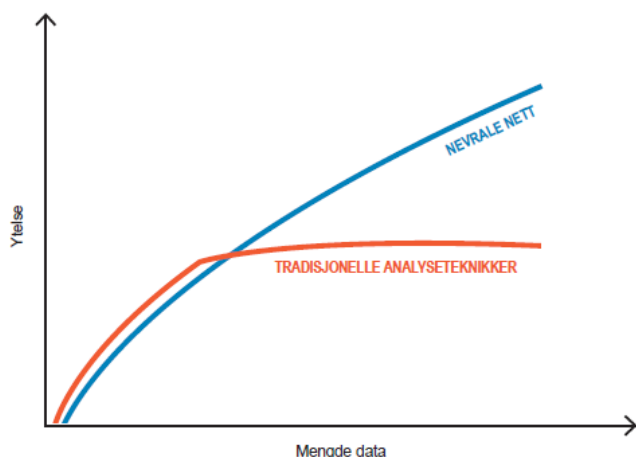
STORE MENGDER DATA

Maskinlæring, og spesielt nevrale nett, lærer ved at de blir matet med store mengder treningsdata fra den virkelige verden. Digitalt innhold har blitt produsert jevnt og trutt de siste tiårene, men de seneste årene har takten eksplodert. Hver dag produseres vi 2,5 trillioner bytes med data, og 90 prosent av den digitale informasjonen i verden i dag har blitt produsert i de siste to årene.¹⁴ Disse enorme datamengdene er med på å gjøre maskinlæring markant mer treffsikker.

Signaler fra sensorer på smarttelefoner og industrielt utstyr, digitale bilder og videoer, en kontinuerlig strøm av oppdateringer i sosiale medier, og det gryende tingenes internett (IoT), vil gi langt mer digitalt råstoff å jobbe med de neste årene.

Mens de tradisjonelle maskinlæringsalgoritmene bare klarer å bli bedre opp til et visst nivå før ytelsen flater ut, er nevrale nett imidlertid en metode der resultatene blir bedre etterhvert som de får tilgang til mer treningsdata. Jo dypere nettene er, jo mer kan de nyttiggjøre seg store datamengder (illustrert i Figur 3.) Dette er grunnen til at det er nevrale nett, og spesielt dyplæringsnett, som nå dominerer området. Det er viktig at dataene er av god kvalitet og representerer et generelt og mangfoldig bilde av problemet som skal løses. Ellers kan læringsmodellen gi upresise, eller gale anbefalinger.

¹⁴ IBM 2018.



Figur 3: Nevrale nett skalerer godt med økende tilgang til data sammenlignet med tradisjonelle analyseteknikker.¹⁵

TILGANG PÅ REGNEKRAFT

Opplæring av maskinlæringsalgoritmer krever stor regnekraft, både fordi de baserer seg på store datamengder og fordi algoritmene justeres underveis gjennom prøving og feiling. Tilgang til nødvendig regnekapasitet har økt betraktelig de siste årene, noe som også har vært viktig for utviklingen av KI-feltet. Følgende faktorer har vært viktige for regnekraften:

- *Moore's lov.* Kapasiteten til den generelle prosesseringsbrikken (*computing processing unit, CPU*) har i gjennomsnitt fordoblet seg jevnt hver 24. måned de siste 50 årene.
- *Nye, kraftige databrikker,* som f.eks. grafiske prosesseringsenheter (*graphical processing unit, GPU*) og prosessorer som er spesielt beregnet for nevrale nett, kan oppnå hastigheter opp til flere ganger raskere enn de generelle CPUene.¹⁶
- *Nettskyen.* Kraftige maskinlæringsinfrastrukturer som er optimalisert for å håndtere nevrale nett blir tilbudt som skytjenester. Disse kan brukes, kjøpes eller leies ved behov, uten at en må gjøre dyre investeringer selv.

¹⁵ Inspirasjon fra foil 30, Ng 2015.

¹⁶ Eksempler er Googles TensorFlow og Intels Nirvana Neural Network Processor.

Til sammen har dette gjort det mulig å komme i gang med å eksperimentere med, utvikle og ta i bruk maskinlæring enkelt, raskt og rimelig. For at fremgangen skal fortsette, er det imidlertid behov for nye og mindre datakrevende algoritmer. Den eksponentielle utviklingen av regnekraft er i ferd med å møte fysiske grenser, og det er hard konkurranse om tilgang til prosesseringskapasitet på grunn av utbredelsen av Bitcoin.¹⁷

¹⁷ Tassev 2018.

SLIK LÆRER MASKINER

Maskinlæringsalgoritmer lærer hovedsakelig på tre måter: ved å veiledes av erfaringer fra historiske datasett, finne nye mønstre og sammenhenger, eller ved å prøve og feile.

VEILEDET AV DATA

Den mest suksessfulle typen maskinlæring i de senere årene har vært veiledet læring (*supervised learning*), som lærer av erfaringer i datasett med eksempler fra den virkelige verden.

Hvert eksempel har egenskaper, også kalt inngangsverdier. Inngangsverdiene kan være pikselverdiene til et bilde, lydbølgene i en lydstrøm eller verdier som boareal, tomtestørrelse og antall soverom i en bolig. Datasettet er også merket med en utgangsverdi, for eksempel det dyret et bilde representerer, ordene i lydstrømmen eller salgsprisen på boligen.



Et godt eksempel på veiledet læring innenfor medisin er klassifisering av føflekker. Algoritmen tar utgangspunkt i et treningssett med medisinske bilder av føflekker (inngangsverdier) som er merket som godartede eller ondartede (utgangsverdier). Den lager så en såkalt prediktiv modell som predikerer om et nytt bilde er ondartet eller godartet med en viss sannsynlighet. I en test viste det seg at modellen kunne klassifisere føflekkreft på nivå med de beste dermatologene.¹⁸

Veiledet læring kan også brukes for å forutsi en fremtidig hendelse. I et tenkt eksempel ønsker en selger å vite hvilke brukere som kommer til å kansellere et abonnement, så han kan sette i gang en målrettet kampanje for å beholde dem før de sier opp. Selgeren vet imidlertid ikke hvordan han kan finne hvilke brukere som vil kansellere. Anta at selskapet har 10.000 kundeforhistorier, der halvparten har kansellert og halvparten fremdeles er kunder. En veiledet læringsalgoritme kan trene en prediktiv modell som lærer seg egenskaper ved de som har sagt opp og ved de som har vært lojale kunder. Når modellen er trent opp kan den forutsi hvilke av de nåværende kunder som har en høyeste sannsynlighet for å si opp, slik at kundebehandleren kan prioritere å sette inn tiltak mot dem.

Slik vurderes modellens kvalitet

Målet med å lære opp en prediktiv modell er at den skal være best mulig tilpasset eksempler fra den virkelige verden. Kvaliteten vurderes etter hvor godt modellen predikerer riktig for hver ny observasjon, også observasjoner som den aldri har sett før.

En vanlig måte å vurdere kvaliteten til en prediktiv modell er å sette til side et testsett fra treningsdataene. Modellen blir matet med inngangsverdiene i test-

¹⁸ Esteva m. fl. 2017.

sett, og kvaliteten blir vurdert ut ifra hvor godt svarene som modellen kommer frem til passer med de riktige svarene. Vanlige mål på kvaliteten til en prediktiv modell er sensitivitet (følsomhet) og spesifisitet (treffsikkerhet):¹⁹

- *Sensitivitet* er den andelen av de som faktisk har en sykdom som blir korrekt fanget opp som syke (ekte positive). Høy sensitivitet betyr at det er få syke som ikke blir fanget opp (få falske negative).
- *Spesifisitet* er den andelen av de som faktisk er friske som blir korrekt fanget opp som friske (ekte negative). Høy spesifisitet betyr at få friske blir uriktig klassifisert som syke (få falske positive).

I eksempelet med bilder av føflekker ble det nevrale nettet trent med et merket datasett på 129.450 kliniske bilder som bestod av 2.032 ulike sykdommer. Ytelser til algoritmen ble sammenlignet med 21 sertifiserte dermatologer på to typer diagnoser, den vanligste føflekkreften *aktinisk keratose* og den farligste kreftformen *malignt melanom*. For hver test ble dermatologene og algoritmen presentert for hhv 135 og 130 bilder de ikke har sett før, og der riktig tilstand var blitt påvist med biopsi (altså korrekt merket). Den prediktive modellen oppnådde både bedre sensitivitet og spesifisitet enn de fleste dermatologene.

Over- og undertilpasning

Et godt treningssett må representere virkeligheten bredt nok. Søppel inn gir også søppel ut. Et treningssett inneholder imidlertid sjelden alle tenkelige verdier og observasjoner fra den virkelige verden, og det er dermed alltid en risiko for at treningssettet gir et skjevt bilde av virkeligheten.

Algoritmer kan være over- eller undertilpasset de dataene de er trent på, noe som fører til at en maskinlæringsalgoritme yter dårlig.

- En algoritme som er overtilpasset (*over-fitted*) er for godt tilpasset treningsdataene, og vil ikke klare å predikere godt med nye observasjoner den ikke har sett før. Den lærer seg rett og slett for mange detaljer fra treningssettet. I boligeleksempelen kan det bety at en eller flere av egenskapene boareal, tomtestørrelse og beliggenhet ikke er så viktige for å kunne forutsi boligpris, eller at algoritmen ikke har fått nok treningsdata å lære fra.

¹⁹ Andre måter å måle kvalitet er *noyaktighet* (accuracy), *feilrate* (error-rate), *F1-målet* (F1-score), *MCC* (Matthews Correlation Coefficient), *positiv prediktiv verdi* og *negativ prediktiv verdi*.

- En algoritme som er undertilpasset (*under-fitted* eller *biased*) er ikke veldig godt tilpasset treningsdataene, og vil heller ikke klare å predikere godt med nye observasjoner. I boligeksempelen kan det bety at egenskapene boareal, tomtestørrelse og beliggenhet ikke er tilstrekkelig til å predikere salgspris for en bolig generelt sett, og at man trenger flere egenskaper.

Det er altså ikke bare mengden data som er avgjørende for hvor presis en modell er. Egenskapene i datasettet man benytter seg av kan i mange sammenhenger være vel så viktige. De som utvikler læringsalgoritmer vil ofte prøve seg frem med flere egenskaper før de kommer frem til den endelige modellen.

Statistiske og dynamiske læringsmodeller

En statistisk modell endrer seg ikke ved bruk. Opplæring av modellen skjer i kontrollerte omgivelser i et testmiljø, og endringer skjer ved at modellen erstattes med en ny versjon. Dette gir utviklerne full kontroll på modellen.

En dynamisk modell kan forbedre modellen kontinuerlig med nye inngangsverdier og brukes gjerne når omgivelsene stadig endrer seg. Et eksempel er overvåking for å identifisere forsøk på datainnbrudd. Den kontinuerlige læringen kan gjøre læringsmodellen mer presis, men ulempen er at endringene får umiddelbar effekt og at utviklerne dermed har mindre kontroll.

Uheldig utvikling av virtuell assistent

Microsoft sin virtuelle assistent Tay lærte kontinuerlig av samtaler den hadde med internettbrukere. Den ble imidlertid utsatt for systematisk feil-læring av brukerne og utviklet seg til å bli en nazistisk sexrobot. Microsoft fjernet Tay 24 timer etter lansering.²⁰

IKKE-VEILEDET

Mennesker lærer utmerket uten veiledning, og tilegner seg mesteparten av kunnskapen om verden ved å assosiere og identifisere mønstre. Ikke-veiledet læring (*unsupervised learning*) er en lignende tilnærming, der maskiner lærer

²⁰ Wakefield 2016.

ved å gjenkjenne mønstre og sortere data uten å kjenne kategoriene på forhånd. Ikke-veiledet læring kan identifisere mønstre som ikke engang mennesker kan oppdage, og har potensielt større nøyaktighet og skalerbarhet enn veiledet læring.²¹

Et viktig gjennombrudd kom i 2012 da Google og forskere ved Stanford-universitetet klarte å identifisere katter i digitale videoer, å bli fortalt hva som er katter.²²



Figur 4: Illustrasjon av selvkjørende bil (TechCrunch)

Selskapet Cortica lærer opp selvkjørende biler til å forstå omgivelsene sine ved å klassifisere og organisere bildene som bilen tar kontinuerlig. Et stoppskilt er for eksempel åttekantet med hvite kanter og rødt i midten. KI-systemet kan lære underveis at noen ganger er det røde falmet og noen gang kan den hvite kanten bli skjult av en

grein fra et tre. Systemet kan likevel gjøre de nødvendige endringene for å kunne klassifisere et stoppskilt som et stoppskilt. Selskapet mener at ikke-veiledet læring vil gjøre det mulig for fremtidens autonome kjøretøy å tilpasse seg bedre til nye situasjoner på veien.²³

En forskningsgruppe på Mount Sinai sykehuset i New York har brukt ikke-veiledet dyplæringsnett til å trekke ut egenskaper fra pasientjournalene til 700.000 individer i et system som kalles DeepPatient. Disse egenskapene ble deretter brukt som inngangsverdier til andre maskinlæringsalgoritmer for å gjøre klassifisering. Uten ekspertinstruksjoner og bare basert på data, har

²¹ Fagella 2016.

²² Hof 2018.

²³ Hall-Geisler 2017. Merk at selskapet som omtales ikke bruker nevrale nett, men en enkel klyngeanalyse.

DeepPatients læringsalgoritmer oppdaget nye mønstre og laget en modell som kan plassere pasienter i den gruppen vedkommende ligner mest på.²⁴

Ved å sammenligne med pasienter som ligner, er håpet at man kan forutsi det fremtidige sykdomsbildet for pasienten eller gi henne den behandlingen som viser seg å ha hjulpet best for denne typen pasienter. For å evaluere modellen brukte forskerne 76 000 test-pasienter som dekket 78 sykdommer. Forskerne mener at deres metode kan forutsi sykdom bedre enn de vanligste statistiske metodene. Metoden viser seg å være spesielt god til å forutsi diabetes, schizofreni og ulike typer kreft.²⁵

Tekstanalyse av juridiske dokumenter sparer tid

JPMorgan Chase introduserte i 2016 et system COiN (*Contract Intelligence*), en plattform for å analysere kontrakter, som bruker ikke-veiledet maskinlæring for å analysere juridiske dokumenter og trekke ut viktige bestemmelser. I et pilotforsøk fant de 150 relevante egenskaper fra 12 000 låneavtaler i løpet av sekunder. Dette ville tatt så mye som 360 000 timer for hver årlige manuelle gjennomgang. Denne kapasiteten kan få store konsekvenser tatt i betraktning at omtrent 80 prosent av feil på lån er skyldes tolkningsfeil av kontrakter.²⁶

Ikke-veiledet læring er fremdeles et umodent fagfelt. Foreløpig krever de fleste systemer noe trening eller tilbakemeldinger fra mennesker. Hvis og når vi lærer å bygge robuste ikke-veilede systemer som lærer uten involvering av mennesker, kan det åpne seg mange muligheter. De kan se på komplekse problemer på nye måter for å hjelpe oss med å oppdage skjulte mønstre i eksempelvis hvordan sykdommer sprer seg, prisutvikling for verdipapirer i et marked og kunders kjøpsatferd.²⁷

LÆRING MED FORSTERKNING

Teknikker for læring med forsterkning (*reinforcement learning*) baserer seg på at maskinen lærer ved å prøve og feile, og blir belønnet eller straffet avhengig av om oppførselen fører den nærmere eller lengre fra målet.

²⁴ Miotto m. fl. 2016.

²⁵ Miotto m. fl. 2016.

²⁶ Zames 2016.

²⁷ Brynjolfsson og McAfee 2017b.

Teknikken fikk et gjennombrudd i 2016, da programmet AlphaGo klarte å slå verdensmesteren i brettspillet Go, Lee Sedol. Go har kompliserte regler, og spillerne baserer seg i stor grad på intuitiv kunnskap som er vanskelig å uttrykke i dataprogrammer. Spillet har også et meget stort utfallsrom, dvs. antall trekk som er mulige. For å lære seg hvordan mennesker spiller spillet, fikk læringsalgoritmen vite spillets regler og studerte deretter 30 millioner posisjoner fra tidligere spill av amatører og profesjonelle spillere. Deretter spilte programmet mot ulike versjoner av seg selv tusenvis av ganger. For hver omgang lærte programmet av egne feil og forbedret seg gradvis, til den til slutt ble så god at den slo mesteren.²⁸



AlphaGo gjorde noen uvanlige trekk som man opprinnelig trodde var feil, men disse har gitt menneskelige Go-spillere ny innsikt i spillet. Lee Sedol har for eksempel vunnet alle sine spill siden han spilte med AlphaGo og har uttalt at AlphaGo lærte ham å spille mer kreativt.²⁹

Nøkkelen til gjennombruddene for læring med forsterkning har vært å bruke dype nevrale nett.³⁰ Takket være dyplæring har man en effektiv måte å gjenkjenne mønstre i data, som posisjoner på Go-brettet. Hver gang programmet feiler eller gjør noe riktig, kalkulerer det en verdi som blir lagret i store tabeller som blir oppdatert etter hvert som programmet lærer. For store og kompliserte oppgaver krever dette store regneressurser.

²⁸ Gibney 2016.

²⁹ House of Commons (i Storbritannia) 2017.

³⁰ Knight 2018.

AlphaGo-teknikken gjenbrukes til å effektivisere energibruk i datasentre

En videreutvikling av AlphaGo-teknikken kan gjøre Googles datasentre mer energieffektive. Kjølssystemene lærer fra data fra sensorer. Mens systemet tidligere gav råd til operatørene, justerer systemet seg selv, kontinuerlig og i sanntid. Ved å kontinuerlig lære fra nye data kan systemet nå spare 30% energi årlig, og DeepMind forventer ytterligere forbedringer.³¹

Læring med forsterkning kan også gjøres helt uten veiledning fra data. DeepMinds neste store gjennombrudd kom med *AlphaGo Zero*. Programmet lærte seg spillet Go helt av seg selv, med bare Go-reglene som inngangsverdier, og helt uten treningseksempler. Denne måten å lære på viste seg å være svært effektiv, og AlphaGo Zero slo i løpet av 40 dager alle tidligere AlphaGo-versjoner. Regnekraften som ble brukt var mindre enn de tidligere versjonene, samtidig som ytelsen var mye bedre.³²

Senere har en videreutvikling av algoritmen, Alpha Zero, lært seg å spille sjakk på egenhånd, med bare sjakk-reglene som inngangsverdier. Etter en opplæring på 4 timer slo den verdens høyest rangerte sjakkprogram Stockfish.³³

Teknikken læring med forsterkning egner seg for anvendelser der mennesker kan spesifisere målet, uten nødvendigvis å kunne uttrykke hvordan man kommer til målet. I tillegg til spill brukes teknikken for å lære opp hvordan selvkjørende biler manøvrerer, som å finne mest optimale trekk for å unngå ulykke. Teknikken har også et potensial i medisin, bl.a. for å identifisere rekkefølgen av medisiner som vil lede til det beste resultatet for pasienten.³⁴

BLANDINGSMODELLER

DELVIS VEILEDET

Delvis veiledet læring benytter seg av en kombinasjon av en liten mengde merkede data og en større mengde umerkede data.

