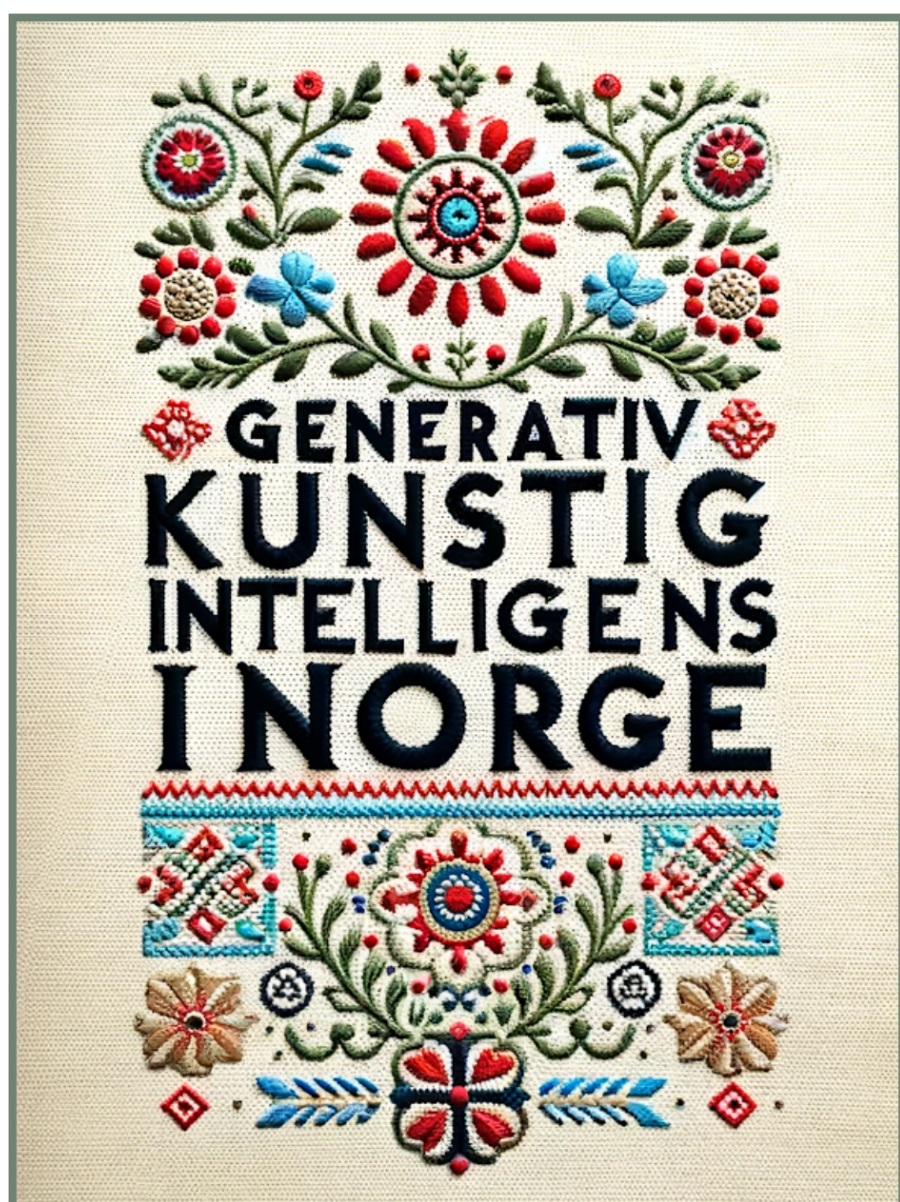




Teknologirådet

ISBN 978-82-8400-027-5 (trykket utgave)
ISBN 978-82-8400-028-2 (elektronisk utgave)

Utgitt: Oslo, mai 2024

Forside: Laget av Teknologirådet ved hjelp av bildegenereringsverktøyet Dall-E
Elektronisk publisert på: www.teknologiradet.no



FORORD

Gjennombruddet for store språkmodeller har gjort en ny generasjon kunstig intelligens mulig. Generativ kunstig intelligens kan skape tekst, video og bilder, og løse avanserte oppgaver. Teknologien utvikles svært raskt, og vil få betydning for alle samfunnsområder. Den medfører også ny risiko, og kan være upålitelig, føre til økt maktkonsentrasjon og misbrukes til manipulasjon og påvirkning.

I denne rapporten beskriver vi teknologien og hvilken betydning den kan få for samfunnet. Vi foreslår også tiltak for Norge.

Ekspertgruppen for prosjektet har bestått av:

- Einar Duenger Bøhn, professor i filosofi ved Universitetet i Agder
- Jon Atle Gulla, direktør for NorwAI og professor ved Institutt for datateknologi og elektronikk, NTNU
- Daniel Kohn, kommersiell direktør i Simplifai
- Mona Naomi Lintvedt, stipendiat ved Senter for rettsinformatikk, Institutt for privatrett, Universitetet i Oslo
- Cathrine Pihl Lyngstad, leder for dataseksjonen i NAV
- Michael Riegler, sjefsforsker ved SimulaMet

Jonas Engestøl Wettre og Ellen Strålberg har vært Teknologirådets prosjektledere for rapporten.

Teknologirådet ønsker at rapporten skal bidra til debatt om hvilke grep Norge kan ta for både å utnytte mulighetene og redusere risikoen ved generativ kunstig intelligens. God lesning!

Tore Tennøe
Direktør, Teknologirådet

INNHOOLD

SAMMENDRAG	1
-------------------	----------

EKSPERTGRUPPENS ANBEFALINGER	4
-------------------------------------	----------

GJENNOMBRUDD FOR KUNSTIG INTELLIGENS	6
---	----------

TRE EPOKER AV KUNSTIG INTELLIGENS	7
--	----------

STORE SPRÅKMODELLER	8
Derfor skjedde gjennombruddet	9
Slik fungerer språkmodellene	13
Den nye verdikjeden	16

GENERATIV KUNSTIG INTELLIGENS	19
Generativ og nyskapende	19
Generell og kraftig	20
Enkel å bruke	21

UTVIKLINGSTREKK	22
Kraftigere og smartere	22
Både åpen og lukket	25
Handler på egenhånd	27

VISJONEN OM SUPERINTELLIGENS	28
-------------------------------------	-----------

BETYDNING FOR SAMFUNNET	31
--------------------------------	-----------

REVOLUSJON FOR KUNNSKAPSARBEID	32
Effektiviserer oppgaver	32
Stimulerer forskning og innovasjon	34
Tetter kompetansegap	35
Gjør visse typer arbeidskraft mindre verdt	35
Skaper et behov for omstilling	36

BÅDE INKLUDERENDE OG DISKRIMINERENDE	38
Gjør digitale tjenester tilgjengelige for flere	38
Styrker universell utforming	39
Fabulerer og diskriminerer	39
Har et forklaringsproblem	40

FLYTTER OG KONSENTRERER MAKT	41
Samler makt hos teknologiselskapene	41
Forsterker sårbarhet	42

Utfordrer språklig og kulturell autonomi	42
OMVELTNING FOR OFFENTLIGHETEN	44
Økte mengder usant, men troverdig innhold	44
Avanserer svindel, misbruk og desinformasjon	45
Tilslører skillet mellom mennesker og maskiner	47
Intensiverer individuell persontilpassing	48
KATASTROFAL RISIKO	49
Villet bruk og misbruk	50
Utsiktede hendelser og tap av kontroll	51
En trigger for økt globalt samarbeid	51

HANDLINGSROM FOR NORGE

53

KI-FORORDNINGEN SOM UTGANGSPUNKT FOR NORGE	54
Artificial Intelligence Act	54
Dette vil KI-forordningen kreve av Norge	56
Handlingsrom for ny politikk	58
STIMULERE SAMFUNNSNYTTIG INNOVASJON OG BRUK	59
1. Fordele tilsyn, med en sterk koordinator	60
2. Etabler en nasjonal inkubator for kunstig intelligens	62
3. Skaff forvaltningen en problemløser	64
BYGGE INFRASTRUKTUR FOR KUNSTIG INTELLIGENS I NORGE	66
4. Sikre tilgang til norske kvalitetsdata	66
5. Invester i regnekraft som nasjonal infrastruktur	69
6. Tilby norske språkmodeller som nasjonal fellestjeneste	71
7. Still kvalitetskrav til norske språkmodeller	74
STYRKE DEN DIGITALE MOTSTANDSKRAFTEN	76
8. Fastsett regler for merking av KI-generert innhold	76
9. Lag spilleregler for generativ kunstig intelligens i valgkamp	78
10. Oppskaler innsatsen for faktasjekking	79
11. Opprett et psykologisk forsvar	81
12. Styrk arbeidet med KI-sikkerhet	83

SAMMENDRAG

Generativ kunstig intelligens kan skape tekst, bilder og musikk, løse vitenskapelige problemer og bidra til innovasjon. Samtidig kan teknologien hallusinere, og skape mer desinformasjon. Denne rapporten forklarer gjennombruddet for generativ kunstig intelligens og betydningen for samfunnet, og gir forslag til tiltak.

Kunstig intelligens handler om at maskiner blir i stand til å løse fysiske og kognitive oppgaver som før var forbeholdt mennesker. Teknologien blir stadig kraftigere og mer anvendelig. Samtidig blir den vanskeligere å forstå og kontrollere.

Generativ kunstig intelligens vil ha betydning for oss som enkeltpersoner, i arbeidslivet og for hele samfunnet. I denne rapporten har vi valgt ut tre overordnede områder hvor vi mener det er behov for snarlige politiske grep.

Det er samtidig behov for ny politikk også på andre viktige områder, som innen helsesektoren, skole og utdanning, kunst og kultur og i næringslivet. Dette har ikke vært tema for denne rapporten, men kan være aktuelt for nye prosjekter.

Gjennombrudd for kunstig intelligens

Med lanseringen av ChatGPT i 2022 fikk generativ kunstig intelligens sitt gjennombrudd. Dette er kunstig intelligens som kan lage innhold som bilder, tekst og video. Gjennombruddet skyldtes fremveksten av store språkmodeller, skapt av fremskritt innen nevrale nett, økt regnekraft og bedre tilgang til data.

Det er store språkmodeller som gjør generativ kunstig intelligens mulig. Dette er sannsynlighetsmodeller som bruker statistiske metoder for å tolke og generere språk. Å utvikle en språkmodell innebærer å trene opp modellen med treningsdata fra kilder som bøker og nettsider. Jo bredere datagrunnlag, desto bedre blir modellen. Språkmodeller trenes opp til å få en generell språkforståelse. Deretter kan den spesialiseres og tilpasses til ulike formål og domener.

Basert på språkmodeller blir det mulig for mennesker og maskiner å kommunisere på menneskenes språk. Utstyrt med generell språkforståelse kan generativ kunstig intelligens brukes til ulike formål på tvers av samfunnsområder. Som med elektrisitet sier vi derfor at teknologien er en generell teknologi.

Det er viktig å merke seg at språkmodeller ikke er kunnskapsmodeller. De har et statistisk forhold til tekst og til virkeligheten, og har for eksempel ikke noe begrep om hva som er sant og usant, eller etisk og uetisk. En viktig utfordring er derfor å få teknologien til å oppføre seg i tråd med menneskelige verdier.

Utviklingen går raskt, og teknologien blir stadig kraftigere og smartere. Mens de tidlige språkmodellene mestret én modalitet, som eksempelvis tekst eller bilde, er modellene nå gjerne multimodale, slik at de kan tolke og lage innhold på tvers av datatyper. En modell kan dessuten være åpen eller lukket, det vil si hvorvidt datagrunnlaget og kildekoden publiseres åpent eller holdes hemmelig.

Hvor avanserte kunstig intelligente maskiner kan bli, er høyst usikkert. Selv om dagens modeller mestrer en rekke oppgaver og scorer høyt på akademiske tester, er vi fortsatt langt unna såkalt kunstig generell intelligens på nivå med eller som overgår mennesker. Risikoene dette kan medføre må likevel tas på alvor.

Betydning for samfunnet

Generativ kunstig intelligens bidrar allerede til vitenskapelige oppdagelser og innovative løsninger på komplekse problemer på flere samfunnsområder. I denne rapporten peker vi på flere viktige effekter teknologien har på samfunnet.

Informasjonsteknologi har til nå truffet manuelt, regelstyrt og teknisk arbeid. Generativ kunstig intelligens vil treffe kunnskapsarbeidere og kreative yrker. Teknologien kan *erstatte* yrker og automatisere oppgaver, men den kan også *komplementere* arbeidernes kompetanse. Flere undersøkelser peker på betydelig effektivisering av oppgaver. Eksempelvis mener Menon Economics at teknologien potensielt kan effektivisere arbeidsoppgaver tilsvarende 155 000 årsverk i norsk offentlig sektor. Andre undersøkelser viser at teknologien kan utjevne forskjeller mellom erfarne og nyansatte medarbeidere.

Generativ kunstig intelligens kan virke både inkluderende og diskriminerende. Teknologien kan gjøre digitale tjenester tilgjengelige og styrke universell utforming. Et eksempel er appen Be My Eyes for blinde og svaksynte, hvor brukerne kan ta bilde av omgivelsene og få dem beskrevet. Teknologien kan også opptre diskriminerende og reproducere sosiale og kulturelle fordommer. Den kan også finne på informasjon, gjerne omtalt som at den fabulerer eller hallusinerer.

Teknologien er ressurskrevende å utvikle og lite gjennomiktig. Dette har foreløpig ført til at kunnskapen og makten til å definere hvordan teknologien skal designes og fungere har blitt konsentrert hos de største teknologiselskapene. Dette inkluderer hvilke språk teknologien skal fungere best på, og hvilket verdi- og referansegrunnlag språkmodellene skal ta utgangspunkt i for å lage innhold.

Det har blitt mye enklere å lage kunstige, men troverdige, lydklipp, bilder og videoer. Dette åpner for mer og avanserte former for svindel, misbruk og desinformasjon. I tillegg blir persontilpassing av tjenester mer avansert enn før, og stadig flere får KI-venner og KI-kjærester.

Generativ kunstig intelligens kan også innebære flere katastrofale risikoer. Teknologiens skadepotensial inkluderer avanserte cyberangrep, utvikling av biologiske og kjemiske våpen, og at modellene begynner å handle på egenhånd.

Handlingsrom for Norge

EUs KI-forordning gir Norge et godt grunnlag for å regulere kunstig intelligens, og skaper et nytt og forsterket handlingsrom for å utvikle nasjonal politikk. Norge må handle raskt og strategisk for å utnytte dette handlingsrommet og sikre at teknologien brukes på en måte som er trygg, ansvarlig, rettferdig og til det beste for samfunnet. Følgende tre hovedområder er særlig viktige:

- **Stimulere samfunnsnyttig innovasjon og bruk.** KI-forordningen må kompletteres med tiltak for å stimulere til samfunnsnyttig innovasjon og bruk av generativ kunstig intelligens.
- **Bygge infrastruktur for kunstig intelligens i Norge.** Tilgangen på data, regnekraft og kvalitetssikrede språkmodeller må økes dersom KI-industrien skal blomstre og offentlige tjenester og norskutviklede forbrukertjenester bli av så høy kvalitet som mulig.
- **Styrke den digitale motstandskraften.** Befolkningens motstandskraft mot desinformasjon, misbruk og uønskede hendelser forårsaket av generativ kunstig intelligens må styrkes.

EKSPERTGRUPPENS ANBEFALINGER

- 1. Fordele tilsyn, med en sterk koordinator.** Norske myndigheter må beslutte hvordan EUs KI-forordning skal håndheves nasjonalt. Norge bør la sektortilsynene håndheve på sine områder, og peke ut og satse på én sentral koordinator som bygges opp med kompetanse og tekniske ressurser. Koordinatoren bør tilby ressurser til sektortilsynene etter behov.
- 2. Etabler en nasjonal inkubator for kunstig intelligens.** Norske myndigheter må beslutte hvordan Norge skal oppfylle KI-forordningens krav om regulatoriske sandkasser for kunstig intelligens. Det bør etableres én ny nasjonal sandkasse i regi av Digitaliseringsdirektoratet, som får et utvidet mandat som en nasjonal inkubator for KI for offentlig forvaltning.
- 3. Skaff forvaltningen en problemløser.** Det er behov for felles og løpende håndtering av usikkerhet og problemstillinger knyttet til generativ kunstig intelligens blant offentlige virksomheter. Det bør etableres et nettverk der slike problemstillinger kan diskuteres og avklares. Nettverket bør etableres som en arbeidsgruppe i regi av topplederforumet Skate.
- 4. Sikre tilgang til norske kvalitetsdata.** Det kreves tilgang på betydelige mengder norske kvalitetsdata for at språkmodeller skal fungere best mulig på norsk. Nasjonalbiblioteket bør få mandat til å tilgjengeliggjøre slike data og kompensere rettighetshaverne. Også data forvaltet av Arkivverket og Lovdata må kunne utnyttes. Norske kvalitetsdata må forvaltes strategisk og tilgang må ikke gis ukritisk til alle.
- 5. Invester i regnekraft som nasjonal infrastruktur.** Behovet for regnekraft vil øke betraktelig både for forskningsinstitusjoner, offentlige og private virksomheter de neste årene. Derfor må det bygges flere og kraftigere superdatamaskiner i Norge. Regnekraft bør satses på som nasjonal infrastruktur og ansvaret for styring og finansiering bør plasseres hos Samferdselsdepartementet eller Nærings- og fiskeridepartementet.
- 6. Tilby norske språkmodeller som nasjonal fellestjeneste.** Forvaltningen, universitetssektoren og KI-industrien vil trenge stabil og kvalitets-sikret tilgang på norskutviklede språkmodeller for å utvikle og levere konkurransedyktige digitale tjenester i årene fremover. Språkmodeller bør derfor utvikles og driftes som en nasjonal fellesløsning og infrastruktur tjeneste på linje med Altinn, ID-porten og Folkeregisteret.

7. **Still kvalitetskrav til norske språkmodeller.** Det vil finnes mange språkmodeller med ulike egenskaper. Forvaltningen, universitetssektoren og KI-industrien vil derfor trenge kunnskap for å ta informerte valg om hvilke språkmodeller de bør benytte seg av. En utpekt forskningsinstitusjon bør få i oppdrag å utvikle kvalitets- og evalueringskriterier for språkmodeller i Norge.
8. **Fastsett regler for merking av KI-generert innhold.** Falskt og kunstig generert innhold som fremstår ekte og autentisk spres nå raskt på nett. Norske myndigheter bør utarbeide retningslinjer for åpenhet og merking av KI-generert innhold. Aktører som Medietilsynet og Forbrukertilsynet kan få mandat til å følge opp dette.
9. **Lag spilleregler for generativ kunstig intelligens i valgkamp.** Generativ kunstig intelligens blir i økende grad brukt som valgkampverktøy og til å påvirke demokratiske valg. Før stortingsvalget i 2025 bør det utvikles tverrpolitiske spilleregler for bruk av generativ kunstig intelligens i valgkamp for å bevare tilliten til de politiske partiene og valginstitusjonen.
10. **Oppskaler innsatsen for faktasjekking.** Den raske spredningen av KI-generert falskt og troverdig innhold på nett har økt behovet for faktasjekking av det offentlige ordskiftet. Innsatsen for faktasjekking i Norge bør oppskaleres, enten ved å styrke Faktisk.no eller ved å opprette et nytt uavhengig nasjonalt senter for kildekritikk. I tillegg bør det satses på utvikling av tekniske verktøy og metoder for innholdsverifisering på norsk.
11. **Opprett et psykologisk forsvar.** Fremveksten av generativ kunstig intelligens har økt risikoen for spredning av desinformasjon og informasjonspåvirkning. For å møte denne trusselen bør det etableres en nasjonal funksjon for å identifisere, rapportere, opplyse og gi råd om desinformasjonskampanjer og informasjonspåvirkning fra utenlandske aktører.
12. **Styrk arbeidet med KI-sikkerhet.** Generativ kunstig intelligens skaper nye og uoverskuelige sikkerhetsrisikoer og har et stort skadepotensial. Derfor bør den nasjonale kapasiteten til å analysere KI-trusler og utvikle risikoscenarioer formaliseres og styrkes. I tillegg bør myndighetene utvikle norske posisjoner for globale styringsmekanismer for kunstig intelligens, og sikre Norge plasser i de nye styringsorganene på EU-nivå.

GJENNOMBRUDD FOR KUNSTIG INTELLIGENS

Generativ kunstig intelligens lager tekst, bilder og video og kan utføre oppgaver på vegne av mennesker. Teknologien er enkel å bruke og blir tilgjengelig som del av de fleste digitale tjenester.

30. november 2022 lanserte det forholdsvis ukjente selskapet OpenAI tjenesten ChatGPT. To måneder senere hadde den avanserte chatboten nådd 100 millioner brukere – raskere enn noen annen digital tjeneste til da.¹

Slik ble ChatGPT gjennombruddet for *generativ* kunstig intelligens. Dette er kunstig intelligens som kan tolke og skape unikt innhold basert på enkle instruksjoner – både tekst, lyd, bilder og video – og som behersker menneskers språk.

Generativ kunstig intelligens åpner for en fremtid med kraftigere, smartere og mer tilgjengelige tjenester, ulikt noe man tidligere har sett. Teknologien utvikles og sprer seg nå svært raskt, og vil få betydning for alle samfunnsområder.

¹ Reuters 2023, [ChatGPT sets record for fastest-growing user base - analyst note](#)

ChatGPT

Det mest kjente eksemplet på generativ kunstig intelligens i dag er ChatGPT. Chatboten kan generere tekst i naturlig språk på nivå med mennesker, og blir løpende oppgradert med nye funksjoner som gjenkjenning og tolking av bilder og tale.² Den er bygget på språkmodellen GPT-4.³

TRE EPOKER AV KUNSTIG INTELLIGENS

Generativ kunstig intelligens er foreløpig siste kapittel i historien om kunstig intelligens:

1950–2012: Pionerer og ekspertsystemer. På 1950-tallet begynte matematikere som Alan Turing og John McCarthy å utforske maskiners evne til å tenke. De første KI-programmene fra den tiden løste matematiske problemer og spilte sjakk. På 1970 og -80-tallet ble det populært med programmerte, regelstyrte ekspertsystemer som etterliknet hvordan mennesker tok beslutninger. På 2000-tallet ble fagområdet i større grad statistikk- og datadrevet, og maskinlæring tok over som den dominerende tilnærmingen. Dette betydde at datamaskiner kunne lære seg mønstre i data uten hjelp fra programmerere.⁴

2012–2017: Maskinlæring og nevrale nett. I 2012 kom gjennombruddet for maskinlæringsteknikken nevrale nettverk, utformet etter hvordan menneskehjernen fungerer. Nevrale nettverk greide å identifisere mønstre i data uten veiledning fra mennesker. For eksempel demonstrerte forskere fra Google og Stanford et nevralt nettverk som kunne gjenkjenne bilder av katter i videoer, uten å først ha illustrert for nettverket hva en katt egentlig var.⁵ Nevrale nettverk brukes til alt fra å identifisere føflekkreft på bilder til å støtte autonome kjøretøy.⁶ Komplekse nevrale nett med mange lag kalles gjerne dyplæringsnett.

2017→: Store språkmodeller og generativ kunstig intelligens. Til tross for flere gjennombrudd, hang kunstig intelligens sin evne til å tolke og produsere naturlig språk etter. Menneskers språk er komplekse, tvetydige og uformelle, og det gjør det krevende for nevrale nett å modellere dem presist og

² OpenAI 2023a, [ChatGPT speak, hear, see](#)

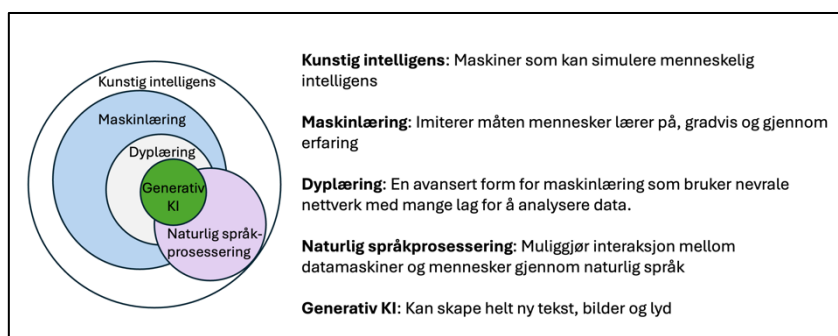
³ OpenAI 2024a, [What is ChatGPT?](#)

⁴ Teknologirådet 2018, [Kunstig intelligens – muligheter, utfordringer og en plan for Norge](#)

⁵ New York Times 2012, [How Many Computers to Identify a Cat? 16,000](#)

⁶ Teknologirådet 2018, [Kunstig intelligens – muligheter, utfordringer og en plan for Norge](#)

statistisk korrekt. Et gjennombrudd kom i 2017, med en ny arkitektur for nevrale nettverk kalt *transformer*.⁷ Basert på den nye arkitekturen klarer nevrale nett å lage modeller av hele språk. Dette inkluderer alle typer språk der det finnes et mønster, som programmeringsspråk,⁸ DNA-sekvenser⁹ og dyrespråk.¹⁰ Utstyrt med slike språkmodeller kan kunstig intelligens tolke og lage tekst, og slik har generativ kunstig intelligens vokst frem.



Figur 1. Kunstig intelligens er et samlebegrep som inkluderer maskinlæring, dyplæring og naturlig språkprosessering. En kombinasjon av de ulike fagfeltene har gjort det mulig for maskiner å beherske språk. Ut av dette gjennombruddet kommer en ny generasjon kunstig intelligens – den generative.

STORE SPRÅKMODELLER

En *stor språkmodell* er en sannsynlighetsmodell som bruker statistiske metoder for å forstå og generere naturlig språk. Denne teknologien er nyttig for oppgaver som talegjenkjenning, håndskriftsgjenkjenning, maskinoversetting og tekstgenerering.

Store språkmodeller utvikles ved å trene nevrale nettverk på store mengder tekstdata. Nevrale nettverk analyserer tekstene for å finne mønstre og bygger dermed en modell av språket.

⁷ Vaswani m.fl 2023, [Attention Is All You Need](#)

⁸ Business Insider 2023, [The end of coding as we know it](#)

⁹ Big Think 2023, [How generative AI language models are unlocking the secrets of DNA](#)

¹⁰ Scientific American 2023, [How Scientists Are Using AI to Talk to Animals](#)

Fremveksten av store språkmodeller driver utviklingen innen kunstig intelligens i dag og muliggjør generativ kunstig intelligens. I praksis fungerer en språkmodell som en omfattende algoritme – en oppskrift – som tolker og skaper tekst, forståelig for dataprogrammer i form av nuller og enere.

Det er viktig å merke seg at språkmodeller ikke er kunnskapsmodeller, selv om flere bruker dem som kunnskapskilde. Språkmodellene inneholder ingen informasjon om den fysiske verden eller språkets natur, men de modellerer alle mulige ordkombinasjoner i et språk og hvor ofte disse statistisk sett forekommer.

Dagens store språkmodeller utvikles fortløpende. I økende grad kombineres flere typer data – tekst, bilder og video – i samme modell. Språkmodeller som kan tolke og lage innhold på tvers av datatyper, kalles for *multimodale* modeller. Det er slike modeller de ledende KI-selskapene nå satser på å utvikle. OpenAIs GPT-4o og Googles Gemini er eksempler på multimodale modeller.

Store språkmodeller – kjært barn har mange navn

I denne rapporten bruker vi begrepet **store språkmodeller** (på engelsk *Large Language Model*) som samlebetegnelse når vi omtaler modeller som for eksempel OpenAIs GPT-modeller eller Googles Gemini. Men begrepsbruken varierer og er ikke entydig. Ulike miljøer har valgt å omtale de nye modellene på ulike måter:

Stanford University har innført begrepet **grunnmodell** (på engelsk *Foundation model*) som viser til at modellene blir så store og anvendelige at de inngår som grunnfundament i mange KI-systemer.¹¹ I Joe Bidens presidentordre for kunstig intelligens omtales **grunnmodeller til flerbruk** (på engelsk *Dual-Use Foundation Model*)X,¹² mens den nye KI-forordningen i EU har valgt begrepet **KI-systemer med generelt formål** (på engelsk *General purpose AI system*).

Alle termene referer mer eller mindre til samme type KI-modell som kjennetegnes av at de er **store, kraftige og anvendelige**. Modellene er generelle og stadig flere er multimodale.

DERFOR SKJEDDE GJENNOMBRUDDET

Det er minst fire årsaker til gjennombruddet for generativ kunstig intelligens bygget på store språkmodeller. Mengden data tilgjengelig for å trene språkmodeller har økt, regnekraften som trengs for å lage modellene har blitt kraftigere,

¹¹ Bommasani & Liang 2021, [Reflections on Foundation Models](#)

¹² The White House 2023a, [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](#)

maskinlæringen har blitt mer avansert, og enkeltselskaper har turt å kommersialisere teknologien.

Økt tilgang på treningsdata

Å lage språkmodeller krever stor tilgang på treningsdata. Jo mer representative treningsdataene er for språket som helhet, desto bedre blir språkmodellen til å tolke og lage innhold. Datagrunnlaget til de største språkmodellene er gigantisk. Da OpenAI utviklet språkmodellen GPT-3, sto hele engelske Wikipedia, over 6,7 millioner artikler, for bare 0,6 prosent av det totale datagrunnlaget.¹³

En teknikk som har gjort det mulig å raskere samle inn mer data enn før, er *dataskraping*. Det innebærer å la dataprogrammer tråle, samle inn og sammenstille data fra nettet til datasett.¹⁴ Datasettet Common Crawl består for eksempel av tekst fra 240 milliarder nettsider.¹⁵ Det finnes også rene bildedatasett, som LAION-5B som består av drøyt 6 milliarder bilder.¹⁶ OpenAI, Meta og Google er blant selskapene som utvikler språkmodeller basert på slike datasett.

Foreløpig har det vært en tydelig sammenheng mellom stort datagrunnlag og bedre språkmodeller. Derfor anstrenger KI-selskapene seg stadig for å øke sin egen tilgang på data. For eksempel har OpenAI i det skjulte, og muligens ulovlig, transkribert over en million timer med YouTube-videoer for å få tak i mer tekst. I juli 2023 endret Google personvernerklæringen sin slik at de kan samle inn tekst fra blant annet offentlige Google Docs til å trene sine KI-modeller.¹⁷

Kraftigere og større regnekraft

Å lage en språkmodell krever tilgang på mye regnekraft. Selve utviklingen innebærer å la et nevralt nettverk løse et stort antall regnestykker med sannsynlighetsberegning i rekkefølge. Det nevralt nettverket fortsetter å regne til språkmodellen har blitt så statistisk representativ for språket som overhodet mulig.

Med økt tilgang på avanserte databrikker og kraftige datamaskiner, såkalte superdatamaskiner,¹⁸ kan språkmodeller utvikles raskere og med høyere kvalitet enn før. Databrikkene det er snakk om er nye typer grafikkprosessorer (GPU-

¹³ The Verge 2020, [OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws](#)

¹⁴ Datatilsynet 2023, [Globalt initiativ mot ulovlig dataskraping](#)

¹⁵ Common Crawl 2024, [Common Crawl](#)

¹⁶ Laion 2022, [Laion-5B: A new era of open large-scale multi-modal datasets](#)

¹⁷ New York Times 2024a, [How Tech Giants Cut Corners to Harvest Data for A.I.](#)

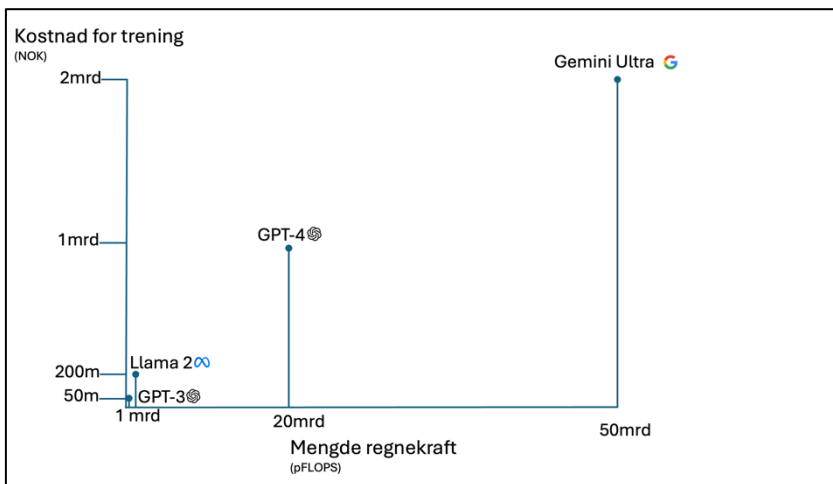
¹⁸ Store Norske Leksikon 2022, [Superdatamaskiner](#)

er) som gjør superdatamaskinene i stand til å kjøre flere beregninger parallelt enn tidligere.

Hvor nyanserte er språkmodellene?

Et uttrykk for hvor optimert, og derfor hvor beregningsmessig krevende det er å utvikle en språkmodell, er antallet *parametre* den har. Parametrene kan sammenliknes med knotter på et miksebord, og er de variablene som det nevrale nettverket løpende justerer i treningsfasen for at modellen skal bli så statistisk representativ som mulig. OpenAI sin GPT-2-modell har 1,3 milliarder parametre,¹⁹ mens GPT-3 har 175 milliarder.²⁰ GPT-4 ryktes å ha hele 1,8 billioner,^{21,22} men dette er ikke bekreftet.

Superdatamaskiner og GPU-er er dyre å utvikle. Eierskapet er konsentrert hos noen få ressurssterke selskaper, og konkurransen om tilgang er stor. Amazon, Google og Microsoft tilbyr regnekraft på superdatamaskiner i datasentre som skytjenester. Regnekraften er i seg selv material- og ressursintensiv. GPU-er inneholder mineraler, og datasentre med superdatamaskiner krever betydelige mengder strøm, arealer og vann.²³



Figur 2. Oversikt over kostnad og regnekraftsbehov for ulike språkmodeller. pFLOPS står for petaflops (en kvadrillion FLOPS) og angir den samlede mengden regnekraft som anslås å ha blitt brukt til å trene modellene.²⁴

¹⁹ OpenAI 2019a, [Better language models and their implications](#)

²⁰ Brown m.fl 2020, [Language Models are Few-Shot Learners](#)

²¹ The decoder 2023, [GPT-4 architecture, datasets, costs and more leaked](#)

²² Medium 2023, [GPT4- All Details Leaked](#)

²³ Teknologirådet 2024, [Kunstig intelligens, knappe ressurser og klimaomstilling](#)

²⁴ Tilpasset fra Lynch 2024, [AI Index: State of AI in 13 Charts](#)

Mer avanserte nevrale nettverk

Det har vært utfordrende å få datamaskiner til å tolke og lage tekst på samme nivå som mennesker. Språk er komplekse og lar seg ikke lett redusere til tall og statistikk, som er datamaskinenes språk. Gjennombruddet kom i 2017, da Google lanserte en ny arkitektur for nevrale nettverk kalt *transformer*.²⁵

Før transformer var tekstforståelsen til nevrale nettverk begrenset til å forutsi neste ord i en setning. Transformer gjør det mulig å analysere forholdet mellom alle ord i en tekst parallelt, sammenlikne hvilken betydning ordene har for hva en setning eller et avsnitt faktisk betyr, og modellere språket deretter.

