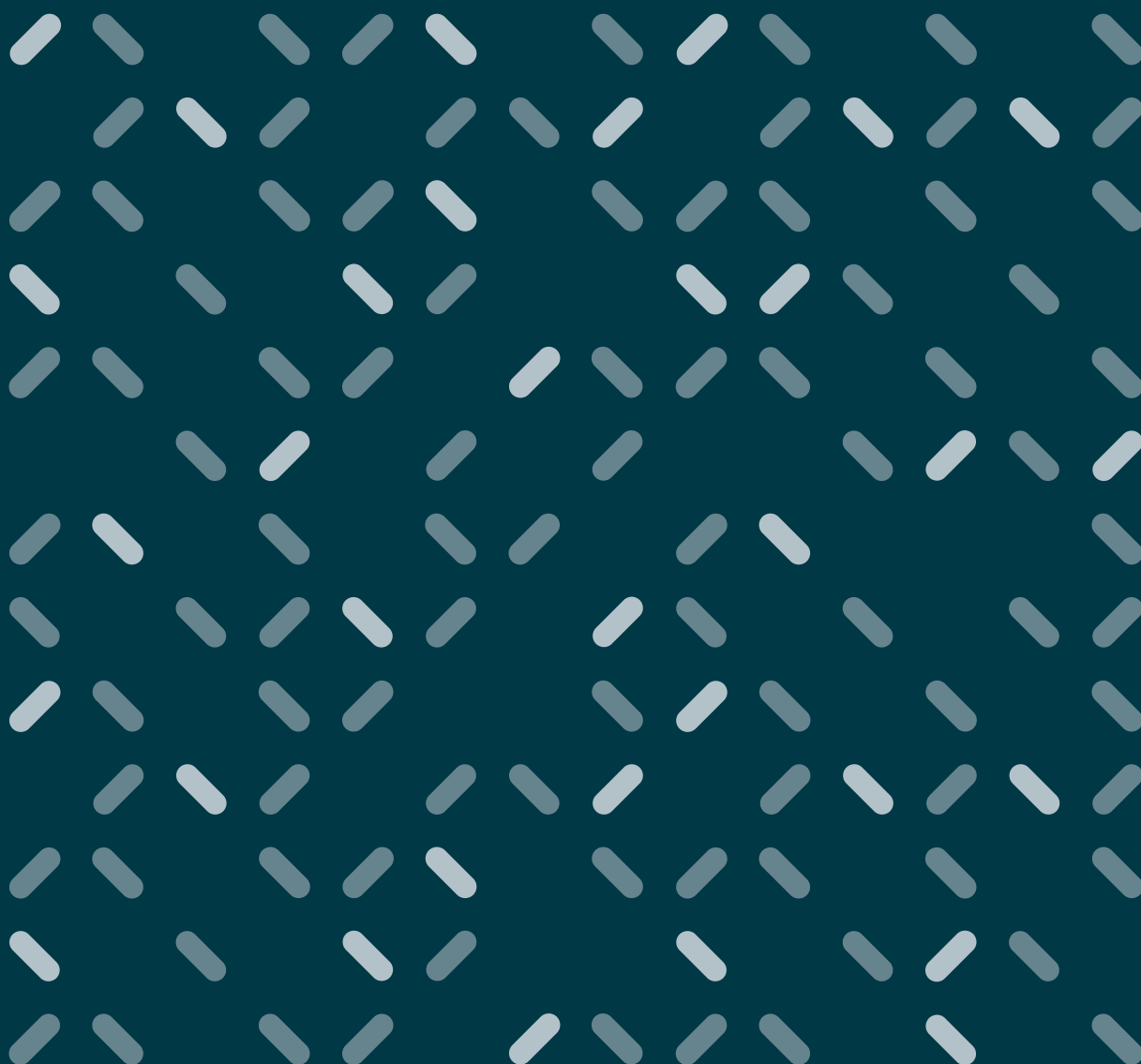


Personvernforemmende teknologier for bruk av kunstig intelligens i helse- og omsorgstjenesten

Makhlysheva A., Bakkevoll P.A., Yigzaw K.Y.



Personvernforemmede teknologier for bruk av kunstig intelligens i helse- og omsorgstjenesten

Rapportnummer

03-2023

Prosjektleder

Per Atle Bakkevoll

Forfattere

Alexandra Makhlysheva

Per Atle Bakkevoll

Kassaye Yitbarek Yigzaw

ISBN

978-82-8242-109-6

Dato

24.10.2023

Antall sider

35

Emneord

Personvern, personvernforemmede teknologi, helsedata, kunstig intelligens, føderert læring, syntetiske data

Oppsummering

Innsikt i helsedata kan gi bedre forsknings- og behandlingsgrunnlag, øke diagnostisk nøyaktighet, forbedre behandlingsresultater og effektivisere ressursbruk. Samtidig er bruken av helseopplysninger strengt regulert. Personvernforemmede teknologier muliggjør dataanalyser samtidig som de ivaretar personvern. Denne rapporten inneholder kunnskapsoppsummeringer om to personvernforemmede teknologier: føderert læring og syntetiske data. Begge teknologiene har sine bruksområder, fordeler og utfordringer. Det er behov for mer forskning, videreutvikling og praktiske erfaringer før teknologiene kan brukes i helsetjenesten.

Utgiver

Nasjonalt senter for e-helseforskning

Postboks 35

9038 Tromsø

E-post: mail@ehealthresearch.no

Internett: www.ehealthresearch.no

Det kan fritt kopieres fra denne rapporten hvis kilden oppgis. Brukeren oppfordres til å oppgi rapportens navn, nummer, samt at den er utgitt av Nasjonalt senter for e-helseforskning og at rapporten i sin helhet er tilgjengelig på www.ehealthresearch.no.

© 2025 Nasjonalt senter for e-helseforskning

Innholdsfortegnelse

1	Sammendrag	4
2	Introduksjon	5
3	Bakgrunn	6
3.1	Personopplysninger og personvern	6
3.2	Personvernregulering.....	6
3.3	Personvern fremmende teknologier.....	7
3.3.1	Eksempler på personvern fremmende teknologier.....	7
3.3.2	Kryptografiske metoder for personvernbevaring	8
4	Føderert læring	9
4.1	Hva er føderert læring?.....	9
4.2	Hvilke fordeler og utfordringer har føderert læring?.....	10
4.2.1	Fordeler.....	10
4.2.2	Utfordringer	10
4.3	Føderert læring og europeisk personvernregulering.....	13
4.4	Eksempler på prosjekter som bruker føderert maskinlæring på helseområdet.....	15
4.4.1	Internasjonale prosjekter.....	15
4.4.2	Prosjekter med norsk deltakelse.....	16
4.5	Oppsummering	17
5	Syntetiske data	18
5.1	Hva er syntetiske data?	18
5.2	Mulige bruksområder for syntetiske data	18
5.2.1	Trene, validere og teste KI-modeller.....	18
5.2.2	Tilgjengeliggjøre mer data til forskning	19
5.3	Hvilke fordeler og utfordringer har syntetiske data?	20
5.3.1	Styrkes personvern ved bruk av syntetiske data?	20
5.3.2	Kan syntetiske data forbedre kvaliteten til treningsdata?	21
5.3.3	Er bruk av syntetiske data effektiv?	21
5.4	Metoder og verktøy for generering av syntetiske data.....	22
5.5	Hvordan kan vi evaluere datakvalitet og personvern i syntetiske datasett?.....	22
5.5.1	Evaluering av datakvalitet.....	22
5.5.2	Evaluering av personvern.....	23
5.5.3	Behov for bedre og mer standardiserte evalueringsmetoder	24
5.6	Oppsummering	24
6	Konklusjon	25
7	Referanseliste	27

Tabelliste

Tabell 1. Potensielle angrep på et føderert konsortium	11
Tabell 2. Beskyttelsesmekanismer mot potensielle angrep på et føderert konsortium.....	12
Tabell 3. Samsvar med personvernprinsipper: føderert læring contra maskinlæring på sentraliserte data.....	13

Figurliste

Figur 1. Føderert læring med den sentrale serveren i helsevesenet. Kilde (27)	9
Figur 2. Eksempel på generering og bruk av syntetiske data i helse- og omsorgstjenesten. Kilde (87)	19

1 Sammendrag

Helsetjenesten produserer store mengder data. Innsikt fra helsedata kan gi bedre grunnlag for forskning og kvalitetsforbedring, som kan føre til bedre diagnostisk nøyaktighet, bedre behandlingsresultater og mer effektiv ressursbruk. Maskinlæring (ML) kan brukes til å analysere store datamengder raskt og effektivt. ML er et underfelt av kunstig intelligens (KI) som oppdager og tolker mønstre i data og gjør prediksjoner på nye data som kommer inn i KI-systemet. Både mengden og kvaliteten på data påvirker resultatet fra en maskinlæringsmodell.

Innsamling, annotering, tilrettelegging og vedlikehold av data til gjenbruk er imidlertid tid- og ressurskrevende. Helseopplysninger er sensitiv informasjon og bruken er regulert av både det generelle personvernregelverket og særlovgivning i helsesektoren. Begrenset tilgang til data som er nødvendige for trening av ML-modeller kan derfor være en betydelig flaskehals for utbredelsen av kunstig intelligens i helse- og omsorgstjenesten.

Det nasjonale koordineringsprosjektet «Bedre bruk av kunstig intelligens» (1) har hatt som formål å hjelpe og veilede helsetjenesten slik at den kan lykkes med å ta i bruk kunstig intelligens (KI) på en trygg måte. Som en del av dette prosjektet har Direktoratet for e-helse publisert rapporten «Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten» (2). Rapporten anbefaler blant annet å etablere samarbeid med forsknings- og fagmiljøer om å utarbeide kunnskapsgrunnlag om personvern fremmende teknologier som tillater å samle, behandle, analysere og dele data, mens datasikkerhet og personvern ivaretas. Med bakgrunn i dette har Nasjonalt senter for e-helseforskning (NSE), i samarbeid med Direktoratet for e-helse, utarbeidet denne kunnskapsoppsummeringer om føderert læring og syntetiske data.

Føderert læring (*federated learning* (FL)) er en type maskinlæring som muliggjør distribuert dataanalyse uten å dele data mellom organisasjoner. Teknologien ivaretar personvern ved at de lokale dataene ikke forlater den organisasjonen de tilhører og at organisasjonen har kontroll over dataenes bruk. Bruk av teknologien gir mulighet til å få større og mer representative datagrunnlag som kan bidra til høyere kvalitet på kliniske beslutninger og bedre helsehjelp til pasienter, uavhengig av behandlingssted og sykdom. Mange helserelaterte prosjekter har brukt føderert læring med lovende resultater. Fødererte konsortier må imidlertid ta hensyn til en rekke forhold, som variasjoner i data og tekniske spesifikasjoner i organisasjonene, eventuell økning i ressursbruk, potensielle kommunikasjonsflasker og datasikkerhetstrusler. Det trengs også metoder for å beregne statistiske parametere på distribuerte datasett uten at personvernet trues. Kryptografiske og andre personvern fremmende teknikker må videreutvikles. Det kreves i tillegg mer forskning på hvordan behovet for nøyaktige dataanalyseresultater kan balanseres med hensynet til personvern.