³¹ Evans og Gao 2016, Lardinois 2018 og Sverdlík 2018.

³² Hassabis og Silver 2017.

³³ Dockrill 2017.

³⁴ Zhao 2011.

I et tenkt eksempel har vi få merkede bilder av katter og hunder, og mange bilder av katter og hunder som ikke er merket. Gjennom en ikke-veiledet læringsprosess kan vi gruppere bildene i klynger. Antakelig vil katte- og hundebildene befinne seg i to ulike grupper. Siden vi også vet hvordan en katt og hund ser ut fra det merkede datasettet, så kan systemet merke den gruppen som ligner mest på en hund som hunder og tilsvarende for katter.

Et annet område der delvis veiledet læring egner seg er tekstklassifisering og analyser av naturlige språk, der tilfanget av merkede data er lite.

Aktiv læring

Teknikken aktiv læring er en type delvis veiledet læring der modellen velger selv hvilke umerkede data som vil være mest informative, og deretter ber et menneske om å merke (kategorisere) dem.

Teknikken kan oppnå en bedre ytelse enn man ville fått ved å gjøre veiledet læring kun på de merkede dataene, eller ved å gjøre ikke-veiledet læring kun på de umerkede dataene.

OVERFØRT LÆRING

Teknikker for overført læring (*transfer learning*) gjør at vi ikke trenger å finne opp hjulet på nytt for hvert problem vi ønsker å løse, men bygger på tidligere kunnskap. Teknikkene gjør det mulig å overføre kunnskap fra domener der vi har rikelig med merkede data, til nye og lignende områder der datagrunnlaget er mer sparsomt, kostbart eller farlig å få tak i. Kunnskap fra opplæring i å gjenkjenne bilder av biler kan for eksempel brukes for å lære opp systemer til å gjenkjenne lastebiler.

Overført læring kan egne seg i det medisinske domenet fordi de ofte mangler treningsdata, noe som en overført modell kan kompensere for. I eksempelet med føflekkreft ble modellen lært opp ved å bruke teknikker for overført læring på et eksisterende nevralt nett som kalles ImageNet³⁵. Det er trent på en mengde bilder av ulike generelle objekt-kategorier, og kan gjenkjenne hvor det er objekter i bilder, formen på objektene osv. Dette nettet kan dermed brukes for å identifisere hvor i bildet det kan være føflekker og formen på disse mulige

³⁵ Se Image-net.org.

føflekkene. Den spesifikke føflekkalgoritmen kan bygge videre på denne kunnskapen, og konsentrere seg om å lære seg å gjenkjenne om det faktisk er føflekker og å kunne skille ondartede og godartede føflekker.

Filtrering av søppel-e-post

Mange merker ikke tilstrekkelig antall meldinger som søppelpost til at individuelle epostfiltre blir effektive nok. Samtidig vil et generelt epostfilter, som er likt for alle, ikke bli presist nok. En hybrid generell, individuell filtreringsløsning som benytter seg av overført læring kan bli effektiv hvis den kan lære av alle de brukerne som merker søppelpost flittig og samtidig lære av hvert individ som bare merker noen få e-poster som søppel.³⁶

Simulering er en teknikk for overført læring som bruker data på en enklere og mindre risikabel måte. For eksempel er det nødvendig med data fra kollisjoner og ulykker for å lære opp selvkjørende biler, og det er nødvendig med data fra personer som faller for å lære opp systemer som automatisk oppdager personer som faller. Disse type data er vanskelig å få tak i, derfor er det en forutsetning å kunne gjøre simuleringer i opplæringen av systemene.

Eksempel på simulator for å lære opp selvkjørende biler

Udacity bruker en simulator for å lære opp selvkjørende biler. Selskapet har også gjort simulatoren tilgjengelig som fri programvare slik at andre kan bruke den til å lære opp selvkjørende biler.³⁷ Videospill kan også brukes som simulator for å lære opp algoritmer for selvkjørende biler, og Open API Universe har gjort tilgjengelig kode for flere videospill.³⁸

Dersom man har et sparsomt datasett, men har tilgang til en læringsmodell som er lært opp på et lignende domene, viser der seg at teknikker for overført læring kan gi bedre ytelse.³⁹

³⁶ Multi-task learning 2018.

³⁷ Etherington 2017.

³⁸ Mannes 2016.

³⁹ Gupta 2017.

GENERERENDE KONKURRERENDE NETTVERK

Genererende konkurrerende nettverk (*Generative Adversarial Networks*, *GANs*) er en måte å generere eller klassifisere data. GANs kan for eksempel brukes til å generere mer data til et treningssett hvis man mangler data, eller lage musikk eller kunstverk som ligner stilen til en gitt komponist eller kunstner.

Slike nettverk er en god metode for å lære av ikke-merkede data, noe som kan være nøkkelen til å gjøre datamaskiner mer intelligente i årene som kommer.⁴⁰

GANs-metoden består av to typer nettverk som konkurrerer med hverandre. Det ene nettverket, G-nettverket, skal lære seg å generere syntetiske data som er mest mulig lik virkelige bilder. Det andre nettverket, D-nettet, skal lære seg å detektere hvilke bilder som er ekte og falske. For å illustrere kan vi se på G- og D-nettverkene som henholdsvis skurker som prøver å forfalske penger og politiet som vil avsløre pengeforfalskere. Skurkene må lære seg å forfalske penger slik at politiet ikke klarer å oppdage dem, mens politiet må lære seg å gjenkjenne falske penger. Ved å konkurrere blir begge parter tvunget til å forbedre seg hele tiden.

La oss si at vi trenger flere bilder av katter for å gjøre en læringsalgoritme bedre til å gjenkjenne katter. G-nettet starter med et bilde som består av helt tilfeldige piksler. D-nettverket tar imot bildet, og vurderer om det er et realistisk bilde av en katt eller ikke. I neste runde lager G-nettverket et nytt bilde, ved å justere litt på den forrige versjonen basert på hint den får fra D-nettverket. Og sånn fortsetter det. Tilbakemeldingene fra D-nettverket gjør at G-nettverket blir bedre til å generere realistiske bilder av katter. Ved å jobbe sammen, kan disse nettverkene både produsere veldig realistiske syntetiske data og bli bedre til å detektere ekte data.

⁴⁰ Knight 2017a.

Oppdage sykdom i tarmen med høy sensitivitet og spesifisitet

Angiodysplasi⁴¹ er en misdannelse av blodkar i veggene i fordøyelseskanalen, og er en av de vanligste årsakene til blødninger i tarmen. Diagnose stilles ved å tolke bilder av tarmen. En metode er at en pasient svelger en kamerapille, som tar opp til 60 000 bilder fra tarmen. En undersøkelse viser at leger som undersøker slike bilder oppdager bare 69 prosent av tilfellene (spesifisitet). Forskere ved Simulasenteret og Universitetet i Oslo har utviklet GANs som kan oppdage angiodysplasi i slike bilder av tarmen med en presisjon og spesifisitet opp mot 100 prosent og sensitivitet på 98 prosent. Dette ligger også høyt over andre tilnærminger med maskinlæring.⁴²

GANs-metoden kan også lære seg hva som kjennetegner Beethovens musikk og lage nye musikkstykker som høres ut som de er komponert av Beethoven, eller på samme vis lære seg å lage malerier som ligner på Munch. GANs kan også være mer praktisk nyttige ved å fylle inn manglende data i et ufullstendig bilde, automatisk generere scener i et videospill, gjøre bilder skarpere, eller generere simulerte data til å lære opp selvkjørende biler.⁴³

⁴¹ Angiodysplasi, 2009.

⁴² Pogorelov m. fl. 2018.

⁴³ Goodfellow m. fl. 2014 og Goodfellow 2017.

BRUKSOMRÅDER – FRA LEGEVAKT TIL SELVKJØRENDE BILER

Kunstig intelligens og maskinlæring gir mulighet for å tolke tale, tekst, tall, bilder og video og forutse fremtidige hendelser basert på mønstre i data. Dette er nyttig innenfor mange områder – fra kreftdiagnostikk og tilpasset undervisning til energi-effektivisering, klimaanalyser og selvkjørende biler.

Kunstig intelligens forandrer forholdet mellom maskiner og mennesker på flere måter. Ved hjelp av maskinlæring kan datamaskinene gjøre oppgaver mer effektivt enn mennesker, det vil si rimeligere eller på kortere tid, som å stille en bildediagnose. Algoritmene kan også gjøre noen oppgaver bedre enn mennesker, som å finne nye mønstre i helsedata eller optimalisere energiforbruk. I tillegg kan maskinlæringssystemer skaleres raskt og rimelig, slik at oppgavene kan gjøres svært mange ganger med marginale kostnader.

ORDNE OG FORUTSI

Maskinlæring brukes til å lage prediksjoner. Enkelt sagt så handler prediksjoner om å fylle ut manglende informasjon. Prediksjoner tar den informasjonen

man har, det vil si data, og bruker den til å generere informasjon man ikke har.⁴⁴ Det kan være prediksjoner om fortiden, nåtiden eller fremtiden, som f.eks. å oppdage om en kredittkorttransaksjon var falsk, bestemme om en føflekk er ondartet eller å forutsi hva været blir i morgen, med en viss sannsynlighet.

Det finnes flere teknikker for å predikere, og de vanligste teknikkene er beskrevet nedenfor.

KLASSIFISERING

Klassifisering er den mest brukte maskinlæringsteknikken og brukes til å bestemme hvilken kategori en ny observasjon tilhører. Det kan for eksempel dreie seg om å identifisere hva som befinner seg på et bilde. Teknikker for veiledet læring egner seg godt til klassifisering, og nevrale nett har vist seg å være veldig effektive.

Det skilles gjerne mellom sortering i to klasser (binær) eller i flere klasser (multi-klasse). Et spam-filter er et eksempel på binær klassifisering som predikerer om en e-post er «spam» eller «ikke spam». Diagnoseverktøy som predikerer den eller de mest sannsynlige diagnosene på grunnlag av symptomene til en ny pasient vil bygge på multi-klassifisering.

KLYNGEANALYSER

Klyngeanalyser (*clustering*) brukes for å utforske nye datasett uten å kjenne til sammenhengene på forhånd. Klyngeanalyse finner nye strukturer og mønstre i umerkede data og deler dem inn i ulike grupper eller klynger basert på egenskaper som ligner.

Teknikken kan brukes for å gruppere filmkunder som ligner på hverandre slik at de kan få tilpassede filmanbefalinger. Eller teknikken kan brukes til å identifisere pasienter med lignende symptomer og hvordan behandlinger har virket på ulike grupper, slik at nye pasienter kan få bedre tilpasset behandling.

Klyngeanalyser kan også brukes for å generere merkede datasett (som ofte er mangelvare), ved å identifisere klynger og deretter få noen til å merke klyngene.

⁴⁴ Agrawal m. fl. 2018

IDENTIFISERE AVVIK

Teknikker for å identifisere avvik (*anomalideteksjon*) oppdager hendelser som ikke samsvarer med et forventet mønster i et datasett. Avvikene kan være forsøk på banksvindel, datainnbrudd eller forstyrrelser i økosystemet. I helsesammenheng kan slike teknikker brukes for å følge med på utviklingen av pasienters helsetilstand, og oppdage en eventuell farlig eller uønsket utvikling.

Teknikken kan også brukes for å forbedre et datasett ved å fjerne avvikende data, noe som kan føre til en statistisk signifikant økning i nøyaktighet.

FORUTSEENDE ANALYSER

Forutseende analyser (*forecast*) handler om å kunne forutsi, noe som vil kunne skje i fremtiden, basert på en serie med historiske data. Slike analyser kan brukes for å lage en risikoprofil for en person. Det kan være sannsynligheten for at en person vil falle ut av skolen, være i stand til å nedbetale en gjeld av en viss størrelsesorden eller utvikle en gitt type sykdom. Forutseende analyser er nyttige når det er viktig å kunne forberede seg på en mulig utvikling eller forebygge at det skjer.

Ett aktuelt område er å bedre forstå værfenomener. Meteorologer bruker i økende grad maskinlæring for å gi prognoser for hvor lenge en storm kan vare eller hvorvidt den vil lage skadelig hagl.⁴⁵ Maskinlæringsalgoritmer som er trent på data fra ekstreme klimahendelser har klart å identifisere tropiske sykloner og atmosfæriske «elver» (søyler med vanndamp som beveger seg sammen med været). Det siste kan føre til farlig mengder med nedbør i et område, men er ikke alltid lett for mennesker å oppdage.⁴⁶

⁴⁵ Reilly 2017.

⁴⁶ Lui m. fl. 2016 og Jones 2017.

TALE- OG LYDFORSTÅELSE

Teknologier for tale- og lydforståelse brukes for å oversette tale til tekst og omvendt. Taleteknologi er nå i daglig bruk på mobiltelefonen i virtuelle assistenter⁴⁷ som Siri på Apples iPhone, Google-assistenten på Android-telefoner og Amazons taleassistent Alexa.

Dype nevrale nett egner seg godt til å kunne lære å kjenne igjen ord i en lydstrøm og er den viktigste årsaken til at talegjenkjenning har hatt store forbedringer på kort tid. Feilraten falt fra 8,5% til 4,9% fra sommeren 2016 til 2017.⁴⁸

Bruksområdene for taleforståelse spenner fra å forenkle rutineoppgaver, gjøre det enklere å håndtere datasystemer og finne mønstre i digitale lydssignaler.

OVERSETTE TALE TIL TEKST

Tale-til-tekst diktering på smarttelefonen gjør det tre ganger å raskt å lage en tekstmelding som å taste den inn for hånd.⁴⁹ Leger kan diktere informasjon direkte inn i en pasientjournal.

NATURLIG BRUKERDIALOG

Tale-til-tekstteknikker kan gjøre dialogen med datasystemer mer naturlig slik at de blir enklere å bruke. Talegrensesnitt kan hjelpe eldre og funksjonshemmede som har vanskeligheter med å bruke tastatur eller berøringsskjerm, slik at de likevel kan bruke digitale tjenester og velferdsteknologi. Leger kan styre datasystemer i operasjonsstuen med stemmen, slik at hendene er fri.

⁴⁷ Virtuell assistent (engelsk: chat bots, virtual assistant): en applikasjon som veileder og svarer på spørsmål, ved bruk av naturlige språk, skriftlig eller muntlig.

⁴⁸ Brynjolfsson og McAfee 2017a.

⁴⁹ Carey 2016.

OPPDAGE FARESIGNALER I HELSEMÅLINGER

Flere helsemålinger er digitale lydsignaler fra f.eks. digitale EKG-apparater eller -stetoskop. Teknikker for taleforståelse kan tolke og oppdage unormale signaler i slike lydstrømmer.

Kardia er et lite apparat som måler 2-punkts EKG i tomlene. Brukerne kan måle EKG jevnlig og registrere målingene automatisk i en mobil-app. Appen bruker maskinlæring for å bygge hjerteprofil for hver pasient fra målingene. Dersom en senere måling ikke passer med profilen, fanger Kardia opp dette og varsler pasienten og/eller helsepersonell.⁵⁰

CliniCloud er et digitalt stetoskop som brukes for å måle hjertelyd. Brukere kan ta opp hjertelyden selv og få hjelp av datamaskin eller lege til å tolke den. Selskapet bak utstyret planlegger å bruke maskinlæring for å tolke hjertelydopptakene, og de håper å kunne tolke like godt som leger. Ambisjonen er å kunne oppdage avvik i lydstrømmen selv om leger ikke har oppdaget dem.⁵¹

TEKSTFORSTÅELSE

Teknikker for tekstforståelse (bl.a. *Natural Language Processing*) finne mening i ustrukturert, skriftlig informasjon. Teknikken hadde et gjennombrudd da IBMs Watson i 2011 vant over stormestrene i spillet Jeopardy.⁵² Siden dette har programmer for tekstforståelse blitt brukt i mange ulike sammenhenger.

OVERSETTE SPRÅK

Både Google og Facebook har gått over til å bruke dyplæringsteknikker for oversettelse. Da Googles metode ble publisert i 2016 rapporterte de at den reduserte feil med 60 prosent.⁵³ Google oversetter nå nesten alle språk til og fra engelsk

⁵⁰ AliveCOR 2017.

⁵¹ Samtale med CEO Andrew Lin i CliniCloud 24.10.2017 og Niesche 2015.

⁵² Ferrucci 2018. I januar 2018 ble Alibabas program for kunstig intelligens det første til å slå mennesker i en kunnskapskonkurranse utformet av Stanford-universitetet. Programmet fra Alibaba scoret 82.44 prosent mot de menneskelige deltagernes 82.304 prosent. Fenner 2018.

⁵³ Castelvocchi 2016.

på denne måten. Facebook bruker dette for mer enn 4,5 milliarder oversettelser daglig.⁵⁴

KUNDESTØTTESYSTEMER

Tekstanalyse kan hjelpe kundebehandlerne å bli mer effektive eller automatisere deler av spørsmål-og-svartjenester, slik de har gjort i Skatteetaten og i Udacity (se faktaboks).