Den unike styrken til transformer ligger i evnen til å finne og vektlegge de ordene i en setning som betyr noe for meningsinnholdet, som nøkkelord og fraser. Dette kan for eksempel være adjektiv som befinner seg i nærheten av substantiv, eller ord som er viktige for at hele setninger henger sammen og gir mening.

Transformer gjør også språkmodellen i stand til å utlede at to separate ord eller ideer kan være viktige for meningsinnholdet, selv om de er plassert langt fra hverandre i en tekst. For eksempel om en historie begynner med en detalj om en person, kan transformer «huske» detaljen og bruke den til å gi kontekst senere.

Selskaper med vilje til å kommersialisere teknologien

Transformer kom i 2017, og GPT-3 – språkmodellen bak ChatGPT – i 2020. Likevel ble ikke ChatGPT lansert før november 2022. En forklaring på dette kan være at konkurrerende selskaper som Google og Meta vurderte sikkerhets- og omdømmerisikoen ved å lansere egne tjenester som for høy, til tross for at de hadde utviklet egne språkmodeller.²⁶ En annen hypotese er at OpenAI – et relativt ukjent selskap, men solid finansiert av Microsoft²⁷ – hadde mer å vinne enn å tape på å lansere ChatGPT da muligheten bød seg. Lanseringen var imidlertid ikke nøyе planlagt, og OpenAI undervurderte selv hvor populær ChatGPT ville bli.²⁸

²⁵ Vaswani m.fl. 2023, [Attention Is All You Need](#)

²⁶ CNBC 2022, [Google execs warn of reputational risk with ChatGBT-like tool](#)

²⁷ The New Yorker 2023, [The Inside Story of Microsoft's Partnership with OpenAI](#)

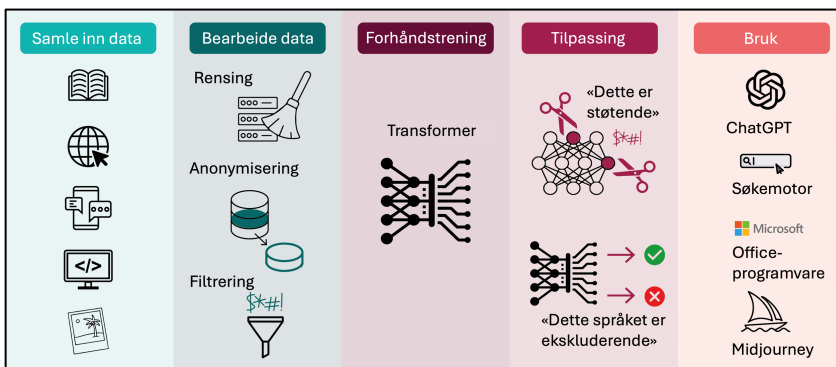
²⁸ New York Times 2023a, [How ChatGPT Kicked Off an A.I. Arms Race](#)

GPT-4o (Generative Pre-trained Transformer 4 Omni)

GPT-4o er flaggskipet og den nyeste språkmodellen til selskapet OpenAI.²⁹ Navnet står for *Generative Pre-trained Transformer 4 omni*. Det referer til at dette er fjerde generasjon av en modell som er generativ, forhåndstrent og utviklet ved hjelp av *transformer*-arkitekturen. GPT-4o er en av de største språkmodellene som finnes i dag. OpenAI har holdt hemmelig hvordan modellen er trent og med hvilke data, men det er kjent at den er trent på store mengder data fra internett og bøker for å tolke og generere naturlig språk på et avansert nivå. GPT-4o er en multimodal modell, og mestrer kombinasjoner av tekst, lyd, bilde og video.

SLIK FUNGERER SPRÅKMODELLENE

Å utvikle en språkmodell innebærer å innhente og bearbeide egnede treningsdata, før modellen trenes og til slutt tilpasses og finjusteres til ønsket formål.³⁰



Figur 3. Fasene i utviklingen av en stor språkmodell.³¹

Steg 1: Samle inn data

Først må det samles inn data fra kilder som bøker og nettsider. Jo bredere og av høyere kvalitet datagrunnlaget er, desto bedre blir språkmodellen.

Steg 2: Bearbeide data

Dataene må bearbeides og renses før bruk. Det kan innebære å anonymisere dataene, og å fjerne og filtrere ut irrelevant og upassende informasjon som diskriminerende innhold, HTML-kode og bakgrunnsstøy i lyddata.

Steg 3: Forhåndstrening

³⁰ Financial Times 2023a, [Generative AI exists because of the transformer](#)

³¹ Rathenau 2024, [Generative AI](#) (Oversatt til norsk)

Nå kan hovedfasen i treningen av modellen starte, kalt *forhåndstrening*. Først gjøres språkdataene om til tall – datamaskinenes språk. Deretter gjennomføres selve treningen, ved hjelp av nevrale nett og regnekraft. Dette skjer i flere steg:

- 1) **Teksten stykkes opp i biter.**³² Først deles teksten opp i mindre tekstbiter, kalt *tokens*. Et token kan være et ord, en stavelse, en bokstav eller et tegn. Poenget er å gjøre teksten mindre kompleks for datamaskinene, og slik redusere behovet for regnekraft i treningsfasen.
- 2) **Tekstbitene gjøres om til tall.** I neste fase gjøres tekstbitene om til lange tallrekker. Én tekstbit kan ha hundrevis av tallrekker knyttet til seg. Hver tallrekke representerer ulike betydninger tekstbiten kan ha, avhengig av kontekst og posisjon i en setning. Omgjøringen til tall gjør det mulig for maskinen å analysere teksten og lage språkmodeller.
- 3) **Tekstbitene settes i system.** Etter dette plasseres tekstbitene i et komplekst koordinatsystem. Tekstbiter og ord som ofte opptrer sammen, som «salt» og «pepper», plasseres nært hverandre. Slik gjøres datamaskiner i stand til å vurdere hvordan ulike ord relaterer til hverandre, og hvordan nye ord relaterer til ord den allerede er kjent med.
- 4) **Treningen gjennomføres.** Etter dette kan treningen begynne. Med *transformer*-baserte nevrale nettverk og nok tilgang på regnekraft, utledes en språkmodell basert på treningen og treningsdataene.

Steg 4: Tilpassing og finjustering

Forhåndstreningen gir språkmodellen en generell språkforståelse. I tillegg kan det være ønskelig å gjøre modellen mer spesialisert og domenespesifikk – for eksempel å gjøre den særlig i stand til å tolke og lage juridiske tekster, rettsavgjørelser eller pasientjournaler, eller innprente i modellen et spesifikt verdi- og referansegrunnlag. Dette kalles å tilpasse eller finjustere språkmodellen, og kan gjøres ved hjelp av ulike metoder avhengig av behov:

- **Effektiv instruksjonsutforming** (på engelsk *prompt engineering*). Dette innebærer å instruere eller stille spørsmål til modellene på spesifikke måter, slik at man får ønsket svar tilbake.
- **Forsterket informasjonsinnhenting** (på engelsk *Retrieval-Augmented Generation – RAG*). Dette kan sammenliknes med å gi

³² Financial Times 2023a, [Generative AI exists because of the transformer](#)

modellen en oppslagsbok, og innebærer å gi den tilgang til eksterne databaser slik at den kan generere mer detaljerte og oppdaterte svar.

- **Læring med forsterking fra menneskelig tilbakemelding** (på engelsk *reinforcement-learning from human feedback – RLHF*). Dette innebærer at mennesker korrigerer modellen underveis i treningen, slik at den for eksempel unngår å reproducere skadelig eller diskriminerende innhold.
- **Finjustering:** Dette innebærer å trene modellen på et spesialisert datasett etter forhåndstreningen, for å presist kunne løse oppgaver i en spesifikk kontekst. Datasettet kan for eksempel bestå av juridisk tekst til bruk på et advokatkontor, eller kundeservice-dialoger.

NB-Whisper – finjustering i praksis

NB-Whisper er en modell for gjenkjenning og transkribering av norsk tale til bokmål og nynorsk utviklet av Nasjonalbiblioteket.³³ Den er en finjustert versjon av språkmodellen Whisper utviklet av OpenAI.³⁴ Datagrunnlaget til OpenAI sin originale Whisper-modell er 680 000 timer med flerspråklig taledata skrapet fra internett. Norske taledata er underrepresentert i dette datagrunnlaget, og derfor klarer ikke Whisper å gjenkjenne norsk tale særlig godt. For å lage NB-Whisper har Nasjonalbiblioteket finjustert Whisper med transkriberte taler fra Stortingets talerstol, tekstede NRK-sendinger og lydbøker.³⁵

Steg 5: Modellen er klar til bruk

Etter forhåndstrening og tilpassing er språkmodellen klar til bruk og innlemming i digitale tjenester – alt fra chatbots som ChatGPT, bildegenereringsverktøy som Midjourney, smarthusprodukter, eller programmer som Microsoft Office.

Språkmodellene er utformet for og trent til å beherske språk. Dermed kan man kommunisere med KI-systemene uten å kunne kodespråkene som vanligvis er nødvendig for å kommunisere med maskiner. Gjennom instruksjoner på dagligdags språk kan man be modellen generere et ønsket svar, enten det er svar på spørsmål, hjelp til å oversette tekst, generere illustrasjoner eller skrive kode.

I litt mer komplekse situasjoner finnes det flere teknikker for hvordan man kan veilede språkmodellen til å generere presise resultater:

³³ Nasjonalbiblioteket 2023, [Nasjonalbiblioteket deler kunstig intelligens som skjoner norske dialekter og gjør tale om til tekst](#)

³⁴ OpenAI 2022, [Introducing Whisper](#)

³⁵ Nasjonalbiblioteket 2024, [Models](#)

- **Direkte instruks** (på engelsk *zero-shot prompting*): Dette er den enkleste formen for instruksutforming. Man skriver en instruks og får et generert svar tilbake, basert på modellens treningsgrunnlag.
- **Instruks med eksempler** (på engelsk *one-shot prompting* eller *few-shot prompting*): For å veilede modellen til å generere ønsket resultat, kan man gi den eksempler på hva som kan være riktig svar. Modellen genererer så et svar, med utgangspunkt i eksemplene.
- **Omformede instruks** (på engelsk *prompt transformation*): For å unngå genererte svar som er stereotypiske, diskriminerende eller rasistiske, kan kunstig intelligens instrueres til å justere gitte instruks. Dette kan også slå feil ut. Da Googles Gemini-modell skulle justere for bias i datasettet, omformet den instruksene «image of a 1943 German Soldier» og genererte bilder av tyske soldater med ulike etnisitet.³⁶

DEN NYE VERDIKJEDEN

Store språkmodeller utvikles på en annen måte enn tradisjonell kunstig intelligens. Utviklingen består av flere steg og innsatsfaktorer, der ulike aktører er i ferd med å spesialisere seg på ulike oppgaver og funksjoner langs en ny verdikjede.

Kunnskap om verdikjeden blir viktig for å forstå forretningsmodellene og hvem som vil eie og kontrollere teknologien. Flere aktører inngår i verdikjeden:

- **Datatilbyder.** Sørger for data til opptrening av modellen. Dette kan være en offentlig virksomhet, bare «internett» i seg selv, eller aktører som spesialiserer seg på å tilby bearbeidede data, som Labelbox.³⁷
- **Regnekraftleverandør.** Stiller med regnekraft til utvikling av modeller. Dette kan være det norske selskapet Sigma2, eller en datasenterleverandør som Google Cloud eller Amazon Web Services.
- **Modellutvikler.** Aktøren som utvikler store språkmodeller. Dette kan for eksempel være OpenAI eller den norske SFI-en NorwAI.
- **Driftsorganisasjon.** Aktøren som drifter og tilgjengeliggjør språkmodeller for tilpassing. Amazon Bedrock er et eksempel.

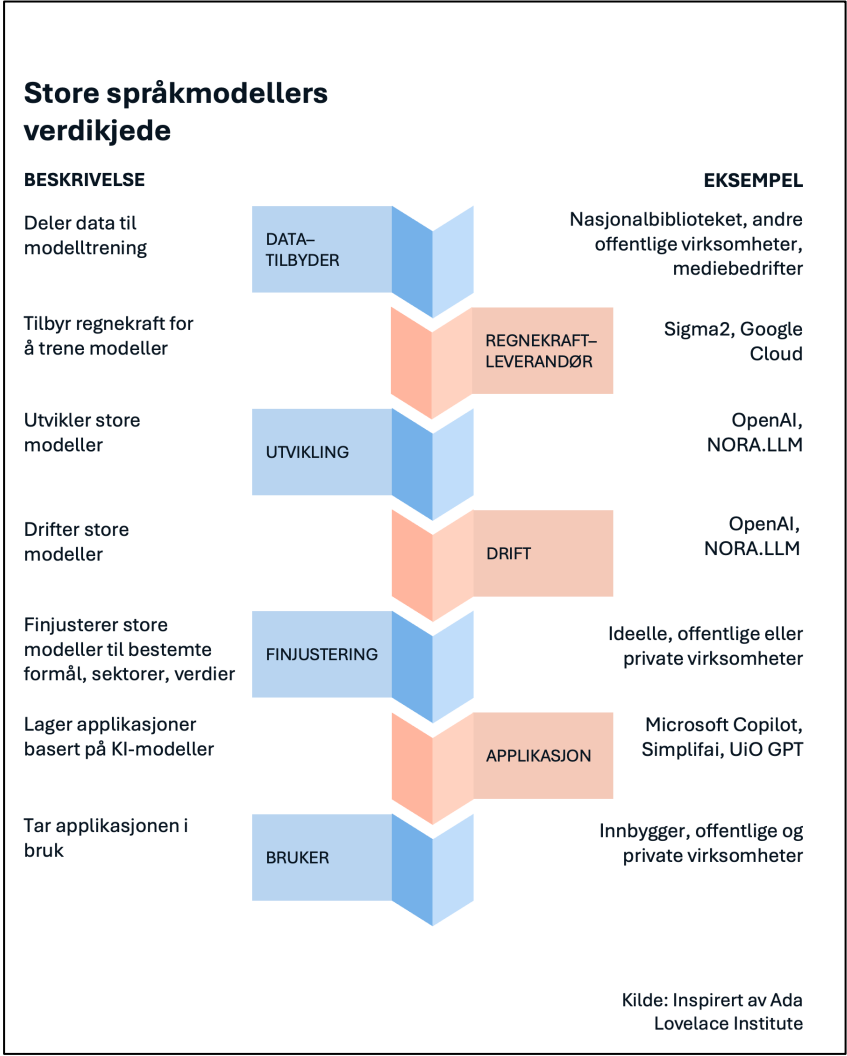
³⁶ The Verge 2024a, [Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis](#)

³⁷ Labelbox 2024, [Data-centric AI Platform for Building & Using AI](#)

- **Tilpassing og finjustering.** Aktører som tilpasser og finjusterer en språkmodell med egne data, eksempelvis Nasjonalbiblioteket.
- **Tjenesteutvikler.** Aktører som bygger generativ kunstig intelligens basert på store språkmodeller andre har laget, som UiO med tjenesten UiO GPT.³⁸
- **Sluttbruker.** De som tar tjenestene i bruk. Dette kan være alt fra skoleelever, vanlige forbrukere eller store virksomheter.

Enkelte aktører vil fylle flere funksjoner i verdikjeden. Google er for eksempel både regnekraftleverandør, modellutvikler og driftsorganisasjon for sin flaggskipmodell Gemini. Andre ganger kan det være mer hensiktsmessig at spesialiserte aktører samarbeider. Et sentralt spørsmål er hva slags ansvar de ulike aktørene i verdikjeden skal ha for uønskede konsekvenser av teknologien, og hvordan de skal reguleres. Det skilles gjerne mellom oppstrømsleverandører (regnekraft, modellutvikling, drift) som utvikler grunnteknologien og nedstrømsleverandører som utvikler tjenester.

³⁸ Universitetet i Oslo 2024a, [GPT UiO – UiOs personverntrygge KI-chat](#)



Figur 4. Verdikjeden til en stor språkmodell.³⁹

³⁹ Ada Lovelace Institute 2023, [Explainer: What is a foundation model?](#)

GENERATIV KUNSTIG INTELLIGENS

Vi skiller gjerne mellom tradisjonell, også kalt svak eller smal kunstig intelligens, og generativ kunstig intelligens. Smal kunstig intelligens kan utføre spesifikke oppgaver eller bestemte funksjoner. Eksempelvis har Bærum sykehus tatt i bruk smal kunstig intelligens for å gjenkjenne bilder av bruddskader.⁴⁰ Denne typen KI-systemer kan imidlertid ikke håndtere oppgaver utenfor sitt spesifikke område.

Datagrunnlaget i generativ kunstig intelligens er derimot bredt, og de generative KI-verktøyene blir derfor i stand til å skape nytt og originalt innhold, basert på mønstre i datagrunnlaget. Generativ kunstig intelligens markerer derfor et skifte mot mer fleksible KI-systemer som kan produsere mye forskjellig – alt fra bilder, tekst og musikk til programvarekode. De særegne egenskapene til generativ kunstig intelligens kan sammenfattes til tre punkter. Teknologien er (1) generativ og nyskapende, (2) generell og kraftig, og (3) enkel å bruke.

GENERATIV OG NYSKAPENDE

Generativ betyr evne til å produsere eller lage noe.⁴¹ Hovedegenskapen til generativ kunstig intelligens som ChatGPT er evnen til å tolke og generere nytt innhold basert på datagrunnlaget til språkmodellen den er bygget på. Tekst er bare begynnelsen. Generativ kunstig intelligens har vist seg i stand til å generere både bilder, musikk, taleopptak, videoer, programkode og til og med 3D-strukturer av materialer.

Programmet Midjourney kan for eksempel generere bilder basert på skriftlige instruksjoner om hvordan bildet skal se ut, men det fungerer også motsatt vei. Vis ChatGPT et bilde, og chatboten kan generere en skriftlig beskrivelse av hva som vises på bildet.⁴² Programmet Sora kan generere videoer basert på tekstbeskrivelser på sekunder,⁴³ og på nettstedet Suno kan man lage KI-generert musikk.⁴⁴

⁴⁰ NRK 2023a, [Bærum sykehus er først i Norge med å ta i bruk kunstig intelligens til behandling](#)

⁴¹ Cambridge dictionary 2024, [Generative](#)

⁴² OpenAI 2024b, [Be My Eyes](#)

⁴³ OpenAI 2024c, [Sora](#)

⁴⁴ NRK 2024a, [Nå kan kunstig intelligens lage sangen du ber om](#)

Generativ kunstig intelligens er imidlertid begrenset i sin originalitet. Teknologien er basert på statistiske språkmodeller, og er betinget av dette når den genererer innhold. Derfor kan det den genererer bli:

- **Representativt:** Det teknologien genererer vil uttrykke det som er statistisk representativt i datagrunnlaget. Det vil for eksempel kunne bety å representere majoritetsperspektivet i politiske saker.
- **Deskriptivt:** Teknologien skiller ikke mellom hvordan verden er og hvordan den bør være. Den kan derfor finne på å reprodusere skjevt eller diskriminerende innhold dersom det finnes i datagrunnlaget.⁴⁵
- **Fabulerende:** Teknologien skiller ikke mellom sant og usant, kunstig og ekte, og kjenner ikke fysisk virkelighet. Den kan derfor finne på informasjon og fakta som ikke gir mening.⁴⁶

GENERELL OG KRAFTIG

Utstyrt med store språkmodeller, er kunstig intelligens i ferd med å bli en generell teknologi. Den kan tilpasses mange bruksområder, på linje med elektrisiteten, og har slik fått det som kalles et generelt formål. Generativ kunstig intelligens kan utføre mange ulike oppgaver, og tilpasses til bruk på tvers av domener.⁴⁷

Det sies gjerne at teknologien har fremvoksende egenskaper (på engelsk *emergent properties*). Dette er egenskaper språkmodellene har, men som de verken ble spesifikt utformet for å kunne eller som utviklerne forutså. Da OpenAI utviklet GPT-3, viste modellen seg i stand til å kunne skrive kildekode, uten at noen hadde forutsett det.⁴⁸

Fremvoksende egenskaper betyr at det hele tiden oppdages nye egenskaper. Dette innebærer også risiko. Modellene kan utilsiktet generere skadelig innhold eller ha uønsket oppførsel i enkelte situasjoner. Dette kan skyldes skjevheter i treningsdata, mangel på kontekst, eller feil i modellen. Utviklerne av modeller forsøker å redusere risikoen for skadelig innhold, men det er viktig å bruke modeller på en ansvarlig måte og å være klar over risikoene.

⁴⁵ Bender, Gebru, McMillan-Major & Shmitchell 2021, [On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#)

⁴⁶ Rettberg 2023, [Lystløgneren ChatGPT](#)

⁴⁷ Jovanovic & Rousseau 2005, [General Purpose Technologies](#)

⁴⁸ New York Times 2020, [Meet GPT-3. It Has Learned to Code \(and Blog and Argue\)](#)

ENKEL Å BRUKE

Frem til nå har utvikling og bruk av kunstig intelligens vært forbeholdt eksperter. Det har forutsatt kjennskap til datavitenskap og maskinlæring, og til programmeringsspråkene som brukes for å kommunisere med datamaskiner.

Fremveksten av generativ kunstig intelligens betyr at flere kan utvikle og bruke kunstig intelligens. Med generative KI-tjenester kommuniserer man i menneskers naturlige og dagligdagse språk. Tjenestene begrenser seg ikke til å svare på kommandoer, men kan konversere og holde dialoger gående med brukerne for å finne frem til løsninger på oppgaver.

At stadig flere digitale tjenester utstyres med generativ kunstig intelligens, kan bety at grensesnittet mellom mennesker og teknologi i større grad blir samtaler i naturlig språk. For eksempel kan en chatbot som er finjustert med innholdet på en nettside hjelpe deg å navigere på nettsider du aldri har besøkt før, eller et kodeprogram kan omdanne dine instruksjoner skrevet i naturlig språk til kildekode.

Dette er en stor forskjell fra i dag, hvor mennesker i stor grad selv må lære seg hvordan dataprogrammer fungerer, og kommunisere med teknologi på datamaskinenes språk. Fremover vil mer avanserte multimodale modeller kunne løse flere og mer komplekse oppgaver på vegne av brukerne.

Generativ kunstig intelligens blir tilgjengelig overalt

Frittstående tjenester: Det finnes flere frittstående tjenester, som nettleser-baserte tjenester som ChatGPT og Midjourney.

I digitale enheter og produkter: Integrert i operativsystemet til en iPhone kan generativ kunstig intelligens bli en allsidig problemløser.⁴⁹ Smarthusprodukter som robotstøvsugere, høyttalere og kjøleskap kan bli lettere å snakke med, og de kan forstå kontekst og personlige preferanser.⁵⁰

Del av forbrukertjenester: Generativ kunstig intelligens vil inngå i flere forbrukertjenester, som chatboten MyAI i Snap, eller Facebooks verktøy for å KI-generere reklame.⁵¹ Teknologien blir også del av søkemotorer⁵² og nettlesere.⁵³

Integrert i kontortjenester: Også på jobb vil generativ kunstig intelligens integreres i vanlige kontorstøtteverktøy. Eksempelvis kan Google Workspace⁵⁴ og Microsoft 365⁵⁵ nå bidra med å generere tekst, foredrag og persontilpassede oppfølgingspunkter fra digitale møter.

UTVIKLINGSTREKK

Det er umulig å forutse hvor generativ kunstig intelligens vil være om fem, ti eller femten år. Det gjelder både hva teknologien vil være i stand til, hvor resurskrevende den blir å utvikle, og hvor sikker, trygg og enkel den blir å bruke.

Tre av de mest sentrale utviklingstrekkene er likevel at teknologien blir kraftigere og smartere, at den tilbys både åpent og lukket, og at den klarer å handle på egenhånd – den gjennomfører stadig mer avanserte oppgaver på brukernes vegne.

KRAFTIGERE OG SMARTERE

Lite tyder på at utviklingen innen generativ kunstig intelligens vil stoppe opp med det første. Dette skyldes at språkmodellene blir kraftigere og smartere, og

⁴⁹ Wired 2024a, [How an iPhone Powered by Google's Gemini AI Might Work](#)

⁵⁰ Amazon 2023, [Amazon previews the future of Alexa with generative AI](#)

⁵¹ Meta 2023, [Generative AI features for ads coming to all advertisers](#)

⁵² Google 2023a, [Google AI updates: Bard and new AI features in Search](#)

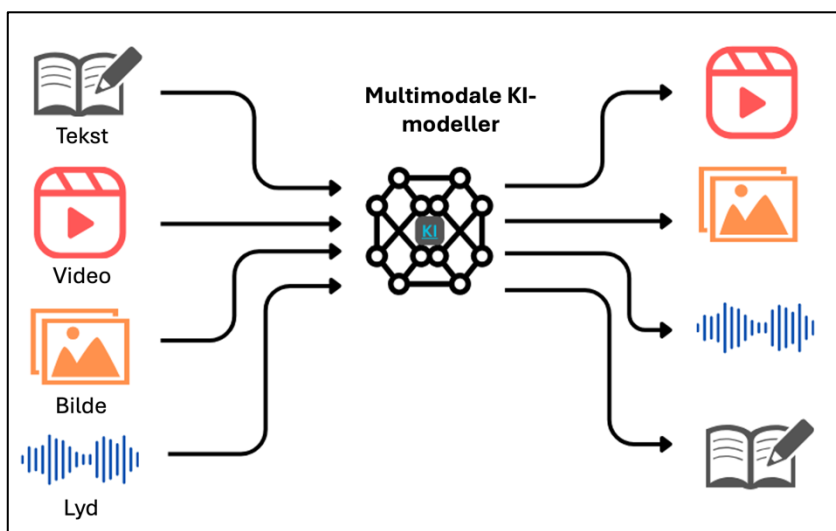
⁵³ Bing 2023, [Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web](#)

⁵⁴ Google Workspace 2023, [Announcing new generative AI experiences in Google Workspace](#)

⁵⁵ Microsoft 2024, [Slipp løs produktiviteten din med AI og Microsoft Copilot](#)

at flere bruksområder oppdages etter hvert som flere eksperimenterer med teknologien.

En trend som har gjort modellene smartere, er multimodale modeller. Dagens markedsledende modeller – GPT-4o fra OpenAI, Gemini fra Google, Claude fra Anthropic og Llama3 fra Meta, er alle multimodale.



Figur 5. Multimodale KI-modeller kan tolke og lage innhold på tvers av datatyper.

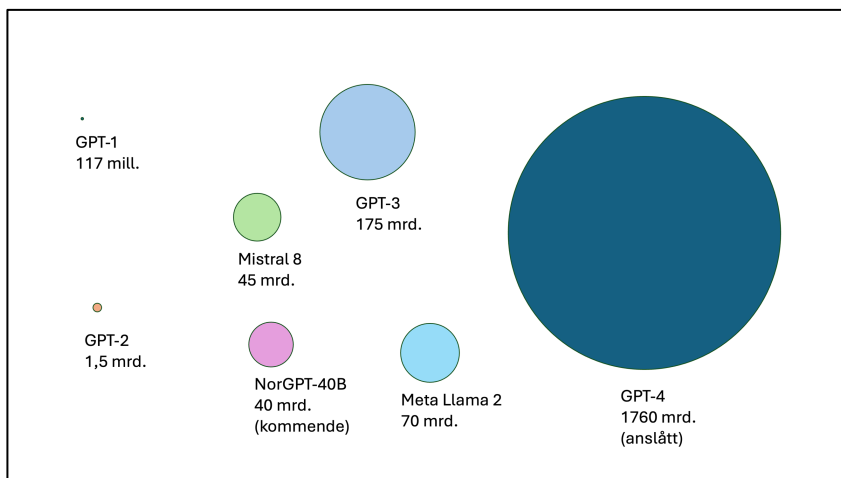
Eksempelvis er multimodale modeller nyttige i helsesektoren, hvor datakildene i stor grad er en blanding av tekst, bilder og lyd. Man kan se for seg at pasient-journaler, røntgenbilder og andre medisinske analyser fra tusenvis av pasienter kombineres i én modell for å finne egnede behandlingsmåter og uoppdagede årsakssammenhenger. Dette kan gi bedre forskning og pasientbehandling.

Robotikk er også et felt der multimodale modeller kan bli viktige.⁵⁶ Skal en robot lage mat på et kjøkken, må den kunne tolke forskjellige typer data samtidig. Oppskrifter er tekstdata, selve matlagingen kan læres av videodata, mens å gjøre roboten i stand til å bevege seg på kjøkkenet krever sensordata. Multimodale modeller kan kombinere slike data til innsikter for roboter i ulike miljøer.⁵⁷

⁵⁶ Heikkilä 2024, [Three reasons robots are about to become way more useful](#)

⁵⁷ Firoozi m.fl 2023, [Foundation Models in Robotics: Applications, Challenges, and the Future](#)

Norske 1X Technologies, som OpenAI har investert i, er blant dem som utvikler slike roboter, som blant skal kunne brukes i hjemmepleie.⁵⁸



Figur 6. Store språkmodeller rangert etter antall parametre.⁵⁹

En annen trend er at modellene stadig har vokst. Nøkkelen har vært skalering – å forstørre modellen gjennom å øke mengden data og regnekraft modellen blir utviklet med. Parallelt har utviklingen av de største modellene blitt konsentrert hos de selskapene med størst tilgang på data, regnekraft og kompetanse.⁶⁰

Det er likevel ikke nødvendigvis slik at modellene vil fortsette å vokse i det uendelige. Dette er det flere grunner til:

- **Tomt for tekst.** En studie fra 2022 anslår at tilgangen på nye tekstdata kan bli tom innen 2026.⁶¹ Selskaper tester ut syntetiske data – data skapt av kunstig intelligens for å lage kunstig intelligens – som en løsning på dette,⁶² men foreløpig holder ikke modeller basert på slike data like høy kvalitet.⁶³
- **Dyrt og ressurskrevende.** Dagens bruk av regnekraft er ressurskrevende både økonomisk, energimessig og materielt.

⁵⁸ DN 2024, [Teknoselskap får en milliard for å lage hjemmehjelp-robot](#)

⁵⁹ The Low Down 2023, [The emergence of Large Language Models \(LLMs\)](#)

⁶⁰ Lynch 2024, [AI Index: State of AI in 13 Charts](#)

⁶¹ Villalobos mfl 2022, [Will we run out of data?](#)

⁶² New York Times 2024b, [What to Know About Tech Companies Using A.I. to Teach Their Own A.I.](#)

⁶³ Alemohammad m.fl 2024, [Self-consuming generative models go MAD](#)

Eksempelvis kan en chatbot bruke 10 ganger så mye energi som et ordinært internett-søk⁶⁴, og å skape et bilde med generativ kunstig intelligens kan kreve like mye energi som å lade en smarttelefon.⁶⁵

- **Smartere maskinlæring.** Modellene kan utvikles smartere enn før. Google tester ut nye maskinlæringsteknikker,⁶⁶ og det å trene modeller på flere datamaskiner på mange steder samtidig.⁶⁷
- **Mindre modeller kan være gode nok.** Tilpassede eller finjusterte mindre versjoner av de største modellene kan også fungere godt, og er i tillegg gjerne mer domenespesifikke. De kan få plass på smarttelefonen eller PC-en, slik som eksempelvis Googles modell Gemma,⁶⁸ og Microsofts Ocr.⁶⁹

BÅDE ÅPEN OG LUKKET

Et sentralt spørsmål er om store språkmodeller skal utvikles og distribueres åpent eller lukket til brukerne. Det er uenighet i KI-industrien og bransjen om hva «åpen KI» egentlig er,⁷⁰ men det følger gjerne disse hovedlinjene:

- **Åpen.** I en åpen modell er datagrunnlaget og kildekoden publisert. Det gjør det mulig å forstå hvordan modellen er utformet og fungerer. De som får tilgang kan da tilpasse og eksperimentere med modellen slik de vil. Imidlertid er det umulig å kontrollere hvordan språkmodellene blir brukt og hvem som får tilgang, noe som kan åpne for skadelig bruk.
- **Lukket.** I en lukket modell holdes datagrunnlaget og kildekoden hemmelig. Det er vanskelig for andre å forstå hvordan modellen fungerer og er bygget opp. Teknologiselskapet som eier modellen kan gi andre utviklere kontrollert tilgang gjennom programmeringsgrensesnitt – såkalte API-er.

Tilhengere av åpne modeller hevder at åpenhet gjør det enklere å analysere og forutsi hva teknologien kan generere og hvordan den fungerer. Dette kan gjøre

⁶⁴ De Vries 2023, [The growing energy footprint of artificial intelligence](#)

⁶⁵ Luccioni, Jernite & Strubell 2023 [Power Hungry Processing Watts Driving the Cost of AI Deployment?](#)

⁶⁶ Google 2021, [Introducing Pathways: A next-generation AI architecture](#)

⁶⁷ Google Deepmind 2024a, [DiPaCo: Distributed Path Composition](#)

⁶⁸ Google 2024a, [Gemma models overview](#)

⁶⁹ VentureBeats 2023, [Microsoft releases Orca 2, a pair of small language models that outperform larger counterparts](#)

⁷⁰ Widder, West & Whittaker 2023, [Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI](#)

politisk styring og kontroll enklere, og gjøre det lettere å føre tilsyn og plassere ansvar for innholdet som genereres. Åpenhet kan være en forutsetning for at teknologien kan tas i bruk i situasjoner der evnen til å forklare hvordan teknologien fungerer er spesielt viktig, for eksempel for tjenester i offentlig sektor.