Syntetiske data er kunstig genererte data som produseres ved å trene en generativ maskinlæringsmodell med reelle data. Syntetiske data beholde de statistiske egenskapene til det originale datasettet, uten å inneholde direkte personidentifiserbar informasjon. Kvaliteten til syntetiske data avhenger av kvaliteten til det originale datasettet. Ved å bruke syntetiske data kan datagrunnlaget utvides på områder der det ikke finnes tilstrekkelig med ekte treningsdata. Dette kan gi et datagrunnlag som er mer rettferdig og har færre systematiske skjevheter. I kombinasjon med andre personvern fremmende teknikker kan syntetiske data styrke personvernet. Risikoen for re-identifisering i syntetiske data er mindre enn for anonymiserte eller aidentifiserte ekte data, men informasjonslekkasje fra det originale datasettet er fortsatt mulig. Det er flere initiativer som benytter teknologien innen helsefeltet.

Både føderert læring og syntetiske data har sine fordeler og utfordringer. Vi ser behov for mer forskning, videreutvikling og praktiske erfaringer for å realisere disse teknologienes fulle potensial i helse- og omsorgssektoren. Dette vil bidra til å sikre forsvarlig og nyttig bruk av KI i helsetjenesten.

2 Introduksjon

Det nasjonale koordineringsprosjektet «Bedre bruk av kunstig intelligens» har som formål å hjelpe og veilede helsetjenesten slik at den kan lykkes med å ta i bruk kunstig intelligens på en trygg måte. Prosjektet er ledet av Helsedirektoratet som samarbeider med Direktoratet for e-helse, Statens legemiddelverk, Helsetilsynet, Folkehelseinstituttet, Kommunesektorens organisasjon og alle de fire regionale helseforetakene. Som en del av prosjektet publiserte Direktoratet for e-helse rapporten «Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten». Rapporten tydeliggjør utfordringer knyttet til datatilgang til forskning, utvikling, validering og bruk av KI-løsninger i helse- og omsorgstjenesten og kommer med anbefalinger for hvordan løse dem. Det anbefales blant annet å etablere samarbeid med forsknings- og fagmiljøer om å utarbeide kunnskapsgrunnlag for bruk av personvern fremmende teknologier som kan være nyttige i helse- og omsorgstjenesten. Med bakgrunn i dette, i samarbeid med Hilde Margrethe Lovett og Inger Anne Tøndel fra Direktoratet for e-helse, har NSE skrevet en kunnskapsoppsummering om de to personvern fremmende teknologiene: føderert læring og syntetiske data.

For å utvikle gode maskinlæringsmodeller er det avgjørende med tilgang til data av høy kvalitet og i tilstrekkelig mengde for trening, validering og testing. Helseinstitusjoner genererer og lagrer mye helsedata. Disse dataene har et stort potensial for å gi ny innsikt i sykdomsutvikling, øke diagnostisk nøyaktighet og forbedre behandlingsresultater. Innsamling, annotering, tilrettelegging og vedlikehold av data av høy kvalitet er tidkrevende og krever betydelige menneskelige ressurser. I tillegg er tilgang til og/eller deling av helsedata utenfor helseinstitusjoner ofte svært begrenset, blant annet på grunn av personvernhensyn. Helsedata regnes som sensitive personopplysninger og bruken er strengt regulert. Bruk av helsedata som treningsdata regnes som sekundærbruk, noe som innebærer at det kreves samtykke fra pasientene eller fritak fra taushetsplikten dersom opplysningene ikke er anonyme. På grunn av den tidkrevende og omfattende prosessen med å sikre at regelverket for personvern og krav til informasjonssikkerhet etterleves, representerer tilgang til treningsdata en betydelig flaskehals for utbredelsen av KI i helse- og omsorgstjenesten. Bruk av syntetiske data kan gjøre tilgangen til treningsdata enklere og raskere, mens føderert læring fremstår som en lovende teknikk for å benytte distribuerte data og beregningsressurser og samtidig ivareta datasikkerhet og personvern.

I neste kapittel introduserer vi personopplysninger, personvernregulering og forklarer behovet for personvern fremmende teknologier. I kapittel 4 snakker vi om føderert læring: hva teknologien innebærer, dens fordeler og utfordringer, beskyttelsesmekanismer mot ulike sikkerhetsangrep, samsvar med personvernprinsippene, samt eksempler på helse relaterte prosjekter som benytter teknologien. I kapittel 5 fokuserer vi på syntetiske data: hva dette handler om, hvilke fordeler og utfordringer bruken av syntetiske data innebærer, mulige bruksområder, eksempler på norske miljøer som jobber med syntetiske data, ser på hvordan vi kan evaluere datakvalitet og personvern i syntetiske datasett, samt metoder og verktøy for generering av syntetiske data. Til slutt oppsummerer vi kunnskapen om begge teknologiene og konkluderer om de kan tas i praksis i helse- og omsorgstjenesten.

3 Bakgrunn

3.1 Personopplysninger og personvern

Personopplysninger (3) er alle opplysninger og vurderinger som kan knyttes til en enkeltperson. Eksempler på personopplysninger er navn, adresse, telefonnummer, e-post eller fødselsnummer. Biometriske data er også personopplysninger. I tillegg kan et bilde regnes som en personopplysning hvis personene på bildet kan gjenkjennes (3). I noen tilfeller kan registreringsnummeret på en bil, en dynamisk IP-adresse eller lydopptak bli definert som personopplysninger (3).

I dagens digitale verden legger vi igjen mange digitale spor: hva og hvor vi handler registreres ved betaling med kort, hva vi søker etter på nettet, vår geolokalisering osv. Disse og andre opplysninger om personlige atferdsmønstre regnes også som personopplysninger (3).

Personvern handler om retten til et privatliv og til å bestemme over egne personopplysninger. Hver enkelt av oss har rett til å kontrollere eller påvirke hvordan informasjonen om oss skal samles inn, brukes og lagres. Samtidig bør personopplysninger ikke kunne spores tilbake, direkte eller fra statistiske utdata (4). Det finnes særlige kategorier av personopplysninger, også kalt sensitive personopplysninger, som er ekstra strengt regulert. Dette gjelder blant annet helse- og genetiske opplysninger, biometriske data (f.eks. fingeravtrykk, irismønster og hodeform), samt informasjon om religion, politisk oppfatning og fagforeningsmedlemskap (5).

Det sentrale regelverket som gjelder personopplysninger (5), er personopplysningsloven som består av nasjonale regler på enkelte områder og EUs personvernforordning (GDPR) (6).

3.2 Personvernregulering

EUs personvernforordning (*General Data Protection Regulation (GDPR)*) er et sett regler som gjelder for alle EU/EØS-land. GDPR trådte i kraft i 2018 og har som formål å beskytte personopplysninger (6). Personvernforordningen gir en rekke rettigheter til enkeltpersoner. Hver enkelt skal ha tilgang på nødvendig informasjon om hvorfor og hvordan deres opplysninger skal samles inn, brukes og lagres. De skal også ha mulighet til å gi aktivt samtykke til datainnsamling eller reservere seg mot den og få all sin informasjon slettet dersom de ønsker det (7).

Forordningen definerer i tillegg en rekke regler for virksomheter for innsamling og bruk av data, samt tiltak som skal iverksettes for å beskytte innsamlede personopplysninger. Dette innebærer bl.a. følgende (8):

- All bruk av personopplysninger må ha et behandlingsgrunnlag (et rettslig grunnlag) (9). Det betyr at for å bruke informasjonen om en enkeltperson, må virksomheten kunne vise til med hvilken begrunnelse de skal gjøre det.
- Formålet med datainnsamlingen må være tydelig og enkelt å forstå, og innsamlede data skal kun brukes til dette spesifikke formålet.
- Manglende samtykke til innsamling av opplysninger bør ikke være til hinder for å få tilgang til en tjeneste. Personen skal også kunne trekke tilbake sitt samtykke, og dette skal man informeres om. Samtykker skal loggføres med eventuelle endringer og tilbaketrekkinger.
- Mengden på personopplysninger som skal samles inn, må minimeres til det som er nødvendig for å realisere formålet med innsamlingen. Når personopplysninger ikke lenger er nødvendige for formålet de ble innhentet for, skal de slettes eller anonymiseres.
- Behandlingen av personopplysninger skal være forståelig for de registrerte og i respekt for deres interesser og ikke foregå på skjulte eller manipulerende måter.
- Bruk av personopplysninger skal være oversiktlig og forutsigbar slik at registrerte blir i stand til å bruke sine rettigheter og ivareta sine interesser.

- Innsamlede personopplysninger skal være korrekte og skal oppdateres om nødvendig (slettes eller rettes).
- Virksomheten må sørge for å iverksette tiltak mot utilsiktet og ulovlig ødeleggelse, tap og endringer av personopplysninger.
- Virksomheten må kunne dokumentere iverksatte nødvendige organisatoriske og tekniske tiltak for å sikre at regelverket etterleves til enhver tid.

De virksomhetene som ikke følger GDPR-retningslinjene, risikerer å miste kundenes tillit som en seriøs og troverdig aktør, i tillegg til å bli ilagt relativt store bøter.

Felles europeisk helsedataområde (*European Health Data Space (EHDS)* på engelsk) er et rammeverk for deling av helsedata i EU (10). EHDS ble fremmet som et reguleringsforslag av Europakommisjonen i 2022 med målet om å gi EU-borgere bedre kontroll over sine personlige helseopplysninger (10). Ifølge Europakommisjonen skal EHDS samsvare med GDPR (6), dataloven (11), lov om datastyring (12), og NIS2-direktivet om cybersikkerhet (13), noe som skal fremme personvern og datasikkerhet for europeiske helsedata.

3.3 Personvern fremmende teknologier

Bruk av kunstig intelligens til helsedataanalyse kan forbedre helse- og omsorgstjenester ved å gi bedre datagrunnlag for helseforskning og innovasjon og effektivisere ressursbruk. For pålitelige data-analyseresultater trenger KI data av høy kvalitet og i tilstrekkelig mengde. Samtidig er bruk av helsedata ofte begrenset: å få tilgang til helsedata er en tidkrevende og omfattende prosess som skal sikre at personvern regelverket og krav til informasjonssikkerhet etterleves (14). Personvern fremmende teknologier (*privacy-enhancing technologies* på engelsk), også kalt personvernbevarende teknologier (*privacy-preserving technologies* på engelsk), er et bredt spekter av teknologier for å samle, behandle, analysere og dele data mens datasikkerhet og personvern ivaretas (15). Nedenfor kommer det eksempler på de mest brukte personvern fremmende teknologier.