Automatisert hjelp om skattespørsmål

Skatteetaten holder på å utvikle en virtuell assistent som svarer på skattespørsmål fra publikum. De startet med å undersøke om det ville være nyttig og lønnsomt med et verktøy som hjelper kundebehandlere å svare raskere og mer presist på spørsmål. De fant imidlertid ut at de enkle og ofte stilte spørsmålene svarte de allerede raskt på, mens de sjeldne og mer kompliserte spørsmålene var det også vanskelig for maskiner å forstå og svare på. De konkluderte med at de heller vil lage en virtuell assistent som kan hjelpe publikum direkte med de enkleste spørsmålene. De bruker maskinlæring, dvs. tekst- og språkanalyse, på et treningssett med ulike typer spørsmål og intensjoner for å forstå hva det er brukeren lur på. Svarene følger faste regler så det trenger de ikke maskinlæring for å komme frem til.⁵⁵

Opplæring ved å lære av de beste selgerne

Selskapet Udacity oppdaget at noen av selgerne var mye mer effektive enn andre når de svarte på spørsmål på en chatte-kanalen. De ville prøve å gjøre alle selgerne bedre ved at de kunne lære av de beste. Loggen til chatte-kanalen representerer i realiteten et sett med merkede treningsdata, som er nettopp hva et veiledet maskinlæringssystem trenger. Samtaler som førte til et salg ble merket som suksesser, og alle andre ble merket feil. De brukte dataene til å predikere hvilke svar vellykkede selgere sannsynligvis ville gi som svar på de vanligste henvendelsene, og delte dem med de andre selgerne for å gi dem et hint om hva de burde svare på akkurat de henvendelsene. Etter 1000 treningssykluser hadde selgerne økt effektiviteten med 54% og kunne betjene dobbelt så mange kunder.⁵⁶ I stedet for å bygge en virtuell assistent som kunne overta alle samtaler, så lagde de heller et system som hjelper alle selgerne å yte bedre.⁵⁷

DIGITAL FØRSTELINJE

I helsetjenesten kan tekstanalyser hjelpe førstelinjen på for eksempel legevakten eller fastlegekontoret til å gi raskere og bedre tilpassede tilbakemeldinger

⁵⁴ Ong 2017.

⁵⁵ Presentasjon av VAKI prosjektet, Skatteetaten, 8. desember 2017 og SkLNytt 2017.

⁵⁶ Ng 2017.

⁵⁷ Brynjolfsson og McAfee 2017b.

til de som ringer inn. Førstelinjen kan etter hvert automatiseres, noe som prøves ut i Storbritannia.⁵⁸

GJENNOMGANG AV PASIENTJOURNALER

En pasientjournal kan bestå av opptil flere hundre dokumenter med ustrukturert tekst, som er tidkrevende å gå gjennom manuelt. I mange tilfeller er det kritisk at detaljer ikke blir oversett. Tekstanalyse kan raskt trekke ut mening og sentral informasjon av journalen og dermed lette arbeidet til helsepersonell.

Før hver operasjon bruker sykehuset i Agder et system som varsler om eventuelle allergier og gjør dette på kortere tid enn en vanlig lege bruker på å gå gjennom papirene til pasienten.⁵⁹ Dette kan spare tid, noe som er spesielt viktig i tidskritiske situasjoner.

BILDE- OG VIDEOANALYSE

Bildeanalyse- og videoanalyse gjenkjenner objekter i bilder og video. Facebook og andre sosiale medier gjenkjenner nå ansikter i bilder og spør vennene deres om de vil hjelpe til med å merke dem med navn. I siste versjon av Apples iPhone brukes ansiktsgjenkjenning for å låse opp telefonen, og som identifikasjon i tjenester på telefonen.

Teknikkene har hatt store fremskritt de senere årene. Feilraten for å gjenkjenne bilder har falt fra høyere enn 30% i 2010 til omtrent 2,25%, i 2017 for de beste systemene. Til sammenligning er den menneskelige feilraten omtrent 5%.⁶⁰ Videogjenkjenningssystemer, som bl.a. brukes i selvkjørende biler, tok tidligere feil så ofte som én gang hvert 30. videobilde, mens de beste systemene i dag gjør feil sjeldnere enn én gang på 30 millioner videobilder.⁶¹

Fremskrittene på feltet gjør at både trivielle og mer avanserte og ressurskrevende oppgaver kan bli lønnsomt å automatisere.

⁵⁸ Murgia 2017.

⁵⁹ Christiansen 2017.

⁶⁰ Echersley og Nasser 2018.

⁶¹ Brynjolfsson og McAfee 2017a.

GJENKJENNE OBJEKTER

Teknikkene kan oppdage og gjenkjenne objekter som ansikter og tekster i bilder. Det kan f.eks. brukes til å identifisere personer raskt i digitale fotoalbum, eller lage bildetekster automatisk.

Ved å kombinere teknikker for tekstoversettelse med teknikker for å gjenkjenne bokstaver i bilder kan man oversette f.eks. reklamebilder fra ett språk til et annet.⁶²

DIAGNOSTISERING

Kommersielle radiologiske deteksjonssystemer er under utvikling for flere typer medisinske bilder, inkludert MR, CT, ultralyd og patologiske bilder.

Forskere på Simulasenteret har utviklet algoritmer basert på kunstig intelligens som kan identifisere de åtte vanligste magesykdommene i bilder av tarmen med minst 93 prosent nøyaktighet.⁶³ Slike systemer kan hjelpe spesialister med raskere og mer treffsikre diagnoser.

AUTONOME KJØRETØY

Analysere av video sammen med data fra sensorer på bilen og i omgivelsene gjør at kjøretøy kan manøvrere i trafikken. Teknikkene er nå så gode at de kan gjøre analysene i tilnærmet sann tid slik at bilene kan manøvrere og reagere umiddelbart hvis noe uventet skulle skje.⁶⁴

FORBEDRE BILDER OG VIDEO

Bilde- og videoanalyser kan også forbedre eller generere bilder og video. Bilder kan gjøres skarpere og scener i videospill kan genereres automatisk. Svart-hvit-filmer kan også automatisk fargelegges⁶⁵.

⁶² Brownlee 2016.

⁶³ Haugnes 2017.

⁶⁴ Hawkins 2018.

⁶⁵ Brownlee 2016.

ANBEFALINGSSYSTEMER

Anbefalingssystemer (*recommender systems*) brukes for å gjøre risikovurderinger og lage persontilpassede tjenester.

«Netflix-prisen», som ble arrangert fra 2006 til 2009, viste seg å være viktig for å stimulere utviklingen av nye og bedre algoritmer for anbefalingssystemer. Selskapet offentliggjorde et datasett på over 100 millioner filmrangeringer, og ga en premie på \$ 1.000.000 til de som kunne lage mer nøyaktige anbefalinger enn selskapets eget system. Vinneren i 2007 var 8 prosent mer nøyaktig og brukte en samling av 107 forskjellige algoritmiske tilnærminger.⁶⁶

Ulike måter å lage anbefalingssystemer

Samarbeidsfiltrering bygger en modell fra en brukers tidligere adferd (ting som brukeren tidligere har kjøpt eller klikket på, eller har rangert (f.eks. med smilefjes)), i tillegg til lignende beslutninger av andre brukere. Modellen brukes deretter for å predikere produkter som brukeren antakelig er interessert i.

Innholdsbasert filtrering bruker en serie med kjennetegn på en ting, for å kunne anbefale andre ting med lignende egenskaper.

Hybride anbefalingssystemer bruker en kombinasjon av samarbeidsfiltrering og innholdsbasert filtrering. En typisk anbefaling kan være at «vi tror denne boken er interessant for deg fordi andre som ligner på deg har kjøpt den og fordi du har kjøpt en lignende bok tidligere».

SKREDDERSY TILBUD

De fleste nettjenester som selger varer og tjenester bruker nå anbefalingssystemer i en eller annen form for å anbefale produkter som du sannsynligvis vil like, og dermed også kjøpe. For eksempel får man ofte anbefalinger av typen «personer som ligner på deg kjøpte også dette».

Ved å analysere hva kunder har gjort og beregne sannsynligheten for hva de kommer til å gjøre, kan nettbutikker på egenhånd tilpasse tilbudene til hver enkelt besøkende.⁶⁷

⁶⁶ Bell 2018.

⁶⁷ Mystore 2017.

FOREBYGGE TRAFIKKULYKKER

Vegdirektoratet har brukt maskinlæring på detaljerte veidata koblet med åpne offentlige data til å analysere ulykkesrisiko på Europa-, riks- og fylkesveier i Norge. Algoritmene har kommet frem til egenskaper ved veien og dens omgivelser som gir økt ulykkesrisiko, som kan brukes for å forebygge ulykker.⁶⁸

FORUTSI SYKDOM

Det er større sannsynlighet for å bli frisk jo tidligere sykdom oppdages. I Horsens i Danmark viser innledende studier at algoritmer kan forutsi med 90 prosent sannsynlighet hvem som vil bli lagt inn med for eksempel akutt blodpropp i løpet av de neste 100 dager.⁶⁹

TILPASSET UNDERVISNING

Adaptive læringssystemer kan hjelpe lærere til å tilpasse undervisningen til hver enkelt elevs ferdigheter og modenhet. Et forskningsprosjekt hadde som mål å holde alle elevene i flytsonen, der balansen mellom det som er for vanskelig og det som er for lett er tilpasset den enkeltes nivå. Prosjektet fant ut at frafallet kunne mer enn halveres og hele klasser kunne øke sin prestasjon med bortimot én karakter i snitt.⁷⁰

PERSONTILPASSET BEHANDLING

Personalisering gjør det mulig å skreddersy behandling til hver pasient. Selskapet Petuum har utviklet metoder for å prosessere pasientjournaler, inkludert diagnoseresultater, og bruker det til å anbefale kombinasjoner av medisiner som høyst sannsynlig vil gi de beste resultatene⁷¹.

⁶⁸ Mandaric og Axelsen 2017.

⁶⁹ Fischer og Olhoff-Jacobsen 2017.

⁷⁰ Bjørkeng 2015.

⁷¹ <http://www.petuum.com> samt e-postutveksling med Eric Xing i Petuum, 30. august 2017.

INDIVIDUELLE TRENINGSPROGRAM

Treningsappen UA Record benytter seg av data om kosthold og fysiske og psykologiske adferdsdata som de sammenstiller med resultater fra folk med lignende helse- og treningsprofiler, for å utarbeide individuelle treningsprogram.⁷²

LISTE OVER BRUKSOMRÅDER

Til sammen kan de ulike teknikkene brukes på et utall bruksområder. Her har vi satt sammen en liste av anvendelser som vi har kommet over i arbeidet med denne rapporten. Listen er ikke uttømmende, men er ment å gi et bilde av bredden til mulighetsrommet. I den elektroniske versjonen av dokumentet er det lenker til eksemplene.

⁷² <http://underarmour.com>.

BRUKSOMRÅDE	EKSEMPEL	KI-ELEMENT
GENERELLE VERKTØY		
Talegjenkjenning	<i>Apple Siri, Google assistent, Amazon Alexa</i>	Klassifisering Tale- og lydforståelse
Oversettelse	<i>Google oversetter, Facebook oversetter</i>	Klassifisering Tekstanalyse
Automatisk kundebehandling	<i>Kongsberg kommune, skatte-spørsmål, bankspørsmål</i>	Klassifisering Tekstforståelse
Rekruttering	<i>Bedre balanse mellom kvinner og menn, LinkedIn</i>	Klassifisering
HANDEL og BANK		
Automatisk lånesøknad	<i>Automatisk lånesøknad</i>	Anbefalingssystemer
Personalisert netthandel	<i>Amazon</i>	Anbefalingssystemer
Tolke kontrakter	<i>JPMorgan</i>	Tekstanalyse
IT-OPPGAVER og -SIKKERHET		
Sortere e-post	<i>Søppel e-postfilter</i>	Klassifisering og klyngeanalyse. Tekstforståelse
Ansiktsgjenkjenning for å håndtere identitet	<i>Identifikasjon på mobiltelefon, merke seg selv i bilder, oppdage når andre bruker bilde av deg,</i>	Klassifisering Bildeanalyse
TRANSPORT		
Veiplanlegging	<i>Forebygge trafikkulykker</i>	Forutseende analyser
Opplæring av førerløse biler	<i>Fra omgivelsene, oppførsel i trafikken, fra simuleringer, av videospill</i>	Klassifisering Klyngeanalyse Bildeanalyse
Parkering	<i>Hvor finnes en passende parkeringsplass</i>	Klassifisering
UNDERHOLDNING		
Personaliserte anbefalinger	<i>Film</i>	Anbefalinger
Spille spill	<i>Jeopardy, AlphaGo, AlphaGo Zero, Alpha Zero, Stanford kunnskapskonkurranse</i>	Forutseende analyser Tekstanalyser
Utvikle spill	<i>Automatisk generere scener</i>	GANs, Klassifisering Bildeanalyse
ENERGI og MILJØ		
Optimalisere energiforbruk	<i>Google driftssentre</i>	Forutseende analyser
Forebygge feil	<i>Predikere behov for vedlikehold</i>	
Kontrollere fiskefelt	<i>Oppdage ulovlig fiske</i>	
VÆR OG KLIMA		
Forutse ekstremvær	<i>Hvor lenge varer en storm, tropiske sykkloner og atmosfæriske elver</i>	Forutseende analyser
SKOLE		
Hjelp til å løse oppgaver	<i>Engelsk stil</i>	Forutseende analyser Tekstanalyse
Persontilpasset oppfølging av elever	<i>Adaptiv læringsplattform</i>	Anbefalingssystemer
Treningscoach	<i>UA Record</i>	

Tabell 1: Generelle bruksområder

BRUKSOMRÅDE	EKSEMPLER	KI-ELEMENT
FØRSTELINJEHJELP	<i>Symptomsjekker, Automatisk triagering</i>	Klassifisering Tekstanalyse
TOLKE BILDER	<i>Føflekkreft, lungekreftknuter, tarmkreft, brystkreft, øyesykdommer, lungebetennelse, prostata,- tarm- og lunge- kreft, hjertesykdom</i>	Klassifisering Klyngeanalyser Bildeanalyse
IDENTIFISERE AKUTTE HENDELSER	<i>Akutt nyreskade, sykehusinfeksjoner</i>	
DIAGNOSEHJELP	<i>Diabetisk retinopati, sjelden kreftsykdom, unormal hjertelyd,</i>	Tekstanalyse Anomalideteksjon
BEDRE TILPASSET BEHANDLING	<i>Forutsi sykdomsutvikling av lungekreft, optimal kombinasjon av medisiner, forutsi sykdomsutvikling av kreftsvulst, identifisere nye undergrupper av diabetes, optimal medisinerings av lungekreft</i>	Tekst- og bildeanalyse Anbefalingssystemer Anbefalingssystemer Bildeanalyse Anomalideteksjon Anbefalingssystemer Læring med forsterkning
EGENHÅNTERING AV KRONISKE SYKDOMMER	<i>Diabetes, astma, kognitiv adferdsterapi, mental helse</i>	Anbefalingssystemer Klassifisering
ETTERLEVELSE AV MEDISINERING	<i>Motivere og varsle, astmamedisin</i>	Forutseende analyse
OPPDAGE UHELDIG UTVIK- LING AV HELSETILSTAND	<i>Fall, egenmålinger, unormal hjertelyd, hjertesvikt</i>	Bildeanalyse Lydforståelse Forutseende analyser
OPTIMALISERE RESSURSBRUK	<i>Pasientforløp, pakkeforløp for prostatakreft</i>	Ikke-veiledet og semi-vei- ledet læring
RISIKO FOR SYKDOM	<i>Forutsi sykdomsutvikling, risiko for blodpropp</i>	Forutseende analyser

Tabell 2: Eksempler på bruksområder innen helse

KAN VI STOLE PÅ MASKINENE?

Kunstig intelligens påvirker allerede mange valg som tas av enkeltmennesker og organisasjoner. Desto viktigere blir det at vi kan stole på og forstå de anbefalingene algoritmene gir oss.

Muligheten for å skape maskiner som tenker og beslutter reiser mange etiske spørsmål. På lang sikt kan dette handle om hvorvidt maskiner skal få en moralsk status, eller at maskinene oppnår superintelligens ved å forbedre seg selv i en positiv feedback-syklus, en såkalt «intelligens-eksplosjon».⁷³ Dette vil i så fall være en eksistensiell trussel for mennesket. Men slike perspektiver forutsetter algoritmer og fysiske forutsetninger som ikke finnes i dag.

Vi skal her ta for oss viktige utfordringer som allerede finnes i dagens teknologi, nemlig domenespesifikk maskinlæring basert på nevrale nett. De ulike måtene maskinene lærer på gir flere utfordringer:

- *Veiledet læring* bruker historiske data som kan speile skjeve forhold i samfunnet og føre til fordomsfulle beslutninger.

⁷³ Bostrom og Yudkowsky 2016.

- *Ikke-veiledet læring* identifiserer nye mønstre og sammenhenger, men kan være lite gjennomsiktede, vanskelige å forstå og forklare, og gjøre ansvaret utydelig.
- *Læring med forsterkning* gjør at maskiner kan komme frem til optimale strategier for nå sine mål, men samtidig overkjøre andre hensyn.
- I tillegg vil maskinlæring også kunne brukes som kraftige verktøy av aktører med *ondsinnete hensikter*.

FORDOMSFULLE ALGORITMER

Amerikaneren Kevin Johnson hadde en ryddig økonomi og en høy kredittscore, men fikk brått beskjed om at kredittgrensen var redusert med nesten 65 prosent. Årsaken var ikke mislighold eller forsinket betaling fra Johnsons side, men fordi handlemønsteret hans lignet mønsteret til kunder som har problemer med å betale.⁷⁴

Johnson var et offer for såkalt «atferdsprofilering», der likhetstrekkene mellom ens egen oppførsel og oppførselen til en større gruppe brukes for å styre beslutninger. På lignende vis lager maskinlæringsmodeller prediksjoner om sannsynlige handlinger eller kvaliteter på gruppenivå.

Veiledet læring gjør at maskiner kan klassifisere eller predikere utfall ganske presist, men prediksjonene kan bare bli så pålitelige og nøytrale som dataene de er basert på. I den grad det er ulikheter i samfunnet, ekskludering eller andre spor av diskriminering, så vil det også gjenspeile seg i dataene. Algoritmene kan dermed bidra til fordomsfulle beslutninger.

Et eksempel er systemer som vurderer jobbsøknader og peker ut de beste kandidatene. Når algoritmene trenes på data fra tidligere ansettelser, vil de kunne preges av skjeve valg og praksis fra intervjurundene. De kan utilsiktet fortsette å lære av fordommer som gjelder f.eks. rase, kjønn eller etnisitet. Slik profilering basert på skjeve data kan bidra til selvoppfyllende profetier og stigmatisering av grupper, selv om dette ikke var tilsiktet fra utviklerens side.