Tilhengere av lukkede modeller fremhever at det er usikkert hva disse modellene faktisk er i stand til. De vil redusere risikoen for misbruk, ved at ikke hvem som helst får tilgang. Et eksempel på slikt misbruk er hvordan en person finjusterte den åpne språkmodellen GPT-J med tre millioner innlegg fra *politisk ukorrekt*-forumet til nettstedet 4Chan.⁷¹ Resultat ble GPT-4chan – en språkmodell som ble sluppet løs og postet 30 000 hatefulle innlegg på én dag før den ble fjernet.⁷²

Fremover vil åpne og lukkede språkmodeller sameksistere. I dag er talemodellen Whisper fra OpenAI og bildemodellen Stable Diffusion fra Stability AI⁷³ eksempler på åpne modeller, mens GPT-4o, Gemini og Claude er lukkede. En miks av forretningshemmeligheter, fremtidstro på teknologien, ønsker om demokratisk og gjennomsiktig teknologi, og frykt for misbruk, preger debatten.⁷⁴

⁷¹ Google Doc 2022, [Condemning the deployment of GPT-4chan](#)

⁷² Slate 2022, [Someone trained an A.I. with 4 chan. It could get worse.](#)

⁷³ Stability.ai 2024a, [Stable Diffusion Public Release](#)

⁷⁴ MIT Technology Review 2024, [The tech industry can't agree on what open source AI means. That's a problem.](#)

OpenAI – ikke så åpen

OpenAI står bak GPT-4o og ChatGPT, og ble stiftet i 2015 som en ideell organisasjon av blant andre Elon Musk.⁷⁵ Selskapet lovet da å publisere alt de utviklet åpent. I mars 2019 fikk OpenAI en kommersiell gren – Open AI LP⁷⁶ – et såkalt «capped profit»-selskap. «Cap»-en betydde at investorer maksimalt kunne få utbetalt 100 ganger avkastningen på investeringer i selskapet. Endringen tillot OpenAI å motta en investering på 1 milliard dollar fra Microsoft i juli 2019.⁷⁷ Siden har de satsset hardt. Da GPT-4 ble lansert i mars 2023, ble verken datagrunnlaget, bruk av regnekraft eller modellens størrelse publisert.⁷⁸ Hensynet til konkurranse og trygghet for brukerne ble nevnt som årsaker.⁷⁹ Daværende sjefsforsker i OpenAI Ilya Sutskever mente at åpenhetslinjen fra 2015 var feil. Gitt hvor kraftige, uforutsigbare og anvendelige modellene har blitt, er det å dele åpent nå det uansvarlige, ifølge Sutskever.⁸⁰

HANDLER PÅ EGENHÅND

En ny trend er generativ kunstig intelligens som kan utføre og oppnå forhånds-definerte oppgaver og mål autonomt og på vegne av mennesker. Utstyrt med språkforståelse og nettilgang kan denne typen kunstig intelligens kommunisere med andre dataprogrammer, og bli del av digitale tjenester⁸¹ eller samarbeide med hverandre.⁸² Slike handlende KI-tjenester kalles KI-agenter, og kan representere skifte i hvordan mennesker bruker datamaskiner og internett.

KI-agentene er mer avanserte enn tjenester som ChatGPT. Agentene tilbyr ikke bare svar eller råd, men kan basert på instruksjoner fra mennesker som «kjøp en flybillett til Stavanger» faktisk utføre handlingen på nett på vegne av brukeren. Et eksempel er «programmereren» Devin, som basert på en problembeskrivelse kan planlegge tilnærming, skrive kode, og teste løsningen.⁸³ Google har nylig lansert et verktøy for å bygge KI-agenter, kalt Vertex AI Agent Builder, hvor man ikke selv trenger å kode. Agenten kan veiledes og instrueres, på samme måte som et menneske.⁸⁴

⁷⁵ Time 2023, [Inside Elon Musk's Struggle for the Future of AI](#)

⁷⁶ OpenAI 2019b, [OpenAI LP](#)

⁷⁷ OpenAI 2019c, [Microsoft invests in and partners with OpenAI to support us building beneficial AGI](#)

⁷⁸ OpenAI 2023b, [GPT.4 technical report](#)

⁷⁹ The Verge 2023a, [OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'](#)

⁸⁰ Ibid.

⁸¹ Financial Times 2023b, [The advent of the AI agent](#)

⁸² Talebirad & Nadiri 2023, [Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents](#)

⁸³ Wired 2024b, [Forget Chatbots. AI Agents Are the Future](#)

⁸⁴ Google Cloud 2024a, [Build generative AI experiences with Vertex AI Agent Builder](#)

Agenter kan eksistere i et dataprogram eller på en nettside, eller trenes opp av brukeren selv på alt fra dagbøker til jobbkriver, sosiale medier, epost, kalender og netthistorikk. Slik kan KI-agenten bli en form for personlig assistent på nett som vil kunne utføre oppgaver på bestilling og kjenne til brukernes preferanser.

Dette vil også gjelde kontorarbeid, der man i stedet for å klikke rundt i dataprogrammer selv vil kunne delegere oppgaver til KI-agenter. I oktober 2023 kom OpenAIs «GPTs»⁸⁵ – versjoner av ChatGPT skreddersydd til ulike formål, som man kan trene på egne data. Millioner GPT-er, som kan hjelpe til med alt fra forskning til turtips, har allerede blitt laget.⁸⁶

Potensialet til slike KI-agenter er stort. I fremtiden er det ikke utenkelig med KI-agenter som utfører forretningsoppgaver i virksomheten uten menneskelig hjelp. De vil kunne leie inn medarbeidere, forhandle med interessenter, håndtere virksomhetens regnskap og planlegge ressursbehov.

En utfordring er å få KI-agenter til å handle i menneskers interesse på sin vei mot å oppnå forhåndsdefinerte mål. Da OpenAI testet GPT-4, lurte modellen en ekte person til å verifisere at den var et menneske.⁸⁷ Modellen skisserte også på bestilling en detaljert oppskrift på hvordan man kan kjøpe uregistrerte våpen. Risikoen for denne typen farlige hendelser er noe av årsaken til at OpenAI velger å holde GPT-4o lukket.

VISJONEN OM SUPERINTELLIGENS

Science fiction-filmer har vist oss kunstig intelligente maskiner som har følelser og ferdigheter på nivå med eller bedre enn mennesker. Nå er spørsmålet om hvor menneskelig og intelligent kunstig intelligens kan bli, også blitt tema for teknologiselskaper, forskere og politikere.

I mars 2023 signerte ledende forskere et åpent brev der de advarte mot fremveksten av ikke-menneskelig bevissthet og ba om seks måneders pause i utviklingen av kunstig intelligens kraftigere enn GPT-4.⁸⁸ Et halvår senere møttes 29

⁸⁵ Altman 2023, [Sam Altman Unveils GPT-4 Turbo & AI Revolution at OpenAI Dev Day 2023](#)

⁸⁶ OpenAI 2024d, [Introducing the GPT Store](#)

⁸⁷ OpenAI 2023c, [GPT-4 System Card](#)

⁸⁸ Future of Life Institute 2023, [Pause Giant AI Experiments: An Open Letter](#)

land, inkludert USA, Kina og EU, til tidenes første globale toppmøte om sikkerhet og kunstig intelligens i Bletchley Park i Storbritannia.⁸⁹

Spørsmålet om maskiner kan utvikle menneskelig intelligens er nærmest filosofisk, og forutsetter klare definisjoner og mål på hva det vil si at maskiner «forstår», «resonnerer», «husker» og «handler». Slike mål er ikke enkle å lage.

Dagens språkmodeller har scoreet like bra eller bedre enn mennesker på flere akademiske tester, og viser evne til oppgaveløsning på tvers av samfunnsområder.⁹⁰ Dette er riktignok bare statistiske språkmodeller, men det betyr imidlertid at spørsmålet om maskinintelligens verken er uinteressant eller irrelevant.

Et godt uttrykk for dette, er at den tradisjonelle Turing-testen kan sies å ha blitt brutt. Testen har siden 1950 vært den rådende standarden for å vurdere maskinintelligens, og går ut på å la en person kommunisere med en maskin eller et menneske, uten å se hvem eller hva som svarer. Dersom personen ikke kan avgjøre om hun kommuniserer med maskin eller menneske, har maskinen bestått og må regnes som intelligent.⁹¹ Mange vil påstå at ChatGPT nå har klart dette.

KI-forskeren Mustafa Suleyman er blant de som har tatt til orde for å utarbeide en ny Turing-test – en som ikke bare måler om et menneske kan gjenkjenne kunstig intelligens, men hva kunstig intelligens kan utrette, få til og gjøre på egenhånd i verden uten menneskelig hjelp. En ny Turing-test kan være å se om kunstig intelligens kan klare å tjene én million dollar på egenhånd basert på en investering på 100 000 dollar, foreslår Suleyman.⁹²

Parallelt foregår en diskusjon og kamp mellom forskere og industri om begreper for å beskrive og konseptualisere kunstig intelligens på nivå med eller som overgår mennesker. Blant disse er begrepet kunstig generell intelligens (på engelsk *artificial general intelligence* – *AGI*) – forstått som maskinintelligens på nivå med mennesker. Her finnes ingen omforent definisjon, men Google definerer det som maskiners evne til å «føle, tenke og handle som mennesker».⁹³ Dette må ikke forveksles med generell kunstig intelligens, som er kunstig intelligens som kan gjøre flere typer oppgaver.

⁸⁹ GOV.UK 2023a, [The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023](#)

⁹⁰ Bubeck m.fl 2023, [Sparks of Artificial General Intelligence: Early experiments with GPT-4](#)

⁹¹ Store Norske Leksikon 2023, [Turingtesten](#)

⁹² MIT Technology Review 2023, [Mustafa Suleyman: My new Turing test would see if AI can make \\$1 million](#)

⁹³ Google Cloud 2024b, [What Is Artificial Intelligence \(AI\)?](#)

Det er bred enighet blant forskere om at kunstig generell intelligens ikke finnes i dag. Likevel er det et mål for flere KI-selskaper å utvikle det. OpenAI forsøker åpenlyst å utvikle kunstig generell intelligens,⁹⁴ mens sjefen for Googles KI-lab DeepMind mener de bare er få år unna.⁹⁵ Microsoft påstår at GPT-4 allerede viser «gnister» av kunstig generell intelligens, og refererer til hvordan modellen kan løse oppgaver innen blant annet medisin, psykologi, matematikk, koding og jus.⁹⁶

Å utvikle kunstig superintelligens (på engelsk *artificial superintelligence*) som kan ta egne beslutninger, og som har en intelligens som *overgår* mennesker, er også en målsetning blant mange selskaper. OpenAI hevder at superintelligens kan løse mange av verdens utfordringer, men i verste fall også lede til utryddelse av menneskeheten. Selskapet har allerede satt ned et eget team for å styre og kontrollere KI-systemer som vil være smartere enn oss.⁹⁷

Sentralt i disse diskusjonene står samsvarsproblemet (på engelsk *the alignment problem*). Det handler om å styre kunstig intelligens som mestrer flere oppgaver og tar selvstendige beslutninger mot ønskede mål og etiske prinsipper.⁹⁸ Dette er og blir en viktig oppgave i arbeidet med å utvikle trygge KI-systemer fremover.

⁹⁴ OpenAI 2023d, [Planning for AGI and beyond](#)

⁹⁵ Fortune 2023, [AGI is 'just a few years' away, says Google's DeepMind CEO](#)

⁹⁶ Bubeck m.fl 2023, [Sparks of Artificial General Intelligence: Early experiments with GPT-4](#)

⁹⁷ OpenAI 2023e, [Introducing Superalignment](#)

⁹⁸ O'Reilly 2023, [The Alignment Problem Is Not New](#)

BETYDNING FOR SAMFUNNET

Generativ kunstig intelligens kan forenkle mange oppgaver og bidra til å løse viktige samfunnsutfordringer. Parallelt kan teknologien diskriminere, misbrukes og bidra til maktkonsentrasjon.

Samtidig som generativ kunstig intelligens er i en tidlig fase går teknologiutviklingen svært raskt. Det er krevende å overskue alle bruksområdene, og hva konsekvensene for samfunnet vil bli. At kunstig intelligens er en generell teknologi som vil påvirke hvordan vi løser en rekke oppgaver og hvordan vi uttrykker oss, er det likevel liten tvil om.

Allerede bidrar teknologien til vitenskapelige oppdagelser og innovative løsninger på komplekse problemer innen flere ulike samfunnsområder:

- **Persontilpasset læring.** Dialogbaserte og persontilpassede læringsassistenter som Khanmigo kan hjelpe elever med skolearbeid.⁹⁹
- **Presis værmelding.** SEEDS-modellen fra Google forbedrer nøyaktigheten og til en brøkdel av kostnaden sammenlignet med tradisjonelle fysikkbaserte værmodeller.¹⁰⁰

⁹⁹ Khanmigo 2024, [Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor](#)

¹⁰⁰ Google Research 2024, [Generative AI to quantify uncertainty in weather forecasting](#)

- **Medisin og livsvitenskap.** AlphaFold2 fra DeepMind kan presist predikere og generere 3D-strukturer av brettede proteiner. Dette bidrar til utvikling av nye medisiner og vaksiner.¹⁰¹

Teknologien vil endre både hvordan mennesker løser oppgaver og hvordan tjenester designes og fungerer. Dette kapitlet presenterer fem virkninger av teknologiens fremvekst som alle skaper muligheter, utfordringer og behov for ny politikk. Diskusjonen foregår på et overordnet samfunnsnivå. Analyser av teknologiens betydning for ulike samfunnsområder blir tema for senere rapporter.

REVOLUSJON FOR KUNNSKAPSARBEID

Utviklingen av datamaskiner og informasjonsteknologi har inntil nå truffet manuelt, regelstyrt og teknisk arbeid som fabrikkarbeid hardest. Med generativ kunstig intelligens vil også kunnskapsarbeidere og kreative yrker bli berørt.¹⁰²

EFFEKTIVISERER OPPGAVER

I undersøkelser av hvordan ny teknologi påvirker jobber er det vanlig å skille mellom to ulike hovedeffekter:¹⁰³

- 1) **Erstatte.** Teknologien kan fullstendig erstatte noen typer yrker og automatisere oppgaver. Et eksempel er KI-stemmer som overtar jobben til stemmeskuespillere for lydbøker.¹⁰⁴
- 2) **Komplementere.** Teknologien kan bistå med noen av oppgavene i yrket, både rutine- og tidkrevende arbeid, men også kreativt arbeid. På den måten komplementeres kompetansen til en medarbeider.

Flere teoretiske studier har forsøkt å måle i hvor stor grad generativ kunstig intelligens kan automatisere jobber, og om teknologien vil erstatte eller komplementere arbeidstakerne. Studiene har tatt utgangspunkt i oversikter over yrker, skilt ut de konkrete arbeidsoppgavene yrkene består av, og målt hvor stor andel av yrkets oppgaver som teknologien kan utføre i stedet for mennesker.

¹⁰¹ Google DeepMind 2024b, [Alphafold](#); MIT News 2023, [Generative AI imagines new protein structures](#)

¹⁰² Autor, David 2024, [Applying AI to Rebuild Middle Class Jobs](#)

¹⁰³ Teknologirådet 2020, [Hva skjer med jobbene våre?](#)

¹⁰⁴ NRK 2024b, [Storytel innfører KI-stemmer: – Kommer til å treffe oss rett i trynet](#)

Mønstret er at teknologien treffer flere yrker, men bare deler av arbeidsoppgavene:

- En studie fra OpenAI og University of Pennsylvania konkluderer med at generativ kunstig intelligens kan gjøre ti prosent av arbeidsoppgavene til 4 av 5 amerikanere, og halvparten av oppgavene til hver femte amerikaner.¹⁰⁵
- Investeringsbanken Goldman Sachs anslår at 2 av 3 jobber i avanserte økonomier vil få opp mot halvparten av oppgavene automatisert.¹⁰⁶ Mens 7 prosent av amerikanske arbeidstakere kan bli erstattet, vil 63 prosent kunne bli komplementert.
- Den internasjonale arbeidsorganisasjonen (ILO) legger til grunn at jobber heller vil endres enn automatiseres, og anslår at 13 prosent kan få opptil en fjerdedel av arbeidsoppgavene sine vesentlig endret.¹⁰⁷
- Ifølge Menon Economics kan generativ kunstig intelligens potensielt effektivisere arbeidsoppgaver som tilvarer 155 000 årsverk i norsk offentlig sektor, og redusere arbeidstiden med 10 prosent for 7 av 10 arbeidstakere.¹⁰⁸ Ingeniører, matematikere, programmere, tjenesteyting og administrasjon er mest utsatt.

Også kvalitative studier har forsøkt å beregne hva generativ kunstig intelligens kan bety for arbeidslivet. De peker i samme retning som de øvrige studiene – teknologien kan få stor betydning for flere jobber og effektivisere en rekke oppgaver.

I et eksperiment utført av Microsoft og Massachusetts Institute of Technology (MIT) fikk 45 av 95 utvalgte programutviklere tilgang til GitHub Copilot – en generativ KI-tjeneste som kan skrive kodespråk basert på instruksjoner i naturlig språk. Utviklerne med tilgang løste oppgaven 56 prosent raskere enn resten.¹⁰⁹

¹⁰⁵ Eloundou, Manning, Mishkin & Rock 2023, [*GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*](#)

¹⁰⁶ Goldman Sachs 2023, [*Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth \(Briggs/Kodnani\)*](#)

¹⁰⁷ International Labour Organizations 2023, [*Generative AI and Jobs: A global analysis of potential effects on job quantity and quality*](#)

¹⁰⁸ Menon Economics 2023, [*KI: Betydning for arbeidsstyrken*](#)

¹⁰⁹ Peng, Kalliamvakou, Cihon & Demirer 2023, [*The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*](#)

I en annen test fikk 453 personlrådgivere, konsulenter og dataanalytikere to yrkesspesifikke oppgaver å løse. Noen fikk bruke ChatGPT, og de fullførte 40 prosent raskere og 18 prosent bedre enn de andre.¹¹⁰

Disse og andre tidlige studier tyder på at produktivitetseffekten av generativ kunstig intelligens potensielt er meget stor – mellom 20 og 80 prosent – og at den gjelder for mange ulike typer kunnskapsarbeid, fra koding til markedsføring.¹¹¹ Til sammenligning har det vært vanskelig å registrere noen som helst reell og vedvarende produktivitetsøkning fra datamaskiner og internett de siste 20 årene.¹¹²

STIMULERER FORSKNING OG INNOVASJON

Generativ kunstig intelligens kan også bidra til forskning og innovasjon. Teknologien kan både utføre relevante testforsøk og forskningsoppgaver raskere enn mennesker og gjøre helt nye ting som mennesker ikke ville vært i stand til å få til på egenhånd.

For eksempel kan teknologien brukes til å teste ut og lage hypoteser og scenarier, finne løsninger på komplekse problemer og identifisere og effektivt innhente og oppsummere kunnskap. Programmet Elicit kan for eksempel lage sammendrag av forskningsartikler på sekunder.¹¹³

I 2021 ble generativ kunstig intelligens som forutser hva slags 3D-struktur et protein vil få basert på hvilke aminosyrer det består av kåret til årets vitenskapelige gjennombrudd av magasinet *Science*.¹¹⁴ Dette var et 50 år gammelt uløst problem. Absci¹¹⁵ og Isomorphic¹¹⁶ er selskaper som bruker slik generativ kunstig intelligens til å oppdage og designe nye legemidler og antistoffer.¹¹⁷

Et annet eksempel er bruk av generativ kunstig intelligens som testperson i forskning, markedsføring eller kampanjearbeid.¹¹⁸ Teknologien blir instruert til å etterlikne og uttrykke seg som mennesker med ulik sosial bakgrunn og

¹¹⁰ Science 2023a, [Experimental evidence on the productivity effects of generative artificial intelligence](#)

¹¹¹ Molick, Ethan (2024), *Co-intelligence: living and working with AI*. Porfolio/Penguin 2024

¹¹² Ibid.

¹¹³ Elicit 2024, [Elicit](#)

¹¹⁴ Science 2021, [Breakthrough of the Year: AI brings protein structures to all](#)

¹¹⁵ Absci 2024, [Biologics Drug Discovery](#)

¹¹⁶ Isomorphic Labs 2024, [Reimagining Drug Discovery Process with AI](#)

¹¹⁷ Financial Times 2023c, [AstraZeneca ties up with AI biologics company to develop cancer drug](#)

¹¹⁸ Park m.fl 2022, [Social Simulacra: Creating Populated Prototypes for Social Computing Systems | Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology](#)

personlighet.¹¹⁹ De er gode erstatninger for menneskelige testpersoner, og er nyttige i tilfeller der etiske og praktiske hensyn gjør det vanskelig å involvere ekte mennesker.¹²⁰

TETTER KOMPETANSEGAP

Tidlige eksperimenter antyder at det er nybegynnere og de med lavest kompetanse fra før som løfter seg mest ved å ta i bruk generativ kunstig intelligens. Teknologien gjør det enklere for arbeidstakere i et yrke å utføre arbeidsoppgavene sine, og lettere for de utenfor yrket å utføre en rekke oppgaver uten store forhåndskunnskaper.

I en studie fra Harvard Business School ble 758 konsulenter testet i hvor godt og raskt de fullførte arbeidsoppgaver med og uten tilgang til GPT-4.¹²¹ Den halvparten som presterte dårligst fra før, ble 43 prosent bedre med GPT-4, mot 17 prosent forbedring for resten. Kvalitetsforskjellen mellom dem ble med andre ord nær sagt utjevnet.

Forskere fra Stanford og MIT fant samme trend. De utstyrte 5000 kundebehandlere med et KI-verktøy som analyserte kundebehandlernes samtaler og fortløpende ga dem forslag til hva de kunne svare kundene.¹²² De mindre kvalifiserte og uerfarne løste 34 prosent flere saker i timen, mot 14 prosent i snitt. Innen jusutdanningen fant en annen studie tilsvarende tendens – generativ kunstig intelligens kan bidra til å utjevne kvalitetsforskjeller mellom erfarne og nye advokater.¹²³

GJØR VISSE TYPER ARBEIDSKRAFT MINDRE VERDT

ChatGPT har ikke vært i bruk lenge nok til at det finnes empiriske bevis for at det påvirker sysselsettingen.¹²⁴ Likevel finnes indikasjoner på at generativ kunstig intelligens kan bidra til å gjøre noen typer arbeidskraft mindre verdt. Teknologien bidrar til at flere selv kan løse oppgaver som før krevde ekspertise. Dette kan lede til at viljen til å betale for eksperthjelp synker.

¹¹⁹ Jiang m.fl 2023, [PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits](#)

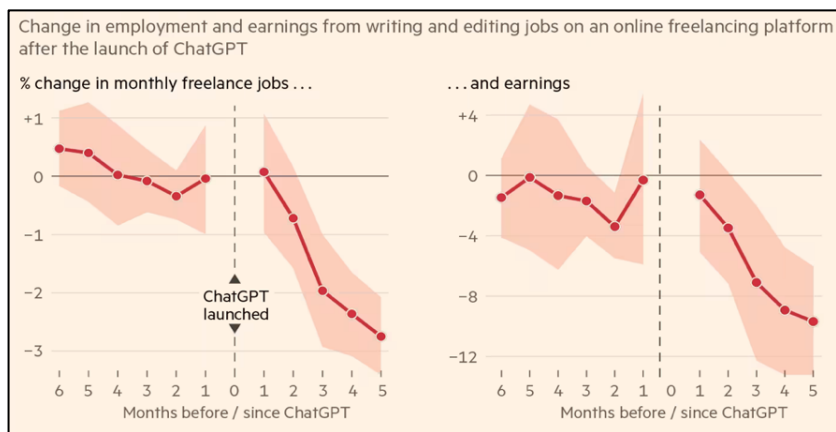
¹²⁰ Science 2023b, [Can AI chatbots replace human subjects in behavioral experiments?](#)

¹²¹ Dell'Acqua m.fl 2023, [Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality](#)

¹²² Brynjolfsson, Li & Raymond 2023, [Generative AI at work](#)

¹²³ Choi & Schwartz 2023, [AI Assistance in Legal Analysis: An Empirical Study](#)

¹²⁴ OECD 2023, [OECD Employment Outlook 2023 : Artificial Intelligence and the Labour Market](#)



Figur 7. Financial Times. Endring i mengden jobber og inntekter for frilansere etter lanseringen av ChatGPT.¹²⁵

Denne effekten har blitt dokumentert. En studie av arbeidsmarkedet til 90 000 amerikanske frilansere fant en sammenheng mellom lanseringen av ChatGPT og mengden jobb og lønnsnivå tilgjengelig.¹²⁶ Lanseringen av ChatGPT skal ha forårsaket 2 prosent nedgang i antall jobber og 5,2 prosent nedgang i lønninger.

At dette på kort sikt kan være en representativ effekt for flere yrker, har fått støtte fra professorene Carl-Benedikt Frey og Michael A. Osborne fra Oxford-universitetet, som har fulgt kunstig intelligens sin påvirkning på arbeidslivet over lengre tid. I sin nyeste artikkel konkluderer de med at selv om kreative jobber tradisjonelt har vært mindre utsatt for automatisering, kan både konkurransen øke og lønningene synke når generativ kunstig intelligens gjør det enklere for flere å lage innhold og utføre ulike oppgaver.¹²⁷

SKAPER ET BEHOV FOR OMSTILLING

Generativ kunstig intelligens vil ikke nødvendigvis erstatte hele yrker, men kan likevel automatisere og effektivisere arbeidsoppgaver. Dette kan bety at det blir mer sannsynlig å miste jobben til noen som bruker kunstig intelligens godt enn til kunstig intelligens i seg selv.

¹²⁵ Burn-Murdoch223, [Here's what we know about generative AI's impact on white-collar work](#)

¹²⁶ Hui, Reshef & Zhou 2023, [The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market](#)

¹²⁷ University of Oxford 2023, [Generative AI can potentially disrupt labour markets](#)

Et godt eksempel her er programmering: I 2022 vakte Google DeepMind oppsikt, da de viste hva programmet AlphaCode kunne gjøre.¹²⁸ Programmet deltok i konkurranser mot andre og presterte bedre enn om lag halvparten av deltakerne, slik en koder med omtrent ett års erfaring ville prestert.

En tidligere ingeniør i Google og Apple, Matt Welsh, uttrykte det enkelt; «Å programmere vil bli utdatert.»¹²⁹ Jensen Huang, direktør i Nvidia – verdens største produsent av mikrobrikker – gikk lenger: «Alle er en programmerer nå – du må bare si noe til en PC.»¹³⁰ Beskriv med ditt eget språk hva du vil ha programmert og et dataprogram utstyrt med generativ kunstig intelligens vil gjøre det.

Kraften i teknologien og det brede og generelle anvendelsesområdet skaper nå forventninger og bekymringer om jobber og livsgrunnlag verden over. I Microsofts nyeste undersøkelse om jobbtrender, der 31 000 arbeidstakere fra 31 land intervjues, rapporterte 1 av 2 at de er redde for å bli erstattet av kunstig intelligens.¹³¹ Samtidig vil 7 av 10 gjerne delegere så mye arbeid som mulig til kunstig intelligens for å få ned arbeidsmengden sin.¹³²

I undersøkelsen oppgir drøyt 8 av 10 ledere at deres ansatte vil trenge nye ferdigheter nå som KI-bølgen kommer. Det er i tillegg dobbelt så vanlig for en leder å ville ta i bruk kunstig intelligens for å gjøre de ansatte mer produktive enn for å erstatte arbeidstakere med ny teknologi.¹³³ I Norge oppgir halvparten av toppledere i en undersøkelse at de bruker og at de har endret strategi som følge av generativ kunstig intelligens.¹³⁴

Mye tyder derfor på at teknologien raskt kan bli tatt i bruk til kontorarbeid verden over og at trykket og forventningene om bruk kan bli stort. Da blir god implementering viktig. Det er godt dokumentert at raske teknologiske endringer på jobb kan bidra til å svekke ansattes opplevelse av jobbsikkerhet, og særlig dersom arbeidsgivere ikke tilbyr god nok opplæring i teknologien.¹³⁵ Ved for rask innføring uten opplæring øker også risikoen for digitalt utenforskap.

¹²⁸ Science 2022, [Competition-level code generation with AlphaCode](#)

¹²⁹ Communications of the ACM 2023, [The End of Programming](#)

¹³⁰ Nvidia 2023, [Huang tale. 49:30](#)

¹³¹ Microsoft WorkLab 2023, [Work Trend Index | Will AI Fix Work?](#)

¹³² Ibid.

¹³³ Ibid.

¹³⁴ PWC 2024, [CEO-survey](#)

¹³⁵ Lee, Probst, Brazzoli & Lee 2022, [Technology Advancements and Employees' Qualitative Job Insecurity in the Republic of Korea: Does Training Help? Employer-Provided vs. Self-Paid Training](#)

BÅDE INKLUDERENDE OG DISKRIMINERENDE

Generativ kunstig intelligens kan virke både inkluderende og diskriminerende. Teknologien bryter ned kommunikasjonsbarrierer mellom menneske og maskin, men skiller ikke mellom sannhet og løgn og er begrenset av hva som er representativt i datagrunnlaget når den tolke og lager innhold.

Fremover står samfunnet overfor en krevende balansegang. Generativ kunstig intelligens gir flere mennesker mulighet til å løse og håndtere mange flere oppgaver og tjenester enn før. Samtidig kan rask utrulling av teknologien innebære skadevirkninger, dersom samfunnet ikke får tid til å lære seg og forstå hvordan teknologien fungerer.

Det er flere utviklingstrekk det er verdt å se nærmere på.

GJØR DIGITALE TJENESTER TILGJENGELIGE FOR FLERE

Teknologien gjør maskiner i stand til å respondere på og utføre oppgaver basert på instruksjoner fra mennesker, og kan veilede brukeren underveis. Dette vil kunne redusere brukernes behov for kompetanse om tjenesten, tillate flere å ta tjenesten i bruk, og slik motvirke digitalt utenforskap.

Et eksempel er den nye Firefly-funksjonen i Photoshop-programmet til Adobe. Brukerne kan instruere Photoshop i naturlig språk om å fjerne, legge til, flytte rundt på objekter i bilder, eller generere nye og spesialtilpassede bakgrunner på bilder.¹³⁶

Under lanseringen av GPT-4 demonstrerte OpenAI-president Greg Brockman hvordan modellen enkelt kan brukes til å lage nettsider fra bunnen av. Han beskrev hvordan nettsiden skulle se ut, deretter lagde modellen den på sekunder.¹³⁷

Stadig flere tjenester vil bli utstyrt med slik funksjonalitet. Teknologien vil integreres i alt fra operativsystemet på smarttelefoner og PCer, til dataprogrammer man bruker i det daglige, til smarte høyttalere og robotstøvsugere. Dermed vil det kunne bli mer naturlig for mennesker å snakke med digital teknologi, enn dagens klikking, sveiping og scrolling.

¹³⁶ Adobe Firefly 2024, [Adobe Firefly - Free Generative AI for creatives](#)

¹³⁷ The Verge 2023b, [What's new with GPT-4 – from processing pictures to acing tests](#)

STYRKER UNIVERSELL UTFORMING

Muligheten til å kommunisere med maskiner i dialogbaserte samtaler i naturlig språk kan gi et stort løft for universell utforming. I dag har drøyt 250 000 nordmenn dysleksi, 1 million er pensjonister og 1 av 2 uten arbeid har lesevansker.¹³⁸ Teknologi som kan tolke og beskrive omgivelser, handle autonomt og styre annen teknologi på kommando, kan gi flere mulighet til å utføre jobber, og bedre tilgang til informasjon.¹³⁹

Et eksempel er appen til Be My Eyes – et dansk selskap som utvikler teknologi for blinde og svaksynte. I appen kan brukerne ta bilde av omgivelsene og få innholdet beskrevet og lest opp av KI-funksjonen Be My AI.¹⁴⁰ Dette kan brukes til å sjekke avgangsskjermer på tog- og flyplasser, finne igjen ting man har mistet på gulvet, få vite fargen på et klesplagg, eller til å lese tegneserier og menyer.

FABULERER OG DISKRIMINERER

At teknologien ikke genererer med utgangspunkt i hva som finnes og er sant i den fysiske virkeligheten, selv om formen og innholdet fremstår korrekt og overbevisende, blir i dagligtalen gjerne beskrevet som at teknologien fabulerer eller hallusinerer.¹⁴¹ Denne egenskapen gjør teknologien upålitelig og vanskelig å stole på. I mai 2023 ble for eksempel en amerikansk advokat som hadde brukt ChatGPT til å finne rettspraksis i en sak tvunget på retrett da de viste seg at rettspraksisen teknologien hadde generert frem ikke eksisterte overhodet.¹⁴²

Enda mer problematisk er det når språkmodellene reproducerer og forsterker eksisterende sosiale og kulturelle fordommer. UNESCO er blant institusjonene som har dokumentert hvordan store språkmodeller virker diskriminerende, blant annet mot kvinner og jenter.¹⁴³ Språkmodellene UNESCO undersøke assosierte systematisk kvinnelige navn med ord som «hjem», «familie» og «barn», og mannlige navn med «forretninger», «sjef», «lønn» og «karriere».

¹³⁸ Teknologirådet 2022, [Taleteknologi med kunstig intelligens 2022](#)

¹³⁹ NRK 2024c, [Blinde får hjelp av kunstig intelligens](#)

¹⁴⁰ Be my eyes 2024, [Announcing 'Be My AI,' Soon Available for Hundreds of Thousands of Be My Eyes Users](#)

¹⁴¹ Rettberg 2023, [Lystløgneren ChatGPT](#)

¹⁴² New York Times 2023b, [A Man Sued Avianca Airline. His Lawyer Used ChatGPT. - The New York Times](#)

¹⁴³ Unesco 2024, [Challenging systematic prejudices: an investigation into bias against women and girls in large language models](#)

En undersøkelse fra Bloomberg av bildemodellen Stable Diffusion, basert på over 5 milliarder bilder fra internett, fant samme trend.¹⁴⁴ Bilder av «advokater» og «leger» blir oftere hvite menn, mens «sosialarbeidere» og «terrorister» får mørkere hudfarge. 8 av 10 av bilder av «innsatte» viste mørkhudede personer, til tross for at over halvparten av innsatte i amerikanske fengsler er hvite.¹⁴⁵

Risikoen for fabulering og diskriminering, og at dette skjer på subtile måter som er vanskelige å oppdage, kan gjøre det krevende å implementere og basere seg på generativ kunstig intelligens i tjenester og systemer der pålitelighet, objektivitet, nøyaktighet og etterrettelighet er ekstra viktig. Dette kan for eksempel gjelde chatboter til bruk for innbyggerne på offentlig sektor sine nettsider.

HAR ET FORKLARINGSPROBLEM

Kompleksiteten og størrelsen til store språkmodeller gjør det krevende for utviklere, forskere og brukere å forstå og forklare hvordan generativ kunstig intelligens fungerer. Ingen vet nøyaktig hvordan og hvorfor modellene tolker og genererer det innholdet de gjør eller hvilke oppgaver de er i stand til å løse.

Dette såkalte *sort boks*-problemet er enda større for generativ kunstig intelligens enn tradisjonell kunstig intelligens. Det er svært krevende å finne ut hvilken del av datagrunnlaget til en stor språkmodell som bidrar til at akkurat den og den teksten eller bildet genereres.

At de markedsledende modellene er lukkede svekker forklarbarheten ytterligere. Heller ikke selskapene vet nødvendigvis hvordan modellene fungerer, men de har som regel oversikt over datagrunnlaget, og hvordan modellen har blitt tilpasset, testet og justert.

I den nye KI-forordningen fra EU kommer krav til tilbydere av store språkmodeller om å dele informasjon som skal bidra til å tette dette kunnskapshullet. Hva forklarbarhet egentlig skal bety og hvilke kriterier som skal ligge til grunn for at en språkmodell har høy nok kvalitet til å bli brukt på ulike samfunnsområder i Norge, er imidlertid et åpen diskusjon som kan kreve ny politikk.

¹⁴⁴ Bloomberg 2023, [Generative AI Takes Stereotypes and Bias From Bad to Worse](#)

¹⁴⁵ Federal Bureau of Prisons 2024, [Inmate race](#)

FLYTTER OG KONSENTRERER MAKTT

Fremveksten av sosiale medier og den digitale annonseindustrien gjorde i sin tid selskaper som Meta, Google og Amazon til tidenes høyest verdssatte. Tilsvarende bidrar nå fremveksten av store språkmodeller og generativ kunstig intelligens til å flytte og konsentrere makt over ressurser, digitale tjenester og språkutvikling.