3.3.1 Eksempler på personvern fremmende teknologier

Føderert læring

Føderert læring (16) er en maskinlæringsteknikk som gjør det mulig for individuelle enheter eller systemer å samarbeide om å trene en maskinlæringsmodell på lokale data som ikke forlater sine organisasjoner for denne felles analysen. Hver enhet eller organisasjon laster ned en gjeldende sentral modell, forbedrer den ved å lære av egne data og laster opp kun aggregerte (oppsummerte) endringer til den sentrale serveren som koordinerer hele læringsprosessen. Koordinatoren beregner endringene i den sentrale modellen basert på alle opplastede lokale oppdateringer.

Identifisering av atferdsmønstrene til brukere kan gjøres ved å analysere brukernes handlinger uten å sende individuelle data til en ekstern server. Læring på enheten (*edge machine learning* på engelsk) (17) kan brukes til å forbedre algoritmer, for eksempel autokorrektur på mobiltelefoner.

Pseudonymisering

Pseudonymisering (18) erstatter eller skjuler sensitiv informasjon ved å bytte ut sensitive opplysninger med fiktive data. Dette er en vanlig mekanisme for å beskytte brukernes sensitive data og overholde personvernregelverket. Likevel kan noen anonymiseringsteknikker, som sletting av kolonner som inneholder personlig identifiserbar informasjon eller datamaskering, være utsatt for re-identifisering.

Generering av syntetiske data

Syntetiske data er kunstige data som er generert fra et rådatasett og har samme statistiske egenskaper som det opprinnelige datasettet. Syntetiske data kan være nyttige for å redusere behovet for datadeling og mengden av reelle data som er nødvendige. En måte å generere syntetiske data på er å

bruke generative adversarial networks (GAN) (19). GAN-modeller trenes ved å bruke to nevralt nettverk, der det ene produserer syntetiske data og det andre lærer seg å skille mellom syntetiske og ekte data. Denne metoden kan produsere store mengder syntetiske data av høy kvalitet. GAN-er kan gjenkjenne komplekse mønstre i data og brukes derfor blant annet også til å finne anomalier i medisinske data.

De ovennevnte personvern fremmende teknologiene brukes ofte sammen med ulike kryptografiske metoder.

3.3.2 Kryptografiske metoder for personvernbevaring

Differensielt personvern

Differensielt personvern (*differential privacy* på engelsk) (20) legger til tilfeldig støy i dataene, slik at egenskapene til et datasett bevares på gruppenivå, men ikke på individnivå. Teknologien kan benyttes f.eks. når store datasett tilgjengeliggjøres for offentlig forskning.

Homomorfsk kryptering

Homomorfsk kryptering (*homomorphic encryption* på engelsk) (21) muliggjør beregningsoperasjoner på krypterte data. Resultatene av analyser forblir kryptert, og bare dataeieren kan dekryptere dem. Denne metoden gjør det for eksempel mulig å analysere krypterte data i skylagring.

Sikker flerpartsberegning

Sikker flerpartsberegning (*secure multiparty computation* på engelsk) (22) er en kryptografisk teknikk som tillater å utføre distribuerte beregninger på tvers av systemer og flere krypterte datakilder. Denne teknikken sikrer at deltakende parter får kun aggregert informasjon nødvendig til å utføre beregninger, men aldri hele datasettet.

Zero-knowledge proof

Zero-knowledge proof (23) er et sett av kryptografiske metoder for å bevise gyldigheten av informasjon uten å avsløre selve informasjonen. Det gjøres ved å løse en utfordring som gir brukere muligheten til å bevise kunnskap om informasjonen uten å avsløre detaljene om informasjonen. Metoden brukes blant annet til identitetsautentisering.

Flere detaljer om personvern fremmende teknikker kommer videre i teksten (se *Tabell 2*).

Rapporten fokuserer ellers på to personvern fremmende teknologier som kan være nyttige i helse- og omsorgstjenesten: føderert læring og syntetiske data.

4 Føderert læring

Helseinstitusjoner produserer mye helsedata. Disse dataene har et enormt potensial til å transformere helse- og omsorgstjenesten på mange måter, til fordel for både enkeltpersoner, tjenesten selv og samfunnet som helhet. Samtidig regnes helsedata som sensitive personopplysninger og bruken er strengt regulert; noe som begrenser tilgang til og deling av helsedata utenfor helseinstitusjonene. Anonymisering av data er ofte ikke nok for å ivareta personvernet. Det er for eksempel mulig å rekonstruere pasientens ansikt fra CT-, MR- eller PET-data (24–26). Føderert læring ser ut som en potensiell løsning på denne problemstillingen.

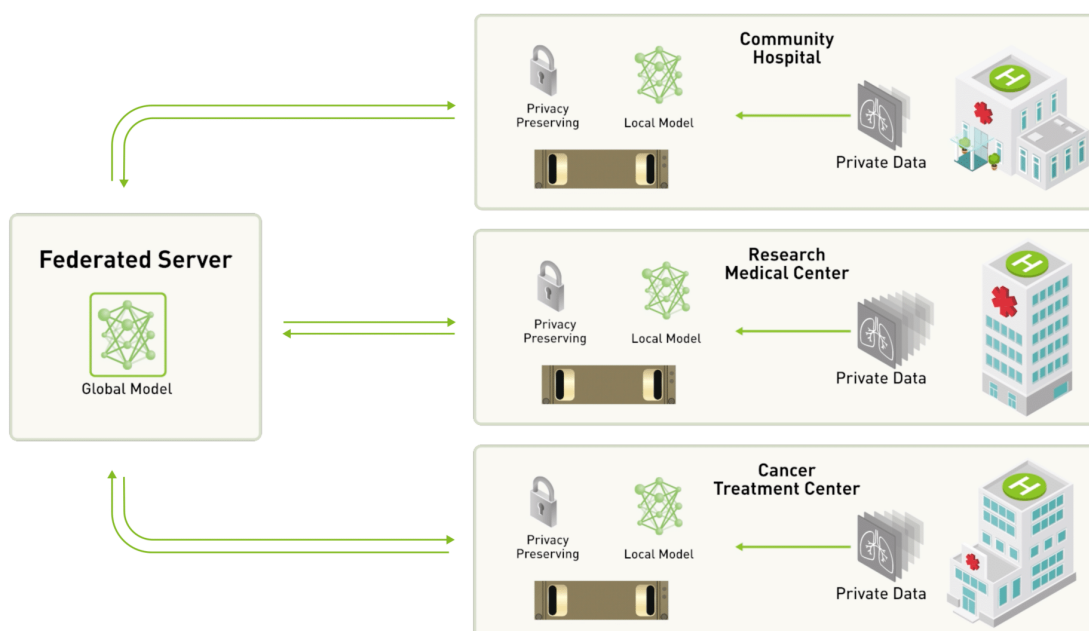
4.1 Hva er føderert læring?

Føderert læring gjør det mulig å bruke maskinlæringsmetoder for å *analysere dataene der de ligger lagret og unngå at de blir synlige for eller deles med eksterne aktører*. Dette gjør teknologien mer personvernvennlig.

Helseinstitusjoner kan etablere et føderert konsortium (et desentralisert nettverk av organisasjoner) for å *trene en ML-modell uten å utveksle sensitive data med hverandre*. Hver organisasjon kan ha ulike roller i konsortiet: delta i selve modelltreningen, validere ferdig trente modeller eller kun støtte enkle spørringer.

En modelltreningsrunde i et føderert konsortium kan se ut som følger (se *Figur 1*). Først lastes den nyeste versjonen av en global modell ned av alle organisasjonene i konsortiet. Deretter trenes modellen opp lokalt i hver organisasjon basert på lokale data. Så sendes lokale modelloppdateringer til en sentral server. Der beregnes det et gjennomsnitt av alle mottatte lokale modelloppdateringer for å forbedre den globale modellen. Deretter er den oppdaterte globale modellen klar til å lastes ned av alle deltakerne. Denne prosessen er *iterativ* og avsluttes når modellen er ferdig trent.

Figur 1. Føderert læring med den sentrale serveren i helsevesenet. Kilde (27)



Verdens økonomiske forum har publisert en veiledning som beskriver hvordan et føderert konsortium for deling av sensitive helsedata kan etableres. Her peker de på suksessfaktorene for å skape et solid konsortium som tar sikte på å analysere distribuerte helsedata og bevare personvern, datasikkerhet og -integritet (28). Når helseinstitusjoner etablerer et slik konsortium, er det en forutsetning at de må ta hensyn til ulikheter i utstyr, infrastruktur og data i organisasjonene, samt hvordan disse ulikhetene skal håndteres. Etablerte «kjøreregler» gjør det relativt enkelt å legge til flere helseinstitusjoner i det fødererte konsortiet.

4.2 Hvilke fordeler og utfordringer har føderert læring?

Når man vurderer å ta i bruk føderert læring, må både fordeler og utfordringer med teknologien vurderes.

4.2.1 Fordeler

Styrket personvern, økt kontroll over egne data og redusert risiko for datainnbrudd

Forskning viser at ML-modeller trent med føderert læring kan være like gode som de som er trent på sentraliserte data (29). Samtidig er fødererte data generelt mindre sårbare for angrep enn store sentraliserte databaser. FL-tilnærmingen er mer personvernvennlig fordi dataprosessering og håndtering av personvernregler og andre typer retningslinjer skjer lokalt. Sykehus og klinikker beholder kontroll over pasientdata ved at de bestemmer hvordan dataene kan brukes og de kan spore bruken. Pasienter kan være trygge på at dataene forblir hos helseinstitusjonen og at datatilgang som er gitt kan trekkes tilbake. Dette kan bidra til at innbyggere blir mer villige til å dele sine data til forskning og utvikling.

Større og mer representativt datagrunnlag – mer presis helseforskning og bedre helsehjelp

Føderert læring gir mulighet til å analysere større datasett (dvs. både flere pasientdataposter og flere datavariabler tilgjengelige for analyse). Derfor blir dataene ofte mer representative, som kan være spesielt aktuelt ved sjeldne medisinske tilstander eller andre tilfeller der pasientgruppene er små. Tilgang til mer data på tvers av ulike pasientgrupper og over tid muliggjør bl.a. større og bedre helseforskningsstudier og mer presise ML-baserte beslutningsstøttesystemer. Føderert læring kan øke nøyaktigheten og robustheten til KI-systemer i helsevesenet og hjelpe med å ta KI i bruk i stor skala i helsetjenesten (30). Det kan utvide klinikernes ekspertise, sikre mer konsistente kliniske beslutninger av høy kvalitet og gi pasienter lik kvalitet på helsehjelpen, uavhengig av behandlingssted og sykdom.