⁷⁴ Cuomo m. fl. 2009.

Dette er ikke et uløselig problem. Med økt bevissthet fra utviklernes side, kan algoritmer programmeres til å motvirke fordommer eller møte mål om likestilling. Det norske ikt-selskapet Evry har eksempelvis oppnådd sitt mål om flere kvinnelige ansatte etter at selskapet bruker et KI-system som del av rekrutteringsprosessen. Kvinneandelen blant de rundt 600 nyansatte i 2017 var 33 prosent, og 40 prosent blant nyutdannede, mot bare 20 prosent et par år tidligere. Selskapet mener dette skyldes at systemet baserer utvelgelsen på mer objektive kriterier.⁷⁵

SORT BOKS-PROBLEMET

I 2013 ble Eric Loomis dømt til seks års fengsel for å ha prøvd å rømme fra politiet i en bil som tidligere hadde blitt brukt i en skyteepisode i Wisconsin. Dommeren baserte den strenge dommen ikke bare på Loomis' rulleblad, men også på COMPAS-algoritmen, som beregnet at risikoen for en ny kriminell handling var høy.⁷⁶

Eric Loomis anket dommen til høyesterett, med henvisning til at dommeren brukte en algoritme som han verken kunne undersøke eller utfordre. Hvilke faktorer som er med i vurderingene og hvor stor vekt de har, regner selskapet bak COMPAS som en bedriftshemmelighet. Loomis fikk ikke medhold i anken. Dommerne mente at han ville fått den samme dommen uansett basert på de vanlige faktorene, som den kriminelle handlingen og hans kriminelle forhistorie. Domstolen påpekte imidlertid at de syntes var problematisk å bruke en hemmelig algoritme til å sende noen i fengsel.⁷⁷

Denne typen problemstillinger vil bli viktige i tiden fremover. Maskinlæringsalgoritmer gir råd og tar i økende grad beslutninger på områder som er av stor betydning for folks livskvalitet og utfoldelse, som låne- og jobbsøknader, medisinske diagnoser og i politiet. Da blir det problematisk om de ansvarlige ikke lenger kan eller vil forklare hvordan og hvorfor en beslutning er tatt. Algoritmene blir «sorte bokser» som i stedet skjuler vurderingene, usikkerhetene og valgene beslutningen hviler på.

⁷⁵ Dagens næringsliv 2017.

⁷⁶ Smith 2016. Se også Garber 2016.

⁷⁷ Liptak 2017.

Vi kan skille mellom to hovedtyper av sort-boks-problemer:

- *Innsyn i algoritmen begrenses bevisst* på grunn av kommersielle hensyn, nasjonal sikkerhet eller personvern hensyn. Dommen mot Eric Loomis er et eksempel på dette.
- *Algoritmen er komplisert og vanskelig å forklare* i menneskelig forståelige termer. Et eksempel er DeepPatient-modellen, som kan forutsi svingninger i schizofreni bedre enn legene, men uten verktøy for å forklare hvordan modellen kom frem til prediksjonene.⁷⁸

Den siste typen har blitt spesielt aktuelt på grunn av utviklingen innen maskinlæring. Tradisjonelle, regelbaserte maskinlæringsmetoder var utviklet av mennesker, og er dermed også lettere å tolke for mennesker. Dette er forskjellig fra dype nevrale nett, som kan ha hundrevis av millioner av tilkoblinger, som hver bidrar litt til den endelige beslutningen.

Ikke-veiledet læring gjør at maskiner kan identifiserer nye mønstre og sammenhenger i dataene, men de kan ikke nødvendigvis forklare den kausale sammenhengen. Algoritmene kan være lite gjennomsiktede og vanskelige å forstå, og mangel på forklaring gjør det både vanskelig å klage på en beslutning og å påta seg ansvar for beslutningene.

Teknikker for å forklare algoritmer

Flere forskningsprosjekt utvikler teknikker for å forklare eller begrunne anbefalinger fra algoritmer. Nedenfor er noen eksempler:

- *Universitetet i Oslo og Simulasenteret* utvikler et verktøy som skal hjelpe leger med rapportering og forklaring av hvordan algoritmer som analyserer video av tarmen, har kommet frem til sine anbefalinger. Systemet velger ut bilder som har vært viktige i avgjørelsen og legger til visualiseringer som viser hva i bildet som har vært avgjørende for anbefalingen.⁷⁹
- *XAI (Explainable AI)* kjøres i regi av det amerikanske forskningsbyrået for forsvaret (DARPA). Forsvaret ønsker hjelp til både automatiske varsler når f.eks fly eller satellitter oppdager noe mistenkelig, i tillegg til forklaring på

⁷⁸ Knight 2017b.

⁷⁹ GitHub 2018.

hvorfor algoritmer flagger noe. På den måten kan operatørene ignorere falske varsler.⁸⁰

- *LIME (Local Interpretable Model-Agnostic Explanations)* viser hvilke elementer i en modell som har vært relevant for en prediksjon, f.eks hvilke symptomer som har vært viktigst for en modell som predikerer om en person har influensa.⁸¹
- *En forskningsgruppe ved Berkeley universitetet i USA* har utviklet et program som er trent til å oppdage ulike fuglearter i fotografier, og å gi en forklaring på hvorfor. Systemet får hjelp fra et annet nevralt nett som har blitt trent til å koble egenskaper i et bilde med setninger som beskriver hva folk ser i bildet. Svaret fra algoritmen kan f.eks være *Dette er en «Svanedykker» fordi fuglen har en lang, hvit nakke, et gult, spisst nebb og røde øyne.*⁸²

RETTE TIL EN FORKLARING

Kunstig intelligens brukes både i automatiske beslutninger og som støtte til mennesker i delvis automatiserte beslutninger. I begge tilfeller vil den som beslutningene gjelder ha behov for forklaring. Da er det et problem om algoritmene er vanskelige å vurdere, kontrollere og rette opp.

I arbeidet med den nye europeiske personvernforordningen (*GDPR*) har en *rett til forklaring* av beslutninger basert på algoritmer derfor blitt et viktig tema. Forklaringer bak en beslutning kan være av to typer:

- *Hvordan systemet fungerer*, det vil si logikken, betydningen, forventede resultater og den generelle funksjonaliteten til systemet. Informasjon om logikken kan være om det er beslutningstrær som brukes og hvordan ulike opplysninger vektlegges og kobles. For eksempel kan høy fart automatisk føre til høyere forsikringspremie.
- *Forklaring av en spesifikk beslutning*, det vil si rasjonale, begrunnelser og individuelle omstendigheter som førte til beslutningen. Forklaringen på at en person får en forhøyet forsikringspremie kan f.eks. være at vedkommende har kjørt i gjennomsnitt 10 km over fartsgrensen over en periode.

⁸⁰ Gunning 2018.

⁸¹ Ribeiro, Singh og Guestrin 2016.

⁸² The Economist 2018a.

Før en automatisert behandling starter, må vedkommende få nok informasjon for å kunne gi sitt samtykke eller fremme innsigelser. I det tilfellet er kun systemfunksjonalitet tilgjengelig. Når en beslutning har funnet sted, og man f.eks. ønsker å klage på beslutningen, er også informasjon den spesifikke beslutningen tilgjengelig.

En studie utført av forskere ved Oxford konkluderer med at den nye personvernforordningen (GDPR) ikke gir en tilstrekkelig og meningsfull rett til forklaring etter at en beslutning har funnet sted. Forordningen gir snarere en mer begrenset rett til å bli informert i forkant, slik at man skal kunne gi sitt samtykke.⁸³

I fortalen til personvernforordningen (punkt 71) heter det at nødvendige garantier for personer som er gjenstand for automatiske beslutninger skal omfatte «... spesifikk informasjon til den registrerte og rett til å få kontakt med et menneske, til å uttrykke sitt synspunkt, til å få en forklaring på vedtaket ... og til å utfordre beslutningen.»⁸⁴

Den europeiske personvernforordningen (GDPR) gir den registrerte rett til ikke å være gjenstand for en avgjørelse som utelukkende er basert på automatisert behandling når avgjørelsen i betydelig grad påvirker vedkommende.⁸⁵ Imidlertid er retten til en forklaring av automatiserte beslutninger berørt i fortalen, men ikke nevnt i selve forordningen. Den er dermed ikke juridisk bindende.⁸⁶

Denne uklarheten er uheldig, gitt den spredningen og potensialet maskinlæring har for både helautomatiserte og delvis automatiserte beslutninger.

HVEM KAN HOLDES ANSVARLIG?

Kunstig intelligens flytter stadig grensen for hvilke oppgaver maskiner kan løse og hvilke beslutninger de kan ta. Selvkjørende biler tar løpende selvstendige beslutninger om manøvrering og bildeanalyseprogrammer kan stille diagnoser

⁸³ Wachter, Mittelstadt og Floridi 2017.

⁸⁴ GDPR 2016.

⁸⁵ Proposisjon til Stortinget 2018, s. 68.

⁸⁶ Wachter, Mittelstadt og Floridi 2017.

uten at en menneskelig ekspert er involvert. Dette reiser nye spørsmål knyttet til hvem som har ansvaret.

I følge dagens EU-regelverk kan autonome systemer som roboter og algoritmer ikke bli holdt ansvarlig for handlinger eller mangel på handlinger som fører til skade på tredjeparter. De kan dermed heller ikke stilles til ansvar for å gi kompensasjon og erstatning.⁸⁷ Det betyr at det er mennesker som fremdeles har det endelige ansvaret for beslutninger der maskinlæringsalgoritmer har vært med i prosessen.

For at et menneske skal kunne påta seg ansvar, vil det i mange tilfeller være nødvendig at vedkommende kan forstå hvordan algoritmene kommer frem til anbefalingene. I de situasjonene der en forklaring er viktig, men ikke tilstede, kan det derfor bli nødvendig å velge mindre presise algoritmer, men som til gjengjeld kan gi en forklaring.

ETISKE ALGORITMER

Teknikker for forsterket læring gjør at maskiner kan komme frem til optimale strategier for nå sine mål, med de reglene som mennesker definerer for dem. AlphaGo Zero har vist at algoritmer kan komme frem til bedre strategier enn de beste GO-spillerne. Maskinene vil imidlertid overkjøre hensyn som ikke blir eksplisitt uttrykt i reglene. Det kan være vanskelig for å overskue alle viktige hensyn når reglene skrives, spesielt hvis det er kompliserte systemer.

Et eksempel er knyttet til militær bruk. World Economic Forum har reist en bekymring knyttet til bruk av algoritmer som AlphaGo i krigføring. Algoritmen vil maksimere sannsynligheten for å vinne snarere enn å optimalisere marginene. Hvis denne spill-logikken brukes i autonome våpen, ville det kunne føre brudd på prinsippet om proporsjonalitet, fordi algoritmen ikke ville se noen forskjell på om den vil drepe 1.000 eller 1 fiender. Dette kan føre til en mer offensiv krigføring,⁸⁸

Dette gir en spesiell utfordring når det gjelder etiske valg. Et dilemma kan belyses med selvkjørende biler. De kan potensielt bli veldig dyktige til å unngå

⁸⁷ Europaparlamentet 2017.

⁸⁸ World Economic Forum 2017, s. 49.

ulykker. Men hva hvis maskinens mest optimale valg for bilen og dens passasjerer samtidig øker sannsynligheten for at noen andre blir skadet? Hvilken vei skal bilen kjøre dersom det står et stort dyr i veien, som potensielt vil gi bilen og passasjerene store skader, og det går et lite barn på fortauet?

Eventuelle etiske veivalg må uttrykkes i reglene dersom systemet skal ta hensyn til dem. Det kan være ubehagelig og vanskelig å uttrykke etiske valg i klare og entydige regler som maskinen forstår.

Tyske myndigheter har laget retningslinjer for hvilken etikk som skal gjelde og hvor ansvaret skal plasseres når selvkjørende biler kommer opp i valgsituasjoner. Noen av prinsippene er at mennesker skal prioriteres over eiendom og dyr, og det skal ikke gjøres forskjell på mennesker med bakgrunn i alder, kjønn eller annet. Målet med retningslinjene er å skape forutsigbarhet og tydelig ansvarsfordeling, slik at utviklere og produsenter ikke får byrden med å ta vanskelige etiske beslutninger. I tillegg skal det hjelpe til at de som befinner seg i trafikken kan føle seg sikrere på at bilene som kjører i trafikken ikke gjør uventede valg.⁸⁹

ONDSINNET BRUK

Teknologi kan brukes på flere måter og med ulike hensikter. Ekkoloddet var for eksempel opprinnelig utviklet for å oppdage og uskadeliggjøre u-båter, men ble senere viktig for fiskeriene. Motsatt kan forskning på virus også brukes for å utvikle farlige våpen for bio-terror eller krigføring. Dette kalles ofte *dual-use* eller «dobbelbruk» av teknologi.

Utfordringer med «dobbelbruk» gjelder i høyeste grad også kunstig intelligens. Autonome droner som leverer forbruksvarer kan for eksempel også levere sprengstoff. Noen generelle trekk ved maskinlæring gjør utfordringen presserende:

⁸⁹ Federal Ministry of Transport and Digital Infrastructure (i Tyskland) 2017.

- *Mer effektive og skalerbare.* KI-systemer kan gjøre oppgaver mer effektivt samtidig som de kan skaleres raskt og rimelig. Angrep, som for eksempel *phishing*, kan dermed bli vanskeligere å forsvare seg mot.⁹⁰
- *Bedre.* KI-systemer kan løse oppgaver langt bedre enn mennesker. De kan klassifisere medisinske bilder bedre enn eksperter, og de er flinkere enn topp-rangerte spillere i sjakk og GO.
- *Spres raskt.* Hvis man først har utviklet en smart algoritme, kan den reproduseres og spres raskt. Det er lave terskler for å ta i bruk teknikkene, siden feltet har en høy grad av åpenhet.
- *Psykologisk distanse.* Slike systemer kan fremme anonymitet og psykologisk distanse. For eksempel kan en soldat som bruker autonome våpen unngå både å være til stede, og dermed å selv bli tatt, og å se på offeret.
- *Nye og uløste sårbarheter* i dagens KI-systemer. De kan trenes opp til å gjøre feil ved å manipulere treningsdata og de kan lures med eksempler som er laget for å bli klassifisert feil. Selv om kunstig intelligens kan overstige menneskelig yteevne på mange måter, så kan de også feile på måter som mennesker aldri ville gjort.

En rapport utarbeidet av forskere og eksperter fra flere land mener maskinlæring kan true innbyggerne, organisasjoner og stater på spesielt tre områder:⁹¹

- *Digital sikkerhet:* f.eks. å trene maskiner til å gjøre dataangrep mer målrettet.
- *Fysisk sikkerhet:* f.eks. å manipulere selvkjørende biler til å kolliderer eller utstyre droner med våpen.

⁹⁰ Phishing er et konsept der angripere lurer et offer til å gjøre noe (som å oppgi konfidensiell informasjon eller overføre penger) ved å sende vedkommende en e-post og utgi seg for å komme fra en organisasjon eller en person hen stoler på. Til nå har phishingangrep stort sett vært basert på likelydende e-poster til en stor mengde mottakere. KI kan gjøre phishing mye mer målrettet og effektivt ved å granske hvert offer først, som å studere aktiviteten i sosiale medier, og persontilpasse e-postene og angivelige avsendere. Dermed er det mer sannsynlig at offeret blir lurt. Slik kan angrep fremover bli både mer effektive, og gjøres med større frekvens og i større skala.

⁹¹ Brundage m. fl. 2018.

- *Politisk sikkerhet*: f.eks. å utforme desinformasjon (*fake news*) og spre den mer målrettet.

Optisk illusjon

Maskinlæringsmodeller kan lures med en slags optisk illusjon til å gjøre feil. Ved å legge til støy i bilder, som ikke synes med det blotte øyet, kan en modell lures til å klassifisere feil. Et bilde som ser helt ut som en panda har f.eks. blitt klassifisert som en ape ved å legge til usynlig støy i bildet.⁹² Slike angrep kan få dramatiske konsekvenser. Algoritmer kan læres opp til å tolke bilder av f.eks. et fartsgrenseskilt som viser 50 km/t som 150 km/t. Hvis man skriver ut det falske bildet og limer det på vanlige trafikkskilt, vil autonome kjøretøy kjøre mye raskere enn den faktiske fartsgrensen, med de konsekvensene det fører med seg. (Dette kan forhindres ved at systemet er innrettet for å håndtere unormale sensorverdier, f.eks. ved at det også er koordinert opp mot et digitalt kart.) Forsøk har også klart å lure algoritmer til å tolke bilder av en person med noen spesielle briller til å klassifisere dem som en annen person.⁹³ Folk kan dermed skifte identitet ved å ikle seg slike briller, og f.eks. snike seg urettmessig gjennom passkontroller.

Det jobbes med teknikker på mange plan for å forsvare seg mot ondsinnede angrep med kunstig intelligens; økt bevissthet hos forbrukerne, systematisk arbeid med å oppdage og avdekke svakheter i systemer, sentraliserte løsninger (som man gjør med spam-filtre) som klarer å holde seg oppdatert på nye trusler, sertifisering av autentisitet for å bevise at bilder og videoer ble kringkastet live (og ikke generert kunstig) og krav om registrering av roboter.⁹⁴

Det kan bli en spenning mellom hensynene til åpenhet og sikkerhet. Åpenhet rundt algoritmer vil være viktig for å redusere risiko for sårbarheter og misbruk, samtidig som åpenhet kan eksponere algoritmene for ondsinnet bruk.

⁹² OpenAI.com 2017.

⁹³ Margolin 2016.

⁹⁴ Brundage m. fl. 2018.

14 FORSLAG FOR NORGE

Kunstig intelligens gir store muligheter både for verdiskaping og bedre velferdstjenester, samtidig som teknologien kan påvirke innbyggernes rettigheter og skape ulikhet. En nasjonal strategi bør adressere kompetanseutfordringen, behovet for data og en ansvarlig utvikling.

NORGE TRENGER EN STRATEGI

Kunstig intelligens gir store muligheter i mange sektorer, har viktige implikasjoner for enkeltmennesket og samfunnet, og utvikler seg raskere enn noen annen teknologi. Det er derfor presserende at norske myndigheter utvikler en strategi for å høste gevinstene og møte utfordringene ved denne potente teknologien.