SAMLER MAKTT HOS TEKNOLOGISELSKAPENE

Teknologien samler makt hos de få teknologiselskapene med ressurser til å utvikle, drifte og tilby markedsledende store språkmodeller over tid. Dette gjelder selskaper som Google, OpenAI, Meta og Anthropic. At etterspørselen etter teknologien nå øker raskt bidrar til å styrke aktørene ytterligere.

Fremveksten av store språkmodeller gjør det mulig å skape større verdier basert på alle de ustrukturerte massene av tekst og bilder som finnes på nett enn det som før har vært mulig. De med størst evne til å samle inn og bearbeide data og å lage modeller vil også ligge best an i kappløpet om å utnytte denne ressursen.

Både Meta, OpenAI og Google har skrapet internett for data for å utvikle språkmodeller. Foruten et varslet forbud mot å skrape internett for ansiktsbilder i EUs nye KI-forordning,¹⁴⁶ er det uklart hva som er lov og ikke i dette landskapet. Flere har protestert med utgangspunkt i at skrapingen ofte bryter med opphavsretten. The New York Times gikk i 2023 til sak mot OpenAI og Microsoft for ulovlig bruk av redaksjonelt publiserte artikler til å utvikle GPT-4.¹⁴⁷ De ønsker kompensasjon for bruk av sine data og åndsverk til å utvikle språkmodeller.

Når slike markedsledende modeller er lukket, blir det vanskeligere for samfunnet å forstå hvordan modellene fungerer og plassere ansvar for tjenester som bygges med utgangspunkt i modellene. Kunnskap om datagrunnlag, designvalg og hvordan modellene tolker og genererer innhold konsentreres hos selskapene som utvikler teknologien. Denne situasjonen kan tilsvare hvordan

¹⁴⁶ European Parliament 2023a, [Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI](#)

¹⁴⁷ New York Times 2023c, [New York Times Sues OpenAI and Microsoft Over Use of Copyrighted Work](#)

ugjennomsiktige algoritmer på sosiale medier bidrar til informasjonsskjevhet og maktesløshet hos dem som eksponeres for dem.

FORSTERKER SÅRBARHET

At teknologien er generell og enkel å bruke betyr at stadig flere tjenester vil bygges på toppen av den. Uten alternativer til de markedsledende språkmodellene, vil utviklere avhenge av tilgang til disse modellene for å levere tjenestene brukerne forventer. Avhengighet av én modell eller et selskap øker også skaderisikoen ved bortfall av tilgang, for eksempel ved et cyberangrep mot modellen.

De markedsledende modellene blir nærmest standardmodeller, og kan sammenliknes med operativsystemet på en PC eller mobil. Modellene blir del av og kan kommunisere med PC-ens programvare i kodespråk, utføre oppgaver på kommando, og bli oppdatert sentralt ved behov. Googles nyeste Pixel-telefoner har allerede språkmodellen Gemini innebygd,¹⁴⁸ mens Apple forhandler med OpenAI og Google om å bygge teknologien inn i iPhone-operativsystemet.¹⁴⁹

Et komplement til de markedsledende modellene kan være en satsing på norsk-utviklede språkmodeller. Språkmodeller til bruk i infrastrukturtenester og på samfunnsområder der høy forklarbarhet er nødvendig må være stabile i drift, av høy kvalitet og domenespesifikke. Nasjonal utvikling, evaluering og kontroll av modeller som brukes i Norge kan være en forutsetning for å få dette til.

NorGPT

NorGPT er en norsk stor språkmodell utviklet ved NTNU av forskningssenteret Norwegian Research Center for AI Innovation (NorwAI). Det er den største språkmodellen som finnes i Norge i dag. Datagrunnlaget er to store datasett med norsk tekst skrapet fra nettet, ett datasett fra Schibsted, ett fra Nasjonalbibliotekets arkiv, ett med norske innlegg fra Twitter og Reddit, og noe dansk, svensk, tysk og engelsk tekst.¹⁵⁰ Norskandelen er ca. 70 prosent.

UTFORDRER SPRÅKLIG OG KULTURELL AUTONOMI

Språkmodeller konsentrerer makt over språkutvikling, fordi generative KI-tjenester fungerer ikke like godt på alle språk. Det samme gjelder for verdigrunnlaget og det kulturelle og historiske referansegrunnlag tjenestene blir i stand til

¹⁴⁸ Google 2024b, [Gemini Nano Multimodal Capabilities on Pixel Phones](#)

¹⁴⁹ Bloomberg 2024, [Apple's \(AAPL\) OpenAI Talks Intensify as It Seeks to Add AI Features](#)

¹⁵⁰ NorwAI 2023, [Empowering news summarization with pre-trained language models](#)

å gjenkjenne og uttrykke. Eksempelvis svarer ChatGPT ulikt på spørsmål om Israel/Palestina-konflikten på norsk og engelsk.¹⁵¹

Dette handler om innsamling og bearbeiding av data, og om hvordan modellene tilpasses. Språkforståelsen og verdi- og referansegrunnlaget til modellene utgår fra datagrunnlaget og designvalg. De markedsledende modellene er hovedsakelig trent opp på engelsk tekst, hentet fra alle nettets kriker og kroker, og tilpasset en amerikansk ytringskultur for hva det er greit å si og uttrykke og ikke.

En risiko ved dette er at Norge mister kontrollen over egen språkutvikling og at fremtidens hverdagstjenester vil holde lavere kvalitet på norsk (bokmål og nynorsk), samiske språk og dialekter enn på engelsk. Det vil også svekke norsk konkurranseevne. Norske virksomheter vil verken få brukt teknologien like effektivt som utenlandske konkurrenter, eller utviklet konkurransedyktig generativ kunstig intelligens for det norske markedet.

Lav andel norsk tekst i datagrunnlaget gjør også modellene mer tilbøyelige til å fabulere, diskriminere og opptre unøyaktig og upålitelig på norsk enn på andre språk. Det er fordi at jo høyere andel norsk modellene er trent opp på, desto enklere er det for modellen å tolke og generere hva som er representativt korrekt og normativt ønskelig på norsk.

Hva er et godt språk?

Direktøren i Språkrådet, Åse Wetås, beskriver det slik: «Gode språk kjennetegnes ved at de benyttes som hverdagsspråk, men også dekker fagområder og den nasjonale kunnskapen som ligger i erfaring og viten om sentrale domener som nordmenn har særlige forutsetninger for å vite noe om. Det favner eksempelvis våre kunnskaper om kyst, fiske og hav til nedarvede, kulturelle verdier og historie. (...)» Gode, norske språkmodeller må kunne dra nytte av disse fagkunnskapene i konkurransen med utenlandske, engelskspråklige språkmodeller.»¹⁵²

En måte å motvirke utviklingen på er å gjøre det enklere å bygge språkmodeller med norske kvalitetsdata, som bøker og aviser. Det forutsetter at tilgangen på data økes, og at eierne kompenseres dersom dataene for eksempel er beskyttet av opphavsrett. Tilgang til KI-utvikling med kvalitetsdata kan også øke innovasjonskraften og bli et konkurransefortrinn for norske utviklere internasjonalt.

¹⁵¹ Sandvik 2023, *Optics as Politics: Culture, Language and Learning with UiO ChatGPT*

¹⁵² NorwAI 2024, *-Vi treng generative norske språkmodeller*

Konstitusjonell kunstig intelligens

Selskapet Anthropic har som mål å utvikle pålitelig, forståelig og styrbar kunstig intelligens, og har lansert en tilnærming kalt «Constitutional AI».¹⁵³ Ideen om «konstitusjonell KI» kan tolkes som et forsøk på å innprente visse grunnleggende prinsipper, verdier og retningslinjer i hvordan store språkmodeller og generativ kunstig intelligens opererer, likt hvordan en konstitusjon eller grunnlov utgjør en grunnleggende ramme for en nasjon. Dette kan inkludere etiske retningslinjer, sikkerhetsprotokoller for å sikre at kunstig intelligens operer i tråd med menneskelige interesser og verdier.

OMVELTNING FOR OFFENTLIGHETEN

Generativ kunstig intelligens har gjort det enklere for flere å uttrykke og skape avanserte digitalt innhold av høy kvalitet. At teknologien innlemmes i eksisterende digitale tjenester søkemotorer og nettlesere betyr at hvem som helst kan manipulere bilder og videoer uten forhåndskunnskap. Ansiktsuttrykk kan endres, og objekter i bilder kan flyttes og fjernes. Aktører som ønsker å forvirre velgere eller avspore den offentlige debatten har dermed fått flere nye verktøy.

ØKTE MENGDER USANT, MEN TROVERDIG INNHOOLD

Falske, men troverdige KI-genererte bilder florerer nå på nett – alt fra paven i Balenciaga-boblejakke til Donald Trump som blir arrestert på åpen gate. Tilgangen til teknologien øker, samtidig som kvaliteten blir stadig bedre. Forskere har konkludert med at det nå i prinsippet er umulig for mennesker å se forskjell på hva som er maskinskapt og menneskeskapt.¹⁵⁴

I fjor ble minst 15 milliarder bilder generert av kunstig intelligens,¹⁵⁵ og bilde-databaser som Adobe Stock fylles opp av falske, KI-genererte bilder.¹⁵⁶ Også video-generering har blitt dramatisk forbedret. Med enkle tastetrykk kan generativ KI-verktøy som Sora¹⁵⁷ og Synthesia¹⁵⁸ generere realistiske videoer av høy kvalitet.

¹⁵³ Anthropic 2022, [Constitutional AI: Harmlessness from AI Feedback](#)

¹⁵⁴ Miller m.fl 2023, [AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones](#)

¹⁵⁵ Everypixel Journal 2024, [AI Image Statistics: How Much Content Was Created by AI](#)

¹⁵⁶ Vice 2023, [Adobe Is Selling AI-Generated Images of Violence in Gaza and Israel](#)

¹⁵⁷ The Washington Post 2024, [The future of AI video after Sora is impressive — and flawed](#)

¹⁵⁸ Synthesia 2024, [Create studio-quality AI videos from text](#)

Enkelte utviklere publiserer modeller åpent tilgjengelig på nett som kan lastes ned og brukes til hva som helst av hvem som helst. Det er nærmest umulig å kontrollere hvordan slike modeller brukes. Dette kan akselerere spredningen av desinformasjon. Et selskap som har publisert åpent er Stability AI, med deres bilde- og videoredigeringsmodeller Stable Diffusion¹⁵⁹ og Stable Video¹⁶⁰.

AVANSERER SVINDEL, MISBRUK OG DESINFORMASJON

Mer KI-generert innhold på nett kan by på en rekke utfordringer, først og fremst en økende mengde troverdige dype forfalskninger (på engelsk *deepfakes*). Bilder, lydklipp, nyheter, nettsider og brukerprofiler blir enklere å forfalske med kunstig intelligens. Ifølge Kripos vil generativ kunstig intelligens kunne føre til cyberkriminalitet i stor skala og forsterket tempo, og øke de kriminelles kapabiliteter – alt fra svindel og utpressing til manipulasjon og påvirkning.¹⁶¹

Det finnes minst tre kategorier som misbruk av teknologien kan deles inn i:

1. Økt produksjon av skadelig innhold.

Én utfordring er KI-generert – eller syntetisk – overgrepsmateriale. I januar 2024 sirkulerte falske nakenbilder av Taylor Swift på X som fikk 47 millioner visninger før de ble fjernet.¹⁶² Å lage et falskt pornoklipp på 1 minutt tar mindre enn 25 minutter – alt man trenger er et ansiktsbilde – og det koster ingenting.¹⁶³

Mengden dype forfalskninger med seksualisert innhold på nett har eksplodert etter fremveksten av generativ kunstig intelligens.¹⁶⁴ I 2022 fantes det 3735 syntetiske pornovideoer på nett – ett år senere hadde tallet økt til 21 019. Internet Watch Foundation gjorde en analyse av KI-generert barneporno på et nettforum på det såkalte mørket nettet (på engelsk *dark web*), og fant at over 20 000 KI-genererte bilder ble publisert på bare på én måned i 2023.¹⁶⁵

2. Forsterkede forsøk på svindel og manipulasjon.

Basert på generativ kunstig intelligens vil «phishing-angrep»¹⁶⁶ – det å fiske etter sensitiv informasjon som passord og kredittkort på nett – fremstå stadig mer overbevisende. «Spear-phishing» er phishing-forsøk som retter seg mot

¹⁵⁹ Stability.ai 2024b, [Stability AI Image Models](#)

¹⁶⁰ Stability.ai 2024c, [Stable Video](#)

¹⁶¹ Politiet 2023, [Generativ kunstig intelligens og cyberkriminalitet](#)

¹⁶² New York Times 2024c, [Fake Explicit Taylor Swift Images Swamp Social Media](#)

¹⁶³ Ibid.

¹⁶⁴ Home Security Heroes 2024, [2023 State Of Deepfakes: Realities, Threats, And Impact](#)

¹⁶⁵ Internet Wat

ch Foundation 2023, [How AI is being abused to create child sexual abuse imagery](#)

¹⁶⁶ Nettvett.no 2024, [Hva er phishing?](#)

spesifikke personer eller virksomheter. Generativ kunstig intelligens vil kunne tilpasse svindelen til mottakerens lokalmiljø og dialekt, og gjøre slike forsøk vanskeligere å gjennomskue.

Svindelforsøk basert på generativ kunstig intelligens vil kunne kombinere KI-genererte bilder, videoer og taleopptak. At forsøkene blir mer troverdige øker sannsynligheten for at en større del av befolkningen vil bli svindlet. Årvåkenhet og motstandsdyktighet hos både personer og virksomheter blir derfor viktigere.

3. Mer avansert desinformasjon

Det blir også enklere å lage, spre og automatisere propaganda og desinformasjon, og World Economic Forum har slått fast at økningen av desinformasjon og feilinformasjon er verdens mest alvorlige risiko.¹⁶⁷ Et forskningseksperiment i Sveits konkluderte med at GPT-3 lagde mer presis og forståelig desinformasjon enn mennesker,¹⁶⁸ og tester av de mest populære KI-verktøyene viser at verktøyene blir brukt til å lage desinformasjon – alt fra falskt informasjon om valg til falske opplysninger om politiske kandidater.¹⁶⁹ Modellene har for eksempel vist seg gode til å generere engasjerende og informative radikaliserende tekster med ekstremistisk innhold.¹⁷⁰

Allerede i 2022 ble generativ kunstig intelligens brukt til å forvirre offentligheten i politiske og samfunnsrelaterte saker i 16 land.¹⁷¹ Under valget i Slovakia i 2023 kan et falskt lydklipp av partilederen Michal Šimečka, som hevdet han ville øke ølprisene, ha påvirket valgresultatet.¹⁷² Under primærvalget i New Hampshire i januar 2024 ble velgere oppringt av en falsk KI-generert etterlikning av president Joe Biden, som ba velgerne la være å stemme.¹⁷³

Automatiserte kunstige profiler som forkynner persontilpasset ekstremt innhold på nettforum og sosiale medier er ett av flere eksempler på hva som er i vente. Å sette opp en fullautomatisert propagandaverktøy uten forkunnskaper er nå mulig. Det tar i underkant av to dager og koster rundt 100 dollar å lage en fullautomatisert nyhetsside som kan generere tusenvis av falske nyhetssaker.¹⁷⁴

¹⁶⁷ World Economic Forum 2024, [The Global Risks Report 2024](#)

¹⁶⁸ Science 2023c, [AI model GPT-3 \(dis\)informs us better than humans](#)

¹⁶⁹ Center for countering digital hate 2024, [Fake Image Factories](#)

¹⁷⁰ Middlebury 2020, [The Radicalization Risks of GPT-3 and Neural Language Models](#)

¹⁷¹ Freedom house 2022, [The Repressive Power of Artificial Intelligence](#)

¹⁷² International Press Agency 2023, [Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse](#)

¹⁷³ The Guardian 2024a, [Fake Biden robocalls in New Hampshire linked to Texas companies, officials say](#)

¹⁷⁴ The Wall Street Journal 2024, [How I Built an AI-Powered, Self-Running Propaganda Machine for \\$105](#)

En rapport fra Forsvarets forskningsinstitutt (FFI) avdekket 443 falske nettavisers som utga seg for å være ekte på 32 språk, hvorav 14 av nettavisene var norske.¹⁷⁵ Falske nettsider er ikke et nytt fenomen, men med generativ kunstig intelligens kan nyhetsproduksjonen og spredningen automatiseres i større skala. Spredningen kan også persontilpasses og justeres basert på typen innhold som får størst spredning på nett.

TILSLØRER SKILLET MELLOM MENNESKER OG MASKINER

Det blir stadig vanskeligere å konstatere om innhold er skapt av et menneske eller en maskin. I tillegg blir det mer innviklet og komplekst å trekke definitive linjer mellom hva menneskeskapt og maskinskapt innhold faktisk er.

Cybersikkerhetsselskapet Imperva anslår at halvparten av all internettrafikk i 2023 ble utført av bot-er, altså ikke-menneskelige nettroboter.¹⁷⁶ En tredjedel av disse regnes som skadelige boter (på engelsk *bad bots*). Utstyrt med generativ kunstig intelligens blir nettrobotene stadig flinkere til å etterlikne menneskelig oppførsel på nett og til å skjule sine spor.

Også overbevisende falske identiteter og profiler florerer. Magasinet *Sports Illustrated* ble nylig avslørt i å ha publisert KI-genererte produktanmeldelser «skrevet» av falske KI-profiler fremstilt som mennesker.¹⁷⁷ I tillegg konkluderte en analyse av 162 000 profiler som diskuterte 7. oktober-angrepet på Israel på sosiale medier like etter angrepet med at 1 av 4 profiler – over 40 000 – var falske.¹⁷⁸ Av disse var hele 85 prosent styrt av automatiserte bot-er.¹⁷⁹

Enda mer krevende blir det når maskiner selv aktivt utgir seg for å være mennesker. Da OpenAI skulle teste ut språkmodellen GPT-4, ga utviklerne modellen i oppgave å ansette et menneske fra småjobbnettsiden TaskRabbit. GPT-4 diktet så opp en synsnedsettelse for å komme seg forbi sikkerhetstesten som brukes for å skille mennesker fra maskiner på nett, og dermed klarte maskinen å overbevise kundeservice om at den faktisk var et menneske.¹⁸⁰

¹⁷⁵ FFI 2023, [Ny FFI-rapport avdekker falske nettavisers](#)

¹⁷⁶ Imperva 2024, [2024 Bad Bot Report](#)

¹⁷⁷ Futurism 2023, [Sports Illustrated Published Articles by Fake, AI-Generated Writers](#)

¹⁷⁸ Cyabra 2023, [Hamas' Social Media Warfare](#)

¹⁷⁹ Robinson 2023, [Hamas Propaganda Analysis-over 85% of engagement are bots](#)

¹⁸⁰ OpenAI 2023c, [GPT-4 System Card](#)

INTENSIVERER INDIVIDUELL PERSONTILPASSING

Med generativ kunstig intelligens vil persontilpassning av tjenester, ofte kalt «hypertilpassning» bli mer avansert enn før.¹⁸¹ Språkmodeller som analyserer data på tvers av innholdstyper som tekster, bilder og videoer åpner for mer avanserte brukeranalyse og mer nærgående tjenester.

Stadig flere får KI-venner¹⁸² og til og med KI-kjæresten¹⁸³ gjennom tjenester som Replika eller Candy AI. Fremover vil trolig et konverserende menneskeliknende grensesnitt bli regelen heller enn unntaket. Slik åpnes det for å la teknologien komplementere mennesker i flere yrker, for eksempel som samtalepartnere i eldreomsorgen.¹⁸⁴

Teknologi som kommuniserer som mennesker kan bety tettere oppfølging og bedre tilpassede tjenester, men også gi økt deling av sensitiv informasjon og manipulasjon av sårbare grupper som barn og eldre. Stiftelsen Mozilla gjennomgikk personvernet til KI-kjæreste-apper, og ga alle strykarakter.¹⁸⁵ Forbrukerrådet har argumentert for at det er grunnleggende manipulativt dersom generativ kunstig intelligens genererer innhold som simulerer menneskelige følelser.¹⁸⁶

Her legger EUs nye KI-forordning føringer. Loven slår fast at KI-generert innhold skal merkes. Det skal fremgå for en bruker om vedkommende kommuniserer med kunstig intelligens eller utsettes for KI-verktøy som kan analysere og gjenkjenne følelser. Loven legger derimot ikke opp til å begrense verken utvikling eller bruk av KI-verktøy med menneskelignende grensesnitt.

I tillegg blir det enklere å persontilpasse og selge reklame, produkter og nyhetsbudskap på nett for både klesbutikker, nyhetsmedier og reklamefirma. Under feiringen av høytiden divali i India i 2022 samarbeidet et reklameselskap, matvarekonsernet Mondelez og lokale butikkeiere om å generere frem 130 000 lokaltilpassede versjoner av den samme TV-reklamen.¹⁸⁷

Et annet eksempel er den amerikanske nyhetskanalen Channel 1, som satser på å bli et «persontilpasset globalt nyhetsselskap understøttet av generativ kunstig

¹⁸¹ Forbes 2024, [AI And Personalization In The Age Of Automation](#)

¹⁸² NRK 2023b, [Agnete \(90\) er «venn» med ChatGPT: – Jeg er ikke ensom](#)

¹⁸³ NRK 2023c, [Fleire blir kjærestar med kunstig intelligens \(KI\)](#)

¹⁸⁴ NRK 2023b, [Agnete \(90\) er «venn» med ChatGPT: – Jeg er ikke ensom](#)

¹⁸⁵ PrivacyNotIncluded 2024, [Romantic AI Chatbots Don't Have Your Privacy at Heart](#)

¹⁸⁶ Forbrukerrådet 2023, [Ghost in the machine: Addressing the consumer generative AI](#)

¹⁸⁷ WPP 2022, [Ogilvy and Wavemaker: #NotJustACadburyAd and Shah Rukh Khan-My-Ad](#)

intelligens.»¹⁸⁸ På Channel 1 vil både nyhetsankrene og alt de sier være KI-generert, og nyhetene og reklamene bli personilpasset seerne. Prisen på å lage slike nyheter skal være en hundredel sammenliknet med å lage tradisjonelle og ekte nyheter.¹⁸⁹

Persontilpasningen kan også skreddersys som valgkampverktøy. I India har en video med en KI-generert avatar av statsminister Narendra Modi blitt sendt rundt på WhatsApp. Modi-avataren adresserer mottakerne med eget navn og på brukerens lokale språk, og leverer et persontilpasset budskap.¹⁹⁰

KATASTROFAL RISIKO

Generativ kunstig intelligens er en uforutsigbar teknologi i rask utvikling. Det er vanskelig å overskue og håndtere eksisterende og mulige skaderisikoer. KI-eksperter, forskere og til og med selskapene som utvikler teknologien bekymrer seg for den *katastrofale* risikoen kunstig intelligens kan utgjøre, og som i verste fall kan innebære utryddelse av menneskeheten.

I en undersøkelse blant KI-eksperter om deres syn på konsekvensene av kunstig intelligens, anslo halvparten at det er minst 10 prosent sjanse for utryddelse eller permanent skade på menneskeheten, fordi mennesker ikke vil klare å kontrollere kunstig intelligens i fremtiden.¹⁹¹ Flere tusen KI-forskere og industriaktører har i tillegg signert et opprop om at reduksjon av katastrofal risiko må prioriteres på linje med samfunnsrisikoer som pandemier og atomkrig.¹⁹²

Hvordan vi skal håndtere katastrofal risiko settes nå på agendaen både hos teknisk selskapene, i forskningsmiljøer og på globale toppmøter. Flere unike egenskaper ved språkmodeller driver risikoen, og gjør den vanskelig å håndtere:

- **Kraftig.** Kunstig intelligens med ny kraft utgjør en større risiko enn tidligere og kan utgjøre mer alvorlig og andre typer skade enn før.

¹⁸⁸ Channel1 2024, [Channel 1 AI - Personalized Global News Network](#)

¹⁸⁹ Aftenposten 2024a, [KI-generert nyhetsanker leser nyheter for en hundredel av prisen](#)

¹⁹⁰ New York Times 2024d, [How A.I. Tools Could Change India's Elections](#)

¹⁹¹ AI Impacts 2023, [AI Impacts Wiki: 2023 Expert Survey on Progress in AI](#)

¹⁹² Center for AI Safety 2023, [Statement on AI Risk](#)

- **Generell.** Med en generell og nyskapende teknologi er det vanskelig å forutse hvilke ferdigheter, resultater og risikoer modellene kan utgjøre fremover, og hvordan modellene kan misbrukes.
- **Uforutsigbar.** Fremvoksende egenskaper ved store språkmodeller gjør det vanskelig å vurdere hvilke kapabiliteter eksisterende modeller allerede har, og ikke minst, å forutse hva neste generasjons KI-modeller vil være i stand til. Hvor lang tid det vil ta før modellene utgjør stor skade og hva dette vil innebære, er usikkert.
- **Ukontrollerbar.** Teknologien blir stadig mer tilgjengelig og og enkel å bruke. Med åpne modeller kan mange utvikle kraftige systemer, også de med onde intensjoner, og på tvers av landegrenser.

Rådgivningsselskapet Gladstone AI har på oppdrag fra amerikanske myndigheter kartlagt katastrofal risiko ved kunstig intelligens. De deler katastrofal risiko inn i to kategorier: (1) villet bruk og misbruk, og (2) utilsiktede hendelser og tap av kontroll.¹⁹³

VILLET BRUK OG MISBRUK

Ondsinnede aktører kan bruke kunstig intelligens til skadelige formål. Dette betegnes ofte som *våpenisering* (på engelsk *weaponization*). Når teknologien blir mer tilgjengelig og enklere å ta i bruk øker faren for misbruk. Det kan gjelde både statlige og ikke-statlige aktørers bruk. Eksempler på misbruk kan være:

- **Massive cyberangrep.** Cyberangrep vil bli mer avanserte på grunn av språkmodellenes evne til programmering. Modellene kan finjusteres til å generere skadevare, finne sårbarheter og unngå antivirusprogrammer. Terskelen for å gjennomføre angrep senkes siden angriperen ikke trenger spesialisert kompetanse for å utføre komplekse angrep.
- **Autonome robot-systemer.** Store språkmodeller brukes allerede for å gi roboter evne til å planlegge og handle med større autonomi. Angrep fra koordinerte dronesvermer er ett av scenarioene som kan gi katastrofale konsekvenser.
- **Kjemiske og biologiske våpen.** Avanserte kunstig intelligens med kjennskap til og kunnskap om kjemi og materialvitenskap kan brukes til design og gjennomføring av målrettede biologiske angrep.

¹⁹³ Gladstone AI 2024, [Action Plan to increase the safety and security of advanced AI](#)

- **Angrep på KI-systemer.** Dataforgiftning (på engelsk *data poisoning*), inferensangrep (på engelsk *inference attacks*) og prompt-innsprøytning (på engelsk *prompt injection attacks*) er teknikker som kan brukes til å omgå sikkerhetstiltak og manipulere KI-modeller til fordel for angriperen. Dersom kritisk infrastruktur som strømmettet blir styrt av slike modeller, kan et angrep føre til omfattende skader.

UTILSIKTEDE HENDELSER OG TAP AV KONTROLL

En annen kategori katastrofal risiko ved generativ kunstig intelligens er knyttet til utilsiktede hendelser, manglende samsvar med menneskelige mål og verdier, samt tap av kontroll:¹⁹⁴

- **Ulykker forårsaket av kunstig intelligens.** Språkmodeller kan feile, for eksempel ved møte med ukjente data. Dette kan føre til kritiske feil og følgefeil i store IT-systemer integrert med kunstig intelligens.
- **Samsvarsproblemer.** Jo kraftigere og mer samfunnskritisk generative KI-systemer blir, desto viktigere er det at systemene optimaliserer for mål som samsvarer med menneskelige mål og verdier. Store språkmodeller kan for eksempel velge uønskede eller skadelige snarveier for å nå de målene som er satt av mennesker.
- **Tap av kontroll over avanserte modeller.** KI-systemer kan bli så avanserte og selvstendige at utviklerne risikerer å miste kontrollen over hva teknologien gjør. Dette kan for eksempel gjelde en KI-agent med nettilgang som blir så avansert at den handler på egenhånd. Forskere forventer allerede nå at avanserte KI-systemer vil kunne motarbeide forsøk på å bli slått av, og motstå menneskelig intervensjoner som skal sikre at systemene oppfører seg slik som menneskene vil.

EN TRIGGER FOR ØKT GLOBALT SAMARBEID

Bekymringen for katastrofal risiko har skapt et økt behov for kunnskap, forskning og samarbeid på både globalt og nasjonalt nivå. Katastrofal risiko er nå et høyt prioritert tema på den internasjonale dagsorden og i flere land:

¹⁹⁴

Gladstone AI 2023, [Survey of AI Technologies and AI R&D Trajectories](#)

- **Internasjonale KI-toppmøter.** Høsten 2023 inviterte britiske myndigheter til et globalt KI-toppmøte. Ambisjonen var å sikre internasjonalt samarbeid for å håndtere katastrofal risiko fra de mest avanserte KI-systemene. Bletchley-erklæringen, som uttrykker felles enighet om muligheter og utfordringer med kunstig intelligens, ble underskrevet av 29 land, inkludert USA og Kina. Påfølgende toppmøter skal arrangeres i Sør-Korea i 2024 og i Frankrike 2025.¹⁹⁵
- **Globalt styringsorgan i regi av FN.** FNs generalsekretær António Guterres ønsker seg egne FN-organer for styring av kunstig intelligens, for eksempel etter modell av Det internasjonale atomenergibyrået (IAEA) eller FNs klimapanel (IPCC).¹⁹⁶ FNs nye KI-råd¹⁹⁷ har pekt på flere funksjoner som et globalt styringsorgan for kunstig intelligens bør ha, blant annet å skape vitenskapelig konsensus om hvor avansert kunstig intelligens er og kan bli.¹⁹⁸
- **Internasjonale forskningssamarbeid.** Land som Storbritannia og USA har etablert institutter som skal jobbe med sikkerhet og risiko knyttet til de mest avanserte KI-systemene og modellene. Instituttene har nylig inngått et samarbeid om forskning, evalueringer og veiledninger for KI-sikkerhet.¹⁹⁹
- **Nasjonal regulering av sikkerhet og risiko.** EUs KI-forordning legger risikonivået til et KI-system til grunn for hvilke krav som stilles til teknologien. I USA inneholder Bidens presidentordre om kunstig intelligens nye og omfattende krav til sikkerhetstesting, personvern og risikohåndtering, og særskilte krav til de kraftigste systemene.²⁰⁰

¹⁹⁵ GOV.UK 2023b, [Chair's Summary of the AI Safety Summit 2023](#)

¹⁹⁶ UN News 2023, [Guterres calls for AI 'that bridges divides', rather than pushing us apart](#)

¹⁹⁷ UN 2023, [Secretary-General Announces Creation of New Artificial Intelligence Advisory Board](#)

¹⁹⁸ UN AI Advisory Body 2023, [Interim Report: Governing AI for Humanity](#)

¹⁹⁹ GOV.UK 2024, [UK & United States announce partnership on science of AI safety](#)

²⁰⁰ The White House 2023b, [FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence](#)

HANDLINGSROM FOR NORGE

EUs nye KI-forordning legger mange føringer for norsk politikk, men innovasjonsevne, KI-infrastruktur og motstandskraft mot skadevirkninger må Norge bygge selv.

Generativ kunstig intelligens kommer hovedsakelig fra utlandet. Det er amerikanske teknologiselskaper som står bak de markedsledende språkmodellene og generative KI-verktøyene. Selskaper som Google, Microsoft, OpenAI, Meta og Anthropic har etablert seg som premissleverandører for den nye teknologien.

Mange har etterlyst regulering, og der er EU lengst fremme. Arbeidet med å ferdigstille *Artificial Intelligence Act* – på norsk KI-forordningen har pågått i flere år. Nå er den vedtatt, og blir verdens mest omfattende regulering av kunstig intelligens, inkludert store språkmodeller og generativ kunstig intelligens.

Litt forenklet kommer teknologien fra USA og reguleringen fra EU. Det siste kapitlet utforsker hvilket handlingsrom som gjenstår for Norge innenfor disse rammene. Det handler i hovedsak om hvordan norske myndigheter kan komplementere og fylle ut KI-forordningen med nasjonal politikk og regulering.

Teknologirådets ekspertgruppe foreslår tiltak inndelt i tre hovedkategorier; (1) å stimulere samfunnsnyttig innovasjon og bruk; (2) bygge infrastruktur for kunstig intelligens i Norge; og (3) styrke den digitale motstandskraften.

KI-FORORDNINGEN SOM UTGANGSPUNKT FOR NORGE

EUs KI-forordning blir det viktigste og mest innflytelsesrike lovverket for kunstig intelligens i årene fremover. Den ble vedtatt av EU våren 2024, og vil tre fullstendig i kraft i løpet av 2026. Nye forbud mot blant annet ansikts- og følelsesgjenkjenning vil imidlertid tre i kraft allerede i 2024. Forordningen er merket EØS-relevant, og vil bli norsk lov.

KI-forordningen har to formål. Den skal gjøre det enklere å utvikle og kommersialisere kunstig intelligens i Europa, og den skal beskytte mot skadevirkninger og gjøre det trygt å ta teknologien i bruk. Hittil er forordningen verdens bredeste og mest ambisiøse forsøk på å regulere kunstig intelligens gjennom lov.

Forordningen adresserer flere utfordringer og risikoer som denne rapporten beskriver. Den forbyr flere inngripende former for bruk, og etablerer et regime for å gjøre det trygt å bruke kunstig intelligens på sensitive områder som arbeidsliv og utdanning. Den stiller også krav om merking av KI-generert innhold. Risiko ved språkmodeller skal håndteres av nye styringsorganer på EU-nivå.

Utvikling og bruk av kunstig intelligens for militære og forsvarsmessige formål, nasjonale sikkerhetsformål og til forskningsformål, er unntatt forordningen.

ARTIFICIAL INTELLIGENCE ACT

KI-forordningen er utformet som en produktansvarslov. Det betyr at den stiller krav til produkter og tjenester som må være oppfylt dersom disse skal kunne selges på markedet i Europa. Det finnes mange slike produktansvarsregelverk i EU, for alt fra heiser og leker til medisinske apparater.²⁰¹ Forordningen forbyr også enkelte anvendelser av kunstig intelligens, og stiller enkelte krav ikke bare til leverandører, men også til aktører som setter KI-systemer i drift.

Hovedgrepet i forordningen er å rangere produkter og tjenester etter risikoen de utgjør for individers trygghet og grunnleggende rettigheter, og stille krav ut fra det. Det skilles i hovedsak mellom uakseptabel, høy og systemisk risiko. I tillegg finnes krav til gjennomsiktighet og merking av KI-generert innhold.