4.2.2 Utfordringer

Variasjoner i tekniske spesifikasjoner

Lagrings-, beregnings- og kommunikasjonskapasitetene til organisasjoner kan variere (31). En organisasjon kan falle ut i løpet av en treningsrunde på grunn av tilkoblingsbegrensninger eller utilstrekkelig lagringskapasitet. I tillegg kan en organisasjon bestemme seg for å ikke være med i treningsrunden. Derfor må systemer som bruker føderert læring, ha mekanismer for å overvåke og tolerere (a) lav deltakelse i treningsprosessen, (b) variasjoner i maskinvare og (c) at flere deltakere kan falle ut i løpet av treningsprosessen.

Variasjoner i data

Data i helseinstitusjoner kan variere betydelig med hensyn til kvantitet, kvalitet og datastrukturer. Rutiner for å sjekke datakvalitet, samt datastandardisering og -harmonisering er viktig ved utføring av føderert læring (32). Dårlig kvalitet på dataene (dvs. dupliserte, unøyaktige, inkonsistente, manglende eller feilstavete data) vil føre til skjeve modellresultater. I et føderert konsortium er det

vanskelig på forhånd å måle den statistiske heterogeniteten til treningsdataene (dvs. å forutsi hvor ujevnt helsedata er fordelt på ulike institusjoner og i hvilken grad dette kan påvirke resultatet). Det er derfor utfordrende å redusere mulige negative konsekvenser denne heterogeniteten kan medføre, blant annet økt kompleksitet ved datamodellering, -analyse og -evaluering (31). Men det finnes for eksempel tilnærminger som metalæring (33) og multitask-læring (34), som muliggjør trening av flere lokale modeller samtidig og håndterer dataenes statistiske heterogenitet på en transparent og etterprøvable måte.

Kommunikasjonsflaskehals

Føderert læring innebærer intensiv meldingsutveksling mellom deltakende organisasjoner og den sentrale serveren for å dele modelloppdateringer. Kommunikasjonsinfrastrukturen til deltakerne må være motstandsdyktig mot kommunikasjonsfeil og forsinkelser. Det er derfor nødvendig å iverksette metoder som sikrer effektiv kommunikasjon og synkronisering (35) og reduserer antall kommunikasjonsrunder og størrelsen på meldingene i hver runde.

Ressursbruk

Systemer som bruker teknologien, kan være ressursbesparende og ha lavere miljøavtrykk sammenlignet med sentraliserte ML-tilnærminger (36). Likevel kan ressursbruken øke ved intensiv meldingsutveksling og bruk av flere personvernbevarende teknikker. For å delta i et føderert konsortium trengs det i tillegg investeringer i beregningsinfrastruktur, særlig ved deltakelse i modelltreningsprosessen; noe som er ganske kostbart.

Potensielle sikkerhetstrusler

Det finnes utfordringer med informasjonssikkerhet i alle systemer. Selv om de lokale dataene til helseinstitusjonene ikke blir eksponert i systemer som bruker føderert læring, eksisterer det en risiko for informasjonsslekkasjer fra meldingene helseinstitusjonene utveksler ved modelloppdateringene. Det er derfor et krav til slike systemer å ha et sikkert kommunikasjonssystem. I et føderert konsortium av helseinstitusjoner kan deltakerne anses som pålitelige og det er nødvendig at de følger en felles modelltreningsprotokoll og bruker deres ekte data til trening. Det er imidlertid en risiko for at en inn-trenger tar kontroll over den sentrale serveren eller deltakende helseinstitusjoner og forsøker å utføre ulike angrep (37) (se Tabell 1).

Tabell 1. Potensielle angrep på et føderert konsortium

Angrep	Mål	Kilde til angrep	
		Helseinstitusjoner	Den sentrale serveren
Forgiftningsangrep på data og modell (<i>data and model poisoning attacks</i>)	Manipulere lokale data eller en lokal modell for å innføre skjevheter i den globale modellen	+	+
Uttrekkingsangrep (<i>inference attack</i>)	Analysere lokale modelloppdateringer for å få kunnskap om treningsprosessen for å hente ut meningsfull innsikt om lokale data	+	+
«Gratispassasjer»-angrep (<i>free-ride attack</i>)	Laste opp falske lokale oppdateringer for å få tak i den globale modellen uten faktisk å delta i treningsprosessen	+	-

Flere mekanismer kan benyttes for å styrke datasikkerhet og personvern i et føderert system (37) (se Tabell 2).

Tabell 2. Beskyttelsesmekanismer mot potensielle angrep på et føderert konsortium

Beskyttelsesmekanisme	Hvilke angrep kan mekanismen hjelpe mot?	Hvordan fungerer det?
Differensiert personvern (<i>differential privacy</i>) (20)	<ul style="list-style-type: none"> • Forgiftningsangrep • Uttrekkingsangrep 	Tilfeldig støy tilføres til lokale sensitive data før deling av lokale modelloppdateringer med den sentrale serveren.
Sikker flerpartsberegning (<i>secure multiparty computation</i>) (22)	<ul style="list-style-type: none"> • Uttrekkingsangrep 	Lokale modelloppdateringer krypteres før de sendes til den sentrale serveren. Den sentrale serveren ser bare de krypterte lokale modelloppdateringene og beregning av den globale modellen utføres med krypterte data.
Oppdagelse av anomalier (<i>anomaly detection</i>) (38)	<ul style="list-style-type: none"> • «Gratispassasjer»-angrep • Forgiftningsangrep 	Lokale modelloppdateringer analyseres for å identifisere ondsinnede deltakere.
Robust aggregering (<i>robust aggregation</i>) (39)	<ul style="list-style-type: none"> • Forgiftningsangrep • Uttrekkingsangrep 	Før de mottatte lokale modelloppdateringene aggregeres på serveren (f.eks. beregnes et gjennomsnitt), kan ondsinnede lokale modelloppdateringer oppdages. I tillegg måles det potensielle bidraget til hver organisasjon i treningsprosessen.
Kunnskapsdestillasjon (<i>knowledge distillation</i>) (40)	<ul style="list-style-type: none"> • Uttrekkingsangrep 	For mer komplekse modeller kan en i stedet for å dele modellparametere overføre kunnskap fra den ferdig trente komplekse modellen til en enklere modell (med hensyn til antall parametere). Her spesifiseres hva som må gjøres for å trene den enklere modellen og beholde modellens gyldighet.
Klarert utførelsesmiljø (<i>trusted execution environment</i>) (41)	<ul style="list-style-type: none"> • Forgiftningsangrep • Uttrekkingsangrep 	Et sikkert område i prosessoren brukes til å lagre dataene og gjennomføre treningsprosessen i hver deltakende organisasjon.
Blokkjede (<i>blockchain</i>) (42)	<ul style="list-style-type: none"> • Forgiftningsangrep • «Gratispassasjer»-angrep 	Alle handlinger og modelloppdateringer registreres, spores og synliggjøres i en sikker og desentralisert transaksjonslogg.

Ved føderert læring er det viktig å sikre balansen mellom informasjonssikkerhet og personvern på den ene siden og ytelsen og nøyaktigheten til de trente modellene på den andre. Dersom et distribuert system ikke er sikret mot angrep kan en inntrenger manipulere treningsprosessen og stjele informasjon. På den andre siden fører et høyt krypteringsnivå til lav modellytelse og dersom modelloppdateringene tilføres for mye støy, vil modellen ha lav nøyaktighet.

Bruk av føderert læring vil med andre ord (a) medføre utfordringer og risikoer som gjelder alle distribuerte systemer, (b) kreve grundig risikovurdering og eventuelt ytterlige sikkerhetstiltak, og (c)

innebære en avveining mellom nøyaktigheten til trenne modeller/dataanalyser opp mot personvern og sikkerhet.

4.3 Føderert læring og europeisk personvernregulering

Et konsortium som driver med føderert læring i EU, skal samsvare med GDPR (EUs personvernforordning). Det er dataansvarlig som har ansvar for å iverksette nødvendige tiltak for at systemet er i samsvar med personvernprinsippene og øvrige krav i forordningen. Vanlige utfordringer ved å sikre at fødererte konsortier er i henhold til personvernforordningen er å definere og dele ansvar på tvers av flere behandlingsansvarlige, gjennomføre flere DPIAer (vurderinger av personvernkonsekvenser; *Data Protection Impact Assessment (DPIA)*), gjennomføre revisjon av deltakende helseinstitusjoner og sikre at modellene fungerer som forventet.

I *Tabell 3* ser vi på om det er enklere for et system å samsvare med personvernprinsippene (8) ved bruk av FL enn ved tradisjonell ML på sentraliserte data og om eventuelle tiltak bør iverksettes.

Tabell 3. Samsvar med personvernprinsipper: føderert læring contra maskinlæring på sentraliserte data

Personvern-prinsipp	FL contra ML på sentraliserte data	Beskrivelse	Mulige tiltak
Lovlig, rettferdig og gjenomsiktig	Ved bruk av FL er det 1) enklere å få et rettslig grunnlag for databehandling, 2) mer rettferdig databehandling, 3) mindre gjenomsiktige analyseresultater.	Det blir enklere å få et rettslig grunnlag for databehandling siden dataene til en organisasjon ikke deles. Organisasjonene har mer kontroll over at databehandlingen gjøres med respekt for den registrertes interesser fordi dataanalysen utføres lokalt. Samtidig er det vanskeligere å bevise rettferdigheten i beslutninger, sjekke datasett for skjevheter og statistisk heterogenitet siden det ikke er innsyn i dataene i de andre organisasjonene som bidrar til den globale modellen.	Det er behov for mer forskning. Det trengs verktøy for å sjekke statistiske datasettparametere på en distribuert og personvernbevarende måte, samt for ulike visualiseringsverktøy for å bedre beslutningsforklaring.
Formålsbegrensning	FL kan indirekte forenkle samsvar med dette prinsippet.	Ved å unngå sentralisering og påfølgende duplisering av data, kan føderert læring bidra til å begrense risikoen for at dataene gjenbrukes til et formål som er uforenlig med det opprinnelige formålet med datainnsamlingen.	(Introdusere og) følge retningslinjer for datastyring i deltakende helseinstitusjoner (dvs. å ha definerte roller, ansvar og prosesser for å sikre ansvarlighet for og eierskap til dataressurser i organisasjonen). Tiltaket skal øke lokal datasikkerhet.