STORE MULIGHETER

I denne rapporten har vi gitt eksempler på hvordan maskinlæring allerede i dag påvirker mange ulike sektorer, blant annet i

- *Helse*: diagnostisering, etterlevelse, presis medisineri, ressursbruk
- *Transport*: selvkjørende biler, trafikkplanlegging
- *Finans*: lånesøknader, tolke kontrakter

- *Energi*: optimalisere datasentre og energisystemer, forebygge feil i petroleumssektoren
- *Offentlige tjenester*: persontilpassede tjenester, automatisert saksbehandling, rekruttering, oversettelse.

I tillegg vil kunstig intelligens være viktig for industri og verdiskapning. En analyse gjort av Accenture konkluderer med at det økonomiske vekstpotensialet frem mot 2035 blir dobbelt så stort med bruk av kunstig intelligens.⁹⁵

For Norge blir dermed kunstig intelligens viktig både for å sikre konkurransekraften i privat sektor og for å utvikle bærekraftige, bedre og mer effektive velferdstjenester.

VIKTIGE KONSEKVENSER FOR DEN ENKELTE OG SAMFUNNET

Med maskinlæring kan analyser, tolkninger og handlinger som hittil har vært forbeholdt mennesker overtas av maskiner, og gjøres raskere og rimeligere. Dette kan også gi nye problemer.

For det første kan det være vanskelig å forstå hvorfor algoritmene foreslår noe eller handler som de gjør. Dette gjør det vanskelig å protestere mot eller endre beslutninger, som kan være basert på skjeve datasett eller skjulte interesser. Forfatteren og matematikeren Cathy O'Neil har betegnet slike algoritmer som «matteødeleggersvåpen» siden de kan ramme mange, er ugjennomsiktige og får store konsekvenser for eksempelvis personlig økonomi, utdanning eller straffeutmåling.⁹⁶

For det andre kan kunstig intelligens også brukes som et raffinert våpen av kriminelle eller fremmede makter. Slik såkalt «dobbelbruk» kan omfatte dataangrep, manipulering av fysiske objekter som droner eller biler, og målrettet manipulering av politikken gjennom spredning av falske nyheter.

En tredje mulig konsekvens er at arbeidsmarkedet blir forstyrret ved at mange arbeidsoppgaver i løpet av relativt kort tid overtas av intelligente maskiner. En studie fra SSB anslår at én av tre norske jobber har høy risiko for å bli erstattet

⁹⁵ Purdy og Daugherty 2016. Analysen omfatter 12 land, deriblant Sverige, Finland og Storbritannia og hadde en tidshorisont fra 2016 til 2035.

⁹⁶ O'Neil 2016.

av maskiner i løpet av de neste to tiårene.⁹⁷ OECDs beregninger tilsier at færre jobber vil kunne forsvinne helt, men at for omtrent en tredjedel av arbeidstakerne kan store deler av oppgavene overtas av datamaskiner.⁹⁸

Dessuten kan utviklingen av kunstig intelligens bidra til økt ulikhet. Kommerisiell kunstig intelligens domineres i dag av plattform-selskaper som Google, som har tilgang til mye data. I den digitale økonomien gjør nettverkseffekter ofte vinnerne enda sterkere: Jo flere som bruker tjenestene, jo mer data får selskapene, som igjen kan forbedre tjenestene ytterligere.

EN TEKNOLOGI I RASK UTVIKLING – OG NORGE LIGGER BAK

Kunstig intelligens er en teknologi som har gjort et kraftig sprang de siste årene. Nye typer maskinlæring gjør nytte av stadig økende regnekraft og de store mengdene data som produseres i samfunnet. En fersk undersøkelse viser at 85 prosent av USAs befolkning allerede bruker tjenester som baserer seg på kunstig intelligens, som for eksempel navigasjon, strømmetjenester eller transport.⁹⁹ Tallet for Norge er antageligvis høyere.

Dagens situasjon kan oppsummeres slik:

Private amerikanske selskaper dominerer

Verdens mest verdifulle selskaper, som Apple, Amazon, Facebook og Alphabet (morselskapet til Google), bruker maskinlæring for å persontilpasse tjenestene og optimalisere driften. De har mange hundre millioner daglige brukere over hele verden og får dermed også store mengder treningsdata.¹⁰⁰ Spesielt Alphabet ligger godt an, og det anslås at dette selskapet har rundt halvparten av de 100 beste utviklerne innenfor maskinlæring, med satsninger som Google Brain

⁹⁷ Pajarinen m.fl. 2014.

⁹⁸ Nedelkoska og Quintini 2018.

⁹⁹ Gallup 2018.

¹⁰⁰ The economist 2017.

(en ny type KI-operativsystem), Google Cloud og DeepMind.¹⁰¹ Amazon har investert 306 millioner amerikanske dollar i nye KI-stillinger, og er dermed det selskapet som har investert mest den siste tiden.¹⁰²

Kina har store ambisjoner og mange brukere

Kina har flest mobil- og internettbrukere i verden, omtrent tre ganger så mange som i USA eller India. Kineserne bruker dessuten mobilbetaling 50 ganger oftere enn amerikanerne. Disse datamengdene er drivstoff for maskinlæring og har allerede bidratt til at Kina er verdenslende på stemme- og ansiktsgjenkjenning. I tillegg har landet en sterk posisjon innenfor robotikk og automatisering.¹⁰³ Kinesiske myndigheter har lansert en ambisiøs strategi for å bli det globale senteret for utvikling av kunstig intelligens innen 2030. Dette skal oppnås gjennom målrettede planer for forskning og nyskaping på bred front, smarte offentlige tjenester innen transport, byutvikling, utdanning og rettsvesen, og omfattende sivil-militært samarbeid.¹⁰⁴

EU vil bli best på etisk KI

EU-kommisjonen la frem sin satsing på kunstig intelligens 20. april 2018. Kommisjonen ønsker en kraftig økning i private og offentlig investeringer til 20 milliarder Euro før utgangen av 2020, og vil tilpasse lovverket for gjenbruk og deling av data. EU-strategien har en klar sosial og etisk profil, med vekt på å takle utfordringer for arbeidsmarked, utdanning og inkludering, samt utvikling av etiske retningslinjer basert på grunnleggende rettigheter og den nye europeiske personvernforordningen (*GDPR*). Ulike interessenter vil bli samlet i European AI Alliance for å utvikle de etiske retningslinjene i løpet av 2018.¹⁰⁵

Våre naboland er i gang

Mange av EU-landene er i gang med sitt eget strategiarbeid. Både i Sverige og Finland er de første rapportene med forslag til regjeringen publisert.¹⁰⁶ Den franske presidenten Emmanuel Macron presenterte i mars sin plan for kunstig

¹⁰¹ Sinovation Ventures 2018, s. 5.

¹⁰² Paysa.com 2017. Undersøkelsen, fra april til september 2017, viser at Amazon har investert 306 millioner amerikanske dollar i nye KI-stillinger, etterfulgt av Microsoft (124 millioner), Apple (105 millioner) og Google (33 millioner).

¹⁰³ Folkerepublikken Kinas statsråd 2017, s. 3.

¹⁰⁴ Folkerepublikken Kinas statsråd 2017.

¹⁰⁵ EU-kommisjonen 2018a.

¹⁰⁶ Vinnova 2018 og Finlands arbeids- og næringsdepartement 2017.

intelligens, med en forskningssatsing på 1,5 milliarder euro.¹⁰⁷ I Storbritannia har regjeringen og en rekke private virksomheter inngått en *AI Sector Deal* for utviklingen av kunstig intelligens.¹⁰⁸

Norge ligger bak og har ingen nasjonal strategi.

Data fra plattformen Kaggle, som samler 15.000 utviklere av maskinlæring, er en indikasjon på hvor kompetansen befinner seg i verden. I en brukerundersøkelse kommer de fleste respondentene fra USA (4.200), foran India (2.700), Russland (578) og Storbritannia (545). Norge ligger langt nede på listen med kun 53 respondenter. I en rangering av de 100 beste utviklerne, kommer ingen fra Norge.¹⁰⁹

Norge ligger dessuten kun som nummer 15 av 35 land på *Government AI Readiness Index*, som rangerer hvor godt forberedt OECD-landene er for å implementere kunstig intelligens i offentlige tjenester.¹¹⁰

NTNU og flere ledende selskaper i Norge har nylig gått sammen om å etablere Norwegian Open AI Lab og Forskningsrådet nevner kunstig intelligens som ett av flere prioriterte områder i IKTPLUSS-programmet.¹¹¹

Imidlertid nevner regjeringens *Langtidsplan for forskning og høyere utdanning 2015–2024* verken kunstig intelligens eller maskinlæring. Norge har heller ingen nasjonal strategi for kunstig intelligens eller har satt i gang arbeid med dette.

I det følgende vil vi komme med konkrete innspill til en norsk strategi for kunstig intelligens. De viktigste elementene er riktig og tilstrekkelig kompetanse til

¹⁰⁷ Reuters 2018.

¹⁰⁸ Department for Business, Energy and Industrial Strategy og Department for Digital, Culture, Media and Sport (i Storbritannia) 2018.

¹⁰⁹ Kaggle 2017 visualiserer svar på en undersøkelse fra 16.716 brukere på Kaggle, fra 171 land og områder. Scimago 2018 viser at Kina (102.000) og USA (84.000) ligger på topp med to til tre ganger så mange publikasjoner i kunstig intelligens som nestemann Japan (34.000) de siste 20 årene. Norge er på 41. plass med 1.700 publikasjoner.

¹¹⁰ Stirling m.fl. 2018.

¹¹¹ Telenor 2018 og <https://www.forskningsradet.no/no/Utlysning/IKTPLUSS/1254002623262/>.

å utvikle, vurdere og ta i bruk maskinlæring, tilgang til data som balanserer personvern med mulighet for å drive innovasjon og tiltak og prinsipper for en utvikling som både er forsvarlig og ønskelig.

KOMPETANSEUTFORDRINGEN

Regjeringen definerer muliggjørende teknologier som «teknologier som viser seg å bli så gjennomgripende at de fører til store endringer i samfunnet».¹¹² Vi vil hevde at kunstig intelligens er en slik teknologi, som kan sammenlignes med en turbolader for digitaliseringen av samfunnet.¹¹³

1. ET UMIDDELBART FORSKNINGSLØFT

Det er nødvendig med spesialisert, forskningsbasert kompetanse for å utvikle robuste algoritmer for maskinlæring.

EU-kommisjonen peker på at det haster med å styrke Europas forskningsmessige kvalitet og kapasitet innen kunstig intelligens. EU vil derfor øke investeringen i forskningsprogrammet Horizon 2020 med 1,5 milliarder Euro frem mot 2020.¹¹⁴ Norge har forpliktet seg til å bidra til denne satsingen gjennom en samarbeidserklæring. Innen utgangen av 2018 skal EU-kommisjonen presentere en koordinert plan for investeringer og videre samarbeid mellom Norge og 24 EU-land.¹¹⁵

Norge ligger i dag et stykke bak på dette området. Forskningsrådet slår for eksempel fast at det er spesielt utfordrende å dekke kompetansebehovet innen kunstig intelligens og maskinlæring i Norge.¹¹⁶

Regjeringens *Langtidsplan for forskning og høyere utdanning 2015–2024* skal revideres i løpet av 2018. Her bør det skapes rom for en dedikert satsing på kunstig intelligens og maskinlæring.

¹¹² Kunnskapsdepartementet 2014, s. 30.

¹¹³ Finlands arbeids- og næringsdepartement 2017, s. 11.

¹¹⁴ EU-kommisjonen 2018a, s. 7. Se også EU-kommisjonen 2018b.

¹¹⁵ EU-kommisjonen 2018c.

¹¹⁶ NFR 2017, s. 90.

2. ETABLERE EN NØKKELINSTITUSJON

Norge har få og spredte forskningsressurser. For å styrke forskningsinnsatsen og bli attraktivt for rekruttering og internasjonalt samarbeid, kan det være en god idé for norske myndigheter å etablere en nøkkelinstitusjon for forskning på kunstig intelligens og maskinlæring.

Institusjonen bør være tverrfaglig sammensatt og adressere både teoriutvikling, anvendelse og etiske vurderinger knyttet til utvikling av kunstig intelligens. For å favne bredt nok og samtidig ha tilstrekkelig forskningshøyde kan institusjonen omfatte flere forskningsmiljøer og virksomheter i en virtuell organisering.¹¹⁷

I Ontario i Canada har lokale og statlige myndighetene etablert Vector-instituttet sammen med bl.a. universitetet i Toronto og private næringsaktører, som ett av tre nasjonale knutepunkt for utvikling av kunstig intelligens. Satsingen er også et tiltak for å forhindre at delstaten mister eksperter på kunstig intelligens og maskinlæring til de store USA-baserte selskapene.¹¹⁸

3. PEKE UT AMBISIØSE OG KONKRETE MÅL FOR NORGE

Norge har ikke forutsetninger for å satse like bredt som Kina eller Frankrike, men kan bli verdensledende ved å koble domenekunnskap med generell kunnskap om kunstig intelligens.

EU-kommisjonen og Storbritannias statsminister Theresa May har introdusert en ny form for forskningssatsninger, som de kaller *missions*. Dette er dristige, inspirerende og ambisiøse målsettinger for å løse noen av de store samfunnsutfordringene vi har i dag. Samtidig skal de ha realistiske forsknings- og innovasjonsaktiviteter, med krav om målbare og tidfestede resultater.¹¹⁹

En av Mays *missions* har som mål å bruke data, kunstig intelligens og innovasjon for å transformere forebygging, tidlig diagnose og behandling av sykdom-

¹¹⁷ Finlands arbeids- og næringsdepartement 2017, s. 49.

¹¹⁸ Satsningen kommer som følge av den canadiske KI-strategien, se også Universitetet i Toronto 2017 og CIFAR 2017.

¹¹⁹ EU-kommisjonen 2018d og Mazzucato 2018.

mer som kreft, diabetes, hjertesykdom og demens innen 2030. En av ambisjonene er at i løpet av 15 år skal det være mulig å diagnostisere kreft i lunge, tykktarm, prostata og eggstokkene på et mye tidligere tidspunkt hos 50 000 flere pasienter i året, og dermed øke overlevelse over 5 år for 22 000 flere briter hvert år.¹²⁰

Kunstig intelligens kan bidra til å løse flere viktige samfunnsutfordringer. Det vil være fornuftig for Norge å formulere mål innenfor følgende områder, der vi har en kombinasjon av gode treningsdata og store samfunnsmessige behov:

- *Helse:* Norge har et relativt enhetlig helsevesen med gode helsedata og digitalt aktive brukere. Det forventes økt behov for helsetjenester i takt med eldrebolegen.
- *Offentlige tjenester:* Norge har offentlige data i verdensklasse, som et resultat av en godt organisert velferdsstat med digitalt aktive innbyggere. Perspektivmeldingen slår fast at offentlige utgifter vil øke raskere enn offentlige inntekter fra 2030.¹²¹ Det blir derfor nødvendig med en omstilling av hvordan det offentlige leverer sine tjenester.
- *Bærekraftig energi:* Det finnes allerede store mengder sensordata fra olje- og gassinstallasjoner. Equinor har nylig etablert to digitaliserte operasjonssentre i Bergen, og forventer at investeringene på 1-2 milliarder kroner skal gi økt verdiskaping på rundt 15-20 milliarder kroner.¹²² Agder Energi bruker maskinlæring for å optimalisere vannkraftproduksjonen.¹²³ De globale klima- og miljøutfordringene krever omstilling og en overgang til produkter og tjenester som gir betydelig mindre negative konsekvenser for klima og miljø enn i dag. Samfunnet må gjennom et grønt skifte.
- *Rent hav:* Norske institusjoner og selskaper har mye data fra bl.a. satellitter, bøyer og droner, som kan gi viktig kunnskap. Norge har juridisk råderett over store havområder og forvalter enorme ressurser både i havet og i offshore-industrien. Samtidig er havet under betydelig press som følge av bl.a. forurensning, oppvarming og forsuring. Her kan Norge ta et særlig ansvar.

¹²⁰ May 2018.

¹²¹ Melding til Stortinget. 2017.

¹²² Equinor 2018.

¹²³ Moe og Breivik 2018.

4. MASTERGRADER FORSTERKET MED KUNSTIG INTELLIGENS

Maskinlæring vil bli en viktig ingrediens i mange bransjer og yrker, fra produksjon, olje og energi, media og underholdning, jord- og havbruk, medisin, undervisning og offentlige tjenester.

Alle profesjoner og utdanningsløp, både på universitets- og høyskolenivå, bør derfor gi en innføring i kunstig intelligens og maskinlæring som et supplement til tilbud for studenter og forskere. Et eksempel er Medisinsk institutt ved Universitetet i Bergen, som starter opp et nytt kurs for bl.a. legestudenter våren 2019. Kurset skal sette dem i stand til å forstå og vurdere hvordan maskinlæring og datanalyse kan brukes i prediktiv og persontilpasset medisin.¹²⁴

Også andre utdanninger vil ha behov for å inkludere kunstig intelligens i tiden fremover, både for å gi ferdigheter til utvikling av domenespesifikk kunstig intelligens, og for å gi et kritisk kunnskapsgrunnlag for praktikere.¹²⁵

Det bør også etableres dedikerte mastergradsprogram i kunstig intelligens. Programmene kan være modulære og mulige å fullføre samtidig med jobb, og de kan også brukes på mange felt, som f.eks. helse, justis og logistikk.¹²⁶

Masterprogram kan utvikles og finansieres i forpliktende samarbeid mellom industrien, akademien og myndighetene. I britiske *AI Sector deal* forplikter industrien seg til å utvikle industrifinansierte master-program innen KI og i første omgang finansiere 200 studenter per år. De skal også vurdere mulighetene som ligger i andre disipliner.¹²⁷

5. GI ALLE MULIGHET TIL Å LÆRE KUNSTIG INTELLIGENS

Kunstig intelligens vil påvirke livene våre og de valgene vi gjør, både privat og i yrkeslivet. Det blir viktig at flest mulig forstår de viktigste implikasjonene av

¹²⁴ <http://www.uib.no/emne/ELMED219>.

Et annet eksempel er Handelshøyskolen ved NMBU, som tilbyr kurs i maskinlæring for optimalisering av forretningsprosesser. <https://www.nmbu.no/emne/INN355>.