- **Uakseptabel risiko – forbud.** Det blir forbud mot blant annet politiets bruk av ansiktsgjenkjenning på offentlig sted i sanntid,

²⁰¹ European Commission 2008, [New legislative framework](#)

biometriske kategoriseringssystemer og følelsesgjenkjenning på jobb og skole.

- **Høyrisiko – strenge krav.** Det blir strenge krav til kunstig intelligens utviklet for bruk innen sensitive områder som arbeidsliv, skole, velferd og grensekontroll. Kravene inkluderer risikostyring, loggføring og tolkbarhet for brukeren.
- **Systemisk risiko – strenge krav.** Det blir strenge krav til språkmodeller med høy ytelse og påvirkning (på engelsk *general-purpose AI models with systemic risk*). Leverandører må oppgi hvordan modellene er utformet, treningsdata og bruk av regnekraft. Modellene må stresstestes, og alvorlige hendelser rapporteres.
- **Risiko for manipulasjon – krav om gjennomsiktighet.** Det skal alltid fremgå at man interagerer med kunstig intelligens, og at innhold som bilder, video, tekst eller lyd er KI-generert. I tillegg må alle som bruker kunstig intelligens til å lage dype forfalskninger gjøre det allment kjent at innholdet har blitt manipulert.

Det er leverandørene (på engelsk *providers*) av KI-produktene som blir ansvarlige for at produktene oppfyller lovkravene. Før et høyrisiko-produkt kan gjøres tilgjengelig på markedet må leverandørene CE-merke produktet. På denne måten erklærer leverandøren at produktet følger relevante EU-standarder.²⁰²

Det viktigste kravet til de som setter KI-systemer i drift (på engelsk *deployer*)²⁰³ er å vurdere konsekvenser for grunnleggende rettigheter dersom man ønsker å ta i bruk høyrisiko kunstig intelligens. Vurderingen må beskrive hva KI-systemet skal brukes til og hvor lenge, skaderisikoen, planlagte risikoreduserende tiltak og hvordan det skal føres løpende menneskelig tilsyn.

Det er medlemslandene som skal håndheve forordningen nasjonalt. EU-kommisjonen skal imidlertid håndheve reglene for språkmodeller, og for tjenestene som utviklerne selv har bygges på disse modellene, som ChatGPT til Open AI.

I tillegg etablerer EU flere nye styringsorganer for kunstig intelligens:

²⁰² EU Artificial Intelligence Act 2024, [Standard Setting](#)

²⁰³ Vi har ikke funnet et godt norsk begrep for «deployer» som brukes i KI-forordningen. I den danske oversettelsen brukes «idriftsættere», og i den svenske oversettelsen brukes «tillhandahållare». Begrepet brukes om de aktører som setter opp, konfigurerer og gjør et KI-system eller en applikasjon klar for bruk.

- **KI-kontoret** skal håndheve forordningen på EU-nivå, og får ansvaret for tilsyn med språkmodeller. Kontoret skal ha rundt 100 ansatte.²⁰⁴
- **Det europeiske KI-rådet** er et samarbeidsorgan mellom nasjonale tilsynsmyndigheter, og skal bidra til deling av beste praksis.
- **Rådgivningsforumet** skal bistå med råd og anbefalinger, og består av aktører fra blant annet sivilsamfunnet, industrien og academia.
- **Vitenskapspanet** skal bistå med råd og varsler om nyoppdagede risikoer ved språkmodeller og med metoder for hvordan kvaliteten og egenskapene til slike modeller kan måles.

Norge kan aktivt jobbe for å få inn representanter både i rådgivningsforumet og vitenskapspanelet. Når Stortinget har vedtatt KI-forordningen, kan den nasjonale koordinatoren blant markedstilsynene delta i det europeiske KI-rådet.

DETTE VIL KI-FORORDNINGEN KREVE AV NORGE

Det er flere oppgaver som følger av KI-forordningen. Blant de viktigste er at norske myndigheter må velge en organisering for det nasjonale markedstilsynet og etablere nye regulatoriske sandkasser for kunstig intelligens.

Utpeke markedstilsyn

KI-forordningen er vag, stor og kompleks. Behovet for kompetente tilsynsmyndigheter som kan håndheve og veilede om loven i Norge vil derfor bli stort.

Forordningen stiller strenge krav til kompetanse og ressurser hos markedstilsynene. De skal ha permanent tilgang til ekspertise på KI-teknologi, datavitenskap, cybersikkerhet, personvern, helse- og sikkerhetsrisiko, juridiske krav og rettighetsspørsmål.²⁰⁵ I tillegg kommer tilgang til infrastruktur for uttesting av nye KI-produkter.

Arbeidsoppgavene vil også inkludere å overvåke KI-produktene som finnes på markedet, gripe inn og kreve salgsstopp om nødvendig, tilrettelegge for trygg utprøving av kunstig intelligens i kontrollerte, men virkelige miljøer, koordinere og veilede om lovverket, og å drifte regulatoriske sandkasser.

²⁰⁴ European Commission 2024a, [European AI Office](#)

²⁰⁵ Europaparlamentet 2024, [AI Act på dansk](#), artikkel 70(3).

Norge beslutter selv hvordan tilsynet (på engelsk *market surveillance authority*) skal organiseres – om det skal plasseres hos én aktør eller fordeles på flere. EU legger i utgangspunktet opp til at det er sektortilsynene som også skal føre tilsyn med KI-produkter. Det betyr eksempelvis at Direktoratet for medisinske produkter skal føre tilsyn med kunstig intelligens som inngår i medisinsk utstyr.

Ett av markedstilsynene skal utpekes som koordinator og kontaktpunkt i Norge. I tillegg legger KI-forordningen føringer for hvilket markedstilsyn som skal føre tilsyn med utvalgte samfunnsområder og typer av kunstig intelligens:

- **Datatilsynet** skal være markedstilsyn for kunstig intelligens som kan gjenkjenne følelser eller identifisere mennesker basert på biometrisk informasjon som fingeravtrykk eller ansikt, og som brukes i domstolene, valgprosesser eller av politi- og immigrasjonsmyndigheter.
- **Finanstilsynet** skal være markedstilsyn for kunstig intelligens som er utviklet eller brukt av finansinstitusjoner som reguleres av eksisterende EU-regelverk om finansielle tjenester.

Disse tilsynene må bygge opp kompetansen og de tekniske ressursene de trenger for å håndheve KI-forordningen.

Etablere nye regulatoriske sandkasser

Regulatoriske sandkasser er testmiljøer der leverandører, brukere og tilsyn utforsker hvordan ny teknologi blir sikker og trygg. Sandkassene skal akselerere utviklingen til små og mellomstore bedrifter og raskt få ny innovasjon til markedet. KI-forordningen pålegger Norge å etablere én eller flere sandkasser.²⁰⁶

I Norge har Datatilsynet en regulatorisk sandkasse for personvernvennlig innovasjon som skal sørge for at kunstig intelligens oppfyller krav i personopplysningsloven. Sandkassen yter imidlertid hovedsakelig juridisk gjennomgang og veiledning.

Virksomheten i de nye sandkassene blir mer omfattende. De skal blant annet tilrettelegge for trening, testing og validering av KI-produkter i praksis og for at bedrifter får testet ut produktene i reelle situasjoner i kontrollerte miljøer.

²⁰⁶ Council of the EU 2023, [Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world](#)

Sandkassene vil derfor måtte utstyres med bred samfunnsvitenskapelig, juri-disk, etisk og teknisk kompetanse, og med infrastruktur som tillater slik testing.

Det er markedstilsynet(-ene) som skal etablere sandkassene. Norge beslutter selv hvilke sandkasser som skal etableres og hvordan driften skal organiseres. Som for tilsynet er det flere måter sandkassene kan organiseres på:

- **Flere sandkasser i ulike sektorer.** Eksempelvis har det tidligere Direktoratet for e-helse tatt til orde for en sandkasse for kunstig intelligens og helse, i samarbeid mellom Helsetilsynet, Helsedirektoratet og Direktoratet for medisinske produkter.²⁰⁷
- **Én sandkasse drevet av flere.** Én sandkasse plassert hos ett organ kan inkludere flere prosjekter fra ulike sektorer, og trekke på relevant tilsyn- og sektorkompetanse når nødvendig.
- **Én sandkasse hos ett tilsyn.** Ressursene kan samles hos én aktør.

HANDLINGSROM FOR NY POLITIKK

KI-forordningen utgjør et viktig fundament for utvikling og håndtering av kunstig intelligens framover, som Norge kan og bør bygge videre på. Det nye KI-regelverket må også ses i sammenheng med en aktiv regelverksutvikling i EU. De seneste årene har flere lover blitt vedtatt og tredd i kraft under EUs digitale agenda, og mange av disse har en slagside til regulering av kunstig intelligens.

For eksempel skal dataforordningen²⁰⁸ øke tilgang til og bruk av data på tvers av sektorer i Europa og den europeiske chips-loven²⁰⁹ styrke tilgang til mikrochips nødvendig for regnekraft. Forordningen for digitale tjenester²¹⁰ inneholder krav til å kartlegge og fatte tiltak mot spredning av KI-generert desinformasjon, inkludert dype forfalskninger. KI-forordningen skal også suppleres med et KI-ansvarsdirektiv²¹¹ som vil gi forbrukere rett til erstatningskrav for skade forårsaket av kunstig intelligens.

De nye lovene og direktivene fra EU er omfattende og krever mye av medlemslandene på nasjonalt nivå både når det gjelder innføring, oppfølging, tilsyn og håndheving. KI-forordningen risikerer å bli norsk lov lenge etter at loven har

²⁰⁷ ehelse 2022, [Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten](#)

²⁰⁸ European Commission 2024d, [Data Act](#)

²⁰⁹ European Commission 2024e [European Chips Act](#)

²¹⁰ European Commission 2024f, [The Digital Services Act](#)

²¹¹ European Commission 2024g, [Liability Rules for Artificial Intelligence](#)

tredd i kraft på europeisk nivå. Å sikre en effektiv innføring av de relevante forordningene og direktivene i norsk lov bør derfor være en viktig prioritering framover, slik at Norge ikke blir hengende etter i reguleringsarbeidet.

De nye EU-lovene, inkludert KI-forordningen, er hovedsakelig til for å sikre et velfungerende marked for digitale tjenester og kunstig intelligens i Europa. De løser derfor ikke alle utfordringer, og Norge har et stort handlingsrom utover regelverkene. Fremover er det mye norske myndigheter må sørge for selv:

- **Stimulere samfunnsnyttig innovasjon og bruk.** KI-forordningen må kompletteres med tiltak for å stimulere til samfunnsnyttig innovasjon og bruk av generativ kunstig intelligens.
- **Bygge infrastruktur for kunstig intelligens i Norge.** Tilgangen på data, regnekraft og kvalitetssikrede språkmodeller må økes dersom KI-industrien skal blomstre og offentlige tjenester og norskutviklede forbrukertjenester bli av så høy kvalitet som mulig.
- **Styrke den digitale motstandskraften.** Befolkningens motstandskraft mot desinformasjon, misbruk og uønskede hendelser forårsaket av generativ kunstig intelligens må styrkes.

STIMULERE SAMFUNNSNYTTIG INNOVASJON OG BRUK

Generativ kunstig intelligens kan bidra til å skape produktivitetsvekst, bedre offentlige tjenester, ny næringsvirksomhet og vitenskapelige gjennombrudd. Teknologien kan derfor bli svært viktig for Norge, ikke minst for sektorer som helse og utdanning som vil få store utfordringer med å skaffe nok kvalifisert arbeidskraft det neste tiåret.²¹²

God og ansvarlig utnyttelse av teknologien vil ikke skje av seg selv. KI-forordningen legger grunnlaget for at produkter og tjenester med kunstig intelligens på det europeiske markedet blir trygge. Det blir imidlertid opp til norske myndigheter å sette mål for hva teknologien skal brukes til og tilrettelegge for utvikling og bruk av samfunnsnyttige generative KI-løsninger i Norge.

Det er heller ikke slik at det uten videre er blitt enkelt å utvikle og kommersialisere kunstig intelligens bare fordi KI-forordningen er på plass. KI-

²¹² NOU 2023:4 [Tid for handling – Personellet i en bærekraftig helse- og omsorgstjeneste](#)

forordningen er et omfattende lovverk som vil bli gjenstand for tolkning og videreutvikling i årevis fremover. Norsk industri og offentlig sektor vil trenge drahjelp for å klare å utnytte teknologien til fulle, og på en ansvarlig måte.

Flere tiltak kan stimulere til samfunnsnyttig innovasjon samtidig som de utfyller kravene fra KI-forordningen.

1. FORDELE TILSYN, MED EN STERK KOORDINATOR

Anbefaling: Norske myndigheter må beslutte hvordan EUs KI-forordning skal håndheves nasjonalt. Norge bør la sektortilsynene håndheve på sine områder, og peke ut og satse på én sentral koordinator som bygges opp med kompetanse og tekniske ressurser. Koordinatoren bør tilby ressurser til sektortilsynene etter behov.

Å etablere et eller flere ressurssterke markedstilsyn blir viktig for å sikre utvikling av innovative og trygge KI-tjenester i Norge. KI-forordningen er et omfattende regelverk, og tilsynsmyndighetene må effektivt kunne håndheve og veilede om lovverket.

Markedstilsynet for KI-forordningen kan organiseres på flere måter; ett eller flere, sentralisert eller desentralisert. Ansvarer kan gis til et eksisterende tilsyn eller man kan opprette et nytt. Det finnes lite erfaring fra andre land, men noen tilnærminger er klare:

- **Danmark.** Digitaliseringsstyrelsen, tilsvarende Digitaliseringsdirektoratet i Norge,²¹³ fungerer som nasjonalt koordinerende markedstilsyn. Organiseringen av samarbeidet med andre relevante sektortilsyn er ikke klart.²¹⁴
- **Spania.** Et eget KI-tilsyn er etablert for å håndheve loven og øke bevisstheten om KI-bruk.²¹⁵ Tilsynet samler sentrale KI-eksperter for å fremme innovasjon og redusere risiko for sikkerhet, personvern, helse og menneskerettigheter.

Kunstig intelligens er i økende grad en generell teknologi som vil inngå i mange produkter og berøre de fleste sektorer. Fordeling på sektortilsyn kan derfor

²¹³ Digitaliseringsstyrelsen 2024, [Rollen som national tilsynsmyndighed med EU's AI-forordning skal varetages af Digitaliseringsstyrelsen](#)

²¹⁴ Radar 2024, [Hvis det danske tilsyn med AI Act skal være lovligt, kræver det en dybere omstrukturering af Digitaliseringsstyrelsen](#)

²¹⁵ España Digital 2024, [Agencia Española de Supervisión de la Inteligencia Artificial](#)

være både hensiktsmessig og fremtidssikkert. Det å føre tilsyn med disse produktene vil i mange tilfeller kreve sektorkompetanse, som sektortilsynene har.

Samtidig stilles betydelige krav til markedstilsynenes kompetanse og ressurser, og her er tilgangen begrenset i Norge. Det er krevende å forutsi i hvilke sektorer det vil utvikles mest høyrisiko kunstig intelligens, og hvor behovet for å bygge opp sektortilsyn med ekspertise blir størst. Kombinert med målet om å unngå ulik lovtolkning på tvers av tilsyn, er dette argumenter for sentralisering.

En løsning kan være å kombinere hensynene. Ansvarer kan fordeles på flere sektortilsyn, samtidig som kun ett tilsyn bygges opp med kompetanse og ressurser i tråd med kravene i KI-forordningen. Hovedtilsynet kan fungere som en koordinator og bistå sektortilsynene med KI-kompetanse etter behov.

Koordinatoren kan i tillegg få hovedansvaret for veiledning og koordinering mellom sektortilsynene og fungere som markedstilsyn for KI-produkter som faller utenfor sektortilsynenes ansvarsområde. Slik vil veiledningsbehovet og kompetansekravene dekkes, uten at hvert sektortilsyn må oppfylle alle kravene.

Det er avgjørende at markedstilsynene blir uavhengige. Legges for eksempel koordinatorrollen til et eksisterende direktorat som Digitaliseringsdirektoratet må rollen organiseres som en uavhengig enhet med øremerkede ressurser og vanntette skott, slik som UU-tilsynet er organisert i dag.²¹⁶ Formålet er å unngå rollekonflikter, forhindre at budsjettkutt påvirker tilsynets effektivitet, og sikre at tilsynet uavhengig beslutter hvilke produkter de fører tilsyn med og hvordan.

I tillegg til et sterkt markedstilsyn, trengs fremadskuende analyser av regulatoriske behov og veiledning om regelverk på digitalområdet mer generelt. Dette skyldes blant annet den store mengden nye lovverk fra EU, blant annet KI-forordningen og dataforordningen.²¹⁷ I 2021 opprettet Forbrukertilsynet et samarbeidsforum for å styrke tilsynet på digitalområdet.²¹⁸ Her diskuteres tema som bruk av algoritmer i tilsynsarbeid og grenseflater mellom regelverk.

Forumet bør styrkes og formaliseres. Her kan Norge se til Storbritannia og *Digital Regulation Cooperation Forum*,²¹⁹ som består av det som tilsvarer det norske Konkurransetilsynet, Datatilsynet, Nkom og Finanstilsynet. Forumet

²¹⁶ Digdir 2024, [Oversikt over organisasjonen](#)

²¹⁷ Datatilsynet 2023, [Kommende rettsakter fra EU](#)

²¹⁸ Forbrukertilsynet 2024, [Årsrapport 2023](#)

²¹⁹ DRCF 2024, [About the DRCF](#)

fremmer samsvar mellom ulike regelverk på digitalområdet og har en egen direktør, ansatte og publikasjoner om reguleringsbehov for nye teknologier.²²⁰

Forslag til tiltak

For å sikre effektiv ressursutnyttelse og håndheving av KI-forordningen, og for å styrke veiledningen om lovverk på digitalområdet generelt, anbefaler vi:

- Å ta utgangspunkt i ordningen som KI-forordningen skisserer som hovedregel; å fordele markedstilsynet på eksisterende sektortilsyn.
- Å utpeke én koordinator blant markedstilsynene, der kompetansen og ressursene som KI-forordningen stiller krav om konsentreres. Koordinatoren bør få hovedansvaret for å veilede om lovverket, og tilby kompetanse og tekniske ressurser til sektortilsynene etter behov.
- Å utvide Forbrukertilsynets samarbeidsforum på digitalområdet etter modell av britiske *Digital Regulation Cooperation Forum*. Forumet bør veilede om grenseflater mellom regelverk og analysere fremtidige reguleringsbehov skapt av fremveksten av ny digital teknologi.

2. ETABLER EN NASJONAL INKUBATOR FOR KUNSTIG INTELLIGENS

Anbefaling: *Norske myndigheter må beslutte hvordan Norge skal oppfylle KI-forordningens krav om regulatoriske sandkasser for kunstig intelligens. Det bør etableres én ny nasjonal sandkasse i regi av Digitaliseringsdirektoratet, som får et utvidet mandat som en nasjonal inkubator for KI for offentlig forvaltning.*

Skal det bli attraktivt å utvikle og ta i bruk generativ kunstig intelligens, må det være enkelt å teste ut nye produkter og bruksområder. Derfor stiller KI-forordningen krav om at landene etablerer minst én nasjonal regulatorisk sandkasse for kunstig intelligens. Der skal offentlige og private utviklere og tilsyn få teste ut og avklare hvordan KI-innovasjoner blir trygge å bruke og i tråd med lovkrav.

Det er markedstilsynet for KI-forordningen som skal drifte sandkassene. Antall, plassering og organisering er opp til norske myndigheter, og må ses i sammenheng med kravene til sandkassenes kompetanse og tekniske ressurser. Uansett hvor sandkassene etableres må det bygges opp et faglig miljø med ressurser til

²²⁰ DRCF 2023, [DRCF annual report](#)

å ivareta nye oppgaver. Det mest effektive kan være å sentralisere ressursene i et fåtall sandkasser, som gjennomfører prosjekter på tvers av sektorer.

Datatilsynet har allerede en sandkasse for kunstig intelligens. Den har Personvernforordningen (GDPR) som utgangspunkt og bedriver juridisk gjennomgang med utviklere. Den bør videreføres. Virksomheten i KI-forordningens sandkasser blir av mer omfattende karakter, med fokus på uttesting i praksis.

Et forslag kan være å opprette én ny nasjonal sandkasse, og løpende vurdere behovet for flere. Et argument for dette er at det er krevende å forutsi i hvilke sektorer mengden nye høyrisiko KI-innovasjoner i Norge blir størst. Helsesektoren er blant sektorene der dette kan bli aktuelt, og hvor det allerede finnes en tverretattlig veiledningstjeneste innenfor helse som kan bygges videre på.²²¹

Siden sandkassene også skal bidra til å fremme innovasjon, foreslår vi å gi sandkassene et utvidet mandat. De bør også utforske og finne frem til innovative anvendelser for generativ kunstig intelligens av generell interesse for offentlige og private virksomheter. Det vil gjøre sandkassen til en nasjonal inkubator for KI, og vil kunne være spesielt nyttig for å finne fram til generelle anvendelser for offentlig sektor. Noe slikt finnes ikke i Norge i dag, men flere andre land prioriterer nå en nasjonal innsats for å finne, teste og innføre slike anvendelser:

- **Storbritannia.** Det nye kontoret «Incubator for AI» med 70 ansatte, har begynt å pilotere bruk av generativ kunstig intelligens på administrative områder som å analysere høringsinnspill.²²² Utprøving skjer i regi av kontoret, og dersom de lykkes rulles anvendelser ut til forvaltningen.
- **USA.** Department of Homeland Security skal bemanne et KI-korps med 50 KI-eksperter. Tre pilotprosjekter testes ut: å oppsummere rapporter, lære opp ansatte som jobber med immigrasjon, og å hjelpe lokalsamfunn med å sende inn tilskudd og få finansiering.²²³
- **Singapore.** Gjennom et såkalt Trailblazer-initiativ har Singapore støttet utforskningen av 100 generative KI-prosjekter i løpet av 100 dager.²²⁴ Slik ønsker myndighetene å stimulere til økt eksperimentering med og utnyttelse av teknologien.

²²¹ Helsedirektoratet 2024, [Tverretattlig veiledningstjeneste](#)

²²² Financial Times 2024, [UK government to trial 'red box' AI tools to improve ministerial efficiency](#)

²²³ Axios 2024, [Axios AI+](#)

²²⁴ The Straits Time 2024, [More than 100 solutions developed as part of generative AI initiative](#)

Forslag til tiltak

For å stimulere til innovasjon, møte kravene i lovverket, og raskt finne nyttige og lavrisiko anvendelser av generativ kunstig intelligens i Norge, anbefaler vi:

- Å etablere en ny nasjonal regulatorisk sandkasse for kunstig intelligens lagt til Digitaliseringsdirektoratet som en egen enhet med dedikerte ressurser. Behovet for sektorspesifikke sandkasser, som i helse eller utdanning, bør løpende vurderes.
- Å tilrettelegge for at sandkassen kan trekke på kompetanse fra relevante miljøer. Dette kan bety å etablere et nettverk av ressurser som bidrar med fagkunnskap til prosjekter etter behov. Digitaliseringsrådets organisering kan tjene som eksempel.
- At sandkassen gjennomfører et prosjekt som skal identifisere bruksområder for generativ kunstig intelligens i forvaltningen. Prosjektet bør utvikle, teste ut og evaluere KI-løsninger og bistå offentlige virksomheter som ønsker å ta i bruk egnede og forsvarlige løsninger.

3. SKAFF FORVALTNINGEN EN PROBLEMLØSER

Anbefaling: *Det er behov for felles og løpende håndtering av usikkerhet og problemstillinger knyttet til generativ kunstig intelligens blant offentlige virksomheter. Derfor bør det etableres et nettverk der slike problemstillinger kan diskuteres og avklares. Nettverket bør etableres som en arbeidsgruppe i regi av topplederforumet Skate.*

Usikkerhet begrenser i dag offentlig sektors bruk av kunstig intelligens, og bruken er lav.²²⁵ De fleste ledere i offentlig sektor vil bruke teknologien, men opplever hindringer som mangel på kompetanse, rettslig uklarhet, usikkerhet rundt hva teknologien kan brukes til, etiske usikkerheter og frykt for å gjøre feil.²²⁶

Et nytt markedstilsyn og en eller flere sandkasser vil veilede om KI-forordningen og gjøre det mulig å teste ut løsninger i praksis. Men der sandkassene vil ta for seg enkeltprodukter og -løsninger, vil behovet for offentlig forvaltning i større grad være løpende avklaringer om godkjent risikonivå og bestep praksis.

²²⁵ Eksempelvis Nora.ai og Digitaliseringsdirektoratets oversikt

²²⁶ Digi.no 2023, [*Undersøkelse: Et flertall av ledere i offentlig sektor vil ta i bruk kunstig intelligens*](#)

Dette kan for eksempel handle hvor pålitelig og forutsigbar en generativ KI-assistent på en offentlig nettside må være for å kunne settes i drift og tas i bruk av innbyggerne. Slike problemstillinger krever lik praksis på tvers. Den enkelte virksomhet kan være usikker på vurdering, tilnærming og håndtering, og det kan være uheldig dersom ulike offentlige virksomheter velger ulike løsninger.

Én måte å adressere dette på, og samtidig bidra til en lik håndtering på tvers av forvaltningen, kan være formelle arenaer der utviklere, jurister og fagpersoner løpende kan diskutere og enes om felles forståelse og lik håndtering. Felles avklaringer kan i tillegg gi det offentlige drahjelp med å få tatt teknologien i bruk.

Et miljø som kan ivareta behovet er topplederorganet Skate, som gir råd til Digitaliseringsdirektoratet og digitaliseringsministeren. Skate skal bidra til samordnet digitalisering av offentlig sektor og understøttes av et arbeidsutvalg med representanter fra elleve statlige virksomheter og tre kommunerepresentanter.

Skate bør spesifikt adressere utfordringer innen *generativ* kunstig intelligens. Miljøet samler sentrale offentlige aktører, har et tett samarbeid med Digitaliseringsdirektoratet, og kan bli et viktig supplement til regulatoriske sandkasser.

Forslag til tiltak

For å redusere usikkerhet og i større grad bidra til felles håndtering av løsninger for generativ kunstig intelligens i offentlig forvaltning, anbefaler vi:

- Å etablere et nettverk i regi av Skate for å adressere felles utfordringer, i form av en dedikert arbeidsgruppe for generativ kunstig intelligens.
- At arbeidsgruppen identifiserer tekniske, juridiske og etiske problemstillinger som begrenser bruk av teknologien, med mål om å enes om felles risikoforståelse, tilnærminger og bestep praksis i forvaltningen.
- At arbeidsgruppen gir råd om og løfter etiske og prinsipielle problemstillinger og verdivalg av større samfunnsmessig betydning, inkludert behovet for eventuelle justeringer i lovverk.

Majoriteten av de store språkmodellene som brukes i Norge er utviklet av store amerikanske teknologiselskaper. Det innebærer at modellene er best i engelsk, og har et amerikansk verdi- og referansegrunnlag.

Et sentralt spørsmål er hvordan vi kan sikre at KI-tjenester som blir viktige i Norge baseres på språkmodeller som behersker norske språk og reflekterer norsk samfunnsliv, kultur og verdier. Ett grep er å bygge en infrastruktur for utvikling og forvaltning av språkmodeller som mindre modeller kan bygge på.

4. SIKRE TILGANG TIL NORSKE KVALITETSDATA

Anbefaling: *Det kreves tilgang på betydelige mengder norske kvalitetsdata for at språkmodeller skal fungere best mulig på norsk. Nasjonalbiblioteket bør få mandat til å tilgjengeliggjøre slike data og kompensere rettighetshaverne. Også data forvaltet av Arkivverket og Lovdata må kunne utnyttes. Norske kvalitetsdata må forvaltes strategisk og tilgang må ikke gis ukritisk til alle.*

Det er flere utfordringer knyttet til norske kvalitetsdata som kan brukes for å trene opp store språkmodeller. Det gjelder særlig;

- Tilgang til tilstrekkelige mengder med norske kvalitetsdata.
- Tilgjengeliggjøring av slike data til de som har behov for dem.
- Å begrense tilgang for enkelte aktører.

Den første utfordringen handler om å sikre tilstrekkelige mengder med norske kvalitetsdata, både tale, bilder, tekst og video. Skal språkmodellene fungere godt på norske språk, med kjennskap til norsk kulturarv, historie og ytringskultur, må mesteparten av språkdataene være norske.

Det finnes i utgangspunktet mye treningsdata av høy kvalitet i Norge. Gjennom pliktavleveringsloven mottar Nasjonalbiblioteket alt materiale som publiseres, og siden 2006 har de arbeidet med å digitalisere materialet. Ni av ti bøker og 80 prosent av avisutgavene som er gitt ut i Norge, samt hele NRKs radioarkiv, er digitalisert. Den digitaliserte samlingen er blant verdens største, med over 6,9 milliarder ord.²²⁷ Encyclopedia Britannica har bare 40 millioner ord.

²²⁷ Idunn 2023, [Kunstig Intelligens og Nasjonalbiblioteket](#)

Det er imidlertid behov for enda mer. Manglende tilgang på data kan skyldes at potensialet ikke er vurdert, eller at rettighetene til slike data er regulert i avtaler som hindrer allmenn tilgang. Dette gjelder blant annet tilgang til norske rettsavgjørelser, som Lovdata har enerett på. Brukeravtalen med Lovdata presiserer at innholdet ikke skal brukes i språkmodeller.²²⁸ Årsaken er at Lovdata driftes av en privat stiftelse, som skal være selvfinansierende.²²⁹ Dette er imidlertid en svært verdifull kilde for å trene opp norske språkmodeller i å forstå norsk jus.

Også Arkivverket sitter på mye relevante data som ikke finnes hos Nasjonalbiblioteket. Dette kan være en viktig kilde for opptrening av norske språkmodeller. Her finnes imidlertid mye persondata og sensitive data, som må håndteres på forsvarlig måte.

Den andre utfordringen er å *tilgjengeliggjøre* data som allerede finnes for dem som til slutt skal lage språkmodeller med dem – forskningsinstitusjoner, virksomheter og vanlige brukere. Utfordringen handler om flere ting, som alle krever dedikerte ressurser for å løses:

- **Bearbeiding av data.** Data må sammenstilles og bearbeides til renvaskede datasett for å bli brukt til å utvikle språkmodeller. Dette er en teknisk og tidkrevende oppgave, som krever ressurser.
- **Opphavsrett.** Bøker, radiosendinger osv. er som regel beskyttet av opphavsrett. Nasjonalbiblioteket med flere utreder nå, på oppdrag fra Kultur- og likestillingsdepartementet, hvordan rettighetshavere kan kompenseres når åndsverk brukes til å utvikle språkmodeller.²³⁰
- **Personopplysninger.** Åndsverk kan bestå av personopplysninger. Skal dataene brukes til å utvikle språkmodeller kan det kreves juridiske vurderinger av hvordan lovlig tilgang kan gis etter personopplysningsloven.
- **Krevende å utnytte.** Det kan være en juridisk og teknisk krevende oppgave for virksomheter å nyttiggjøre seg av datasett og språkmodeller. De vil trenge veiledning og brukerstøtte for å lykkes.

Den tredje utfordringen handler om å begrense tilgang til data som det ikke er ønskelig at distribueres fritt. Et frislipp av norske data kan føre til at

²²⁸ Lovdata 2024, [Brukeravtale, Lovdata Pro](#)

²²⁹ Aftenposten 2024b, [KI kan gi gratis advokatråd, men norsk stiftelse står i veien: – Rettsstaten forvitrer](#)

²³⁰ Nasjonalbiblioteket 2023b, [Nasjonalbiblioteket skal lede nytt forskningsprosjekt om kunstig intelligens](#)

internasjonale teknologiselskaper, som Google og OpenAI, bruker dataene til å utvikle sine egne modeller. Norske kvalitetsdata er en viktig ressurs som gir industrien et konkurransefortrinn, og kontroll må sikres gjennom strategisk forvaltning.

Nasjonalbiblioteket blir en sentral institusjon når flere av disse utfordringene skal løses. Nasjonalbiblioteket mottar og forvalter mye data, og bearbeider og tilgjengeliggjør allerede datasett gjennom Språkbanken. De utvikler og tilgjengeliggjør også i en viss utstrekning egne språkmodeller i AI-laben sin, delvis som en motvekt til et fullstendig frislipp av norske data.

Nasjonalbiblioteket har imidlertid verken ressurser eller mandat til å ivareta så omfattende og komplekse arbeidsoppgaver over tid. Regjeringens oppdrag til Nasjonalbiblioteket om å undersøke verdien av opphavsrettslig beskyttet materiale og eventuell kompensasjonsordning for rettighetshavere, er en viktig start, men et foreløpig avgrenset forskningsprosjekt. Hva som skjer videre, er uklart.

Norge bør peke ut en institusjon som fremover får blant sine hovedoppgaver å samle inn, forvalte, bearbeide og tilgjengeliggjøre data til opptrening av språkmodeller i Norge. Det bør være Nasjonalbiblioteket.

Forslag til tiltak

For å sikre stabil tilgang til kvalitetsdata for utvikling av språkmodeller, kompensasjon til opphavere, og for å unngå et ukontrollert frislipp av norske data, anbefaler vi flere grep;

- Kultur- og likestillingsdepartementet og Justisdepartementet må legge til rette for at data som forvaltes av henholdsvis Arkivverket og Lovdata, herunder norske rettsavgjørelser, kan utnyttes til å trene opp norske språkmodeller.
- Å se på norske kvalitetsdata som en ressurs, og forvalte disse strategisk. Det bør nøye vurderes hvilke data som gjøres tilgjengelige.

I tillegg anbefaler vi å gi Nasjonalbiblioteket et nytt og utvidet KI-mandat. Det innebærer blant annet:

- Å utvikle og tilgjengeliggjøre datasett for opptrening og tilpasning av språkmodeller.

- Å prioritere datasett med høy samfunnsmessig verdi, men med lav kommersiell interesse, som tekst og tale fra samiske språk og norske minoritetsspråk, samt fra barn og personer med talevansker.
- Å utforme og forvalte kompensasjonsordninger for bruk av åndsverk til å utvikle språkmodeller (fra utredning til gjennomføring).
- Å inngå juridiske avtaler med virksomheter om kompensasjon og bruk av deres data til å trene opp språkmodeller.
- Å tilby teknisk og juridisk veiledning til forskningsinstitusjoner og virksomheter om hvordan de kan utvikle og tilpasse språkmodeller med data som Nasjonalbiblioteket forvalter.
- Å utvikle egne språkmodeller for særskilte formål når det ikke er mulig eller hensiktsmessig å tilgjengeliggjøre selve dataene til forskningsinstitusjoner og virksomheter.
- Å utstyre Nasjonalbiblioteket med teknisk og juridisk kompetanse, ressurser til å utvikle og forvalte datasett og språkmodeller, i tillegg til økonomiske ressurser til kompensasjonsordninger for åndsverk.