Dataminimering	Bedre sikret med FL	Ved føderert læring unngår man overføring av rå (usikrede) treningsdata til sentralisert lagring, og derfor elimineres unødvendig dataduplisering. Dessuten er det enklere å sikre at opplysningene slettes når de ligger lokalt.	(Introdusere og) følge retningslinjer for datastyring i deltakende helseinstitusjoner (se Formålsbegrensning for detaljer).
Riktighet	I beste fall på linje med bruk av ML på sentraliserte data	Lokale modelloppdateringer fra deltakere lagres og behandles i sin opprinnelige form uten endringer og oppdateres etter hver treningsrunde på den sentrale serveren, noe som gjør at dataene er riktige til enhver tid. Data fra flere organisasjoner kan bli mer representative for en populasjon enn data fra en enkelt organisasjon, noe som gir bedre analyse-resultater. Likevel er modeller trent med FL er i beste fall like gode som de som er trent på tilsvarende mengde sentraliserte data.	(Introdusere og) følge retningslinjer for datastyring i deltakende helseinstitusjoner (se Formålsbegrensning for detaljer). Det er også behov for et verktøy for å sjekke statistiske datasettparametere på en distribuert og personvernbevarende måte.
Lagringsbegrensning	Bedre sikret med FL	På den sentrale serveren lagres hverken data fra helseinstitusjoner eller individuelle lokale modelloppdateringer. Det er også enklere å sikre at de lokale dataene slettes når treningen er ferdig.	(Introdusere og) følge retningslinjer for datastyring i deltakende helseinstitusjoner (se Formålsbegrensning for detaljer).
Integritet og konfidensialitet	Bedre sikret med FL	Ved føderert læring lagres dataene lokalt og overføres ikke til sentralisert lagring. Det lagres heller ikke individuelle, lokalt trent modellparametere på den sentrale serveren, men bare de aggregerte resultatene. Den globale modellen lagres, men den er anonym (siden den er aggregert) og regnes ikke som personopplysninger (43–45).	Ulike personvern-fremmede mekanismer (f.eks. sikker flerpartsberegning, differensiert personvern) iverksettes både på den sentrale serveren og i deltakende helseinstitusjoner, samt datatilgangskontroll (dvs. regler for å styre hvem som skal ha tilgang til hvilke opplysninger eller systemer), retningslinjer for datastyring (se Formålsbegrensning for detaljer) og sikre kommunikasjonsprotokoller for en sikker meldingsutveksling.

Ansvarlighet	Bedre sikret med FL	Lokal datalagring gir bedre kontroll over dataene. Samtidig er dataene av mindre interesse for angrep enn sentralisert lagrede store helsedata-sett.	Ha datatilgangskontroll og (introdusere og) følge retningslinjer for datastyring i deltakende helseinstitusjoner (se Formålsbegrensning for detaljer).
---------------------	---------------------	--	--

4.4 Eksempler på prosjekter som bruker føderert maskinlæring på hel-seområdet

Studier som har brukt føderert læring på pasientjournaldata, har vist at det er mulig å finne klinisk lignende pasienter (46,47), forutsi sykehusinnleggelses basert på hjertehendelser (48) og oppholdstid på intensivavdelingen (49). Innen medisinsk bildediagnostikk brukes føderert læring til MR-segmentering av hjernesvulster (29,50,51), patologi- (52) og kreftforskning (53–59). Føderert læring har også blitt brukt for å finne pålitelige sykdomsrelaterte biomarkører (60,61), forebygge smittsomme sykdommer (62), predikere metabolske og hjerte- og karsykdommer (63), oppdage medisinske implantater hos pasienter før MR-undersøkelser (64), analysere genomiske og molekylærbiologiske data (65), i sammenheng med COVID-19 (66,67), legemiddeloppdagelse (68) og -bivirkninger (64).

Nedenfor beskrives kort noen store prosjekter innenfor helse som har benyttet føderert læring.

4.4.1 Internasjonale prosjekter

London Medical Imaging & Artificial Intelligence Centre for Value-Based Healthcare (AI4VBH)

AI4VBH (69) er et stort prosjekt der Kings College London (70), NVIDIA (71) og OWKIN (72) samarbeider om å utvikle et personvern fremmende føderert læringssystem for MR-segmentering av hjernesvulster (51). De skal koble sammen fire undervisningssykehus i London før det utvides til hele Storbritannia. Ambisjonen er å tilby kunstig intelligens-tjenester innen et bredt spekter av terapeutiske områder, blant annet kreft, hjertesvikt og hjernesykdommer. Dette er et pågående prosjekt.

Samarbeid mellom Moorfields øyesykehus og Bitfount om øyesykdomsrelaterte biomarkører

Bitfount (73) samarbeider med forskere fra Moorfields øyesykehus (74) i London og Universitetet i Surrey om å bruke føderert læring for å evaluere biomarkørmodeller som kan forutsi visse øyetilstander (60). Disse modellene flagger pasienter for rekruttering til kliniske studier, uten å påvirke personvernet og vil samtidig redusere rekrutteringskostnadene. Dette er et pågående prosjekt.

Bigpicture

Bigpicture (52) er et europeisk offentlig-privat partnerskap innen patologiforskning som består av bl.a. akademiske institusjoner, små og mellomstore bedrifter, offentlige organisasjoner og farmasøytiske selskaper. De tilbyr en infrastruktur (både maskin- og programvare) til å lagre, dele og prosessere patologibilder som er i samsvar med GDPR. Dette er et pågående prosjekt som varer til 2027.

HealthChain

HealthChain (53) var et fransk konsortium av sykehus, universiteter og teknologipartnere for klinisk forskning. De trente ML-modeller på histologiske og dermaskopibilder for å forutsi behandlingsrespons hos brystkreft- og melanom pasienter. Prosjektet var ferdig i 2021.

Melloddy (Machine Learning Ledger Orchestration for Drug Discovery)

MELLODDY (68) var et europeisk konsortium av farmasøytiske, teknologiske og akademiske partnere som brukte føderert læring på dataene fra farmasiselskaper for å øke effektiviteten i legemiddeloppdagelse. Prosjektet varte i 3 år og ble ferdigstilt i 2022.

Federated Tumor Segmentation (FeTS)

FeTS (54) er et pågående internasjonalt konsortium av helseinstitusjoner og en plattform for tumor-grensegjenkjenning i MR-bilder fra store og varierte pasientpopulasjoner.

Trustworthy Federated Data Analytics (TFDA)

TFDA (55) er et pågående tysk prosjekt innenfor kreftforskning som anvender føderert læring, med stråleterapi som et brukstilfelle.

Joint Imaging Platform

Joint Imaging Platform (56) er et pågående strategisk initiativ innenfor det tyske kreftkonsortiet som sikter til å etablere en teknisk infrastruktur for distribuert medisinsk bildeforskning.

OPTIMA (Optimal Treatment for Patients with Solid Tumors in Europe Through Artificial intelligence)

OPTIMA (57) er et pågående europeisk prosjekt og en plattform som skal fremme kreftbehandlinger og enklere beslutningstaking for leger og pasienter med prostata-, bryst- og lungekreft. Plattformen vil være i samsvar med GDPR.

Epiverse

Epiverse(62) er et globalt samarbeid mellom akademia, regjeringer og andre organisasjoner som utvikler et pålitelig dataanalyse-økosystem av standardiserte epidemiologiske programvareverktøy for smittsomme sykdommer. Prosjektet er pågående og får støtte av Rockefeller-stiftelsen (75) og Wellcome-veldedighetsstiftelsen (76).

4.4.2 Prosjekter med norsk deltakelse

HealthData@EU

Dette er et 2-årig pilotprosjekt (varer til 2024) for EHDS, som tar sikte på å fremme sikker tilgang til og utveksling av helsedata på tvers av landegrensene i EU (63,77). Norge deltar i tre brukstilfeller, i tillegg til deltakelse i de fleste av de øvrige arbeidspakkene i piloten. Det ene brukstilfellet handler om sammenligning av hendeshistorikk for å evaluere interoperabilitet av helsedata på tvers av EU innenfor hjerte- og karsykdommer og metabolske sykdommer. Målet er å (a) vurdere grad av samsvar i data, kodeverk og standarder, (b) studere prevalens av hjerte- og karsykdommer og metabolske sykdommer på tvers av EU, og (c) utvikle og validere nye og forbedrede prediksjonsmodeller for metabolske og hjerte- og karsykdommer. Frankrike som leder dette brukstilfellet, skal utvikle en KI-prediksjonsmodell. Deretter skal modellen i en iterativ prosess testes på distribuerte finske, danske, ungarske og norske data og rekalkibreres. Dette brukstilfellet skal teste modenheten på tvers av Europa for sekundærbruk av helsedata i sammenheng med stordata og modenheten for bruk av ML på distribuerte helsedata. Det tas sikte på å få kunnskap og erfaringer for å øke sannsynligheten for hensiktsmessig videre bruk av KI på stordata.

FederatedHealth: A Nordic Federated Health Data Network

FederatedHealth (64) er et pågående 3-årig prosjekt finansiert av Nordic Innovation, som analyserer kliniske notater i pasientjournaler i nordiske land (Finland, Sverige, Danmark, Estland, Norge). De skal bruke føderert læring for å utvikle en personvernbevarende virtuell infrastruktur for dataanalyse og forskningsformål. Infrastrukturen skal testes på to brukstilfeller: (a) oppdage om man har medisinske implantater før MR-undersøkelser og (b) finne ut pasientens mulige legemiddelbivirkninger. Kunnskapen og erfaringer fra dette prosjektet kan utnyttes videre i arbeidsprosesser i EHDS.

Elixir

Elixir (65) er et europeisk prosjekt og et rammeverk for sikker innsending, arkivering, deling og analyse av genomiske og molekylærbiologiske data. Universitetet i Oslo drifter den norske delen av EGA-nettverket (European Genome-phenome Archive) (78).

Samarbeid innen kreftforskning med Nederland

Kreftregisteret i Norge samarbeidet med kreftregisteret i Nederland for å sammenligne indikatorer for ulike pasientgrupper med brystkreft, uten å utveksle pasientinformasjon (58).

FLORENCE

FLORENCE (59) (Federated learning using OMOP modelling of health data for elevating colorectal cancer care in the Nordic countries) er et 3-årig interregionalt prosjekt som utvikler et beslutningsverktøy for klinikere innen kolorektalkreft. Dette er et samarbeid mellom Kreftregisteret, Center for Surgical Science (CSS) ved Sjælland Universitetssykehus i Danmark, Lunds Universitet i Sverige, Computerome ved Danmarks Tekniske Universitet og enheten for forskningsprosjekter ved Sjællands Universitetssykehus. De bruker sykehus- og registerdata som er harmonisert til OMOP-datamodellen for å dele modellendringer istedenfor helseopplysninger og samtidig ha et like godt beslutningsverktøy på alle deltagende institusjoner. OMOP står for Observational Medical Outcomes Partnership og er utviklet for å standardisere data til et felles format (79). Prosjektet er finansiert av Interreg Øresund-Kattegat-Skagerrak (ØKS) frem til 2025.