¹²⁵ Eksempler på områder kan være energi og miljø, hav- og jordbruk, politi og justis, helsefag og lærere.

¹²⁶ Et lignende opplegg er nå foreslått i Finland. Finlands arbeids- og næringsdepartement 2017, s. 52.

¹²⁷ Department for Business, Energy and Industrial Strategy og Department for Digital, Culture, Media and Sport (i Storbritannia) 2018.

kunstig intelligens, slik at de kan tenke kritisk om temaet, være med å utforme bruken i på arbeidsplassen, og at de kan delta i og forme debatten.

Helt konkret kan norske myndigheter la seg inspirere av Finland og sette et ambisiøst mål om at f.eks. én prosent av befolkningen skal lære seg grunnleggende kunnskaper om kunstig intelligens hvert år. Finske myndigheter lanserte i mai 2018 *Elements of AI*, som er et gratis, nettbasert og poenggivende grunnkurs i kunstig intelligens.¹²⁸

For det andre er det et stort behov for opplæring på arbeidsplassen. OECD vurderer at omtrent 1/3 av norske jobber vil få radikalt endret innhold fremover på grunn av automatisering og kunstig intelligens. Om lag 850 000 nordmenn vil dermed få behov for omfattende kompetansehevede tiltak, noe dagens etter- og videreutdanning ikke er rigget for.¹²⁹

Dette krever at Norge legger om dagens etter- og videreutdanning og gjør den tilpasset den enkelte med nye insentiver. Singapore tilbyr *SkillsFuture for Digital Workplace* som gir opplæring i digitale ferdigheter tilpasset ulike yrkesgrupper.¹³⁰ Alle borgere over 25 år får *SkillsFuture Credit*, som er 500 singaporske dollar til å bruke på kurs hvert år.¹³¹

NORGES FORTRINN: DATA

Maskinlæringsalgoritmer trenes på data. Uten data som råmateriale, er det dermed vanskelig å utvikle kunstig intelligens. Norge har offentlige data i verdensklasse, som et resultat av en godt organisert velferdsstat med digitalt aktive innbyggere. Dette har utvilsomt stor verdi for private aktører som ønsker å utvikle kommersielle tjenester, men en strategi bør også skape verdier for fellesskapet og gi tilstrekkelig kontroll over egne data til den enkelte innbygger.

¹²⁸ Kurset er rettet mot folk flest, og har ingen krav om forkunnskaper i matematikk eller programmering. Bestått kurs gir studiepoeng for finske innbyggere og sertifikat for alle som gjennomfører. Ambisjonene er at én prosent av den finske befolkningen gjennomfører kurset det første året. <https://course.elementsofai.com/>.

¹²⁹ Nedelkoska og Quintini 2018.

¹³⁰ SkillsFuture 2017.

¹³¹ <http://www.skillsfuture.sg/Credit>

6. ÅPNE OFFENTLIGE DATA

Åpne, offentlige data kan bidra til innovasjon og nye tjenester i mange sektorer.

Offentlige institusjoner bør som hovedregel dele data. Norge ligger på tiende plass av 114 land på et barometer som viser i hvilken grad offentlige myndigheter publiserer og bruker åpne data.¹³² Offentlig sektor i Norge bør likevel ha ambisjoner om å publisere mer offentlige data, og gjøre det på et åpent format som er lett å finne frem i og gjenbruke i maskinlæring.¹³³

I *Sector Deal* i Storbritannia har myndighetene forpliktet seg til å publisere mer åpne data, selv om landet allerede er rangert som nummer 1 på det samme barometeret.¹³⁴

7. DATADELING SOM TJENER FELLESSKAPET

Hvis data fra norske sykehus, skoler og smarte byer skal deles med tredjeparter, bør felleskapet få en merverdi i form av bedre offentlige tjenester, ny næringsutvikling, arbeidsplasser eller skatteinntekter.

Ved maskinlæring er det ikke alltid forutsigbart hvordan dataene skaper verdi og for hvem. Maskinene kan lære på egenhånd og komme frem til nye sammenhenger som ikke var kjent tidligere, og læringen kan overføres fra ett system til et annet. Dette gjør det komplisert å regulere partenes ansvar og rettigheter. Hvor mye skal for eksempel det offentlige betale for tjenester trent på egne data?

I tillegg er innbyggerne part i saken, siden offentlige data er om og fra dem. Dette innebærer ekstra krav til forsvarlig bruk og innsyn dersom tilliten til data-delingen skal opprettholdes.

¹³² https://opendatabarometer.org/?_year=2016&indicator=ODB

¹³³ For eksempel bør registre som Norsk pasientregister og Reseptregisteret dele aggregerte data og syntetiske datafiler (ikke personidentifiserbare) som alle kan bruke.

¹³⁴ Department for Business, Energy and Industrial Strategy og Department for Digital, Culture, Media and Sport (i Storbritannia) 2018.

Det er derfor nødvendig at myndighetene etablerer juridiske rammer som gjør det mulig å utveksle data trygt og som sikrer at rettigheter, verdier og ansvar fordeles rettferdig og likeverdig, også i fremtiden.

Her kan norske myndigheter hente inspirasjon fra Storbritannia, som etablerer såkalte *Data Trusts*.¹³⁵ Dette er langsiktige juridiske rammeverk for deling av data mellom offentlige virksomheter og private selskaper som skal utvikle kunstig intelligens. Rammeverket inkluderer konkrete verktøy, avtalemaler og mekanismer for fordeling av gevinstene.¹³⁶

Tillitsavtalene må utvikles og justeres løpende, i tett samarbeid med både de som deler og de som bruker data. Det bør derfor etableres en kompetent driftsorganisasjon som leder prosessene og har tilsyn med rammeverket.¹³⁷

Det er også nødvendig med kompetansebygging i hvordan data kan deles, kobles og håndteres på en sikker måte, som samtidig gjør at mange kan få tilgang. For eksempel gjelder dette kompetanse i å lage syntetiske data og kryptering.

¹³⁸

8. GI INNBYGGERNE REELL KONTROLL OVER EGNE DATA

Hvis offentlige data om oss deles for å drive forskning og innovasjon, forutsetter det at innbyggerne også får reell kontroll over hvordan egne data deles og at det gjøres på en trygg måte.

¹³⁵ Department for Business, Energy and Industrial Strategy og Department for Digital, Culture, Media and Sport (i Storbritannia) 2018 og Hall og Presenti 2017. Se også Thornhill 2017 og Artificial Lawer 2017.

¹³⁶ Finansnæringen i Norge har sammen med Brønnøysundregistrene, Skatteetaten, NAV og Politiet samarbeidet om digitalisering i DSOP-samarbeidet (Digital Samhandling Offentlig - Privat). Hensikten er å kunne dele informasjon enkelt, effektivt og digitalt slik at den blir mest mulig produktiv i samfunnet, men samtidig innenfor trygge rammer som ivaretar personvernet for den enkelte. Rammeverket er åpent for alle, og det kan vurderes om elementer herifra kan videreutvikles og tilpasses deling av data for å utvikle KI-systemer. Se <https://www.bits.no/project/dsop/> og Holte 2018.

¹³⁷ I Storbritannia blir det anbefalt at det etableres en driftsorganisasjon for Data Trusts. Det skal opprettes et nytt *Centre for Data Ethics and Innovation*. Hall og Presenti 2017.

¹³⁸ Tyskland har jobbet for å etablere kompetanse om deling av data og om hva det offentlige kan gjøre: Antoni og Schnell 2017.

I Norge har innbyggeren i dag begrenset kontroll over data om seg selv. Innbyggerens data befinner seg i ulike private og offentlige «siloer», med ulik policy for innsamling, deling og bruk. Dermed er det også krevende å forstå, vurdere og håndtere risikoen knyttet til datainnsamling og bruk.

Teknologirådet har tidligere argumentert for å etablere en tydelig digital samfunnskontrakt for samspillet mellom innbyggere og offentlige virksomheter.¹³⁹ Å gi innbyggerne reell mulighet til å kontrollere og forme sin digitale profil og bestemme om og hvordan egne data skal deles vil være en viktig del av en slik samfunnskontrakt.

Myndighetene må derfor legge til rette for at innbyggerne får tilgang til hensiktsmessige verktøy, slik at de reelt sett og på en effektiv måte kan kontrollere informasjon om seg selv.¹⁴⁰ På samme måte som de forvalter sin personlige økonomi i nettbanken, må de få et digitalt grensesnitt som gir enkel og forståelig oversikt over hvordan personlige data forvaltes og brukes av det offentlige. Innbyggeren må også få muligheten til å aktivt gi eller trekke tillatelser for ulike bruksformål.

ANSVARLIG OG ØNSKET UTVIKLING

I Global Risk Report 2017 betegner World Economic Forum kunstig intelligens som en av de fremvoksende teknologiene med størst nytteverdi, men også med størst skadepotensial.¹⁴¹

Lærende maskiner kan bidra til bedre velferdstjenester, raske og presise diagnoser og bedre bærekraft. Samtidig kan kunstig intelligens bety færre jobber, mer overvåkning, større ulikhet og autonome våpen.

¹³⁹ Teknologirådet 2017, s. 55.

¹⁴⁰ Dette påpeker også den franske og finske KI-strategien, som viser til eksempler som Personaldata.ai, Personal Information Management Systems (PIMS) og MyData.org. Se også Villani 2018, s. 31 og Finlands arbeids- og næringsdepartement 2017, s. 44 og 45.

¹⁴¹ World Economic Forum 2017.

Mye står på spill. Det er derfor nødvendig å løpende adressere og diskutere hva som er ansvarlig og ønsket utvikling, og hva som bør gjøres for å forme teknologien. De såkalte *missions* som EU og Storbritannia har introdusert vil kunne være et viktig instrument for å oppnå en ønsket og ansvarlig utvikling.¹⁴²

9. ETISKE RETNINGSLINJER

Regjeringen har erklært at den vil utvikle retningslinjer og etiske prinsipper for bruk av kunstig intelligens.¹⁴³ Dette er en god idé. Muligheten for å skape maskiner som lærer, tolker og beslutter reiser mange etiske spørsmål.

På lengre sikt kan dette handle om hvorvidt maskiner skal få en moralsk status, eller at maskinene oppnår superintelligens og blir en eksistensiell trussel for mennesket.¹⁴⁴ Men slike perspektiver forutsetter algoritmer og fysiske forutsetninger som ikke finnes i dag.

Det er i denne rapporten pekt på etiske problemer som allerede er til stede og vil forsterkes. Tradisjonelle europeiske verdier som verdighet, autonomi, frihet, solidaritet, likeverd, demokrati og tillit blir utfordret i takt med digitaliseringen generelt og utviklingen av kunstig intelligens spesielt.¹⁴⁵ Regjeringen bør begynne å utvikle etiske retningslinjer og praksiser der teknologien allerede setter etablerte verdier under sterkt press:

- *Autonomi* for mennesker i møte med teknologien. Maskinlæring gjør prediksjoner av enkeltpersoners atferd og preferanser mer presise og billigere enn før. Dermed blir det også mulig å påvirke og manipulere handlinger og holdninger.
- *Demokrati*. Potensialet for politisk manipulering ved hjelp av kunstig intelligens er åpenbar. Det britiske analysefirmaet Cambridge Analytica brukte Facebook som dataleverandør for å lage en psykologisk profil av flere millioner enkeltpersoner og som plattform for å tilby politisk påvirkning til kunder. Også manipulering av media ved persontilpassete falske nyheter undergraver demokratiske verdier. Den kinesiske KI-strategien er

¹⁴² Mazzucato 2018 og May 2018.

¹⁴³ Høyre 2018.

¹⁴⁴ Bostrom og Yudkowsky 2016.

¹⁴⁵ EDPS 2018, s. 16.

tydelig i sin målsetning om å bruke maskinlæring for å oppnå sosial kontroll.

- *Rettferdighet.* Det er en økende asymmetri mellom enkeltmennesket og bedrifter eller myndigheter som har mye data og bruker disse til analyse og påvirkning. Dette forsterkes av at nettverkseffekter gir store kommersielle aktører tilnærmet monopol innenfor sine områder.
- *Likeverd.* Maskiner lærer av data som er samlet inn om samfunnet. De kan spille historisk skjeve forhold og dermed sementere fordomsfulle beslutninger og føre til diskriminering. Også persontilpasning er i prinsippet diskriminering.
- *Solidaritet.* Velferdsordninger som helsetjenester og trygder, og ulike former for forsikring bygger på solidarisk deling av risiko. Økende persontilpasning og hyperindividualisering gjennom risikoskåring og prediksjon for hver enkelt innbygger kan undergrave dette.
- *Ansvarlighet.* Det at maskinene får mer autonomi med kunstig intelligens kan tilslore det grunnleggende prinsippet om at et menneske alltid må stå ansvarlig for beslutninger som berører andre mennesker. Algoritmene kan være lite gjennomsiktede og vanskelige å forstå, noe som gjør det både vanskelig å plassere ansvar for beslutningene og klage på den. Det kan også være umulig å vite om man er i kontakt med en maskin eller et menneske. Fremveksten av intelligente våpensystemer med stort potensial for autonomi setter spørsmålet om ansvarlighet på spissen.

10. RETT TIL EN FORKLARING

Maskinlæringsalgoritmer gir råd og tar i økende grad beslutninger på områder som er av stor betydning for folks liv, som låne- og jobbsøknader, medisinske diagnoser og i politiet. Da blir det også viktig å kunne få en forklaring på beslutningen, slik at det blir mulig å klage på eller endre en urettferdig praksis.

Den europeiske personvernforordningen (GDPR) gir den registrerte rett til ikke å være gjenstand for en avgjørelse som utelukkende er basert på automatisert

behandling når avgjørelsen i betydelig grad påvirker vedkommende.¹⁴⁶ Imidlertid er retten til en forklaring av automatiserte beslutninger berørt i fortalen, men ikke nevnt i selve forordningen. Den er dermed ikke juridisk bindende.¹⁴⁷

Norske myndigheter bør derfor knesette og presisere en slik rett til forklaring.¹⁴⁸

Denne retten bør omfatte to typer forklaringer, dvs. hvordan systemet fungerer (formålet, logikken og konsekvensene) så vel som forklaring av de individuelle omstendighetene som førte til en beslutning.¹⁴⁹ I tillegg bør hva som utgjør en tilstrekkelig forklaring i ulike sammenhenger presiseres.

Muligheten til forklaring av algoritmen kan være begrenset fordi algoritmen er komplisert og vanskelig å forklare i menneskelig forståelige termer. Det kan spesielt være utfordrende å gi en forklaring av hvordan spesifikke data har blitt vektet i algoritmer som baserer seg på nevrale nett.

Det må eventuelt vurderes om offentlig sektor bør avstå fra å gjøre automatiserte beslutninger dersom det ikke er mulig å gi en fullgod forklaring. Den franske strategien anser det som utenkelig å akseptere beslutninger som ikke kan forklares på områder som er avgjørende for en persons liv, så som tilgang til kreditt, arbeid, bolig, rettsapparatet og helsetjenester.¹⁵⁰

11. KRAV OM ÅPNE ALGORITMER I OFFENTLIG SEKTOR

Når maskiner overtar oppgaver som tidligere ble gjort av mennesker, er det spesielt viktig å kunne vise at algoritmene ikke gir fordomsfulle anbefalinger. Algoritmer kan i verste fall forsterke sosiale forskjeller, føre til utilsiktet diskriminering og skjule normative valg.

¹⁴⁶ Proposisjon til Stortinget 2018, s. 68.

¹⁴⁷ Wachter, Mittelstadt og Floridi 2017.

¹⁴⁸ Jf. Villani 2018, s. 125.

¹⁴⁹ Formålet kan for eksempel være å bestemme en kredittscore, logikken kan være type data, egenskaper i dataene og kategorier i et beslutningstre og konsekvensene kan være at kredittscoren blir brukt av långivere for å gjøre en kredittvurdering, som vil kunne påvirke rentesatsen. De individuelle omstendighetene kan være den faktiske kredittscoren, hvilke faktiske data eller egenskaper som ble benyttet og hvordan de ble vektet i beslutningstreet eller modellen. Se også Wachter, Mittelstadt og Floridi 2017.

¹⁵⁰ Villani 2018, s. 1156-116.

Norske myndigheter bør derfor som hovedregel kreve at alle algoritmer som brukes av det offentlige, skal være åpne for innsyn, slik at andre samfunnsaktører kan kontrollere at de blir brukt korrekt og etisk forsvarlig. Dette vil dessuten være viktig for tilliten til offentlig sektor.¹⁵¹

12. REVIDERE ALGORITMER

Krav om åpne algoritmer egner seg ikke nødvendigvis i alle sammenhenger. Forretningshensyn, personvernet, eller rikets sikkerhet kan bli skadelidende dersom noen typer algoritmer kan kopieres og distribueres fritt. Spesielt bekymringsverdig er det at åpne algoritmer også potensielt styrker aktører med ondsinnede hensikter.

Algoritmer for maskinlæring som av tungtveiende grunner ikke er åpne for alle, bør likevel bli gjenstand for en vurdering før de kan bli tatt i bruk i stor skala i samfunnet. En mulighet er å kreve at lukkede KI-algoritmer må testes grundig og revideres eller sertifiseres av en uavhengig tredjepart før de kan anvendes i samfunnet.¹⁵² Det bør blant annet vurderes om beslutningene til algoritmen er

- rettferdige,
- korrekte,
- forklarbare,
- etterprøvbare og
- at det er synliggjort hvordan man kan klage på uønskede effekter.

Det er ikke alltid nødvendig, nyttig eller mulig å undersøke kildekoden. Et alternativ er å prøve ut algoritmen. For å vurdere om for eksempel en rekrutteringsalgoritme diskriminerer, kan den testes med et stort antall CV-er til menn og kvinner med like kvalifikasjoner. Det kan derfor være aktuelt å kreve et programmeringsgrensesnitt¹⁵³, for å kunne teste algoritmen på et stort antall fiktive brukere.¹⁵⁴

¹⁵¹ Den franske presidenten har uttalt at Frankrike vil øke presset på private aktører for at også de gjør sine algoritmer åpne for innsyn. Thompson 2018.