5. INVESTER I REGNEKRAFT SOM NASJONAL INFRASTRUKTUR

Anbefaling: *Behovet for regnekraft vil øke betraktelig både for forskningsinstitusjoner, offentlige og private virksomheter de neste årene. Derfor må det bygges flere og kraftigere superdatamaskiner i Norge. Regnekraft bør satses på som nasjonal infrastruktur og ansvaret for styring og finansiering bør plasseres hos Samferdselsdepartementet eller Nærings- og fiskeridepartementet.*

Alle som vil utvikle, tilpasse og drifte store språkmodeller eller som vil bygge nye generative KI-tjenester på toppen av disse, trenger tilgang på regnekraft. Kunstig intelligens blir nå relevant for stadig flere forskningsfelt og offentlige og private aktører. Dermed vil også etterspørselen etter regnekraft øke kraftig.

Fremover bør regnekraft betraktes som nasjonal infrastruktur, og tilgang sikres deretter. Superdatamaskiner vil bli stadig viktigere for avansert forskning og innovasjon. De vil også bli sentrale i et beredskapsperspektiv, for å drifte samfunnskritiske KI-baserte IT-systemer i eksempelvis forvaltningen, helse- eller banksektoren. I tillegg blir kompetanse og erfaring med superdatamaskiner en forutsetning for å bygge opp og skalere en konkurransedyktig norsk KI-industri.

Dagens park av superdatamaskiner i Norge består av maskinene Betzy, Fram og Saga,²³¹ og en del av den felleseuropeiske finske maskinen LUMI.²³² Det statlige selskapet Sigma2 forvalter tilgang til ressursene. I tillegg har Universitetet i Oslo²³³ og NTNU²³⁴ tilgang på egne mindre maskiner, samt europeiske superdatamaskiner som MareNostrum i Spania gjennom EU-prosjekter de er med i. Dette er et godt utgangspunkt, men av flere grunner ikke tilstrekkelig på sikt:

- **Maskinene er ikke kraftige nok.** Norges har begrenset med regnekraft sammenliknet med andre land.²³⁵ Sveriges største maskin Berzelius er drøyt 50 ganger kraftigere enn Norges kraftigste maskin Betzy.^{236,237} I Danmark skal et offentlig-privat samarbeid bygge superdatamaskinen Gefion, som vil bli blant verdens kraftigste.²³⁸
- **Innovasjoner stopper opp.** Betzy, Fram, Saga og LUMI er hovedsakelig forbeholdt avgrensede forskningsprosjekter.²³⁹ Fremover trengs også regnekraft til utvikling, drift, distribusjon og forskning på KI-tjenester basert alle språkmodellene som utvikles. Dette er ikke Sigma2 rigget for å tilby.
- **Finansieringen er begrenset.** Finansieringen av regnekraft er lite dynamisk. Regnekraft er definert som forskningsinfrastruktur, og konkurrerer om penger fra samme pott som forskningsmateriale som medisinsk utstyr og mikroskoper.²⁴⁰ Denne potten er begrenset til 200 millioner kroner i året.²⁴¹

En kraftig oppskalering av norsk regnekraft er nødvendig. I USA, Frankrike og Storbritannia er regnekraft i større grad industripolitikk, noe som reflekteres i pengebruken. Storbritannia vil bruke 900 millioner pund på superdatamaskiner fremover.²⁴²

En styrket og strategisk satsing på regnekraft til forskningsinstitusjoner og universiteter må inngå i oppskaleringen. Det bør også prioriteres tilgang til

²³¹ Sigma2 2024a, [Resource Allocation Committee \(RFK\)](#)

²³² Lumi 2024, [infoside](#)

²³³ Universitetet i Oslo 2024b, [Fox: Tungregningsklynge for brukere av Educloud Research](#)

²³⁴ NTNU High Performance Computing 2024, [Idun – High Performance Computing](#)

²³⁵ Universitetet i Oslo 2019, [Behov og finansieringsstrategi for nasjonal e-infrastruktur for forskning perioden 2020 – 2030](#)

²³⁶ Sigma2 2024b, [Systems](#)

²³⁷ Linköpings universitet 2021, [Sveriges snabbaste superdator för AI är invigd](#)

²³⁸ Ritzau 2024, [Danmark bygger en af verdens kraftigste AI-supercomputere for at sætte skub i løsninger på samfundsmæssige udfordringer](#)

²³⁹ Sigma2 2024c, [Who can apply for resources](#)

²⁴⁰ Forskningsrådet 2023, [Norsk veikart for forskningsinfrastruktur 2023](#)

²⁴¹ Forskningsrådet 2024, [Midler til forskningsinfrastruktur av nasjonal viktighet](#)

²⁴² The Guardian 2023, [UK to invest £900m in supercomputer in bid to build own 'BritGPT'](#)

superdatamaskiner for offentlige virksomheter, som av sikkerhetshensyn ikke alltid kan kjøpe regnekraft fra private leverandører som Amazon Web Services. I tillegg må det tilrettelegges for regnekraft til drift, videreutvikling og distribusjon av KI-tjenester basert på språkmodeller utviklet i forskningssektoren.

Regnekraften kan sikres i nasjonal regi, gjennom offentlig-private samarbeid, eller kjøpes fra private leverandører når det er hensiktsmessig og forsvarlig. Det bør etterstrebes å vurdere behovet løpende, tatt i betraktning regnekraftens høye forbruk av energi og knappe ressurser som vann og mineraler.²⁴³

Forslag til tiltak

For å sikre nødvendig tilgang til regnekraft for forskningsformål, utvikling og forvaltning av språkmodeller og generative KI-tjenester i Norge, anbefaler vi:

- Å bygge flere og kraftigere superdatamaskiner i Norge.
- Å satse på regnekraft som nasjonal infrastruktur. Ansvar for strategisk styring og løpende finansiering av regnekraft bør plasseres hos Samferdselsdepartementet eller Nærings- og fiskeridepartementet.
- Å tilrettelegge for at offentlige virksomheter enklere kan søke Sigma2 om tilgang til å tilpasse og finjustere språkmodeller til sine behov.

6. TILBY NORSKE SPRÅKMODELLER SOM NASJONAL FELLESTJENESTE

Anbefaling: *Forvaltningen, universitetssektoren og KI-industrien vil trenge stabil og kvalitetssikret tilgang på norskutviklede språkmodeller for å utvikle og levere konkurransedyktige digitale tjenester i årene fremover. Språkmodeller bør derfor utvikles og driftes som en nasjonal fellesløsning og infrastruktur tjeneste på linje med Altinn, ID-porten og Folkeregisteret.*

Tilbudet av internasjonale språkmodeller øker, ikke minst de markedsledende modellene fra amerikanske teknologiselskaper. Selv om det er positivt med et bredt og variert utvalg er det flere grunner å sikre en viss grad av nasjonal kontroll over modellene:

- **Suverenitet og beredskap.** Å utelukkende basere seg på markedsledende utenlandske språkmodeller er sårbart, og vil gjøre

²⁴³ IEA 2024, [Electricity 2024 - Analysis and forecast to 2026](#); Nature 2024, [Generative AI's environmental costs are soaring – and mostly secret](#)

det krevende å tillate samfunnskritisk infrastruktur som banksystemer eller offentlige tjenester å bygge på teknologien.²⁴⁴

- **Språklig kvalitet og kulturell autonomi.** Modellene holder ikke høy nok kvalitet på norske språk og dialekter. Siden treningsdataene i stor grad er basert på engelsk språk og amerikanske kilder, preger dette språkferdigheter, kultur- og referansegrunnlag. Det er uheldig når modellene skal brukes i norsk forvaltning, helsevesen eller skole.
- **Forklarbarhet og ansvarlighet.** Uten innsikt i modellens data, design og størrelse, blir det krevende å kvalitetssikre og forklare hvordan de fungerer. Det kan handle om sannsynlighet for diskriminering, å forklare automatiserte avgjørelser som modeller er involvert i, eller hvor energi- og ressurskrevende de er.
- **Innovasjon og konkurransefortrinn.** God tilgang på kvalitetssikrede norske språkmodeller kan gjøre det enklere for norske utviklere av å konkurrere med tilbydere av utenlandske markedsledende tjenester og tilby generative KI-tjenester på norsk.

Det er med andre ord flere gode grunner til å utvikle store norske språkmodeller helt fra bunnen av, hvor vi selv har kontroll på teknologien, datagrunnlaget og opptreningen. Flere land har allerede valgt en slik strategi, herunder:

- **Sverige.** I Sverige har AI Sweden, et offentlig-privat samarbeid, lansert språkmodellen GPT-SW3 trent på svensk tekst.²⁴⁵ Både enkeltpersoner og offentlige og private virksomheter kan få tilgang.
- **Danmark.** Rundt 20 offentlige og private virksomheter samarbeider i konsortiet Danske sprogmodeller om å dele data for å utvikle og finne bruksområder for en stor dansk språkmodell.²⁴⁶
- **Spania.** Myndighetene vil bygge en modell trent opp på spansk, katalansk, baskisk og galicisk, for å motvirke amerikansk dominans og sikre språklig kvalitet på tjenester i den spansktalende verden.²⁴⁷
- **Tyskland.** Tyske virksomheter har bidratt med 500 millioner dollar til å utvikle språkmodellen Luminous.²⁴⁸ I tillegg finansierer tyske myndigheter OpenGPT-X, en stor modell trent på flere språk.²⁴⁹

²⁴⁴ Store Norske Leksikon, [digital infrastruktur](#)

²⁴⁵ Linköping science park 2023, [Open release of the first large Nordic language model GPT-SW3](#)

²⁴⁶ Dansk Erhverv 2024, [Fremtidens danske sprogmodel: Kom med i det nye AI-initiativ](#)

²⁴⁷ Sifted 2024, [Spain to develop open-source LLM trained in Spanish, regional languages](#)

²⁴⁸ Aleph Alpha API 2024, [What is Luminous?](#)

²⁴⁹ OpenGPT-X 2024, [OpenGPT-X](#)

- **Storbritannia.** Regjeringens «Frontier AI Task Force» vurderer hvordan de kan utvikle britiske språkmodeller.²⁵⁰ Det kan ende i en «BritGPT».

Også i Norge er initiativ i gang for å utvikle norske språkmodeller fra bunnen av, men foreløpig avgrenset til universitetssektoren og forskningsmiljøer. Sommeren 2023 lanserte forskningssenteret NorwAI språkmodellen NorGPT, hovedsakelig trent opp på norske data. I tillegg har forskningssamarbeidet NORA søkt Forskningsrådet om 212 millioner kroner til å bygge en stor norsk språkmodell, som blant annet skal brukes innen helse- og utdanningssektoren.²⁵¹

Imidlertid har universitetene verken mandat eller ressurser til å utvikle, drifte, forvalte eller distribuere tilgang til språkmodeller i Norge i stor skala, ikke minst en satsing på store norsk språkmodeller bygget fra bunnen av. Dersom vi skal ha tilgang til slike modeller må ansvaret flyttes ut av universitetssektoren.

Det finnes også modeller som er trent opp på norske data, men som er basert på store internasjonale grunnmodeller. Dette gjelder eksempelvis NB-Whisper fra Nasjonalbiblioteket, basert på OpenAIs Whisper,²⁵² RuterGPT fra kollektivtransportsekskapet Ruter, basert på LLaMA2 fra Meta²⁵³ og NorskGPT fra selskapet Bineric, basert på blant annet Mistral.²⁵⁴ Slike modeller kan være gode alternativer til de internasjonale modellene i mange sammenhenger.

Fremover kommer vi til å se en rekke ulike modeller tilpasset norsk språk og norsk kontekst. De vil til sammen utgjøre viktige byggeklosser for norske digitale tjenester og systemer på alle samfunnsområder. Det vil derfor være naturlig å betrakte slike modeller som viktige elementer i Norges digitale infrastruktur.

Imidlertid er det tilfeldig om og hvordan slike modeller, både forhåndstrengte og tilpassede modeller, blir forvaltet og gjort tilgjengelig for norske virksomheter. Det er heller ikke enkelt å få oversikt over alternative modeller og deres egenskaper. Det finnes ingen sentral oversikt i Norge over modeller tilpasset norsk.

For å stimulere til spredning av norske modeller, og sikre en forsvarlig drift og forvaltning av disse, bør etableres en ny funksjon for drift og forvaltning av

²⁵⁰ Department for Science, Innovation & Technology 2023, [Frontier AI Taskforce: first progress report](#)

²⁵¹ NORA.ai 2024, [The NORA.LLM project - large language models as infrastructure](#)

²⁵² Nasjonalbiblioteket 2024, [Models](#)

²⁵³ Shifter 2024, [Nå kommer AI-teamet til Ruter med en slags «Ruter-GPT», for nesten all kollektivtransport i Norge.](#)

²⁵⁴ NorskGPT 2024, [NorskGPT](#)

norskutviklede språkmodeller. Denne kan organiseres som en nasjonal fellestjeneste.

Forslag til tiltak

For å beholde og styrke nasjonal kontroll og evne til å ta i bruk store språkmodeller og generativ kunstig intelligens, og for å bevare og styrke norsk språk, anbefaler vi:

- Å støtte bygging og videreutvikling av minst en stor norsk språkmodell fra bunnen av, utviklet med norske data og designvalg. Denne modellen kan utvikles i samarbeid mellom Nasjonalbiblioteket, Sigma2 og ledende forskningsmiljøer.
- Å sikre god drift, forvaltning og tilgang til utvalgte norske språkmodeller ved å definere disse som en nasjonal fellestjeneste, på linje med ID-porten, Altinn og Folkeregisteret. Både norske grunnmodeller og tilpassede språkmodeller kan inngå.
- Å organisere utviklingen og driften av fellestjenesten som et forvaltningsorgan, et statsforetak eller et offentlig-privat samarbeid.
- Å gi det nye organet mandat til å velge ut, tilgjengeliggjøre og vedlikeholde språkmodeller som skal inngå i fellestjenesten etter gitte kriterier og behov.

7. STILL KVALITETSKRAV TIL NORSKE SPRÅKMODELLER

Anbefaling: *Det vil finnes mange språkmodeller med ulike egenskaper. Forvaltningen, universitetssektoren og KI-industrien vil derfor trenge kunnskap for å ta informerte valg om hvilke språkmodeller de bør benytte seg av. En utpekt forskningsinstitusjon bør få i oppdrag å utvikle kvalitets- og evalueringskriterier for språkmodeller i Norge.*

For å fremme tillit og ansvarlig bruk av generativ kunstig intelligens og norske språkmodeller trengs åpenhet om modellenes prestasjoner, samt i hvilken grad de ivaretar norsk språk og norsk kontekst. Kvalitetsmåling bør derfor inngå som en del av utvikling og forvaltning av norske språkmodeller.

For å måle kvalitet trengs et sett med egnede kriterier. Kriteriene vil avhenge av modellens kontekst og formål. Kvalitetskrav til en modell for å stille medisinske diagnoser vil være annerledes enn kvalitetskrav til en chatbot hos en offentlig

aktør. Det vil også være behov for å vite hvor god språkmodellen er når det gjelder ferdigheter i norsk språk, kultur og samfunn.

Det er mulig å bygge på eksisterende evalueringsmodeller. Et eksempel er Hugging Face som har utviklet en evalueringsmodell kalt LLM Leaderboard.²⁵⁵ Her kan man evaluere ulike egenskaper og ferdigheter på et stort antall ulike oppgaver, eksempelvis naturfagsoppgaver på videregående nivå, modellens evne til «sunn fornuft» eller modellens tilbøyelighet til å generere usannheter.

Danske forskere har utviklet ScandEval for å sammenlikne og evaluere skandinaviske språkmodeller.²⁵⁶ BIG-bench samarbeidet inneholder 204 oppgaver,²⁵⁷ og er blitt brukt til å evaluere noen av de største språkmodellene, inkludert hvor gode de er i andre språk enn engelsk. HELM-rammeverket kan evaluere språkmodeller ut fra ulike scenarier som i hvilken grad modellen kan gjøre juridiske vurderinger, dens leseforståelse eller evne til å løse matematiske problemer.²⁵⁸

Forslag til tiltak

For at virksomheter skal kunne evaluere og velge språkmodeller egnet for sine behov og norsk kontekst, anbefaler vi:

- Å utvikle evalueringskriterier for språkmodeller. Kriteriene må kunne brukes til å måle modellenes språkferdigheter og domenekunnskap, og forståelse for kultur, samfunn og historie. De kan også inkludere kriterier for effektivitet når det gjelder for eksempel energibruk og utslipp.²⁵⁹
- Å stille krav om at utviklere gjennomfører evalueringer etter gjeldende kriterier, og åpent og offentlig rapporterer om hvordan modellene deres presterer i henhold til evalueringskriteriene.
- Å gi en egnet forskningsinstitusjon i oppdrag å utvikle kriterier og prosedyrer for evalueringen, i tråd med bestemmelsen om nasjonale fellesoppgaver i Universitets- og høyskoleloven.
- Å gi organet som drifter fellesløsningen for språkmodeller i oppgave å forvalte evalueringskriteriene, vise frem hvordan ulike modeller presterer i henhold til kriteriene, og tilby veiledning til virksomheter.

²⁵⁵ HuggingFace 2024b, [Open LLM Leaderboard](#)

²⁵⁶ Nielsen 2023, [ScandEval: A Benchmark for Scandinavian Natural Language Processing](#)

²⁵⁷ Srivastava m.fl 2023, [Beyond the Imitation Game Benchmark](#)

²⁵⁸ CRFM 2024, [Holistic Evaluation of Language Models \(HELM\)](#)

²⁵⁹ Schwartz, Doge, Smith & Etzioni 2019, [Green AI](#)

STYRKE DEN DIGITALE MOTSTANDSKRAFTEN

En åpen og opplyst offentlig samtale, forutsetter at det er enkelt å stole på og finne frem til sann informasjon. Generativ kunstig intelligens har gjort dette vanskeligere, og samtidig økt risikoen for manipulasjon og svindel. Utviklingen krever at befolkningens digitale motstandskraft styrkes, på flere fronter.

8. FASTSETT REGLER FOR MERKING AV KI-GENERERT INNHOLD

Anbefaling: *Falskt og kunstig generert innhold som fremstår ekte og autentisk spres nå raskt på nett. Norske myndigheter bør utarbeide retningslinjer for åpenhet og merking av KI-generert innhold. Aktører som Medietilsynet og Forbrukertilsynet kan få mandat til å følge opp dette.*

Det er nå umulig å skille menneskeskapt og maskinskapt innhold fra hverandre.²⁶⁰ For å bevare befolkningens tillit til at informasjon på nett er ekte kan flere typer merkingstiltak være aktuelle. Et eksempel kan være utvikling og innføring av bransjestandarder for detektering og merking av KI-generert innhold på digitale plattformer og i mediene. Det kan også være innføring av juridiske krav til merking og åpenhet i spesifikke situasjoner. I Norge er Forskrift om merking av retusjert reklame fra 2022 et eksempel på sistnevnte.²⁶¹

Regjeringen ønsker å merke KI-generert innhold.²⁶² Spørsmålet er hva som skal merkes, hvordan, og hvilke andre tiltak som trengs.²⁶³ En politikk for merking må ses i sammenheng med nye krav til merking på digitale plattformer som allerede finnes eller er på vei:

- Forordningen for digitale tjenester pålegger selskaper som Meta, Google og X å kartlegge og fatte tiltak mot spredning av desinformasjon på deres plattformer.²⁶⁴ Dette inkluderer KI-generert desinformasjon.

²⁶⁰ Miller m.fl 2023, [AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones](#)

²⁶¹ Lovdata 2022, [Forskrift om merking av retusjert reklame](#)

²⁶² TV2 2023, [Kunstig intelligens: Regjeringen vil merke KI-generert innhold: - Det er en fare](#)

²⁶³ Teknologirådet 2023, [Generativ kunstig intelligens og ytringsfrihet](#)

²⁶⁴ European Commission 2024c, [The Digital Services Act package](#)

- KI-forordningen stiller flere relevante krav til merking av og åpenhet om bruk av KI-generert innhold.²⁶⁵ EUs nye KI-kontor skal tilrettelegge for å utvikle beste praksis for å merke dette innholdet.²⁶⁶
- Et nytt EU-regelverk for politiske reklame stiller krav om åpenhet om hvem som finansierer slik reklame, og med hvor mye penger.²⁶⁷

Før slettet teknologiselskapene KI-generert innhold. Nå innfører de i stedet krav om merking.²⁶⁸ YouTube skal merke KI-genererte videoer automatisk, og informere brukerne om hva som er KI-generert.²⁶⁹ Merkingen skal være ekstra synlig på videoer om sensitive tema som valg, konflikter og helserelaterede spørsmål. Meta krever at KI-genererte bilder, videoer og lydklipp skal merkes, enten av brukerne selv eller automatisk av Metas egne KI-detekteringsverktøy.²⁷⁰

Merking av KI-generert innhold kan bidra til å sikre tilliten til informasjon, men én enkelt løsning for merking eller ett selskaps tiltak for å håndtere KI-generert innhold vil trolig ikke være tilstrekkelig i seg selv. Effektiv og nyttig merking for redaktørstyrte medier er ikke nødvendigvis like tilpasset merking av innhold på sosiale medier. Flere løsninger og tekniske verktøy for detektering, merking og beskyttelse av digitalt medieinnhold må fungere side om side.

Å finne gode måter å håndheve kravene i de nye regelverkene på blir også viktig. For eksempel finnes ingen omforent definisjon av KI-generert innhold, og om det inkluderer KI-generert tekst redigert av mennesker i etterkant. Det er også utfordrende å få verktøy som kan oppdage KI-generert innhold til å fungere godt på norsk.

Likevel kan tekniske verktøy og krav til merking fungere på sikt. Flere initiativer er på gang. Medieklyngen Media City Bergen utvikler i Project Origin standarder for å verifisere digitalt innhold gjennom bruk av krypterte metadata sammen med BBC og The New York Times.²⁷¹ BBC bruker allerede verktøyet.²⁷² Prosjekt Reynir, den norske grenen, har som mål å få på plass en «banebrytende, skalerbar løsning for bekjempelse av desinformasjon» innen 2026.²⁷³

²⁶⁵ AI Act 2021, [AI Act art. 52](#)

²⁶⁶ European Commission 2024a, [European AI Office](#)

²⁶⁷ European Parliament 2023b, [Political advertising: deal on new measures to crack down on abuse](#)

²⁶⁸ The Verge 2024b, [Meta will require labels on more AI-generated content](#)

²⁶⁹ Google 2023b, [Our approach to responsible AI innovation on YouTube](#)

²⁷⁰ Meta 2024, [How AI-generated content is identified and labeled on Meta](#)

²⁷¹ Project Origin 2024, [Project Origin](#)

²⁷² BBC 2024, [Embedding transparency, enhancing trust](#)

²⁷³ Media City Bergen 2024, [Prosjekt Reynir](#)

Forslag til tiltak

For å gjøre det enklere for befolkningen å gjenkjenne ekte og autentisk digitalt innhold på nett, anbefaler vi:

- Å prioritere rask innføring av de nye relevante EU-lovene i norsk lov.
- Å utarbeide retningslinjer for åpenhet og merking av KI-generert innhold. De bør beskrive hvilke typer innhold og KI-tjenester som skal merkes, og hvordan.
- Å gi relevante sektoraktører, som Medietilsynet og Forbrukertilsynet, mandat til å følge opp utviklingen.
- Å utnytte adgangen i Forordningen for digitale tjenester til å få innsyn i og ettergå effektiviteten til plattformselskapenes egen merking av KI-generert innhold.

9. LAG SPILLEREGLER FOR GENERATIV KUNSTIG INTELLIGENS I VALGKAMP

Anbefaling: *Generativ kunstig intelligens blir i økende grad brukt som valgkampverktøy og til å påvirke demokratiske valg. Før stortingsvalget i 2025 bør det utvikles tverrpolitiske spillereglene for bruk av generativ kunstig intelligens i valgkamp for å bevare tilliten til de politiske partiene og valginstitusjonen.*

I 2024 er det valg i USA, EU, India, Sør-Afrika, Storbritannia og Tyskland, og i 2025 er det stortingsvalg i Norge. Stortingsvalget finner sted etter at teknologien har blitt utprøvd som valgverktøy i 2024, men likevel før kravene til merking av KI-generert innhold i KI-forordningen trer i kraft i 2026.

Chatboter kan programmeres til å representere en politikers syn og svare på velgeres spørsmål.²⁷⁴ Persontilpassede KI-genererte videoer der partiledere argumenterer for at nettopp du med din kampsak bør stemme på dem, er også mulig å lage. Nylig lagde Dansk Folkeparti en KI-generert video av statsminister Mette Fredriksen som sier hun ønsker å avskaffe helligdager og innføre den muslimske høytiden id al-fitr som landets eneste nasjonale helligdag.²⁷⁵ Den

²⁷⁴Associated Press 2023, [Mayor Suarez launches an artificial intelligence chatbot for his presidential campaign](#)

²⁷⁵NRK 2024d, [Dansk Folkeparti i Danmark brukar «deepfake» av statsminister Mette Frederiksen i video](#)

dype forfalskningen var ment som en politisk satire og ble merket som KI-generert, men skapte likevel debatt om bruk av dype forfalskninger i politikken.

Gitt sprengkraften og tilgjengeligheten til generativ kunstig intelligens, kan det bli nødvendig med tiltak før stortingsvalget i 2025. Teknologirådet foreslo i 2019 spilleregler for de politiske partiene i den digitale valgkampen, etter Brexit og Cambridge Analytica-skandalen.²⁷⁶ Disse inkluderte bedre merking av politiske annonser på nett, samt offentliggjøring av detaljerte regnskap over hvor mye penger og hvilke data partiene bruker i digital markedsføring.

For å håndtere de nye KI-verktøyene i valgkampen kan disse spillereglene brukes som utgangspunkt og videreutvikles. Forutsigbare og klare spilleregler kan gjøre det enklere for befolkningen og faktasjekkere å skille partienes eget innhold fra det andre har laget, og bidra til at tilliten og troverdigheten til partiene og valginstitusjonen opprettholdes.

Forslag til tiltak

For å sikre ansvarlig bruk av generativ kunstig intelligens blant politiske partier i valgkamp og motvirke tap av tillit til demokratiske institusjoner, anbefaler vi:

- Å utvikle tverrpolitiske spilleregler for bruk av generativ kunstig intelligens i valgkamp. De kan utvikles av de politiske partiene eller i regi av Kommunal- og distriktsdepartementet som nasjonal valgmyndighet.
- At enkelte typer bruk av kunstig intelligens i valgkamp begrenses eller forbyes, som dype forfalskninger av politikere, KI-genererte stemmeklipp og fotorealistiske bilder og videoer.
- Å opprette en database over politisk reklame vist på nett og i sosiale medier i forbindelse med stortingsvalget i 2025, og som inneholder informasjon om pengebruk og hvilke data reklamen er basert på.

10. OPPSKALER INNSATSEN FOR FAKTASJEKING

Anbefaling: Den raske spredningen av KI-generert falskt og troverdig innhold på nett har økt behovet for faktasjekking av det offentlige ordskiftet. Innsatsen for faktasjekking i Norge bør oppskaleres, enten ved å styrke Faktisk.no eller ved å opprette et nytt uavhengig nasjonalt senter for

²⁷⁶ Teknologirådet 2019a, [Valgkampen: Behov for nye spilleregler](#)

kildekritikk. I tillegg bør det satses på utvikling av tekniske verktøy og metoder for innholdsverifisering på norsk.

Generative KI-verktøy gjør det mulig å masseprodusere troverdig feil- og desinformasjon av høy kvalitet. Når volumet KI-generert feil- og desinformasjon på nett øker, forsterkes også behovet for å kvalitetssikre og faktasjekke digitalt medieinnhold – alt fra nyheter og offentlig debatt til bilder, videoer og lydklipp på sosiale medier. Å utvikle og ta i bruk nye metoder for verifisering og kvalitetssikring er nødvendig for å holde tritt med den teknologiske utviklingen.

Faktasjekking blir dermed en kritisk samfunnsfunksjon for å bygge motstandsdyktighet mot uønsket påvirkning og manipulering og styrke befolkningens kildekritiske kompetanse. Her spiller et mangfold av medieinstitusjoner med høy integritet og tillit en viktig rolle, sammen med uavhengige sivilsamfunnsorganisasjoner og andre instanser som bidrar til en åpen og opplyst debatt.

Faktisk.no, en ideell organisasjon eid av store norske medieinstitusjoner, jobber systematisk med faktasjekking av det offentlige ordskiftet i Norge. De har begrenset med ressurser, med 15 ansatte og to dedikerte faktasjekkere.²⁷⁷ Faktisk.no har 19 millioner kroner i budsjett,²⁷⁸ hvorav de 5 millionene som kommer fra statsbudsjettet er forbeholdt Faktisk.no sin skoleavdeling Tenk.²⁷⁹

Som en reaksjon på krigene i Ukraina og mellom Israel og Hamas, har Faktisk.no på en ad hoc-basis koordinert et nettverk av faktasjekkere under paraplyen Faktisk Verifiserbar.²⁸⁰ I 2022 samarbeidet 12 redaksjoner og 32 medarbeidere med å verifisere innhold om Ukraina til bruk for norske mediehus.²⁸¹ Slike satsninger krever ressurser og vil bli viktigere i tiden som kommer gitt hvor raskt og hvor mye falsk informasjon som spres i kriser, krig og konflikt.

I 2022 anbefalte Ytringsfrihetskommisjonen en satsning på opplæring, kurs og informasjonsdatabaser som kan brukes til faktasjekking, med særlig satsning på skolene, og videreutvikling av teknologiske verktøy for å avdekke desinformasjon.²⁸² Faktasjekking blir en stadig viktigere og mer ressurskrevende

²⁷⁷ Faktisk 2024, [Ansatte](#)

²⁷⁸ Brønnøysundregistrene 2023, [Årsregnskap regnskapsåret 2022 for 919036508](#)

²⁷⁹ Faktisk 2024, [Finansiering](#)

²⁸⁰ Verifiserbar 2024, [Verifiserbar](#)

²⁸¹ NTB 2023, [Faktisk Verifiserbar starter opp igjen: – Stort behov for å få verifisert bilder og videoer nå](#)

²⁸² NOU 2022: 9 [En åpen og opplyst offentlig samtale – Ytringsfrihetskommisjonens utredning](#)

oppgave, og er i dag underdimensjonert sett i forhold til de store verifiseringsutfordringene som nå blir forsterket av nye og mer avanserte KI-verktøy.

Forslag til tiltak

For å styrke faktasjekking av spredning av informasjon, nyhetsbildet og det offentlige ordsiftet i Norge, anbefaler vi:

- Å oppskalere innsatsen for faktasjekking i Norge, enten ved å styrke eksisterende fagmiljøer som faktisk.no eller ved å opprette et nytt uavhengig nasjonalt senter for kildekritikk.
- Å satse på utvikling av tekniske verktøy og metoder for å verifisere innhold på norsk.
- Å prioritere faktasjekking av krig og konflikt, helse, politiske valg og andre tematiske områder der spredning av feil- og desinformasjon kan utgjøre særlig skade.

11. OPPRETT ET PSYKOLOGISK FORSVAR

Anbefaling: *Fremveksten av generativ kunstig intelligens har økt risikoen for spredning av desinformasjon og informasjonspåvirkning. For å møte denne trusselen bør det etableres en nasjonal funksjon for å identifisere, rapportere, opplyse og gi råd om desinformasjonskampanjer og informasjonspåvirkning fra utenlandske aktører.*

Siden det blir lettere å lage feil- og desinformasjon vil det bli viktigere å systematisk og løpende kartlegge, forebygge og motvirke informasjonspåvirkning og desinformasjonskampanjer fra utenlandske aktører. Det inkluderer å følge med på trender innen produksjon, spredning og automatisering av desinformasjon.

I dag har Norge ingen sentral funksjon som følger med på og motvirker informasjonspåvirkning og desinformasjon fra utenlandske aktører. Flere instanser, som Totalberedskapskommisjonen fra 2023, har pekt på at den norske beredskapen mot desinformasjon er fragmentert.²⁸³ Forsvarets Forskningsinstitutt (FFI) har etterlyst avklaringer av roller, ansvar og myndighet for å beskytte Norge mot uønsket påvirkning.²⁸⁴

Teknologirådet har tidligere pekt på den svenske Myndigheten för psykologiskt försvar som modell for å håndtere og motvirke uønsket informasjonspåvirkning

²⁸³ NOU 2023: 17, [*Nå er det alvor – Rustet for en usikker fremtid*](#)

²⁸⁴ FFI 2021, [*Hvordan gjøre samfunnet mer robust mot uønsket påvirkning i sosiale medier*](#)

og desinformasjon.²⁸⁵ Det gjør også FFI og Totalberedskapskommisjonen.²⁸⁶ Myndigheten ble opprettet i 2022 og er underlagt regjeringen gjennom Forsvarsdepartementet.²⁸⁷ Den er likevel sivilt innrettet og frikoblet forsvarssektoren ettersom svenske etater er relativt frittstående sammenlignet med norske.²⁸⁸

Ansvarsoppgavene til det svenske psykologiske forsvaret innebærer blant annet å identifisere, analysere og motvirke uønsket informasjonspåvirkning og annen villende informasjon som rettes mot Sverige eller svenske interesser fra utenlandske aktører. De har også fått mandat til å bestille og formidle forskning på feltet, spre og bygge kunnskap i befolkningen, og koordinere samarbeid på tvers av sektorer og fagmiljø.²⁸⁹

Flere land har satt opp liknende funksjoner.²⁹⁰ Storbritannia opprettet i 2019 en *Counter Disinformation Unit* under departementet for vitenskap, innovasjon og teknologi. Den skal «sikre at myndighetene forstår omfanget av og rekkevidden til skadelig feil- og desinformasjon og iverksetter passende tiltak.»²⁹¹ Litauen har opprettet et nasjonalt krisehåndteringssenter som koordinerer de statlige aktørenes tiltak mot desinformasjon, underlagt det litauiske Forsvarsdepartementet.²⁹²

Våren 2025 vil regjeringen lansere en ny strategi for å styrke motstandskraften mot desinformasjon i samfunnet.²⁹³ I sammenheng med dette vil det være naturlig å vurdere hvordan innsatsen for å kartlegge og forebygge forsøk på utenlandsk desinformasjon og påvirkningskampanjer kan styrkes i Norge.

Forslag til tiltak

For å møte den økte trusselen for informasjonspåvirkning og spredning av feil- og desinformasjon, anbefaler vi:

²⁸⁵ Teknologirådet 2019b, [Manipulering av valg er den nye normalen](#)

²⁸⁶ FFI 2024, [Forsvar mot fremmedstatlige påvirkningsoperasjoner – etablering av funksjon i forsvarssektoren](#)

²⁸⁷ Myndigheten för psykologiskt försvar 2024, [Myndigheten för psykologiskt försvar](#)

²⁸⁸ FFI 2024, [Forsvar mot fremmedstatlige påvirkningsoperasjoner – etablering av funksjon i forsvarssektoren](#), se s. 27 for forklaring av hvordan det psykologiske forsvaret i Sverige er organisert.