PraksisNett

PraksisNett (47,80) er et forskningsnettverk i primærhelsetjenesten og en infrastruktur som muliggjør distribuerte analyser fra pasientjournaler på norske legekontor. Det er 92 allmennpraksiser fra hele landet som deltar i PraksisNett. Disse inkluderer 492 fastleger med nesten 520 000 pasienter.

4.5 Oppsummering

Føderert læring er en personvernvennlig tilnærming til maskinlæring der organisasjoner samarbeider om distribuert modelltrening uten at dataene deres deles eller blir synlige for de andre. Denne teknologien gjør det mulig å benytte de store datamengdene som produseres i helseinstitusjonene, slik at de kan få bedre kunnskapsgrunnlag for medisinske beslutninger og dermed tilby bedre pasientbehandling. Mange helserelevante prosjekter prøver ut teknologien, med lovende resultater. Likevel må helseinstitusjoner som vurderer å ta i bruk føderert læring ta hensyn til flere problemstillinger: variasjoner i data og tekniske oppsett i organisasjoner, ressursbruk, potensielle kommunikasjonsflaskehalser, samt datasikkerhet. Det er i tillegg behov for mer forskning før teknologien kan brukes for fullt. For å sikre rettfærdige beslutninger trengs det metoder for å kunne sjekke statistiske parametere på datasett i de helseinstitusjonene som er involvert i treningsprosessen. Dette må skje på en distribuert, men personvernvennlig måte. Da helsedata er sensitive og har høy verdi er det nødvendig med videreutvikling av effektive kryptografiske og personvernbevarende teknikker for ytterligere sikkerhetstiltak i fødererte konsortier av helseinstitusjoner. Det må også forskes på avveiningen mellom personvern og nøyaktighet i dataanalyser som utføres i slike systemer.

5 Syntetiske data

KI-systemer trenger tilgang til data av høy kvalitet og i tilstrekkelig mengde for trening, validering og testing. Utvikling av KI-modeller til bruk i helse- og omsorgstjenesten krever ofte tilgang til pasientdata, for eksempel fra pasientjournaler. Tilgang til nødvendige helsedata kan være utfordrende, blant annet på grunn av personvern hensyn. Dette kan være en betydelig flaskehals for utbredelsen av KI i helse- og omsorgstjenesten (2,81). Syntetiske data er et av flere mulige virkemidler for å gjøre tilgangen til treningsdata enklere og raskere.

5.1 Hva er syntetiske data?

Syntetiske data er kunstig genererte data som etterligner data fra den virkelige verden, uten å inneholde personidentifiserbar informasjon. Generering av syntetiske data har en historie som strekker seg over flere tiår, men metodene har utviklet seg raskt de senere årene. De opprinnelige matematiske og regelbaserte tilnærmingene er i stor grad erstattet av maskinlæringsmetoder. Maskinlæringsmodeller kan generere syntetiske data av mange ulike typer, for eksempel tabelldata (data organiserte i rader og kolonner), tidsserier, bilder, lyd og tekst. Det siste året har prateroboter som ChatGPT fra OpenAI og Bard fra Google AI nærmest revolusjonert generering av syntetisk tekst (82,83). En generativ maskinlæringsmodell trenes med reelle data (dvs. treningsdatasett) til å generere syntetiske data. Modellen lærer statistiske mønstre og sammenhenger mellom variabler i treningsdatasettet og bruker denne informasjonen til å generere nye, syntetiske data som har tilnærmet de samme egenskapene som de originale, reelle dataene. Syntetiske data kan brukes til blant annet dataanalyser, teste programvaresystemer eller trene maskinlæringsmodeller.

I denne rapporten ser vi på bruken av syntetiske data for trening av maskinlæringsmodeller. Et syntetisk datasett som er produsert av en generativ maskinlæringsmodell brukes til å trene en ny maskinlæringsmodell som kan utføre oppgaver som diagnostikk eller predikere behandlingsutfall.

5.2 Mulige bruksområder for syntetiske data

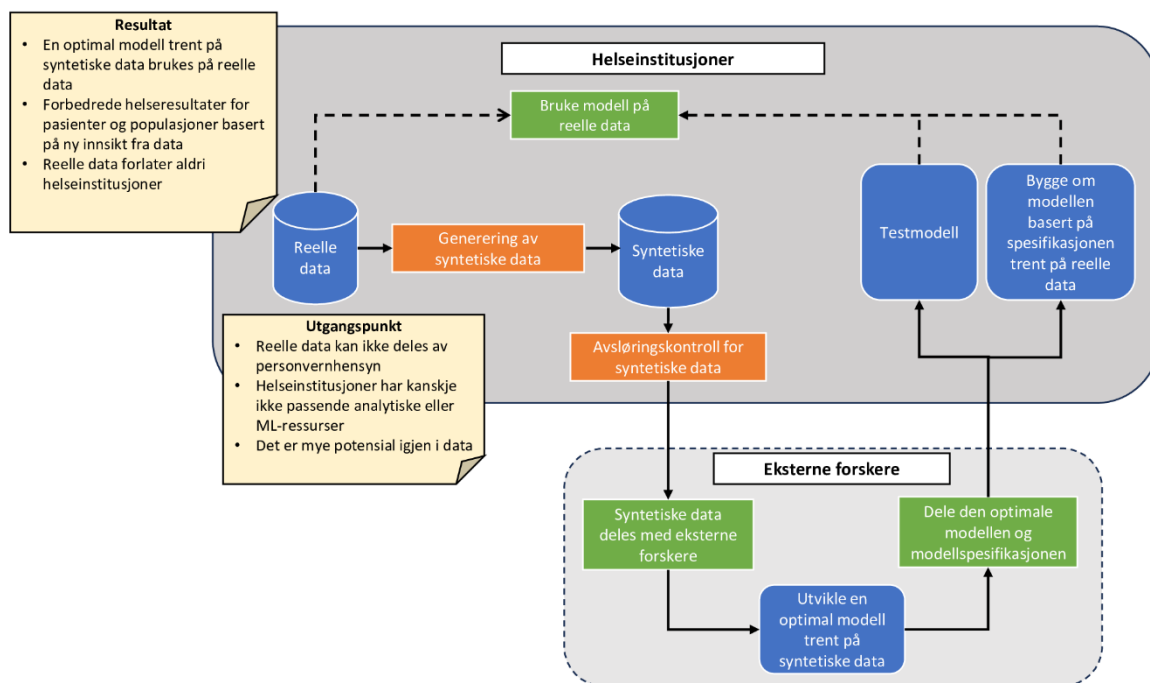
Syntetiske data er en umoden teknologi, og bruken i helse- og omsorgstjenesten er beskjeden. For å forstå hvilke muligheter og begrensninger som ligger i bruk av syntetiske data til å trene maskinlæringsmodeller, er det nødvendig å bygge kunnskap og erfaringer.

5.2.1 Trene, validere og teste KI-modeller

Bruk av syntetiske data i tidlige faser i utviklingen av maskinlæringsmodeller som senere evalueres med reelle data, er en mulig strategi for å få erfaring med syntetiske data i helse- og omsorgstjenesten. Mange studier konkluderer med at maskinlæringsmodeller som er trent med syntetiske data, bør bekreftes med reelle data, siden det kan være usikkert om syntetiske datasett er representative (84). Dersom modellen skal brukes til oppgaver som kan påvirke pasientsikkerheten, er en slik evaluering helt avgjørende.

«*Train on Synthetic, Test on Real*» (TSTR) (85,86) er en tilnærming der en maskinlæringsmodell trenes med syntetiske data, mens ytelsen til den trente modellen evalueres med reelle data. Et eksempel på bruk av TSTR er forskere som planlegger å utvikle en maskinlæringsmodell og har behov for pasientdata til trening og testing av modellen. Dersom bruk av syntetiske data kan gi forskerne raskere tilgang til et anonymt syntetisk datasett basert på et reelt datasett, kan de starte modellutviklingen med dette, mens de venter på tilgang til de ekte dataene. Når de får tilgang til det ekte datasettet bruker de dette til å finjustere maskinlæringsmodellen, eventuelt gjenta hele treningsprosessen med det ekte datasettet dersom modellen presterer dårlig. På denne måten kan mye tid spares i prosjektet. Tilnærmingen er illustrert av Rankin og kollegaer (87) (se *Figur 2*).

Figur 2. Eksempel på bruk av syntetiske data i helse- og omsorgstjenesten. Kilde (87)



Validering

Begrepet «validering» har ulik betydning blant fagfolk innen KI-feltet og medisin og helse. Innen KI er validering et av stegene i treningen av en ML-modell, der modellen finjusteres, som også kalles teknisk validering (88–90). Klinisk validering betyr at KI-systemet evalueres i en ekte klinisk setting for å bedømme pasientsikkerhet og effekt. Ved utvikling av kliniske KI-systemer kan syntetiske data brukes til teknisk validering i den innledende treningen, mens til klinisk validering er det nødvendig med data fra ekte pasienter i en klinisk setting (88). Teknisk validering er ofte ikke tilstrekkelig for å validere den kliniske ytelsen eller generaliserbarheten til modellen, og mangel på grundig klinisk validering er en viktig grunn til at KI-baserte kliniske systemer ikke har blitt tatt i større skala (88). For KI-systemer som har stor innvirkning på pasientsikkerheten, er klinisk validering et mellomsteg før et eventuelt klinisk forsøk (89). Det finnes flere forslag til rammeverk for validering av KI-baserte kliniske systemer (89–92).

5.2.2 Tilgjengeliggjøre mer data til forskning

Mangelen på åpne pasientjournaldata gjør at svært mye forskning gjøres på de få tilgjengelige datasettene som eksisterer, for eksempel MIMIC-III (Medical Information Mart for Intensive Care), som inneholder aidentifiserte data fra intensivavdelinger i USA (88). Slik ensidig bruk av et lite antall datasett kan føre til skjevheter i forskningen (89). Ettersom syntetiske datasett normalt har lavere risiko for re-identifisering av enkeltpersoner enn reelle data kan de bidra til økt tilgang til åpne datasett. Hvor stor risiko for re-identifisering som kan aksepteres, vil variere etter hvilken type tilgang som gis til det syntetiske datasettet og hvor stor spredning det kan få: om dataene er fritt tilgjengelige for nedlasting på nettet, om brukere må inngå en avtale for bruk av dataene før de kan laste det ned, eller om brukere ikke får tilgang til selve det fysiske datasettet, men kan behandle dataene i et sikkert analyserom.