¹⁵² Storbritannia er i gang med å etablere *Centre for Data Ethics & Innovation*, som bl.a. skal evaluere ulike verktøy for å identifisere og håndtere fordomsfulle algoritmer på og gi anbefalinger på verktøy som privat og offentlig sektor bør bruke. House of Commons 2018.

¹⁵³ Application Programming Interface, ofte forkortet API

¹⁵⁴ Se også Villani 2018, s. 117.

Det finnes allerede eksisterende ordninger, for eksempel innen revisjon, som kan utvides til å omfatte algoritmer.¹⁵⁵

13. INNEBYGD ETIKK

Algoritmer kan kontrolleres ved å åpne dem for innsyn eller revisjon, men det mest hensiktsmessige for utviklere så vel som brukere er å bygge inn etiske hensyn fra start. Uønskede hendelser som fordomsfulle eller urettferdige avgjørelser kan føre til svekket tillit, og det er vanskelig og kostbart å rette opp i etterkant.

En slik tenkning er allerede etablert når det gjelder personvern.¹⁵⁶ Innebygd personvern vil si at skal det tas hensyn til personvernprinsipper, -rettigheter og -krav i hele utviklingsløpet, fra design og koding til testing og drift.¹⁵⁷

Kunstig intelligens er nå i ferd med å bli viktig på mange samfunnsområder, og derfor bør også etiske vurderinger bygges inn i utviklingen av algoritmene. Det vil være viktig å vurdere om algoritmen kan føre til diskriminering og om den er robust mot manipulasjon. I Frankrike er det i tråd med dette nå foreslått å utvide plikten til å gjøre personvern vurderinger til å også omfatte diskrimineringsvurderinger.¹⁵⁸

Ved å gjøre fortløpende etiske vurderinger kan utviklerne gjøre endringer underveis, samtidig som de kan demonstrere hvordan de har gjort nødvendige tiltak ved en revisjon. Slike vurderinger forutsetter at utviklere har eller tilegner seg etikk-kompetanse. Etikk bør bli en integrert del av utdanningen slik at utviklere blir i stand til å identifisere og håndtere moralske spørsmål som følge av systemet de lager.

¹⁵⁵ Det finnes allerede selskaper som spesialiserer seg på revisjon av algoritmer, som O'Neil Risk Consulting: <http://www.oneilrisk.com/>.

¹⁵⁶ Datatilsynet 2018b.

¹⁵⁷ Datatilsynet 2018c.

¹⁵⁸ Villani 2018, s. 121.

14. NASJONAL DIALOG OM KUNSTIG INTELLIGENS

Kunstig intelligens er i ferd med å komme tett på livene til nordmenn flest. Algoritmene vil gi anbefalinger om viktige livshendelser og beslutninger i forvaltningen. De har potensial til å persontilpasse, skape filterbobler og påvirke atferd. Teknologien kan dessuten forsterke forskjeller og ekskludering i samfunnet, men også brukes til å redusere disse problemene.

Den kinesiske strategien avslutter med et punkt om å «veilede opinionen» til å akseptere kunstig intelligens.¹⁵⁹ Når rask teknologiutvikling berører folks liv og verdier på denne måten, holder det ikke å jobbe for passiv aksept. Norske myndigheter bør aktivt ta initiativ til å involvere lekfolk og sivilsamfunnet i diskusjonen om kunstig intelligens og ta imot råd om hva som er ønsket utvikling. Dette kan bygge på prinsipper for ansvarlig forskning og innovasjon (RRI):

- *Dialogbasert*: Regjeringen bør invitere til en bred meningsutveksling på tvers av fagfelt og samfunnsgrupper. Her finnes det etablerte metoder som lekfolkskonferanse, åpne høringer, folketoppmøter og borgerpaneler, i tillegg til nettbaserte konsultasjoner. Aktuelle problemstillinger er for eksempel hvilke beslutninger maskiner skal ta, eller hvilke mål regjeringen skal sette for forskningssatsingen.
- *Forutseende*: Bruk av scenarier koblet med åpne fremsynsprosesser er spesielt aktuelt når utviklingen har stor hastighet og er preget av usikkerhet. Regjeringen bør ta initiativ til fremtidsanalyser som også åpner for alternative utviklingsbaner som følger av spørsmål som «hva om?» og «kan vi oppnå det samme på en annen måte?».
- *Responsiv*: Konsultasjon med offentligheten er verdiløs om ikke regjeringen responderer. Regjeringen bør jevnlig revidere den nasjonale strategien og invitere til innspill. Den bør også publisere kunnskapsstatus årlig.

¹⁵⁹ Folkerepublikken Kinas statsråd 2017.

REFERANSER

Agrawal, Ajay, Gans, Joshua og Goldfarb, Avi (2018) *Prediction Machines. The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, 2018.

Ahlqvist m. fl. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes. *The Lancet*, publisert: mars 2018.
Hentet fra: [https://doi.org/10.1016/S2213-8587\(18\)30051-2](https://doi.org/10.1016/S2213-8587(18)30051-2).

Al-Darwish, Maryam (2018) Machine Learning, Contrib.andrew.cmu.edu, 07.04.2018.
Hentet fra: <http://www.contrib.andrew.cmu.edu/~mndarwis/ML.html>.

AliveCOR (2017) AliveCor Unveils First AI-Enabled Platform for Doctors to Improve Stroke Prevention Through Early Atrial Fibrillation Detection. AliveCORs hjemmeside, publisert 16. mars.
Hentet fra: https://www.alivecor.com/press/press_release/alivecor-unveils-first-ai-enabled-platform-for-doctors/.

Angiodysplasi. (2009). Angiodysplasi. I Store Medisinske Leksikon, oppslag 13. februar 2018.
Hentet fra: <https://sml.snl.no/angiodysplasi>.

Antoni, Manfred og Schnell, Rainer (2017) *The Past, Present and Future of the German Record Linkage Center (GRLC)*, Journal of Economics and Statistics, 2017.

Artificial Lawyer (2017) *UK Gov-Backed Report Calls for AI Data Trusts; Praises Legal Sector*, 16. oktober 2017.

Hentet fra: <https://www.artificiallawyer.com/2017/10/16/uk-gov-backed-report-calls-for-ai-data-trusts-praises-legal-sector/>.

Bell, Robert M.; Koren, Yehuda; Volinsky, Chris (2018) *The BellKor solution to the Netflix Prize*, tilgang 7. mars 2018.

Hentet fra: https://www.netflixprize.com/assets/ProgressPrize2007_Kor-Bell.pdf.

Bjørkeng, Per Kristian (2015) *Datasystemer som holder elevene i flytsonen*. Aftenposten, 9. mai 2015.

Hentet fra: <https://www.aftenposten.no/norge/i/qwVL/Datasystemet-som-holder-elevene-i-flytsonen>.

Bostrom, Nick og Yudkowsky, Eliezer (2016) *The Ethics of Artificial Intelligence*, Cambridge Handbook of Artificial Intelligence, Cambridge University Press 2016.

Brownlee, Jason (2016) *8 Inspirational Applications of Deep Learning*, Machine Learning Mastery, 14. juli 2016.

Hentet fra: <https://machinelearningmastery.com/inspirational-applications-deep-learning/>.

Brundage, Miles; Avin, Shahar; Clark, Jack; Toner, Helen; Eckersley, Peter (2018) *The malicious use of artificial forecasting, prevention, and mitigation*. Februar 2018,

Hentet fra: https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf.

Brynjolfsson, Erik; McAfee, Andrew (2017a) *The business of Artificial Intelligence*. Harvard Business Review, 18. juli 2017.

Hentet fra: <https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence>.

Brynjolfsson, Erik; McAfee, Andrew (2017b) *What's Driving the Machine Learning Explosion?* Harvard Business Review, 18. juli 2017,

Hentet fra: <https://hbr.org/2017/07/whats-driving-the-machine-learning-explosion>.

CIFAR (2017) *Pan-Canadian Artificial Intelligence Strategy*. CIFAR. Tilgang 20. august 2018.

Hentet fra: <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>.

Castelvecchi, Davide (2016) *Deep learning boosts Google translate tool*, Nature, 27. september 2016.

Hentet fra: <https://www.nature.com/news/deep-learning-boosts-google-translate-tool-1.20696>.

Carey, Bjorn (2016) *Smartphone speech recognition can write text messages three times faster than human typing*, Stanford News, 24. august 2016.

Hentet fra: <http://news.stanford.edu/2016/08/24/stanford-study-speech-recognition-faster-texting/>.

Christiansen, Atle (2017) *Kunstig intelligens kan hjelpe leger*, Universitetet i Agder, 24. februar 2017.

Hentet fra: <https://www.uia.no/nyheter/kunstig-intelligens-kan-hjelpe-leger>.

Cuomo, Chris; Shaylor, Jay; McGuirt, Mary; Francescani, Chris (2009) *'GMA' Gets Answers: Some Credit Card Companies Financially Profiling Customers*, ABC News, 28. januar 2009.

Hentet fra: <http://abcnews.go.com/GMA/TheLaw/gma-answers-credit-card-companies-financially-profiling-customers/story?id=6747461>.

Dagens næringsliv (2017) *Rekrutterer ved hjelp av robot – ansetter flere kvinner*, Dagens næringsliv, 7. desember 2017.

Hentet fra: Dagens næringsliv 2017. <https://www.dn.no/ta-lent/2017/12/07/0649/Arbeidsliv/rekrutterer-ved-hjelp-av-robot-ansetter-flere-kvinner>.

Datatilsynet (2018a) *Kunstig intelligens og personvern*.

Hentet fra: <https://www.datatilsynet.no/globalassets/global/om-personvern/rapporter/rapport-om-ki-og-personvern.pdf>.

Datatilsynet. (2018b). *Innebygd personvern Datatilsynet*. Tilgang 25. august 2018.

Hentet fra: <https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/innebygd-personvern/>.

Datatilsynet (2018c) *Programvareutvikling med innebygd personvern*, Datatilsynet, tilgang 25. august 2018.

Hentet fra: <https://www.datatilsynet.no/regelverk-og-verktoy/veiledere/programvareutvikling-med-innebygd-personvern/?id=7729>.

Department for Business, Energy and Industrial Strategy og Department for Digital, Culture, Media and Sport (i Storbritannia) (2018) *AI Sector Deal*, Policy Paper, 26. April 2018.

Hentet fra: <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>.

Dockrill, Peter (2017) *Google's AI has mastered all the chess knowledge in history – in just 4 hours*. World Economic Forum, 8. desember 2017.

Hentet fra: <https://www.weforum.org/agenda/2017/12/google-s-ai-has-mastered-all-the-chess-knowledge-in-history>.

Echersley, Peter, Nasser, Yomna (2018) *Measuring the progress of AI research*, Electronic Frontier Foundation. Tilgang 7. mars 2018.

Hentet fra: <https://www.eff.org/ai/metrics#Vision>.

EDPS (2018) *Towards a digital ethics*, European Data Protection Supervisor, Ethics Advisory Group, rapport, 2018.

Hentet fra: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf.

Equinor (2018) *Joint offshore digitalisation*, 7. mars 2018.

Hentet fra: <https://www.equinor.com/en/news/07mar2018-joint-offshore-digitalisation.html>.

Esteva; Andre, Kuprel, Brett; Novoa, Roberto A.; Ko, Justin; Swetter, Susan M.; Blau, Helen M.; Thrun, Sebastian (2017) *Dermatologist-level classification of skin cancer with deep neural networks*, Nature 542 (2. februar 2017): 115-118.

Hentet fra: <https://www.nature.com/nature/journal/v542/n7639/index.html>.

Etherington, Darrell (2017) *Udacity open sources its self-driving car simulator for anyone to use*, TechCrunch.com, 8. februar 2017.

Hentet fra: <https://techcrunch.com/2017/02/08/udacity-open-sources-its-self-driving-car-simulator-for-anyone-to-use/>.

EU-kommisjonen (2018a) *Artificial Intelligence for Europe*. 25. april 2018.

Hentet fra: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625.

EU-kommisjonen (2018b): *Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines*. Pressemelding 25. april. 2018.

Hentet fra: http://europa.eu/rapid/press-release_IP-18-3362_en.htm.

EU-kommisjonen (2018c) *EU Member States sign up to cooperate on Artificial Intelligence*. 10. april 2018.

Hentet fra: <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>.

EU-kommisjonen (2018d) *Bold science to meet big challenges: independent report calls for mission-oriented EU research and innovation*. 22. februar 2018.

Hentet fra: https://ec.europa.eu/info/news/bold-science-meet-big-challenges-independent-report-calls-mission-oriented-eu-research-and-innovation-2018-feb-22_en.

Europaparlamentet (2017) *Civil Law Rules on Robotics*. 16. februar 2017.

Hentet fra: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONGML+TA+P8-TA-2017-0051+O+DOC+PDF+Vo//EN>.

Evans; Richard; Gao, Jim (2016) *DeepMind AI Reduces Google Data Centre Cooling Bill by 40%*. Deepmind.com, 20.juli 2016.

Hentet fra: <https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/>.

Fagella, Daniel (2016) *Unsupervised Machine Learning Could Help Us Solve the Unsolvable*. Huffington Post, 18. mai 2016.

Hentet fra: https://www.huffingtonpost.com/daniel-faggella/unsupervised-machine-lear_b_10010452.html.

Federal Ministry of Transport and Digital Infrastructure (i Tyskland) (2017) *Ethics commission's complete report on automated and connected driving*. 28. august 2017.

Hentet fra: <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf>.

Fenner, Robert (2018) *Alibaba's AI outguns humans in reading test*, Bloomberg, 15. januar 2018.

Hentet fra: <https://www.bloomberg.com/news/articles/2018-01-15/alibaba-s-ai-outgunned-humans-in-key-stanford-reading-test>.

Ferrucci, David (2018) *A computer called Watson*. IBM Icons of Progress, tilgang 7. mars 2018.

Hentet fra: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/watson/breakthroughs/>.

Finlands arbeids- og næringsdepartement (2017) *Finland's Age of Artificial Intelligence*, desember 2017.

Hentet fra: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf.

Fischer, Astrid; Olhoff-Jacobsen, Emil Eusebius (2017) *Computer kan forudsige om du havner akut på sygehuset*. Danmarks Radio, 27. desember 2017.

Hentet fra: <https://www.dr.dk/nyheder/indland/computer-kan-forudsige-om-du-havner-akut-paa-sygehuset>.

Folkehelseinstituttet (2018) *Kreft I Norge*, publisert 24.01.2018.

Hentet fra: <https://www.fhi.no/nettpub/hin/ikke-smittsomme/kreft-i-norge--folkehelserapporten/>.

Folkerepublikken Kinas statsråd (2017) *A Next Generation Artificial Intelligence Development Plan*. China's State Council, 20. juli 2017, basert på oversettelse til engelsk av New America.

Hentet fra: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm og <https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf>.

- Gallup (2018) *Most Americans Already Using Artificial Intelligence Products*. Gallup, 6. mars 2018.
Hentet fra: <https://news.gallup.com/poll/228497/americans-already-using-artificial-intelligence-products.aspx>.
- Garber, Megan (2016) *When Algorithms Take the Stand* The Atlantic, 30. juni 2016.
Hentet fra: <https://www.theatlantic.com/technology/archive/2016/06/when-algorithms-take-the-stand/489566/>.
- GDPR (2016) *Official Journal L119 of the European Union*. Volume 59. Legislation. Mai 2016.
Hentet fra: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL>.
- Geng, Daniel og Shih, Shannon (2017) *Machine Learning Crash Course: Part 3*, Machine Learning a Berkeley, 4. februar 2017.
Hentet fra: <https://ml.berkeley.edu/blog/2017/02/04/tutorial-3/>.
- Gibney, Elizabeth (2016) *Google AI algorithm masters ancient game of go*, Nature 529, 445-336. Publisert: 28.01.2016.
- Github (2018). *Mimir*, juni 2018.
Hentet fra: <https://github.com/Stevenah/mimir>.
- Goodfellow, Ian J; Pouget-Abadie, Jean; Mirza, Mehdi; Zu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014) *Generative Adversarial Networks*. 10. juni 2014.
Hentet fra: <https://arxiv.org/abs/1406.2661>.
- Goodfellow, Ian J; (2017) *NIPS 2016 – Generative Adversarial Networks*, YouTube-film, 1:55:53,
Hentet fra: <https://www.youtube.com/watch?v=AJVyzdorqdc>.
- Gunning, David (2018) *Explainable artificial intelligence (XAI)*, Defence Advances Research Projects Agency, tilgang 7. mars 2018.
Hentet fra: <https://www.darpa.mil/program/explainable-artificial-intelligence>.

Gupta, Dishashree (2017) *Transfer learning & the art of using pre-trained models in deep learning*. Analytics Vidhya, 1. juni 2017.

Hentet fra: <https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/>.

Hall-Geisler, Kristen (2017) *Cortica teaches unsupervised vehicles with unsupervised learning*. TechCrunch.com, 13. juli 2017.

Hentet fra: <https://techcrunch.com/2017/07/13/cortica-teaches-autonomous-vehicles-with-unsupervised-learning/>.

Hall, Wendy; Pesenti Jérôme (2017) *Growing the artificial intelligence industry in the UK*.

Hentet fra: <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>.

Hassabis, Demis; Silver, David (2017) *AlphaGO Zero: Learning from Scratch*. Deepmind, 12. oktober 2017.

Hentet fra: <https://deepmind.com/blog/alphago-zero-learning-scratch/>.

Haugnes, Gunhild M (2017) *Avslører magesykdom med algoritmer*. Titan, 2. mai 2017.

Hentet fra: <https://titan.uio.no/node/2296>.

Hawkins, Andrew J (2018) *Inside Waymo's strategy to grow the best brains for self-driving cars*. The Verge, 9. mai 2018.

Hentet fra: <https://www.theverge.com/2018/5/9/17307156/google-waymo-driverless-cars-deep-learning-neural-net-interview>.

Hof, Robert D (2018) *Deep Learning*. MIT Technology Review.

Hentet fra: <https://www.technologyreview.com/s/513696/deep-learning/>.

Holte, Hans Christian (2018) *Digitalisering – et samarbeid om sammenhenger som burde vært der*. Ukeavisen Ledelse, 22. Juni 2018.

House of Commons (i Storbritannia) (2018) *Algorithms in decisionmaking*, Fourth report of Session 2017-19, 23 mai 2018.

Hentet fra: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>.