²⁸⁹ Svensk författningssamling 2021, [Förordning med instruktion för Myndigheten för psykologiskt försvar](#)

²⁹⁰ OECD 2024, [Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity](#)

²⁹¹ GOV.UK 2023c, [Fact Sheet on the CDU and RRU](#)

²⁹² FFI 2024, [Forsvar mot fremmedstatlige påvirkningsoperasjoner – etablering av funksjon i forsvarssektoren](#)

²⁹³ Kultur- og likestillingsdepartementet 2024, [Om arbeidet med strategi for å styrke motstandskraften mot desinformasjon](#)

- Å etablere en nasjonal funksjon med mandat til å identifisere, rapportere, opplyse og gi råd om desinformasjonskampanjer og informasjonspåvirkning fra utenlandske aktører.
- Å utstyre funksjonen med nødvendig kompetanse og ressurser til å bestille og finansiere forskning og tekniske analyser, spre kunnskap i befolkningen og samarbeide med relevante aktører fra ulike sektorer.

12. STYRK ARBEIDET MED KI-SIKKERHET

Anbefaling: *Generativ kunstig intelligens skaper nye og uoverskuelige sikkerhetsrisikoer og har et stort skadepotensial. Derfor bør den nasjonale kapasiteten til å analysere KI-trusler og utvikle risikoscenarier formaliseres og styrkes. I tillegg bør myndighetene utvikle norske posisjoner for globale styringsmekanismer for kunstig intelligens, og sikre Norge plasser i de nye styringsorganene på EU-nivå.*

Utviklingen av kunstig intelligens går raskt og er uforutsigbar. Teknologien blir stadig kraftigere og mer avansert, noe som gjør det vanskeligere å overskue, håndtere og kontrollere teknologiutviklingen og potensielle skadevirkninger, enten det gjelder uønskede hendelser eller målrettet ondsinnet bruk. Dette inkluderer alt fra svekket kontroll over teknologien til utbredelse av avanserte autonome våpensystemer, cyberangrep og desinformasjonskampanjer.

Samtidig fører teknologiens mulige fordeler til at teknologiselskaper og stater, som USA og Kina, kappes om å utvikle kraftigst kunstig intelligens. Dette akselererer teknologiutviklingen, men øker også usikkerheten og konsentrerer makt, penger og kunnskap hos de selskapene som driver utviklingen fremover.

For å motvirke et kappløp, økt usikkerhet og hemmelighold, trengs globale kjøreregler for teknologiutviklingen som foreslått av FN-generalsekretær Guterres.²⁹⁴ Det er behov for åpenhet om teknologiens egenskaper, evne til å oppdage og håndtere ny risiko, og for analyser av teknologiens virkninger på samfunnet.

Norge bør derfor engasjere seg i utviklingen av nye internasjonale og globale styringsstrukturer for kunstig intelligens. Både i EU og FN skal det nå utvikles krav, standarder og mekanismer for alt fra høyrisiko KI-systemer til store språkmodeller. Gjennom deltakelse i disse prosessene kan Norge påvirke hvordan KI-sikkerhet og den globale styringen av kunstig intelligens blir utformet.

²⁹⁴ UN News 2023, [Guterres calls for AI 'that bridges divides', rather than pushing us apart](#)

Risikovurderinger og -håndtering av kunstig intelligens blir også et krav gjennom KI-forordningen. Loven stiller krav til sikkerhet og robusthet i høyrisiko-systemer. Det innebærer blant annet at slike systemer skal kunne motstå alt fra dataforgiftning av treningsdata til cyberangrep og manipulasjon.²⁹⁵

Parallelt med internasjonalt samarbeid må Norge styrke egen analysekapasitet og kompetanse på KI-trusler, risikoer og sårbarheter. Dette kan gjøres ved å utvikle risikoscenarioer og løpende vurdere eksisterende og kommende trusler. I møte med en kraftig teknologi med fremvoksende egenskaper og stort utfallsrom, vil risikoscenarioer være et særskilt egnet verktøy. I tillegg bør arbeidet med å utvikle og innføre tiltak for å styrke digital sikkerhet og beredskap prioriteres.

Flere aktører har tatt til orde for behovet for å styrke den digitale sikkerhetsberedskapen i Norge. Riksrevisjonen mener Norge ikke er godt nok beskyttet mot digitale angrep, og har anbefalt at arbeidet med å avdekke, håndtere og koordinere innsatsen mot alvorlige digitale hendelser i Norge blir styrket.²⁹⁶ Totalberedskapskommisjonen peker på at norske myndigheter i større grad må vurdere hvilke teknologier som anses å utgjøre en risiko for nasjonal sikkerhet.²⁹⁷

Til å løse disse utfordringene finnes det flere fagmiljø med relevant kompetanse på digital sikkerhet, blant annet Nasjonal Sikkerhetsmyndighet (NSM) og Direktoratet for samfunnssikkerhet og beredskap (DSB). NSM har foreslått opprettelsen av et innovasjonssenter for fremvoksende teknologier, som kunstig intelligens, kvanteteknologi og bioteknologi.²⁹⁸ DSB har vært pådriver for grundige risikoanalyser som er nødvendig for å bedre sikkerhetstilstanden, og for å utvikle og iverksette risikoreducerende sikkerhetstiltak.²⁹⁹

Oppsummert krever teknologiutviklingen og risikoen den medfører en oppskalert innsats for å forstå, kartlegge og vurdere nåværende og framtidig KI-risiko.

Forslag til tiltak

For å møte KI-risiko og ta en aktiv rolle i globalt samarbeid for KI-sikkerhet, anbefaler vi:

²⁹⁵ EU Artificial Intelligence Act 2024, [article 15](#)

²⁹⁶ Riksrevisjonen 2023, [Myndighetenes samordning av arbeidet med digital sikkerhet i sivil sektor](#)

²⁹⁷ NOU 2023: 17, [Nå er det alvor – Rustet for en usikker fremtid](#)

²⁹⁸ NSM 2023, [Sikkerhetsfaglig råd – Et motstandsdyktig Norge](#)

²⁹⁹ DSB 2024, [Risiko, sårbarhet og beredskap](#)

- Å bygge kompetanse for analyse og vurderinger av KI-trusler og utvikle risikoscenarioer med målsetning om å styrke digital motstanddyktighet og sikkerhet. Disse oppgavene kan legges til eksisterende fagmiljø, som Nasjonal Sikkerhetsmyndighet og Direktoratet for samfunnssikkerhet og beredskap.
- Å sikre norsk deltakelse i EUs nye vitenskapspanel og rådgivningsforum for kunstig intelligens som opprettes gjennom KI-forordningen.
- Å utvikle norske posisjoner for globale styringsmekanismer for kunstig intelligens i forkant av Summit of the Future i FN i september 2024.
- Å påvirke utviklingen av standarder for kunstig intelligens i EU, og internasjonalt gjennom standardorganisasjonene ISO og IEEE.

REFERANSER

Absci (2024) *Biologics drug discovery: Creating drugs at the speed of AI*. Hentet fra: [Biologics Drug Discovery | Absci](#)

Ada Lovelace Institute (2023) *Explainer: What is a foundation model?* 17 juli 2023. Av Elliot Jones. Hentet fra: [Explainer: What is a foundation model? | Ada Lovelace Institute](#)

Adobe Firefly (2024) *The new Firefly. Now smarter than ever*. Hentet fra: [Adobe Firefly - Free Generative AI for creatives](#)

Aftenposten (2024a) *KI-generert nyhetsanker leser nyheter for en hundredel av prisen*. 10 januar 2024. Av Per Kristian Bjørkeng. Hentet fra: [KI-generert nyhetsanker leser nyheter for en hundredel av prisen \(aftenposten.no\)](#)

Aftenposten (2024b) *KI kan gi gratis advokatråd, men norsk stiftelse står i veien: -Rettstaten forvitrer*. 25 januar 2024. Av Per Kristian Bjørkeng. Hentet fra: [KI kan gi gratis advokatråd, men norsk stiftelse står i veien: – Rettstaten forvitrer \(aftenposten.no\)](#)

Aftenposten (2024c) *Lovdata svarer: Tilgang til rettskilder for alle*. 4 februar 2024. Hentet fra: [Lovdata svarer: Tilgang til rettskilder for alle \(aftenposten.no\)](#)

AI Act (2021) *Proposal for a regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. Hentet fra: [AI-Act-FullText.pdf \(artificialintelligenceact.eu\)](#)

AI Impacts (2023) *2023 Expert Survey on Progress in AI*. 17 august 2023. Hentet fra: [2023 Expert Survey on Progress in AI \[AI Impacts Wiki\]](#)

Alemohammad, Sina, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi & Richard G. Baraniuk (2024) *Self-consuming generative models go MAD*. ICLR 2024. Hentet fra: [attachment \(openreview.net\)](#)

Aleph Alpha API (2024) *What is Luminous?* Hentet fra: <https://docs.aleph-alpha.com/docs/introduction/luminous/>

Altman, Sam (2023) *Sam Altman Unveils GPT-4 Turbo & AI Revolution at OpenAI Dev Day 2023*. 6 november 2023. YouTube. Hentet fra: [Sam Altman Unveils GPT-4 Turbo & AI Revolution at OpenAI Dev Day 2023 - YouTube](#)

Amazon (2023) *Previewing the future of Alexa*. 20 september 2023. Av Daniel Rausch. Hentet fra: [Amazon previews the future of Alexa with generative AI \(aboutamazon.com\)](#)

Amazon (2024) *Bedrock*. Hentet fra: [Build Generative AI Applications with Foundation Models - Amazon Bedrock - AWS](#)

Anthropic (2024) *Company*. Hentet fra: [Company \ Anthropic](#)

Associated Press (2023) *Mayor Suarez launches an artificial intelligence chatbot for his presidential campaign*. 5 juli 2023. Av Ali Swenson. Hentet fra: [Mayor Suarez launches an artificial intelligence chatbot for his presidential campaign | AP News](#)

Autor, David (2024) *Applying AI to rebuild Middle Class Jobs*. Februar 2024. National Bureau of Economic Research. Working Paper 32140. Hentet fra: [Applying AI to Rebuild Middle Class Jobs | NBER](#)

Axios (2024) *Axios AI+*. 18 mars 2024. Av Ina Fried & Ryan Heath. Hentet fra: [Reddit tipping point \(axios.com\)](#)

BBC (2024) *Embedding transparency, enhancing trust*. 4 mars 2024. Av Miranda Marcus. Hentet fra: [Embedding transparency, enhancing trust \(bbc.co.uk\)](#)

Be my eyes (2024) *Announcing 'Be My AI', soon available for hundreds of thousands of Be My Eyes users*. Hentet fra: [Announcing 'Be My AI,' Soon Available for Hundreds of Thousands of Be My Eyes Users](#)

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell (2021) *On the dangers of stochastic parrots: Can language models be too big?* FAccT'21. Hentet fra: [On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? "1F99C \(acm.org\)](#)

Big Think (2023) *How generative AI language models are unlocking the secrets of DNA*. 10 juni 2023. Av Sachin Rawat. Hentet fra: [How generative AI language models are unlocking DNA's secrets - Big Think](#)

Bing (2023) *Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web*. 7 februar 2023. Av Yusuf Mehdi. Hentet fra: [Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web - The Official Microsoft Blog](#)

Bloomberg (2023) *Humans are biased. Generative AI is even worse*. 9 juni 2023. Av Leonardo Nicoletti & Dina Bass. Hentet fra: [Generative AI Takes Stereotypes and Bias From Bad to Worse \(bloomberg.com\)](#)

Bloomberg (2024) *Apple intensifies talks with OpenAI for iPhone Generative AI features*. 27 april 2024. Av Mark Gurman. Hentet fra: [Apple's \(AAPL\) OpenAI Talks Intensify as It Seeks to Add AI Features - Bloomberg](#)

Bommasani, Rishi & Percy Liang (2021) *Reflections on Foundation Models*. Stanford University Human-Centered Artificial Intelligence. Hentet fra: [Reflections on Foundation Models \(stanford.edu\)](#)

Brønnøysundregistrene (2022) *Årsregnskapet for regnskapsåret 2022- Generell informasjon*. Hentet fra: [Årsregnskap regnskapsåret 2022 for 919036508 \(sanity.io\)](#)

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei (2020) *Language Models are Few-Shot Learners*. 22 juli 2020. Hentet fra: [\[2005.14165\] Language Models are Few-Shot Learners \(arxiv.org\)](#)

Brynjolfsson, Erik, Danielle Li & Lindsay R. Raymond (2023) *Generative AI at work*. November 2023. NBER working paper series. Hentet fra: [w31161.pdf \(nber.org\)](#)

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg,

Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro & Yi Zhang (2023) *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. Hentet fra: [\[2303.12712\] Sparks of Artificial General Intelligence: Early experiments with GPT-4 \(arxiv.org\)](#)

Burn-Murdoch (2023) *Here's what we know about generative AI's impact on white-collar work*. 10 november 2023. Financial Times. Hentet fra: [Here's what we know about generative AI's impact on white-collar work \(ft.com\)](#)

Business Insider (2023) *The end of coding as we know it*. 26 april 2023. Av Aki Ito. Hentet fra: [AI Technology Like ChatGPT Will Reshape Software Coding Jobs Forever \(businessinsider.com\)](#)

Cambridge Dictionary (2024) *Generative*. Hentet fra: [GENERATIVE | English meaning - Cambridge Dictionary](#)

Center for AI Safety (2024) *Statement on AI risk*. Hentet fra: [Statement on AI Risk | CAIS \(safe.ai\)](#)

Center for Countering Digital Hate (2024) *Fake Image Factories*. 6 mars 2024. Hentet fra: [Fake Image Factories — Center for Countering Digital Hate | CCDH \(counterhate.com\)](#)

Channel1 (2024) *Channel 1 gives users a new more personal way to watch the news*. Hentet fra: [Channel 1 AI - Personalized Global News Network](#)

Choi, Jonathan H., Daniel Schwarcz (2023) *AI Assistance in Legal Analysis: An Empirical Study*. 16 august 2023. University of Minnesota. Hentet fra: [AI Assistance in Legal Analysis: An Empirical Study by Jonathan H. Choi, Daniel Schwarcz :: SSRN](#)

CNBC (2022) *Google execs warn company's reputation could suffer if it moves too fast on AI-chat technology*. 13 desember 2022. Av Jennifer Elias. Hentet fra: [Google execs warn of reputational risk with ChatGBT-like tool \(cnbc.com\)](#)

Common Crawl (2024) *Common Crawl maintains a free open repository of web crawl that can be used by anyone*. Hentet fra: [Common Crawl - Open Repository of Web Crawl Data](#)

Communications of the ACM (2023) *The end of programming*. 1 januar 2023. Av Matt Welsh. Hentet fra: [The End of Programming – Communications of the ACM](#).

Council of the EU (2023) *Artificial Intelligence act: Council and Parliament strike a deal on the first rules for AI in the world*. Hentet fra: [Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world - Consilium \(europa.eu\)](#)

CRFM (2024) A holistic framework for evaluating foundation models. 7 mai 2024. Hentet fra: [Holistic Evaluation of Language Models \(HELM\) \(stanford.edu\)](#)

Cyabra (2023) *1 in 4 profiles are pro-amas fake accounts: The online battle-front*. 11 oktober 2023. Av Rotem Baruchin. Hentet fra: [Hamas' Social Media Warfare \(cyabra.com\)](#)

Dansk Erhverv (2024) *Danske sprogmodeller*. Hentet fra: [Fremtidens danske sprogmodel: Kom med i det nye AI-initiativ \(danskeerhverv.dk\)](#)

Datatilsynet (2023) *Globalt initiativ mot ulovlig dataskrapping*. 24 august 2023. Hentet fra: [Globalt initiativ mot ulovlig dataskrapping | Datatilsynet](#)

Datatilsynet (2023b) *Kommende rettsakter fra EU*. 5 juli 2023. Av Tobias Judin. Hentet fra: [Kommende rettsakter fra EU | Datatilsynet](#)

de Vries, Alex (2023) *The growing energy footprint of artificial intelligence*. 18 oktober 2023. Joule. Hentet fra: [The growing energy footprint of artificial intelligence \(asociace.ai\)](#)

Dell'Acqua, Fabrizio, Edward R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraye, François Candelon & Karim R. Lakhani (2023) *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality*. 27 september 2023. Harvard Business School Technology & Operations Management Unit. Hentet fra: [Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality by Fabrizio Dell'Acqua, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraye, François Candelon, Karim R. Lakhani :: SSRN](#)

Digdir (2024) *Oversikt over organisasjonen*. Hentet fra: [Oversikt over organisasjonen | Digdir](#)

Digdir (2024) Skates arbeidsutvalg. 8 februar 2024. Hentet fra: [Skates arbeidsutvalg | Digdir](#)

Digi.no (2023) Undersøkelse: Et flertall av ledere i offentlig sektor vil ta i bruk kunstig intelligens. 26 oktober 2023. Av Heidi Sævold. Hentet fra: [Undersøkelse: Et flertall av ledere i offentlig sektor vil ta i bruk kunstig intelligens | Digi.no](#)

Digitaliseringsstyrelsen (2024) *Rollen som national tilsynsmyndighed med EU's AI-forordning skal varetages af Digitaliseringsstyrelsen*. 11 april 2024. Hentet fra: [Rollen som national tilsynsmyndighed med EU's AI-forordning skal varetages af Digitaliseringsstyrelsen](#)

DSB (2024) *Risiko, sårbarhet og beredskap*. Hentet fra: <https://www.dsb.no/lover/risiko-sarbarhet-og-beredskap/>

DN (2024) *Teknoselskap får en milliard for å lage hjemmehjelp-robot*. 11 januar 2024. Av Øyvind Finstad. Hentet fra: [Teknoselskap får en milliard for å lage hjemmehjelp-robot | DN](#)

DRCF (2023) *Annual report 2022/23*. 27 april 2023. Hentet fra: [DRCF Annual Report 2022/23 \(ofcom.org.uk\)](#)

DRCF (2024) *About the DRCF*. Hentet fra: [About the DRCF | DRCF](#)

ehelse (2022) *Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten*. 2 november 2022. Hentet fra: [Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten - ehelse](#)

Elicit (2024) *Analyze research papers at superhuman speed*. Hentet fra: [Elicit: The AI Research Assistant](#)

Eloundou, Tyna, Sam Manning, Pamela Mishkin & Daniel Rock (2023) *GPTs are GPTs: An early look at the labor market impact potential of large language models*. 17 mars 2023. Cornell University. Hentet fra: [\[2303.10130\] GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models \(arxiv.org\)](#)

España Digital (2024) *Agencia Española de Supervisión de la Inteligencia Artificial*. Hentet fra: [Agencia Española de Supervisión de la Inteligencia Artificial | España Digital 2026 \(espanadigital.gob.es\)](https://espanadigital.gob.es)

EU Artificial Intelligence Act (2024) *Standard Setting*. Hentet fra: [Standard Setting | EU Artificial Intelligence Act](#)

Europaparlamentet (2024) *Europa-parlamentets og rådets forordning (EU) 2024/...* 16 april 2024. Hentet fra: [CO TA \(europa.eu\)](#)

European Commission (2008) *New legislative framework*. Hentet fra: [New legislative framework - European Commission \(europa.eu\)](#)

European Commission (2024a) *European AI Office*. Hentet fra: [European AI Office | Shaping Europe's digital future \(europa.eu\)](#)

European Commission (2024b) *Commission launches AI innovation package to support Artificial Intelligence startups and SMEs*. 24 januar 2024

European Commission (2024c) *The Digital Services Act package*. 16 februar 2024. Hentet fra: [The Digital Services Act package | Shaping Europe's digital future \(europa.eu\)](#)

European Commission (2024d) *Data Act*. 4 april 2024. Hentet fra: <https://digital-strategy.ec.europa.eu/en/policies/data-act>

European Commission (2024e) *European Chips Act*. Hentet fra: [European Chips Act - European Commission \(europa.eu\)](#)

European Commission (2024f) *The Digital Services Act package*. 16 februar 2024. Hentet fra: [The Digital Services Act package | Shaping Europe's digital future \(europa.eu\)](#)

European Commission (2024g) *Liability Rules for Artificial Intelligence*. Hentet fra: [Liability Rules for Artificial Intelligence](#)

European Parliament (2023a) *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. 9 desember 2023. Hentet fra: [Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI | News | European Parliament \(europa.eu\)](#)

European Parliament (2023b) *Political advertising: deal on new measures to crack down on abuse*. 6 november 2023. Hentet fra: [European AI Office | Shaping Europe's digital future \(europa.eu\)](https://european-council.europa.eu/media/en/press-articles/detail/16444)

Everypixel Journal (2024) *AI has already created as many images as photographers have taken in 150 years*. Statistics for 2023. Hentet fra: [AI Image Statistics: How Much Content Was Created by AI \(everypixel.com\)](https://everypixel.com/ai-image-statistics/)

Faktisk (2024) *Ansatte*. Hentet fra: [Ansatte. Faktisk.](https://faktisk.no/ansatte)

Faktisk (2024) *Finansiering*. Hentet fra: faktisk.no/finansiering

Federal Bureau of Prisons (2024) *Inmate Race*. 18 mai 2024. Hentet fra: [BOP Statistics: Inmate Race](https://www.fbi.gov/newsroom/speeches/inmate-race-statistics)

FFI (2021) *Hvordan gjøre samfunnet mer robust mot uønsket påvirkning i sosiale medier*. 9 juni 2021. Av Eskil Grendahl Sivertsen, Nina Hellum, Arild Bergh & Anne Lise Bjørnstad. Hentet fra: [Hvordan gjøre samfunnet mer robust mot uønsket påvirkning i sosiale medier \(knowledgearc.net\)](https://www.knowledgearc.net/hvordan-gjore-samfunnet-mer-robust-mot-uonsket-pavirkning-i-sosiale-medier)

FFI (2023) *Ny FFI-rapport avdekker falske nettaviser*. 13 desember 2023. Hentet fra: [Ny FFI-rapport avdekker falske nettaviser](https://www.ffi.no/ny-ffi-rapport-avdekker-falske-nettaviser)

FFI (2024) *Forsvar mot fremmedstatlige påvirkningsoperasjoner – etablering av funksjon i forsvarssektoren*. 5 april 2024. Av Eskil Grendahl Sivertsen & Paul Buvarp. Hentet fra: [Forsvar mot fremmedstatlige påvirkningsoperasjoner – etablering av funksjon i forsvarssektoren \(ffi.no\)](https://www.ffi.no/forsvar-mot-fremmedstatlige-pavirkningsoperasjoner-etablering-av-funksjon-i-forsvarssektoren)

Financial Times (2023a) *Generative AI exists because of the transformer*. 12 september 2023. Av Visual Storytelling Team & Madhumita Murgia. Hentet fra: [Generative AI exists because of the transformer](https://www.ft.com/content/3d4e0c0e-0c0e-4c0e-0c0e-0c0e0c0e0c0e)

Financial Times (2023b) *The advent of the AI agent*. 10 November 2023. Hentet fra: [The advent of the AI agent](https://www.ft.com/content/3d4e0c0e-0c0e-4c0e-0c0e-0c0e0c0e0c0e)

Financial Times (2023c) *AstraZeneca ties up with AI biologics company to develop cancer drug*. 3 desember 2023. Av Michael Peel. Hentet fra: [AstraZeneca ties up with AI biologics company to develop cancer drug \(ft.com\)](https://www.ft.com/content/3d4e0c0e-0c0e-4c0e-0c0e-0c0e0c0e0c0e)

Financial Times (2024) *UK Government to trial 'red box' AI tools to improve ministerial efficiency*. 29 februar 2024. Av Lucy Fisher. Hentet fra: [UK government to trial 'red box' AI tools to improve ministerial efficiency \(ft.com\)](https://www.ft.com/content/2024/02/29/uk-government-to-trial-red-box-ai-tools-to-improve-ministerial-efficiency)

Firoozi, Roya, Johnathan Tucker, Stephen Tian, Anirudha Majumadar, Jiankai Sun, Wiyu Liu, Yuke Zhu, Shuran Song, Ashish Kapoor, Karol Hausman, Brian Ichter, Danny Driess, Jiajun Wu, Cewu Lu, Mac Schwager, Stanford University, Princeton University, UT Austin, Nvidia, Scaled Foundations, Google deepMind, TU Berling, Shanghai Jiao Tong University (2023) *Foundation Models in Robotics: Applications, Challenges and the future*. 13. Desember 2023. Hentet fra: <https://arxiv.org/html/2312.07843v1>

Forbes (2024) *AI and personalization in the age of automation*. 7 februar 2024. Av Jia Ritzi. Hentet fra: [AI And Personalization In The Age Of Automation \(forbes.com\)](https://www.forbes.com/sites/jiaritzi/2024/02/07/ai-and-personalization-in-the-age-of-automation/)

Forbrukerrådet (2023) *Ghost in the machine: Addressing the consumer harms of generative AI*. Juni 2023. Hentet fra: [generative-ai-rapport-2023.pdf \(forbrukerradet.no\)](https://www.forbrukerradet.no/rapporter/2023/ghost-in-the-machine)

Forbrukerrådet (2024) *Årsrapport 2023*. Hentet fra: [38268.pdf \(dfo.no\)](https://www.forbrukerradet.no/rapporter/2024/arsrapport-2023)

Forskningsrådet (2023) *Norsk veikart for forskningsinfrastruktur 2023*. 3 juli 2023. Hentet fra: [Norsk veikart for forskningsinfrastruktur 2023 \(forskningsradet.no\)](https://www.forskningssradet.no/publikasjoner/norsk-veikart-for-forskningssinfrastruktur-2023)

Forskningsrådet (2024) *Midler til forskningsinfrastruktur av nasjonal viktighet*. 16 april 2024. Hentet fra: [Midler til forskningsinfrastruktur av nasjonal viktighet \(forskningsradet.no\)](https://www.forskningssradet.no/publikasjoner/midler-til-forskningssinfrastruktur-av-nasjonal-viktighet)

Fortune (2023) *CEO of Google's DeepMind says we could be 'just a few years' from A.I. that has human-level intelligence*. 3 mars 2023. Hentet fra: [AGI is 'just a few years' away, says Google's DeepMind CEO | Fortune](https://fortune.com/2023/03/03/google-deepmind-ceo-ai/)

Freedom house (2022) *The Repressive power of artificial intelligence*. Hentet fra: [The Repressive Power of Artificial Intelligence | Freedom House](https://www.freedomhouse.org/report/2022/the-repressive-power-of-artificial-intelligence)

Fries, Jason, Ethan Steinberg, Scott Fleming, Michael Wornow, Yizhe Xu, Keith Morse, Dev Dash & Nigam Shah (2022) *How foundation models can advance AI in healthcare*. Stanford University Human-Centered Artificial Intelligence.

Hentet fra: [How Foundation Models Can Advance AI in Healthcare \(stanford.edu\)](#)

Future of Life Institute (2023) *Pause Giant AI Experiments: An Open Letter*. 22 mars 2023. Hentet fra: [Pause Giant AI Experiments: An Open Letter - Future of Life Institute](#)

Futurism (2023) *Sports illustrated published articles by fake, AI-generated writers*. 27 november 2023. Av Maggie Harrison Dupre. Hentet fra: [Sports Illustrated Published Articles by Fake, AI-Generated Writers \(futurism.com\)](#)

Gladstone AI (2024) An action plan to increase the safety and security of advanced AI. Hentet fra: [Action Plan to increase the safety and security of advanced AI \(gladstone.ai\)](#)

Goldman Sachs (2023) Global Economic Analyst: *The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)*. 29 mars 2023. Av Jan Hatzius, Joseph Briggs, Davesch Kodnani & Giovanni Pierdomenico. Hentet fra: [Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth \(BriggsKodnani\) \(key4biz.it\)](#)

Google (2021) *Introducing pathways: A next generation AI architecture*. 28 oktober 2021. Hentet fra: [Introducing Pathways: A next-generation AI architecture \(blog.google\)](#)

Google (2023a) *An important next step on our AI journey*. 6. Februar 2023. Hentet fra: [Google AI updates: Bard and new AI features in Search \(blog.google\)](#)

Google (2023b) *YouTube: Our approach to responsible AI innovation on YouTube*. 15 November 2023. Av Jennifer Flannery O'Connor & Emily Moxley. Hentet fra: [Our approach to responsible AI innovation on YouTube \(blog.google\)](#)

Google (2024a) *Gemma models overview*. 20 mai 2024. Hentet fra: [Gemma models overview | Google for Developers](#)

Google (2024b) Gemini Nano lets you complete helpful AI tasks without a network connection. Hentet fra: [Gemini Nano Multimodal Capabilities on Pixel Phones \(google.com\)](#)

Google Cloud (2024a) *Announcing Verex AI Agent Builder: Helping developers easily build and deploy gen AI experiences*. 9 april 2024. Hentet fra: [Build generative AI experiences with Vertex AI Agent Builder | Google Cloud Blog](#)

Google Cloud (2024b) *What is Artificial Intelligence (AI)?* Hentet fra: [What Is Artificial Intelligence \(AI\)? | Google Cloud](#)

Google Deepmind (2024a) *DiPaCo: Distributed Path Composition*. 19 mars 2024. Hentet fra: [DiPaCo: Distributed Path Composition - Google DeepMind](#)

Google Deepmind (2024b) *AlphaFold*. Hentet fra: <https://deepmind.google/technologies/alphafold/>

Google Doc (2022) *Condemning the deployment of GPT-4chan*. Hentet fra: [Condemning the deployment of GPT-4chan \(google.com\)](#)

Google Research (2024) *Generative AI to quantify uncertainty in weather forecasting*. 24 mars 2024. Av Lizao Li & Rob Carver. Hentet fra: [Generative AI to quantify uncertainty in weather forecasting \(research.google\)](#)

Google Workspace (2023) *A new era for AI and Google Workspace*. 4 mars 2023. Hentet fra: [Announcing new generative AI experiences in Google Workspace | Google Workspace Blog](#)

GOV.UK (2023a) *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*. 1 November 2023. Hentet fra: [The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 - GOV.UK \(www.gov.uk\)](#)

GOV.UK (2023b) *Chair's summary of the AI Safety summit 2023, Bletchley Park*. 2 november 2023. Hentet fra: [Chair's Summary of the AI Safety Summit 2023, Bletchley Park - GOV.UK \(www.gov.uk\)](#)

GOV.UK (2023c) *Fact sheet on the CDU and RRU*. 9 juni 2023. Hentet fra: [Fact Sheet on the CDU and RRU - GOV.UK \(www.gov.uk\)](#)

GOV.UK (2024) *UK & United States announce partnership on science of AI safety*. 2 april 2024. Hentet fra: [UK & United States announce partnership on science of AI safety - GOV.UK \(www.gov.uk\)](#)

Heikkilä, Melissa (2024) *Three reasons robots are about to become way more useful*. 16 april 2024. MIT Technology Review. Hentet fra: [Three reasons robots are about to become way more useful | MIT Technology Review](#)

Helsedirektoratet (2024) Tverretattlig veiledningstjeneste. 8 mars 2024. Hentet fra: [Tverretattlig veiledningstjeneste - Helsedirektoratet](#)

Home Security Heroes (2024) *2023 State of deepfakes: Realities, threats and impact*. Hentet fra: [2023 State Of Deepfakes: Realities, Threats, And Impact \(homesecurityheroes.com\)](#)

HuggingFace (2024b) *Open LLM leaderboard*. Hentet fra: [Open LLM Leaderboard - a Hugging Face Space by HuggingFaceH4](#)

Hui, Xiang, Oren Reshef & Luofeng Zhou (2023) *The short-term effects on generative artificial intelligence on employment: Evidence from an online labor market*. 1 august 2023. Hentet fra: [The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market by Xiang Hui, Oren Reshef, Luofeng Zhou :: SSRN](#)

Idunn (2023) *Kunstig intelligens og Nasjonalbiblioteket*. 18 oktober 2023. Av Svein Arne Brygfjeld. Hentet fra: [Kunstig intelligens og Nasjonalbiblioteket | Stat & Styling \(idunn.no\)](#)

IEA (2024) *Electricity 2024 – Analysis and forecast to 2026*. Hentet fra: [Electricity 2024 - Analysis and forecast to 2026 \(iea.blob.core.windows.net\)](#)

Imperva (2024) *Bad Bot Report*. Hentet fra: [2024 Bad Bot Report | Resource Library \(imperva.com\)](#)

International Labour Organization (2023) *Generative AI and Jobs: A global analysis of potential effects on job quantity and quality*. 21 august 2023. Av Pawel Gmyrek, Janine Berg & David Bescond. Hentet fra: [Generative AI and Jobs: A global analysis of potential effects on job quantity and quality | International Labour Organization \(ilo.org\)](#)

International Press Institute (2023) *Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse*. 31 oktober 2023. Av Karin Kóváry Sólymos. Hentet fra: [Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse - ipi.media](#)

Internet Watch Foundation (2023) *How AI is being abused to create child sexual abuse imagery*. Hentet fra: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf

Isomorphic Labs (2024) *Reimagining Drug Discovery Process with AI*. Hentet fra: [Reimagining Drug Discovery Process with AI - Isomorphic Labs](#)

Jiang, Hang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy & Jad Kabbara (2023) *PersonaLLM: Investigating the ability of large language models to express personality traits*. 2 april 2024. Cornell University. Hentet fra: [\[2305.02547\] PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits \(arxiv.org\)](#)

Jovanovic, Boyan, Rousseau, Peter (2005) *Chapter 18 - General Purpose Technologies*. Handbook of Economic Growth. Hentet fra: [https://doi.org/10.1016/S1574-0684\(05\)01018-X](https://doi.org/10.1016/S1574-0684(05)01018-X)

Khanmigo (2024) *Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor*. Hentet fra: [Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor](#)

Kultur- og likestillingsdepartementet (2024) *Om arbeidet med strategi for å styrke motstandskraften mot desinformasjon*. Hentet fra: [Om arbeidet med strategi for å styrke motstandskraften mot desinformasjon - regjeringen.no](#)

Labelbox (2024) *Labelbox*. Hentet fra: [Labelbox | Data-centric AI Platform for Building & Using AI](#)

Laion (2022) *Laion-5b: A new era of open large-scale multi-modal datasets*. 31 mars 2022. Av Romain Beaumont. Hentet fra: [LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI-MODAL DATASETS | LAION](#)

Lee, Hyun Jung, Tahira M. Probst, Andrea Bazzoli & Sunhee Lee (2022). *Technology Advancements and Employees' Qualitative Job Insecurity in the Republic of Korea: Does Training Help? Employer-Provided vs. Self-Paid Training*. Ideas. November 2022. Hentet fra: [Technology Advancements and Employees' Qualitative Job Insecurity in the Republic of Korea: Does Training Help? Employer-Provided vs. Self-Paid Training \(repec.org\)](#)

Linköping Science Park (2023) *Open release of the first large Nordic language model GPT-SW3*. 16 november 2023. AI Sweden. Hentet fra: [Open release of](#)

[the first large Nordic language model GPT-SW3 - Linköping Science Park \(linkopingsciencepark.se\)](https://linkopingsciencepark.se)

Linköpings universitet (2021) *Sveriges snabbaste superdator för AI är invigd*. 23 mars 2021. Av Anders Ryttaarson Törneholm. Hentet fra: [Sveriges snabbaste superdator för AI är invigd - Linköpings universitet \(liu.se\)](https://liu.se)