Publisering av syntetiske data som åpne datasett vil kunne gjøre forskning innen helse mer transparent og øke reproduserbarheten. Forskere kan for eksempel generere et syntetisk datasett basert på de originale, potensielt sensitive dataene som de har brukt i en studie og dele dette. Det er imidlertid utfordrende å generere syntetiske data som har høy nytte for ett eller flere formål, samtidig som de har lav risiko for personvernbrudd. Før et syntetisk datasett eventuelt kan publiseres må det evalueres grundig for å sikre at kvaliteten er god og personvernet er ivaretatt. Det må utføres statistiske evalueringer for å sikre at datasettet ikke inneholder skjevheter eller unøyaktigheter som gjør at det ikke er representativt for de originale dataene. For å styrke personvernet bør algoritmene som genererer syntetiske data bruke andre personvern fremmende teknikker i tillegg. Simulering av ulike re-identifiseringsangrep vil også være nyttig. Metodene som brukes, må være godt dokumentert slik at brukere av disse syntetiske dataene kan forstå prosessen. Selv om datasettene er «åpne», må tilgangen være kontrollert slik at kun de som har legitimt grunnlag, får tilgang.

Eksempler på norske miljøer som jobber med syntetiske data

Vi nevner her noen eksempler på miljøer i Norge som har pågående forsknings- og utviklingsprosjekter på syntetiske data i helseområdet.

- *Norsk helsenett* (NHN) har definert syntetiske data som et satsningsområde. De bruker syntetiske produksjonsdata i sitt testmiljø der systemleverandører og helsevirksomheter kan utføre ulike tester av sine IT-systemer (95). De bruker bl.a. syntetiske testdata fra Skatteetatens test-folkeregister, som består av en million testpersoner. På lengre sikt ønsker NHN å tilby en løsning for syntetiske helseregisterdata der det vil det være mulig å sammenstille data fra flere datakilder som sentrale helseregistre, kvalitetsregistre og befolkningsundersøkelser og generere syntetiske data fra det sammenstilte datasettet.
- *Helse Stavanger* er prosjekteier i Pilot SYNdata (96) som utforsker hvordan man kan bruke syntetiske helsedata til datadrevet innovasjon i helsesektoren med fokus på digitale helse-tjenester for barne- og ungdomspsykiatrien.
- Forskere ved *Simula Metropolitan Center for Digital Engineering* (SimulaMet) har i samarbeid med danske forskere brukt et GAN som ble trent på ca. 7.000 ekte normale EKG (med normal hjerterytme) til å generere over 120.000 syntetiske normale EKG. Det syntetiske EKG-datasettet er publisert med åpen tilgang for forskere (97).
- *DIPS* har pågående prosjekter der de bruker syntetiske data både for testing av sine EPJ-løsninger og til andre formål, der de også samarbeider med klinikere og forskere.
- *Kreftregisteret* har pågående forskningsaktiviteter på syntetiske data og andre personvern fremmende teknologier (59).

5.3 Hvilke fordeler og utfordringer har syntetiske data?

Den viktigste grunnen til den økende interessen for syntetiske data ligger i potensialet de har som personvern fremmende teknologi. Syntetiske data kan samtidig bidra til å forbedre datakvaliteten ved å kompensere for eventuelle svakheter og mangler i det originale datasettet. Dette gir muligheter, men det er også viktige utfordringer knyttet til bruk av syntetiske data som treningsdata.

5.3.1 Styrkes personvern ved bruk av syntetiske data?

Syntetiske data har lavere risiko for re-identifisering enn reelle data som er aidentifisert/ anonymisert ved tradisjonelle teknikker (90). Syntetiske data nevnes derfor ofte som et mulig virkemiddel for å gjøre tilgangen til treningsdata enklere. Dette gjelder spesielt dersom et syntetisk datasett vurderes å være anonymt. Personvernlovgivningen gjelder ikke for personopplysninger som har blitt

anonymisert på en måte som gjør at *enkeltpersoner ikke lenger kan identifiseres med hjelpemidler som med rimelighet kan tenkes å bli brukt*» (14). Anonyme data kan derfor brukes uten begrensninger.

Selv om syntetiske data i utgangspunktet ikke er direkte knyttet til en bestemt person, er det likevel en mulig risiko for at informasjon kan lekke fra de reelle treningsdataene de er generert fra. For at de skal være nyttige, må de være tilstrekkelig representative for de ekte dataene de skal etterligne. Det er likevel viktig at de ikke er *for* like de originale dataene da dette kan øke risikoen for re-identifisering av enkeltpersoner. Denne risikoen kan reduseres ved å bruke differensielt personvern (20) (se *Tabell 2* for detaljer om teknikken) for å beskytte det syntetiske datasettet ved å tilføre kontrollerte mengder tilfeldig støy. Støyen gjør syntetiske data mindre lik de originale dataene, slik at re-identifisering blir vanskeligere. Ulempen er at de samtidig blir mindre representative og mindre nyttige som treningsdata. Det er derfor viktig å finne et nivå på støyen som tilføres, slik at en oppnår en god balanse mellom personvern og nytte.

5.3.2 Kan syntetiske data forbedre kvaliteten til treningsdata?

For å utvikle gode maskinlæringsmodeller er det nødvendig med treningsdata av høy kvalitet og tilstrekkelig omfang og variasjon. Disse dataene må være representative for de dataene den ferdig trente modellen skal brukes på, og de må ikke inneholde systematiske skjevheter eller urettferdigheter, som for eksempel underrepresenterte demografiske grupper.

Kvaliteten til syntetiske datasett avhenger av kvaliteten i det originale datasettet det er basert på. Dersom det originale datasettet har skjevheter eller er ufullstendig vil det syntetiske datasettet arve de samme egenskapene. Noen forskere argumenterer likevel for at et syntetisk datasett i enkelte tilfeller kan være bedre enn det originale fordi det er mulig å korrigere for kjente skjevheter, urettferdigheter og mangler i det originale datasettet (91,92). Dette oppnås ved å gjøre endringer i det originale datasettet før det brukes til å trene maskinlæringsmodellen som genererer de syntetiske dataene, eller gjennom å kompensere for disse svakhetene direkte i den generative modellen.

Store og varierte treningsdatasett med få skjevheter vil være å foretrekke fordi de ofte er representative og fører til gode og rettferdige maskinlæringsmodeller. Det finnes imidlertid spesielle tilfeller der det er vanskelig å få tak i gode reelle data. Det kan være fordi pasientgruppene er små og datagrunnlaget derfor er begrenset, som for eksempel for sjeldne medisinske tilstander eller hendelser. Syntetiske data kan bidra til et større datagrunnlag i slike tilfeller.

Det at det er mulig å kompensere for kjente svakheter i det originale treningsdatasettet ved generering av syntetiske data betyr ikke at høy kvalitet er garantert. Unøyaktigheter i maskinlæringsmodellen som brukes til å generere de syntetiske dataene, kan introdusere nye skjevheter som ikke var til stede i det originale datasettet eller forsterke eksisterende skjevheter.

5.3.3 Kan vi spare ressurser ved å bruke syntetiske data?

Bruk av syntetiske data som treningsdata kan gi besparelser knyttet til innsamling og klargjøring av datasettet. Tilgang til reelle data innebærer ofte en tidkrevende søknadsprosess. Størrelsen på datasettet man søker tilgang til kan ha betydning for hvor komplisert søknadsprosessen blir. Det kan derfor være mindre ressurskrevende å bruke et syntetisk datasett.

Samtidig er det nødvendig med tilgang til reelle treningsdata for å trene maskinlæringsmodellen som skal generere det syntetiske datasettet. Mengden reelle data som er nødvendig for å generere syntetiske data, er likevel mindre enn det som kreves for å trene en maskinlæringsmodell direkte med reelle data. Dette kommer av at det er mulig å generere syntetiske datasett som er større enn treningsdatasettet de er generert fra, det er ikke et en-til-en-forhold mellom størrelsen på de to datasettene. Samtidig vil genereringen av det syntetiske datasettet ta tid og ofte bruke stor regnekapasitet, så det er uklart hvor stor en eventuell besparelse vil være.

5.4 Metoder og verktøy for generering av syntetiske data

Det eksisterer mange verktøy for generering av syntetiske data, både kommersielle og med åpen kildekode. De ulike metodene for generering av syntetiske data kan deles i tre kategorier (93):

- *Datadrevne metoder* der syntetiske data genereres fra reelle datasett ved å bruke en maskinlæringsmodell
- *Kunnskapsdrevne metoder* der data genereres ved å bygge en modell fra domenekunnskap. Modellen kan inneholde beslutningsregler, statistiske modeller og matematiske modeller. Bygging av slike modeller er tid- og kompetansekrevende.
- *Hybride metoder* som kombinerer de to hovedtypene.

De fleste nyere verktøyene er basert på generative maskinlæringsmodeller som bruker nevralt nettverksalgoritmer, der generative adversarial networks (GAN) (94,95) og variational autoencoders (96), er de vanligste. GAN-modeller trenes ved å bruke to nevralt nettverk, der det ene produserer syntetiske data, mens det andre lærer seg å skille mellom syntetiske og ekte data. Tilnærminger basert på nevralt nettverk er populære fordi de er gode til å lære mønstre i data og kan generere mer varierte data (97). GAN ble først brukt til å generere syntetiske bilder og tekst, noe de er svært gode til. De har blitt videreutviklet til også å generere tabelldata, det vil si data som typisk er lagret i databaser, for eksempel pasientjournaldata (19,97–99). En ulempe med GAN er at de gir modeller som kan ha varierende kvalitet (100). Dersom treningsprosessen repeteres flere ganger med de samme treningsdataene, kan hvert forsøk resultere i modeller med betydelige forskjeller i resultatene. Dette betyr at kvaliteten til de syntetiske dataene disse modellene genererer kan variere.

Noen verktøy bruker enklere maskinlæringsmodeller, for eksempel modeller basert på beslutnings-trær (101) eller bayesianske nettverk (102). Fordelen med disse modellene er at de krever mindre treningsdata enn nevralt nettverk og er mer transparente. Selv om ML-metoder blir stadig mer brukt til datagenerering er de kunnskapsdrevne metodene fortsatt i bruk, blant annet i Synthea (103), som er et verktøy for å generere syntetiske EPJ-data.

Den store nyheten innen datagenerering det siste året er avanserte prateroboter som ChatGPT og Bard (82,83). Praterobotene bruker store språkmodeller, basert på transformer-arkitekturen, som er en type nevralt nettverk (104). De store språkmodellene er forhåndstrent med enorme mengder tekst og kan deretter finjusteres mot spesifikke oppgaver som spørsmål og svar, oversettelse mellom språk og sammendrag av tekst. De kan dessuten kombineres med GAN-modeller for tekst-til-bilde generering. Praterobotene er fortsatt nye og har svakheter som gjør at teksten som genereres ofte har betydelige feil og mangler (105). For å kunne bruke den genererte teksten til å trene andre ML-modeller må kvaliteten bli bedre og mer forutsigbar. Til tross for disse utfordringene har praterobotene i løpet av kort tid vekket stor interesse, og det diskuteres hvordan de kan brukes innen helse- og omsorgstjenesten (105–108).