House of Commons (i Storbritannia) (2017) *Robotics and Artificial Intelligence: Fifth Report of Sessions 2016-2017*. House of Commons Science and Technology Committee, 13. september 2017.

Hentet fra: <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/896/896.pdf>.

Høyre (2018) *Enighet om blågrønn regjeringsplattform*. 14. januar 2018,

Hentet fra: <https://hoyre.no/aktuelt/nyheter/2018/enighet-om-blaagronn-regjeringsplattform/>.

IBM (2018) *10 Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations*. IBM, tilgang 7. mars 2018.

Hentet fra: <https://public.dhe.ibm.com/common/ssi/ecm/wr/en/wrl12345usen/watson-customer-engagement-watson-marketing-wr-other-papers-and-reports-wrl12345usen-20170719.pdf>.

Jones, Nicola (2017) *Machine learning tapped to improve climate forecasts*, Nature 548 (24. august 2017).

Hentet fra: https://www.nature.com/polopoly_fs/1.22503!/menu/main/topColumns/topLeftColumn/pdf/548379a.pdf.

Kaggle (2017). *Data Science FAQ*. Kaggle. Tilgang 20. april 2018.

Hentet fra: <https://www.kaggle.com/rounakbanik/data-science-faq>.

Karpathy, Andrej (2014) *What I learned from competing against a ConvNet on ImageNet*. Sist endret 2. september 2014.

Hentet fra: <http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/>.

Knight, Will (2017a) *5 big predictions for artificial intelligence in 2017*. MIT Technology Review, 4. januar 2017.

Hentet fra: <https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/>.

Knight, Will (2017b) *The dark secret at the heart of AI*, MIT Technology Review. 11. april 2017.

Hentet fra: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.

Knight, Will (2018) *By experimenting, computers are figuring out how to do things that no programmer could teach them*. MIT Technology Review. Tilgang 7. mars 2018.

Hentet fra: <https://www.technologyreview.com/s/603501/10-breakthrough-technologies-2017-reinforcement-learning/>.

Kreftforeningen (2018) *Slik kan du sjekke føflekkene dine*, oppslag 17. august 2018.

Hentet fra: <https://kreftforeningen.no/om-kreft/a-oppdage-kreft-tidlig/sjekk-foflekkene-dine/>.

Kunnskapsdepartementet (2014) *Langtidsplan for forskning og høyere utdanning 2015–2024*. 2014.

Hentet fra: <https://www.regjeringen.no/contentassets/e10e5d5e2198426788ae4f1ecbbbbc20/no/pdfs/stm20142015000700odddpdfs.pdf>.

Kunnskapsdepartementet (2017) *Nasjonal strategi for tilgjengeligjføring og deling av forskningsdata*. 2017.

Hentet fra: <https://www.regjeringen.no/contentassets/3aoceaa1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf>.

Lardinois, Frederic (2018) *Google gives its AI the reins over its data center cooling systems*, TechCrunch, 17. august 2018.

Hentet fra: <https://techcrunch.com/2018/08/17/google-gives-its-ai-the-reins-over-its-data-center-cooling-systems/?guccounter=1>.

Liptak, Adam (2017) *Sent to prison by a software program's secret algorithms*, The New York Times, 1. mai 2017.

Hentet fra: <https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html>.

Lui, Yunjie; Racah, Evan; Correa, Joaquin Prabhat; Khosrowshahi, Amir; Lavers, David; Kunkel, Kenneth; Wehner, Michael; Collins, William (2016) *Application of deep convolutional neural networks for detecting extreme weather in climate datasets*, 4. mai 2016.

Hentet fra: <https://arxiv.org/pdf/1605.01156.pdf>.

Mandarić, Stefan; Axelsen, Vebjørn (2017) *Forebygging av trafikkulykker ved bruk av avansert dataanalyse*, BearingPoint (2017).

Hentet fra: https://www.vegvesen.no/_attachment/2073052/binary/1219064?fast_title=Forebygging+av+trafikulykker+ved+bruk+av+avansert+dataanalyse.pdf.

Mannes, John (2016) *OpenAI's Universe is the fun parent every artificial intelligence deserves*. TechCrunch.com, 5. desember 2016.

Hentet fra: <https://techcrunch.com/2016/12/05/openai-universe-is-the-fun-parent-every-artificial-intelligence-deserves/>.

Margolin, Madison (2016) *These glasses fool facial recognition into thinking you're someone else*, Motherboard, 2. november 2016.

Hentet fra: https://motherboard.vice.com/en_us/article/pgkxgv/glasses-fool-facial-recognition.

May, Theresa (2018) *PM speech on science and modern Industrial Strategy*, 21. mai 2018

Hentet fra: <https://www.gov.uk/government/speeches/pm-speech-on-science-and-modern-industrial-strategy-21-may-2018>.

Mazzucato, Mariana (2018) *Mission-oriented Research and Innovation in the European Union*. Februar 2018.

Hentet fra: <https://publications.europa.eu/en/publication-detail/-/publication/5b2811d1-16be-11e8-9253-01aa75ed71a1/language-en>.

Melding til Stortinget (2017) *Perspektivmeldingen 2017*. Meld. St. 29 (2016-2017).

Hentet fra: <https://www.regjeringen.no/no/dokumenter/meld.-st.-29-20162017/id2546674/>.

Miotto, Riccardo; Li, Li; Kidd, Brian A.; Dudley, Joel T (2016) *Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records*, Nature Scientific Reports, mai 2016.

Hentet fra: <https://www.nature.com/articles/srep26094>.

Moe, Sigrid og Breivik, Steinar Rostad (2018) *Snart kan kunstig intelligens styre vannkraftverkene*, E24, 5. mars 2018.

Hentet fra: <https://e24.no/energi/vannkraft/snart-kan-kunstig-intelligens-styre-vannkraftverkene/23933863>.

Multi-task learning (2018) Multi-task learning. I *Wikipedia*, 7. mars 2018.

Hentet fra: https://en.wikipedia.org/wiki/Multi-task_learning.

Murgia, Madhumita (2017) *NHS to trial artificial intelligence app in place of 111 helpline*, Financial Times, 4. januar 2017.

Hentet fra: <https://www.ft.com/content/ae0ee3b8-d1d8-11e6-b06b-680c49b4b4c0>.

Mystore (2017) *Kunstig intelligens og dens rolle i netthandel*, Mystore.no. 12. mai 2017.

Hentet fra: <https://www.mystore.no/kunstig-intelligens-netthandel/>.

Nedelkoska, L. og Quintini, G (2018) *Automation, skills use and training*, OECD Social, Employment and Migration Working Papers, No. 202, 2018.

Hentet fra: http://www.oecd-ilibrary.org/employment/automation-skills-use-and-training_2e2f4eea-en.

NFR (2017) *Innspill til revidert langtidsplan 2019 – 2022*. Innspill til Regjeringens langtidsplan for forskning, september 2017.

Hentet fra: <https://www.regjeringen.no/content-tassets/8e11ef7f053e43a0a7ac06f2486e16c7/forskningsradet---innspill-til-revisjon-av-langtidsplanen-for-forskning-og-hoyere-utdanning-2015-2024-signert.pdf>.

Ng, Andrew (2015). *What Data Scientists Should Know About Deep Learning*. Presentasjon på Extract Data Conference, 24. November, 2015.

Hentet fra: <https://www.slideshare.net/ExtractConf>.

Ng, Andrew (2017) *Deep learning's next frontier*, Harvard Business Review, 25. juli 2017,

Hentet fra: <https://hbr.org/2017/07/deep-learning-next-frontier>.

Niesche, Christophe (2015) *Medtech startup turns home into medical clinic*, Australia Unlimited, 18. august 2015.

Hentet fra: <https://www.australiaunlimited.com/technology/medtech-startup-turns-home-into-medical-clinic>.

OECD (2017) *OECD Digital Economy Outlook*. OECD Publishing, oktober 2017.

Hentet fra: <http://www.oecd.org/internet/oecd-digital-economy-outlook-2017-9789264276284-en.htm>.

O'Neil, Cathy (2016) *Weapons of Mass Destruction*. New York: Crown Publishers, 2016.

O'Neil, Cathy (2018) *Audit the algorithms that are ruling our lives. Governments should follow France and move towards algorithmic accountability*, Financial Times, 30. juli 2016.

Hentet fra: <https://www.ft.com/content/879d96d6-93db-11e8-95f8-8640db9060a7>.

Ong, Thuy (2017) *Facebook's translations are now powered completely by AI*. The Verge, 4. august 2017.

Hentet fra: <https://www.theverge.com/2017/8/4/16093872/facebook-ai-translations-artificial-intelligence>.

OpenAI.com (2017) *Attacking machine learning with adversarial examples*, OpenAI.com, 24. februar 2017.

Hentet fra: <https://blog.openai.com/adversarial-example-research/>.

Oren Etzioni (2017) *How to regulate artificial intelligence*, The New York Times, 1. september 2017.

Hentet fra: <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html>.

Pajarinen, M., Rouvinen, P. og Ekeland, A (2014) *Computerization and the Future of Jobs in Norway*. ETLA og SSB 2014.

Hentet fra: <http://nettsteder.regjeringen.no/fremtidensskole/files/2014/05/Computerization-and-the-Future-of-Jobs-in-Norway.pdf>.

Paysa.com (2017) *New Paysa Study Reveals U.S. Companies Across All Industries Investing \$1.35 Billion Dollars in AI Talent*, Paysa.com, 29. november 2017.

Hentet fra: <https://www.paysa.com/press-releases/2017-11-29/11/new-paysa-study-reveals-us>.

Pogorelov, Konstantin mfl. (2018) *Automatic Detection of Angiectasia: Evaluation of Deep Learning and Handcrafted Approaches*.

IEEE Conference on Biomedical and Health Informatics 2018.

Hentet fra: <http://home.ifi.uio.no/paalh/publications/files/bhi-2018.pdf>.

Proposisjon til Stortinget (2018) *Lov om behandling av personopplysninger (personopplysningsloven)* Prop. 56 LS (2017–2018).

Hentet fra: [https://www.regjeringen.no/content-](https://www.regjeringen.no/content-tassets/1a36e88f124d4a1ea92a9c790be2d69a/no/pdfs/prp20172018005600odddpdfs.pdf)

[tassets/1a36e88f124d4a1ea92a9c790be2d69a/no/pdfs/prp20172018005600odddpdfs.pdf](https://www.regjeringen.no/content-tassets/1a36e88f124d4a1ea92a9c790be2d69a/no/pdfs/prp20172018005600odddpdfs.pdf).

Purdy, Mark og Daugherty, Paul (2016). Why artificial intelligence is the future of growth. Accenture, 2016.

Hentet fra: https://www.accenture.com/t20170927To80049Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDF?la=en.

Reilly, Michael (2017) *Climate-change research is getting a big dose of A*. Technology Review, 23. august 2017.

Hentet fra: <https://www.technologyreview.com/the-download/608726/climate-change-research-is-getting-a-big-dose-of-ai/>.

Reuters (2018) *France to spend \$1.8 billion on AI to compete with U.S., China*, Reuters 29. Mars 2018.

Hentet fra: <https://www.reuters.com/article/us-france-tech/france-to-spend-1-8-billion-on-ai-to-compete-with-u-s-china-idUSKBN1H51XP>.

Ribeiro, Marco Tulio, Singh, Sameer og Guestrin, Carlos (2016) *Introduction to local interpretable model-agnostic explanations (LIME)*.

O'Reilly, 12. august 2016.

Hentet fra: <https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime>.

Royal Free (2018) *Our work with DeepMind*, tilgang 18.8.2018.

Hentet fra: <https://www.royalfree.nhs.uk/patients-visitors/how-we-use-patient-information/our-work-with-deepmind/>.

Ruan, Sherry; Wobbrock, Jacob O.; Liou, Kenny; Ng, Andrew; Landay, James (2017) *Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones*. Journal Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive. Volume 1 Issue 4, desember 2017.

Hentet fra: <https://doi.org/10.1145/3161187>.

Scimago (2018) Oppslag på «Computer Science» og «Artificial Intelligence», i *Scimago Journal and Country Rank*, tilgang 14. mars 2018.

Hentet fra: <http://www.scimagojr.com/countryrank.php?area=1700&category=1702>.

Sinovation Ventures (2018) *China embraces AI: A Close Look and A Long View*, 2018.

Hentet fra: https://www.eurasiagroup.net/files/upload/China_Embraces_AI.pdf.

SkillsFuture 2017. *Launch Of SkillsFuture For Digital Workplace - Building Confidence in All Singaporeans for the Digital Economy*, 5.oktober 2017.

Hentet fra: http://www.ssg-wsg.gov.sg/new-and-announcements/05_Oct2017.html.

SkLNytt (2017) *Virkelig noe å være redd for?* SkLNytt nr. 5, 40. årgang 2017.

Hentet fra: https://www.skl.no/sfiles/31/41/1/file/sklnytt_5_17_ferdig_lavoppl.pdf.

Smith, Mich (2016) *In Wisconsin, a backlash against using data to foretell defendants' futures*, The New York Times, 22. juni 2016.

Hentet fra: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>.

Snow, Jackie (2018) *Most Americans are already using AI*, MIT Technology Review, 7. mars 2018. Oxford Insights,

Hentet fra: <https://www.technologyreview.com/the-download/610438/most-americans-are-already-using-ai/>.

Stirling, Miller og Martinho-Truswell (2018) *Government AI Readiness Index*.

Hentet fra: <https://www.oxfordinsights.com/government-ai-readiness-index/>.

Sverdlik, Yevgeniy (2018) *Google is Switching to a Self-Driving Data Center Management System*, Data Center Knowledge, 2. august 2018.

Hentet fra: <https://www.datacenterknowledge.com/google-alphabet/google-switching-self-driving-data-center-management-system>.

Tassev, Lubomir (2018) *GPU shortage hinders scientific research – cryptocurrency miners blamed*, Bitcoin.com, 15. februar 2018.

Hentet fra: <https://news.bitcoin.com/gpu-shortage-hinders-alien-search-cryptocurrency-miners-blamed/>.

Telenor (2018) *Styrker kraftsenter for kunstig intelligens i Norge* Telenor pressemelding, 15. august 2018.

Hentet fra: <https://www.telenor.com/no/pressemelding/styrker-kraftsenter-for-kunstig-intelligens-i-norge/>.

Teknologirådet (2017) *Denne gangen er det personlig*, Oslo, mai 2017.

Hentet fra: https://teknologiradet.no/wp-content/uploads/sites/19/2013/08/Rapport_Denne-gangen-er-det-personlig.-Det-digitale-skiftet-i-offentlig-sektor.pdf.

The Economist (2017) *Google leads race to dominate artificial intelligence*, The Economist, 7. desember 2017.

Hentet fra: <https://www.economist.com/news/business/21732125-tech-giants-are-investing-billions-transformative-technology-google-leads-race>.

The Economist (2018a) *For artificial intelligence to thrive, it must explain itself*, The Economist, 15. februar 2018.

Hentet fra: <https://www.economist.com/news/science-and-technology/21737018-if-it-cannot-who-will-trust-it-artificial-intelligence-thrive-it-must>.

The Economist (2018b) *Humans may not always grasp why AIs act. Don't panic*, The Economist, 15. februar 2018.

Hentet fra: <https://www.economist.com/leaders/2018/02/15/humans-may-not-always-grasp-why-ais-act.-dont-panic>.

Thompson, Nicolas (2018). *Emmanuel Macron talks to Wired about France's AI Strategy*. Wired, 31. mars 2018.

Hentet fra: <https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/>.

Thornhill, John (2017) *Would you donate your data for the collective good?* Financial Times, 30. oktober 2017.

Hentet fra: <https://www.ft.com/content/00390a76-bd4a-11e7-9836-b25f8adaa111>.

Turner, Karen (2016) *Google Translate is getting really, really accurate*, The Washington Post, 3. oktober 2016.

Hentet fra: <https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-getting-really-really-accurate/>.

Tørresen, Jim (2014) *Hva er kunstig intelligens*, Universitetsforlaget, 14. jan. 2014.

Hentet fra: https://issuu.com/universitetsforlaget/docs/hva_er_kunstig_intelligens.

UCLH (2018): *Revolutionising healthcare with AI and data science: UCLH and The Alan Turing Institute announces breakthrough partnership today*, 21. Mai 2018.

Hentet fra: <https://www.uclh.nhs.uk/News/Pages/RevolutionisinghealthcarewithAI.aspx>.

Universitetet i Toronto (2017) *Toronto's Vector Institute officially launched*. 30.mars 2017.

Hentet fra: <https://www.utoronto.ca/news/toronto-s-vector-institute-officially-launched>.

Villani, Cedric (2018) *For a meaningful artificial intelligence. Towards a French and European strategy*, Mars 2018.

Hentet fra: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

Vinnova (2018) *Artificiell intelligens i svenskt näringsliv och samhälle. Analys av utveckling och potential*, Vinnova rapport, 2018.

Hentet fra: https://www.vinnova.se/contentassets/55b18cf1169a4a4f8340a5960b32fa82/vr_18_08.pdf.

Wachter, Sandra; Mittelstadt, Brent; Floridi, Luciano (2017) *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, International Data Privacy Law, Elsevier 2017.

Hentet fra: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469.

Wahed, Muntasir (2018) *What is the relationship between artificial intelligence, machine learning, deep learning, and artificial neural networks?* Quora, januar 2018.

Hentet fra: <https://www.quora.com/What-is-the-relationship-between-artificial-intelligence-machine-learning-deep-learning-and-artificial-neural-networks>.

Wakefield, Jane (2016) *Microsoft chatbot is taught to swear on Twitter*, BBC News, 24. Mars 2016.

Hentet fra: <http://www.bbc.com/news/technology-35890188>.

World Economic Forum (2017) *The global risk report*, 11. januar 2017.

Hentet fra: http://www3.weforum.org/docs/GRR17_Report_web.pdf.

Zames, Matt (2016) *Redefining the Service Industry*, J. P. Morgan Chase & Co, 2016.

Hentet fra: <https://www.jpmorganchase.com/corporate/annual-report/2016/ar-ceo-letter-matt-zames.htm>.

Zhao, Yufan; Zeng, Donglin, Socinski, Mark; Kosorok, Michael R. (2011) *Reinforcement Learning Strategies for Clinical Trials in Non-small Cell Lung Cancer*, Journal of the International Biometric Society, 2011.

Hentet fra: <https://www.ncbi.nlm.nih.gov/pubmed/21385164>.