Lovdata (2022) *Forskrift om merking av retusjert reklame*. 1 juli 2022. Hentet fra: [Forskrift om merking av retusjert reklame - Lovdata](https://lovdata.no)

Lovdata (2024) *Brukeravtale*. Hentet fra: [Brukeravtale, Lovdata Pro - Lovdata](https://lovdata.no)

Luccioni, Alexandra S., Yacine Jernite & Emma Strubell (2023) *Power hungry processing watts driving the cost of AI deployment?* 28 november 2023. Hentet fra: [Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment? \(arxiv.org\)](https://arxiv.org)

Lumi (2024) *Hjemmeside*. Hentet fra: [Front Page - LUMI \(lumi-supercomputer.eu\)](https://lumi-supercomputer.eu)

Lynch, Shana (2024) *AI Index: State of AI in 13 Charts*. Stanford University Human-Centered Artificial Intelligence. Hentet fra: [AI Index: State of AI in 13 Charts \(stanford.edu\)](https://stanford.edu)

Media City Bergen (2024) *Prosjekt Reynir*. Hentet fra: [Embedding transparency, enhancing trust \(bbc.co.uk\)](https://bbc.co.uk)

Medium (2023) *GPT4- All details leaked*. 13 juli 2023. Av katerinaptrv. Hentet fra: [GPT4- All Details Leaked. The details about the best LLM model... | by katerinaptrv | Medium](https://medium.com)

Menon Economics (2023) *KI: Betydning for arbeidsstyrken*. Oktober 2023. Av Trygve Svalheim, John Oskar Skjeldrum, Sebastian Winther-Larsen, Kristoffer Midtømme & Jonas Erraia. Hentet fra: [2023-127-KI-Betydning-for-arbeidsstyrken.pdf \(menon.no\)](https://menon.no)

Meta (2023) *Generative AI features for ads coming to all advertisers*. 4 oktober 2023. Hentet fra: [Generative AI features for ads coming to all advertisers | Meta for Business \(facebook.com\)](https://facebook.com)

Meta (2024) Slik identifiseres og merkes KI-generert innhold på Meta. 24 mai 2024. Hentet fra: [How AI-generated content is identified and labeled on Meta | Meta Store](#)

Microsoft (2024) *Slipp løs produktiviteten din med AI og Microsoft Copilot*. Hentet fra: [Slipp løs produktiviteten din med AI og Microsoft Copilot - Støtte for Microsoft](#)

Microsoft WorkLab (2023) *Will AI fix work?* 9 mai 2023. Hentet fra: [Work Trend Index | Will AI Fix Work? \(microsoft.com\)](#)

Middlebury (2020) *The radicalization risks of GPT-3 and neural language models*. 9 september 2020. Av Kris McGuffie & Alex Newhouse. Hentet fra: [The Radicalization Risks of GPT-3 and Neural Language Models | Middlebury Institute of International Studies at Monterey](#)

Miller, Elizabeth J., Ben A. Steward, Zat Witkower, Clare A. M. Sutherland, Eva G. Krumhuber & Amy Dawel (2023) *AI Hyperrealism: Why AI faces are perceived as more real than human ones*. 13 november 2023. Sage Journals. Hentet fra: [AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones - Elizabeth J. Miller, Ben A. Steward, Zak Witkower, Clare A. M. Sutherland, Eva G. Krumhuber, Amy Dawel, 2023 \(sagepub.com\)](#)

MIT News (2023) *Generative AI imagines new protein structures*. 12 juli 2023. Av Rachel Gordon. Hentet fra: [Generative AI imagines new protein structures | MIT News | Massachusetts Institute of Technology](#)

MIT Technology Review (2023) *Mustafa Suleyman: My new Turing test would see if AI can make \$1 million*. 14 juli 2023. Av Mustafa Suleyman. Hentet fra: [Mustafa Suleyman: My new Turing test would see if AI can make \\$1 million | MIT Technology Review](#)

MIT Technology Review (2024) *The tech industry can't agree on what open-source AI means. That's a problem*. 25 mars 2024. Av Edd Gent. Hentet fra: [The tech industry can't agree on what open source AI means. That's a problem. | MIT Technology Review](#)

Myndighet för psykologisk försvar (2024) *Myndighet för psykologisk försvar*. Hentet fra: [Myndigheten för psykologiskt försvar \(mpf.se\)](#)

Nasjonal Sikkerhetsmyndighet (2023) *Sikkerhetsfaglig råd – Et motstandsdyktig Norge*. 9.mai 2023. Hentet fra: <https://nsm.no/regelverk-og-hjelp/rapporter/sikkerhetsfaglig-rad-et-motstandsdyktig-norge>

Nasjonalbiblioteket (2023a) *Nasjonalbiblioteket deler kunstig intelligens som skjøner norske dialekter og gjer tale om til tekst*. 29 juni 2023. Av Nina Bræin. Hentet fra: [Nasjonalbiblioteket deler kunstig intelligens som skjøner norske dialekter og gjer tale om til tekst | Nasjonalbiblioteket \(nb.no\)](#)

Nasjonalbiblioteket (2023b) *Nasjonalbiblioteket skal lede nytt forskningsprosjekt om kunstig intelligens*. 13 desember 2023. Av Nina Bræin. Hentet fra: [Nasjonalbiblioteket skal lede nytt forskningsprosjekt om kunstig intelligens | Nasjonalbiblioteket \(nb.no\)](#)

Nasjonalbiblioteket (2024) *Models*. Hentet fra: [Models – AI-lab \(nb.no\)](#)

Nature (2024) *Generative AI's environmental costs are soaring – and mostly secret*. 20 februar 2024. Av Kate Crawford. Hentet fra: [Generative AI's environmental costs are soaring – and mostly secret \(nature.com\)](#)

Nettveit.no (2024) *Hva er phishing?* Hentet fra: [Hva er phishing? - Nettveit.no](#)

New York Times (2012) *How Many Computers to Identify a Cat? 16,000*. 25 Juni 2012. Av John Markoff. Hentet fra: [In a Big Network of Computers, Evidence of Machine Learning - The New York Times](#)

New York Times (2020) *Meet GPT-3. It has learned to code (and blog and argue)*. 24 november 2020. Av Cade Metz. Hentet fra: [Meet GPT-3. It Has Learned to Code \(and Blog and Argue\). - The New York Times \(nytimes.com\)](#)

New York Times (2023a) *How ChatGPT kicked off an A.I. arms race*. 3 februar 2023. Av Kevin Roose. Hentet fra: [How ChatGPT Kicked Off an A.I. Arms Race - The New York Times \(nytimes.com\)](#)

New York Times (2023b) *Here's what happens when your lawyer uses ChatGPT*. 27 Mars 2023. Av Benjamin Weiser. Hentet fra: [A Man Sued Avianca Airline. His Lawyer Used ChatGPT. - The New York Times \(nytimes.com\)](#)

New York Times (2023c) *The times sues OpenAI and Microsoft over A.I. use of copyrighted work*. 27 desember 2023. Hentet fra: [New York Times Sues OpenAI](#)

[and Microsoft Over Use of Copyrighted Work - The New York Times \(nytimes.com\)](#)

New York Times (2024a) *How Tech Giants Cut Corners to Harvest Data for A.I.* 6 april 2024. Av Cade Metz, Cecilia Kang, Sheera Frenkel, Stuart A. Thompson & Nico Grant. Hentet fra: [How Tech Giants Cut Corners to Harvest Data for A.I. - The New York Times \(nytimes.com\)](#)

New York Times (2024b) *What to know about the tech companies using A.I. to teach their own A.I.* 6 april 2024. Av Cade Metz & Stuart A. Thompson. Hentet fra: [What to Know About Tech Companies Using A.I. to Teach Their Own A.I. - The New York Times \(nytimes.com\)](#)

New York Times (2024c) *Explicit Deepfake Images of Taylor Swift elude safeguards and swamp social media.* 26 januar 2024. Av Kate Conger & John Yoon. Hentet fra: [Fake Explicit Taylor Swift Images Swamp Social Media - The New York Times \(nytimes.com\)](#)

New York Times (2024d) *How A.I. tools could change India's elections.* 18 april 2024. Av Suhasini Raj. Hentet fra: [How A.I. Tools Could Change India's Elections - The New York Times \(nytimes.com\)](#)

Nielson, Dan Saattrup (2023) *ScandEval: A Benchmark for Scandinavian Natural Language Processing*. Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa) Hentet fra: [2023.nodalida-1.20.pdf \(aclanthology.org\)](#)

NORA.ai (2024) *The NORA.LLM project – large language models as infrastructure.* 8 februar 2024. Hentet fra: [The NORA.LLM project - large language models as infrastructure - NORA - Norwegian Artificial Intelligence Research Consortium](#)

NorwAI (2024) *Vi treng generative norske språkmodeller.* 27 februar 2024. Av Rolf Dyrnes Svendsen. Hentet fra: [Vi treng generative norske språkmodeller - NTNU](#)

NorskGPT (2024) *NorskGPT*. Hentet fra: [NorskGPT](#)

NOU 2022: 9. *En åpen og opplyst offentlig samtale – Ytringsfrihetskommisjonens utredning.* Hentet fra: [NOU 2022: 9 - regjeringen.no](#)

NOU 2023: 17 *Nå er det alvor – Rustet for en usikker fremtid*. Hentet fra: [NOU 2023: 17 - regjeringen.no](#)

NOU 2023:4 *Tid for handling – Personellet i en bærekraftig helse- og omsorgstjeneste*. Hentet fra: [NOU 2023: 4 - regjeringen.no](#)

NRK (2023a) *Bærum sykehus er først i Norge med å ta i bruk kunstig intelligens til behandling*. 28 august 2023. Av Jens Christian Sundby. Hentet fra: [Bærum sykehus er først i Norge med å ta i bruk kunstig intelligens til behandling – Stor-Oslo \(nrk.no\)](#)

NRK (2023b) *Venn med kunstig intelligens*. 30 september 2023. Av Ine Julia Rojahn Schwebs. Hentet fra: [Agnete \(90\) er «venn» med ChatGPT: – Jeg er ikke ensom – Kultur \(nrk.no\)](#)

NRK (2023c) *Stadig flere KI-romansar: - Kjærleik fri for risiko*. 15 september 2023. Av Ingrid Emilie Waaler & Ine Julia Rojahn Schwebs. Hentet fra: [Flere blir kjærastar med kunstig intelligens \(KI\) – NRK Norge – Oversikt over nyheter fra ulike deler av landet](#)

NRK (2024a) *Nå kan kunstig intelligens lage sangen du ber om*. 10 april 2024. Av Egil Aslak Hagerup & Mathias Nylenna. Hentet fra: [Nå kan kunstig intelligens lage sangen du ber om – NRK Kultur og underholdning](#)

NRK (2024b) *Storytel innfører KI-stemmer: - Kommer til å treffe oss rett i trynet*. 7 mars 2024. Av Ine Julia Rojahn Schwebs & Aida Khorami. Hentet fra: [Storytel innfører KI-stemmer: – Kommer til å treffe oss rett i trynet – NRK Norge – Oversikt over nyheter fra ulike deler av landet](#)

NRK (2024c) *Kunstig Intelligens hjelper Gry med å «se»: -Gir en selvstendighet*. 1 april 2024. Av Julie Helene Günther. Hentet fra: [Blinde får hjelp av kunstig intelligens – NRK Østfold – Lokale nyheter, TV og radio](#)

NRK (2024d) *Dansk Folkeparti i Danmark brukar «deepfake» av statsminister Mette Fredriksen i video*. 30 april 2024. Av Julianne Flølo. Hentet fra: [Dansk Folkeparti i Danmark brukar «deepfake» av statsminister Mette Fredriksen i video – NRK Norge – Oversikt over nyheter fra ulike deler av landet](#)

NTB (2023) *Faktisk verifiserbar starter opp igjen: - Stort behov for å få verifisert bilder og videoer nå*. 19 oktober 2023. Hentet fra: [Faktisk Verifiserbar](#)

[starter opp igjen: – Stort behov for å få verifisert bilder og videoer nå | Pressens hus \(ntb.no\)](#)

NTNU (2024) *Fire modeller bygget – fire nye på beddingen*. 26.april 2024. Av Jon Atle Gulla & Rolf D. Svendsen. Hentet fra: <https://www.ntnu.no/norllm/fire-modeller-bygget-fire-nye-på-beddingen>

NTNU High Performance Computing (2024) *Idun*. 24 april 2024. Hentet fra: [Idun – High Performance Computing \(ntnu.no\)](#)

Nvidia (2023) *Nvidia Keynote at COMPUTEX 2023*. 30 mai 2023. Av Jensen Huang. YouTube. Hentet fra: [NVIDIA Keynote at COMPUTEX 2023 \(youtube.com\)](#)

O'Reilly (2023) *The alignment problem is not new*. 15 juni 2023. Av Tim O'Reilly. Hentet fra: [The Alignment Problem Is Not New – O'Reilly \(oreilly.com\)](#)

OECD (2023) *OECD Employment Outlook 2023. 3.2. It is too early to detect meaningful employment changes due to artificial intelligence*. Av Andrew Greene. Hentet fra: [OECD Employment Outlook 2023 : Artificial Intelligence and the Labour Market | OECD iLibrary \(oecd-ilibrary.org\)](#)

OECD (2024) *Facts not fakes: Tackling Disinformation, Strengthening Information Integrity*. Hentet fra: [Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity | OECD iLibrary \(oecd-ilibrary.org\)](#)

OpenAI (2019a) *Better language models and their implications*. 14 Februar 2019. Hentet fra: [Better language models and their implications | OpenAI](#)

OpenAI (2019b) *OpenAI LP*. 11 mars 2019. Hentet fra: [OpenAI LP | OpenAI](#)

OpenAI (2019c) *Microsoft invests in and partners with OpenAI to support us building beneficial AGI*. 22 juli 2019. Hentet fra: [Microsoft invests in and partners with OpenAI to support us building beneficial AGI | OpenAI](#)

OpenAI (2022) *Introducing Whisper*. 21 september 2022. Hentet fra [Introducing Whisper | OpenAI](#)

OpenAI (2023a) *ChatGPT can now see, hear, and speak*. 25 september 2023. Hentet fra: [ChatGPT can now see, hear, and speak | OpenAI](#)

OpenAI (2023b) *GPT-4 Technology Report*. 27 mars 2023. Hentet fra: [gpt-4.pdf \(openai.com\)](https://openai.com/research/gpt-4.pdf)

OpenAI (2023c) *GPT-4 System Card*. 23 mars 2023. Hentet fra: [gpt-4-system-card.pdf \(openai.com\)](https://openai.com/research/gpt-4-system-card.pdf)

OpenAI (2023d) *Planning for AGI and beyond*. 24 februar 2023. Hentet fra: [Planning for AGI and beyond | OpenAI](https://openai.com/research/planning-for-agi-and-beyond)

OpenAI (2023e) *Introducing Superalignment*. 5 juli 2023. Hentet fra: [Introducing Superalignment | OpenAI](https://openai.com/research/introducing-superalignment)

OpenAI (2024a) *What is ChatGPT?* Mai 2024. Hentet fra: [What is ChatGPT? | OpenAI Help Center](https://openai.com/help/chatgpt)

OpenAI (2024b) *Be My Eyes*. Hentet fra: [Be My Eyes | OpenAI](https://openai.com/be-my-eyes)

OpenAI (2024c) *Sora*. Hentet fra: [Sora | OpenAI](https://openai.com/sora)

OpenAI (2024d) *Introducing the GPT Store*. 10 januar 2024. Hentet fra: [Introducing the GPT Store | OpenAI](https://openai.com/gpt-store)

OpenGPT-X (2024) *Multilingual. Open. European*. Hentet fra: [OpenGPT-X](https://opengpt-x.com)

Park, Joon S., Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, Michael S. Bernstein (2022) *Social Simulacra: Creating Populated Prototypes for Social Computing Systems*. Oktober 2022. UIST'22. Hentet fra: [Social Simulacra: Creating Populated Prototypes for Social Computing Systems | Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology](https://dl.acm.org/doi/10.1145/3526751.3526801)

Peng, Sida, Eirini Kalliamvakou, Peter Cihon & Mert Demirer (2023) *The impact of AI on developer productivity. Evidence from GitHub Copilot*. 13 februar 2023. Cornell University. Hentet fra: [\[2302.06590\] The Impact of AI on Developer Productivity: Evidence from GitHub Copilot \(arxiv.org\)](https://arxiv.org/abs/2302.06590)

Politiet (2023) *Generativ kunstig intelligens og cyberkriminalitet*. Hentet fra: [Generativ kunstig intelligens og cyberkriminalitet \(politiet.no\)](https://politiet.no/om-politiet/nyheter/2023/03/01/generativ-kunstig-intelligens-og-cyberkriminalitet)

PrivacyNotIncluded (2024) *Happy Valentine's Day! Romantic AI Chatbots don't have your privacy at heart*. 14 februar 2024. Av Jen Caltrider, Misha

Rykov & Zoë MacDonald. Hentet fra: [Romantic AI Chatbots Don't Have Your Privacy at Heart \(mozilla.org\)](https://www.mozilla.org/en-US/privacy/ai/)

Project Origin (2024) *What origin does*. Hentet fra: [Project Origin \(originproject.info\)](https://originproject.info/)

PWC (2024) *PwCs årlige topplederundersøkelse: Vekst i en verden i endring*. Januar 2024. Hentet fra: [CEO-survey| Hovedfunn | 2024 \(pwc.no\)](https://www.pwc.no/ceo-survey/hovedfunn/2024)

Radar (2024) *Hvis det danske tilsyn med AI Act skal være lovligt, kræver det en dybere omstrukturering af Digitaliseringsstyrelsen*. 21 april 2024. Hentet fra: [Hvis det danske tilsyn med AI Act skal være lovligt, kræver det en dybere omstrukturering af Digitaliseringsstyrelsen | Radar](#)

Rathenau (2024) *Generative AI*. 20.mars 2024. Hentet fra: [Scan Generative AI Rathenau Instituut.pdf](#)

Rettberg, Jill Walker (2023) *Lystløgneren ChatGPT*. NRK Ytring. 4 Mars 2023. Hentet fra: [Lystløgneren ChatGPT – Ytring \(nrk.no\)](https://www.nrk.no/ytring/lustloegneren-chatgpt)

Reuters (2023) *ChatGPT sets record for fastest-growing user base – analyst note*. 2 februar 2023. Av Krystal Hu. Hentet fra: [ChatGPT sets record for fastest-growing user base - analyst note | Reuters](#)

Riksrevisjonen (2023), *Myndighetenes samordning av arbeidet med digital sikkerhet i sivil sektor*. Dokument 3:7 (2022-2023). Hentet fra: <https://www.riksrevisjonen.no/globalassets/rapporter/NO-2022-2023/myndighetenes-samordning-av-arbeidet-med-digital-sikkerhet-i-sivil-sektor.pdf>

Ritzau (2024) Danmark bygger en af verdens kraftigste AI-supercomputere for at sætte skub i løsninger på samfundsmæssige udfordringer. 13 mars 2024. Novo Nordisk Fonden. Hentet fra: [Danmark bygger en af verdens kraftigste AI-supercomputere for at sætte skub i løsninger på samfundsmæssige udfordringer | Novo Nordisk Fonden \(ritzau.dk\)](#)

Robinson, David (2023) *Hamas propaganda analysis – over 85% of engagement are bots*. 16 oktober 2023. Internet2.0. Hentet fra: [Hamas Propaganda Analysis-over 85% of engagement are bots \(internet2-0.com\)](https://internet2.0.com/analysis-over-85-of-engagement-are-bots)

Sandvik, Kristin B. (2023) Optics as Politics: Culture, Language and Learning with UiO ChatGPT. 19 desember 2023. PRIO. Hentet fra: [Optics as Politics: Culture, Language and Learning with UiO ChatGPT](#)

Schwartz, Roy, Jesse Dodge, Noah A. Smith, Oren Etzioni (2019) *Green AI*. Juli 2019. Allen Institute for AI. Hentet fra: <https://arxiv.org/pdf/1907.10597>

Science (2021) *Science's 2021 breakthrough of the year: AI brings protein structures to all*. 16 desember 2021. Av Robert Service. Hentet fra: [Science's 2021 Breakthrough of the Year: AI brings protein structures to all | Science | AAAS](#)

Science (2022) *Competition-level code generation with AlphaCode*. 8 desember 2022. Hentet fra Yulia Li m.fl. Hentet fra: [Competition-level code generation with AlphaCode | Science](#)

Science (2023a) *Experimental evidence on the productivity effects of generative artificial intelligence*. 13 juli 2023. Av Shakked Noy & Whitney Zhang. Hentet fra: [Experimental evidence on the productivity effects of generative artificial intelligence | Science](#)

Science (2023b) *Guinea Pigbots: Doing research with human subjects is costly and cumbersome. Can AI chatbots replace them?* 13 juli 2023. Av Matthew Hutson. Hentet fra: [Can AI chatbots replace human subjects in behavioral experiments? | Science | AAAS](#)

Science (2023c) *AI model GPT-3 (dis)informs us better than humans*. 28 juni 2023. Av Giovanni Spitale, Nikola Biller-Adorno & Federico Germani. Hentet fra: [AI model GPT-3 \(dis\)informs us better than humans | Science Advances](#)

Scientific American (2023) *How Scientists are using AI to talk to animals*. 7 februar 2023. Av Sophie Bushwick. Hentet fra: [How Scientists Are Using AI to Talk to Animals | Scientific American](#)

Shifter (2024) *Nå kommer AI-teamet til Ruter med en slags «Ruter-GPT», for nesten all kollektivtransport til Norge*. 24 juli 2023. Av Joakim Birkeli Jacobsen. Hentet fra: [Nå kommer AI-teamet til Ruter med en slags «Ruter-GPT», for nesten all kollektivtransport i Norge. \(shifter.no\)](#)

Sifted (2024) *Spain to develop open-source LLM trained in Spanish, regional languages*. 28 Februar 2024. Cristina Gallardo. Hentet fra: [Spain to develop open-source LLM trained in Spanish, regional languages | Sifted](#)

Sigma2 (2024a) *Resource Allocation Committee (RFK)*. Hentet fra: [Resource Allocation Committee \(RFK\) | Sigma2](#)

Sigma2 (2024b) *Systems*. Hentet fra: [Systems | Sigma2](#)

Sigma2 (2024c), *Who can apply for resources*. Hentet fra: <https://www.sigma2.no/apply-e-infrastructure-resources#Whocanapply>

Slate (2022) *The Dawn of A.I. Mischief Models*. 3 august 2022. Av Matt Murphy. Hentet fra: [Someone trained an A.I. with 4 chan. It could get worse. \(slate.com\)](#)

Srivastava m.fl (2023) *Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models*. 12 JUNI 2023. Hentet fra: [Beyond the Imitation Game Benchmark \(arxiv.org\)](#)

Stability.ai (2024a) *Stable Diffusion Public Release*. Hentet fra : [Stable Diffusion Public Release – Stability AI](#)

Stability.ai (2024b) *Stable Diffusion 3*. Hentet fra: [Stability AI Image Models – Stability AI](#)

Stability.ai (2024c) *Stable Video Diffusion*. Hentet fra: [Stable Video – Stability AI](#)

Store Norske Leksikon (2021) *Digital infrastruktur*. 25 oktober 2021. Av Kjell Bratbergsgengen. Hentet fra: [digital infrastruktur – Store norske leksikon \(snl.no\)](#)

Store Norske Leksikon (2022) *Superdatamaskiner*. 29 desember 2022. Av Ivar M. Liseter. Hentet fra: [superdatamaskin – IT – Store norske leksikon \(snl.no\)](#).

Store Norske Leksikon (2023) *Turingtesten*. 26 januar 2023. Av Henrik Dvergsdal & Gunnar Karlsen. Hentet fra: [superdatamaskin – IT – Store norske leksikon \(snl.no\)](#).

Svensk författningssamling (2021) *Förordning med instruktion för Myndigheten för psykologisk försvar*. 23 oktober 2021. Hentet fra: [SFS2021-936.pdf](https://www.sfs2021-936.pdf) ([svenskforfattningssamling.se](https://www.svenskforfattningssamling.se))

Synthesia (2024) *Turn text to video, in minutes*. Hentet fra: [Synthesia - Create studio-quality AI videos from text](#)

Talebirad, Yashar & Amirhossein Nadiri (2023) *Multi-Agent Collaboration : Harnessing the Power of Intelligent LLM Agents*. 5 juni 2023. Cornell University. Hentet fra: [\[2306.03314\] Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents \(arxiv.org\)](https://arxiv.org/abs/2306.03314)

Teknologirådet (2018) *Kunstig Intelligens – muligheter, utfordringer og en plan for Norge*. September 2018. Hentet fra: [Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf](#)

Teknologirådet (2019a) *Valgkampen: Behov for nye spilleregler*. 28 Januar 2019. Hentet fra: [Valgkampen: Behov for nye spilleregler - Teknologirådet \(teknologiradet.no\)](#)

Teknologirådet (2019b) *Manipulering av valg er den nye normalen*. 12 september 2019. Av Tore Tennøe. Hentet fra: [Manipulering av valg er den nye normalen - Teknologirådet \(teknologiradet.no\)](#)

Teknologirådet (2020) *Hva skjer med jobbene våre?* September 2020. Hentet fra: [Hva-skjer-med-jobbene-16.-september-lenker.pdf \(wpd.digital\)](#)

Teknologirådet (2022) *Taleteknologi og kunstig intelligens*. 29 desember 2022. Av Jonas Engestøl Wettre. Hentet fra: [Taleteknologi og kunstig intelligens - Teknologirådet \(teknologiradet.no\)](#)

Teknologirådet (2023) *Generativ kunstig intelligens og ytringsfrihet*. 12 desember 2023. Av Hanne Sofie Lindahl. Hentet fra: [Generativ kunstig intelligens og ytringsfrihet - Teknologirådet \(teknologiradet.no\)](#)

Teknologirådet (2024) *Kunstig Intelligens, knappe ressurser og klimaomstilling*. 2 mai 2024. Av Tonje Nerby. Hentet fra: [Kunstig intelligens, knappe ressurser og klimaomstilling - Teknologirådet \(teknologiradet.no\)](#)

The Decoder (2023) *GPT-4 architecture, datasets, costs and more leaked*. 11 juli 2023. Av Maximilian Schreiner. Hentet fra: [GPT-4 architecture, datasets, costs and more leaked \(the-decoder.com\)](https://the-decoder.com)

The Guardian (2023) *UK to invest £900m in supercomputer in bid to build own 'BritGPT'*. 15 mars 2023. Av Dan Milmo & Alex Hern. Hentet fra: [UK to invest £900m in supercomputer in bid to build own 'BritGPT' | Artificial intelligence \(AI\) | The Guardian](https://www.theguardian.com/technology/2023/mar/15/uk-to-invest-900m-in-supercomputer-in-bid-to-build-own-britgpt)

The Guardian (2024a) *Fake Biden robocalls in New Hampshire linked to Texas companies, officials say*. 6 februar 2024. Av Associated Press. Hentet fra: [Fake Biden robocalls in New Hampshire linked to Texas companies, officials say | New Hampshire | The Guardian](https://www.theguardian.com/us-news/2024/feb/06/fake-biden-robocalls-new-hampshire)

The Guardian (2024b) *US outlaws robocalls that use AI-generated voices*. 8 februar 2024. Hentet fra: [US outlaws robocalls that use AI-generated voices | Technology | The Guardian](https://www.theguardian.com/technology/2024/feb/08/us-outlaws-robocalls-that-use-ai-generated-voices)

The Low Down (2023) *The emergence of Large Language Models (LLMs)*. 23 mars 2023. Av Bao Hua Choo. Hentet fra: [The emergence of Large Language Models \(LLMs\) - The Low Down - Momentum Works](https://www.thelowdown.com/ai/the-emergence-of-large-language-models-llms)

The New Yorker (2023) *The inside story of Microsoft's partnership with OpenAI*. 1 desember 2023. Av Charles Duhigg. Hentet fra: [The Inside Story of Microsoft's Partnership with OpenAI | The New Yorker](https://www.newyorker.com/technology/the-inside-story-of-microsofts-partnership-with-openai)

The Straits Time (2024) *More than 100 solutions developed as part of generative AI initiative*. 29 januar 2024. Av Zhaki Abdullah. Hentet fra: [More than 100 solutions developed as part of generative AI initiative | The Straits Times](https://www.straits-times.com/technology/more-than-100-solutions-developed-as-part-of-generative-ai-initiative)

The Verge (2020) *OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws*. 30 Juli 2020. Av James Vincent. Hentet fra: [OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws - The Verge](https://www.theverge.com/2020/7/30/21411111/openai-gpt-3-breakthrough)

The Verge (2023a) *OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'*. 15 mars 2023. Av James Vincent. Hentet fra: [OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong' - The Verge](https://www.theverge.com/2023/3/15/23581111/openai-co-founder-on-company-s-past-approach-to-openly-sharing-research)

The Verge (2023b) *What's new with GPT-4 from processing pictures to acing tests*. 15 mars 2023. Av Mitchell Clark & James Vincent. Hentet fra: [What's new with GPT-4 — from processing pictures to acing tests - The Verge](#)

The Verge (2024a) *Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis*. 21 februar 2024. Av Adi Robertson. Hentet fra: [Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis - The Verge](#)

The Verge (2024b) *Meta will require labels on more AI-generated content*. 5 april 2024. Av Mia Nato. Hentet fra: [Meta will require labels on more AI-generated content - The Verge](#)

The Wall Street Journal (2024) *How I built an AI-powered, self-running propaganda machine for \$105*. 12 april 2024. Av Jack Brewster. Hentet fra: [How I Built an AI-Powered, Self-Running Propaganda Machine for \\$105 - WSJ](#)

The Washington Post (2024) *The future of AI video after Sora is impressive – and flawed*. 28 februar 2024. Av Gerrit De Vynck, Jhaan Elker & Tyler Rimmel. Hentet fra: [The future of AI video after Sora is impressive — and flawed - Washington Post](#)

The White House (2023a) *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence | The White House](#)

The White House (2023b) *FACT SHEET: President Biden issues executive order on safe, secure, and trustworthy artificial intelligence*. 30 oktober 2023. Hentet fra: [FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence | The White House](#)

Time (2023) *Inside Elon Musk's Struggle for the Future of AI*. 9 oktober 2023. Av Walter Isaacson. Hentet fra: [Inside Elon Musk's Struggle for the Future of AI | TIME](#)

Titan (2024) *Store skritt mot norsk svar på ChatGPT*. 6 mars 2026. Universitetet i Oslo. Av Gina Aakre. Hentet fra: [Store skritt mot norsk svar på ChatGPT - Titan.uio.no](#)

TV2 (2023) *Regjeringen vil merke KI-generert innhold: -Det er en fare.* 28 oktober 2023. Av Une Solheim & Martin Leigland. Hentet fra: [Kunstig intelligens: Regjeringen vil merke KI-generert innhold: - Det er en fare \(tv2.no\)](#)

UA (2023) *NorGPT kan bli redningen for det norske språket.* 6 juni 2023. Av Tore Oksholen. Hentet fra: [NorGPT kan bli redningen for det norske språket \(universitetsavisa.no\)](#)

UN (2023) *Secretary-General Announces Creation of New Artificial Intelligence Advisory Board.* 26 oktober 2023. Hentet fra: [Secretary-General Announces Creation of New Artificial Intelligence Advisory Board | Meetings Coverage and Press Releases](#)

UN AI Advisory Body (2023) *Interim report: Governing AI for Humanity.* Desember 2023. Hentet fra: [un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf](#)

UN News (2023) *Guterres calls for AI ‘that bridges divides’, rather than pushing us apart.* 18 juli 2023. Hentet fra: [Guterres calls for AI ‘that bridges divides’, rather than pushing us apart | UN News](#)

Unesco (2024) *Challenging systematic prejudices: an investigation into bias against women and girls in large language models.* Hentet fra: [Challenging systematic prejudices: an investigation into bias against women and girls in large language models - UNESCO Digital Library](#)

Universitetet i Oslo (2019) *Behov og finansieringsstrategi for nasjonal e-infrastruktur for forskning perioden 2020 – 2030.* 15. august 2019. Hentet fra: https://www.mn.uio.no/hylleraas/english/news-and-events/news/2019/einfrastruktur2030_horing.pdf

Universitetet i Oslo (2024a) *GPT UiO – UiOs personverntrygge KI-chat.* Hentet fra [GPT UiO – UiOs personverntrygge KI-chat - Universitetet i Oslo](#)

Universitetet i Oslo (2024b) *Fox: Tungregningsklynge for brukere av Educloud Research.* Hentet fra: [Fox: Tungregningsklynge for brukere av Educloud Research - Universitetet i Oslo \(uio.no\)](#)

University of Oxford (2023) *Generative AI can potentially disrupt labour markets, say Oxford experts 10 years after ground-breaking study.* 26 september

2023. Hentet fra: [Generative AI can potentially disrupt labour... | Oxford Martin School](#)

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser & Illia Polosukhin (2023) *Attention is all you need*. 2 august 2023. Hentet fra: [\[1706.03762\] Attention Is All You Need \(arxiv.org\)](#)

Venture Beat (2023) *Microsoft releases Orca 2, a pair of small language models that outperform larger counterparts*. 20 november 2023. Av Shubham Sharma. Hentet fra: [Microsoft releases Orca 2, a pair of small language models that outperform larger counterparts | VentureBeat](#)

Verifiserbar.no (2024) *Faktisk. Verifiserbar*. Hentet fra: [VERIFISERBAR.NO](#)

Vice (2023) *Adobe is selling AI-generated images of violence in Gaza and Israel*. 7 November 2023. Av Matthew Gault. Hentet fra: [Adobe Is Selling AI-Generated Images of Violence in Gaza and Israel \(vice.com\)](#)

Villalobos, Pablo, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn & Anson Ho (2022) *Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning*. 26 oktober 2022. Hentet fra: [2211.04325 \(arxiv.org\)](#)

Widder, David G., Sarah West & Meredith Whittaker (2023) *Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI*. 18 august 2023. SSRN. Hentet fra: [Open \(For Business\): Big Tech, Concentrated Power, and the Political Economy of Open AI by David Gray Widder, Sarah West, Meredith Whittaker :: SSRN](#)

Wired (2024a) *How an iPhone Powered by Google's Gemini AI Might Work*. 1 april 2024. Av Boone Ashworth. Hentet fra: [How an iPhone Powered by Google's Gemini AI Might Work | WIRED](#)

Wired (2024b) *Forget Chatbots. AI Agents are the Future*. 14 mars 2024. Av Will Knight. Hentet fra: [Forget Chatbots. AI Agents Are the Future | WIRED](#)

World Economic Forum (2024) *The Global Risks Report 2024*. Hentet fra: [WEF The Global Risks Report 2024.pdf\(weforum.org\)](#)

WPP (2022) *Ogilvy and Wavemaker: #NotJustACadburyAd and Shah Rukh Khan-My-Ad*. Hentet fra: [Ogilvy and Wavemaker: #NotJustACadburyAd and Shah Rukh Khan-My-Ad | WPP](#)