5.5 Hvordan kan vi evaluere datakvalitet og personvern i syntetiske datasett?

Empirisk evaluering er nødvendig for å vurdere om syntetiske datasett har høy kvalitet, er egnet til å trene maskinlæringsmodeller og ivaretar personvernet (100). Evalueringene omfatter både statistiske analyser og praktiske tester som sammenligner det syntetiske datasettet med ett eller flere reelle datasett. I tillegg til å sammenligne med det originale treningsdatasettet bør en sammenligne med et valideringsdatasett med reelle data som ikke ble brukt i treningen av den generative modellen.

5.5.1 Evaluering av datakvalitet

For å evaluere datakvaliteten til syntetiske datasett er det vanlig å utføre to ulike typer sammenligninger. Den første typen evaluerer nøyaktigheten til det syntetiske datasettet, det vil si hvor likt det er

det originale datasettet. Den andre typen sammenligner nytten, det vil si hvor godt de ulike datasettene presterer for konkrete brukstilfeller (85).

Nøyaktigheten til det syntetiske datasettet evalueres ved å utføre de samme statistiske analysene på det syntetiske datasettet og treningsdatasettet. Høy nøyaktighet betyr at det er stor grad av statistisk likhet mellom de to datasettene. Det betyr likevel ikke at det syntetiske datasettet nødvendigvis har høy nytte. Nyttene til syntetiske datasett vil variere avhengig av det konkrete bruksområdet. Den må derfor evalueres ved å teste hvor godt det syntetiske datasettet presterer på en bestemt oppgave og sammenligne med hvordan et datasett med reelle data klarer den samme oppgaven (87,109,110).

Flere studier rapporterer at prediksjonene er litt mindre nøyaktige når modellene er trent med syntetiske data versus reelle data, men at forskjellene ofte er relativt små (84,87,111,112). Dette viser potensialet i teknologien, men evalueringer av nytte er spesifikke og kan ikke generaliseres til andre maskinlæringsmodeller eller brukstilfeller.

5.5.2 Evaluering av personvern

Siden syntetiske data er tilfeldig genererte, representerer de i utgangspunktet ikke ekte personer eller hendelser. I tidlige studier var forskerne derfor mest opptatt av nøyaktigheten til de syntetiske dataene mens personvern ble mindre prioritert. Etter hvert ble det klart at syntetiske datasett som er generert fra maskinlæringsmodeller som er trent med ekte personopplysninger, ikke nødvendigvis er anonyme. De er sårbare for angrep som forsøker å utlede informasjon som stammer fra treningsdatasettet, for eksempel avdekking av sensitive attributter om en person (*attribute inference*) og avdekking av medlemskap (*membership inference*), det vil si identifisere om data fra en bestemt person er til stede i treningsdataene (113,114). Det er en generell utfordring ved trening av maskinlæringsmodeller at det kan oppstå en *overtilpasning*, hvor modellen husker informasjon om enkeltpersoner i treningsdataene og gjensker disse i resultatene (115,116). Dette kan redusere modellens evne til å generalisere til nye data, og føre til økt risiko for re-identifisering av enkeltpersoner (117).

Syntetiske data har potensial til å styrke personvernet, men for å oppnå dette kan det være nødvendig å inkludere andre personvern fremmende teknikker, som differensielt personvern (20,118,119) eller føderert læring (*federated learning*) (se *Føderert læring*). Differensielt personvern er en lovende metode for å redusere risiko for re-identifisering, men den er umoden og det er behov for mer forskning, utvikling og evalueringer før den kan bli tatt i bruk i større skala (120).

I den vitenskapelige litteraturen beskrives flere ulike metoder for evaluering av risikoen for re-identifisering i syntetiske datasett (102). Det fins likevel ingen pålitelige og objektive metoder for å evaluere om et syntetisk datasett er tilstrekkelig ulikt det originale datasettet til at det kan regnes som anonymt (121). En utbredt måte å evaluere personvernrisiko i datasett på er å simulere kjente typer angrep som avdekking av attributter eller medlemskap (122,123). Forfatterne av en ikke-fagfelleurdert studie som simulerte personvernangrep mot flere av de nyeste modellene for generering av syntetiske data hevder at mange studier overvurderer fordelene syntetiske data har sammenlignet med tradisjonelle anonymiseringsteknikker og at det ikke er evidens for at syntetiske data gir bedre beskyttelse mot personvernangrep med mindre tap av nytte enn tradisjonelle anonymiseringsteknikker (122).

I tillegg til angrep direkte mot syntetiske datasett kan også maskinlæringsmodellene som brukes til generering av syntetiske data være sårbare for angrep som prøver å trekke ut sensitiv informasjon (85,124). Dersom angriperen har informasjon om arkitektur, parametere og andre egenskaper til en trent modell, kan de bruke dette til å utforme målrettede angrep mot modellen. Interne opplysninger om modellene bør derfor være konfidensielle (90). Dette vil redusere risikoen, men ikke eliminere den (125).

5.5.3 Behov for bedre og mer standardiserte evalueringsmetoder

Ved evaluering av syntetiske datasett er det en utfordring at det ikke eksisterer en standardisert måte å evaluere hverken statistisk likhet med det originale datasettet, nytte eller personvern. En systematisk oversiktsartikkel som studerte generering av syntetiske helsedata, viste at de eksisterende studiene bruker mange ulike rammeverk og indikatorer, noe som gjør det vanskelig å sammenligne ulike tilnærminger for datagenerering på tvers av studier (97). Økt standardisering blir derfor trukket fram som et viktig steg mot mer objektive evalueringer. En annen utfordring er at noen av indikatorene som ofte brukes i evalueringer, kan være vanskelige å forstå for klinikere uten omfattende statistikkunnskaper (126). Måling av skjevheter og rettferdighet i datasett er fortsatt et ubesvart spørsmål (100). Videre forskning trenger standardiserte, objektive og pålitelige målemetoder og referansepunkter for å kunne evaluere disse viktige egenskapene ved syntetiske data på en troverdig måte. Disse må i tillegg bygges inn i verktøy for generering og evaluering av syntetiske data, slik at denne prosessen blir mer effektiv og ikke krever spesialisert kompetanse på området.

5.6 Oppsummering

Tilgang til treningsdata er en flaskehals for utbredelsen av KI i helse- og omsorgstjenesten. Forskning på personvern fremmende teknologier som gjør det enklere å trene ML-modeller uten å kompromittere personvernet, er et viktig område. Dette inkluderer syntetiske data.

Syntetiske data er kunstig genererte data som blir produsert ved bruk av en generativ maskinlæringsmodell som er trent med reelle data. Den generative modellen lærer de statistiske egenskapene til treningsdatasettet og bruker dette til å generere et syntetisk datasett med tilnærmet like statistiske egenskaper. Det syntetiske datasettet kan brukes til trening av nye ML-modeller. Fordelene med å bruke syntetiske treningsdata fremfor reelle data er at de kan redusere risikoen for personvernbrudd og skape større og mer varierte datagrunnlag.

Data, både reelle og syntetiske som skal brukes til trening av ML-modeller, må være representative for den populasjonen den ferdig trente modellen skal brukes på og ikke inneholde systematiske skjevheter eller urettferdigheter. Syntetiske data må derfor ha stor statistisk likhet med de originale dataene. Syntetiske data må i tillegg ha høy nytte, det vil si være egnet for det konkrete formålet de skal brukes til. Det er samtidig et dilemma at de ikke kan være for lik de originale dataene, da dette øker risikoen for re-identifisering. Siden syntetiske data er kunstig generert inneholder de i utgangspunktet ikke personidentifiserbar informasjon. Det har vist seg at det likevel er risiko for lekkasje av informasjon fra de originale treningsdataene. For å redusere risikoen for personvernbrudd i syntetiske datasett kan det være nødvendig å bruke flere personvern fremmende teknologier.

For å sikre at de syntetiske dataene er egnet til formålet og har lav risiko for personvernbrudd er det nødvendig med empiriske evalueringer. Det er en utfordring at det ikke eksisterer standardiserte måter å evaluere statistisk likhet mellom syntetiske og originale datasett, og nytte eller personvern i syntetiske datasett. For videre utvikling av fagområdet er det behov for standardiserte og objektive metoder for evaluering av syntetiske datasett, og som bygges inn i effektive verktøy. Det er også uklart hvor kostnadseffektive syntetiske data er og om fordelene med syntetiske data veier opp for ressursbruken.

Hovedbudskapet i denne kunnskapsoppsummeringen er at syntetiske data kan ha nytte, men de bør brukes med forsiktighet. Det er ikke mulig å gi et generelt svar på nytten av å bruke syntetiske data til å trene maskinlæringsmodeller. Svaret vil variere for ulike brukstilfeller og må avgjøres ved å utføre empiriske evalueringer. Mer forskning på feltet og utvikling av metoder og verktøy for generering og evaluering av syntetiske data vil gi bedre svar på nytteverdien ved bruk av syntetiske data som metode.

6 Konklusjon

I denne rapporten har vi vurdert to personvern fremmende teknologier med stort potensial til å være nyttige verktøy ved bruk av kunstig intelligens i helse- og omsorgstjenesten: føderert læring og syntetiske data.

Føderert læring er en type maskinlæring som gjør det mulig for flere organisasjoner å samarbeide om trening av en maskinlæringsmodell, uten at de må dele dataene sine. De viktigste fordelene er:

- tilgang til større og mer representative datagrunnlag,
- bedre kontroll over egne data,
- styrket personvern,
- redusert risiko for dataeksponering.

Bruk av føderert læring gjør det enklere å oppfylle kravene til rettslig grunnlag for databehandlingen og oppnå samsvar med personvernprinsippene.

Denne teknologien har blitt brukt i flere helse relaterte prosjekter, med lovende resultater. Bruk av føderert læring reiser likevel flere problemstillinger som må håndteres:

- variasjoner i dataegenskaper mellom organisasjonene, f.eks. datakvantitet, -kvalitet og -strukturer, som påvirker dataenes statistiske heterogenitet og dermed gyldigheten av analyseresultater,
- variasjoner i tekniske spesifikasjoner, som lagrings-, beregnings- og kommunikasjonskapasitetene,
- mulig økning i ressursbruk (her gjelder det økonomiske ressurser og miljøavtrykk),
- økende trykk på et kommunikasjonssystem og eventuelle flaskehalsar,
- potensielle datasikkerhetstrusler.

Helsesdata har stor verdi og er svært sensitive. Det er derfor nødvendig med videreutvikling av effektive kryptografiske og personvern fremmende teknikker, med hensyn til både datasikkerhet og ressursbruk. For å unngå diskriminerende resultater trengs det metoder som kan sjekke statistiske parametere på distribuerte datasett uten å bryte personvernet. Dessuten er det behov for mer forskning på hvordan nøyaktigheten til dataanalyseresultater påvirkes ved bruk av personvern fremmende teknikker.

Syntetiske data er kunstig genererte data som er basert på reelle data. De beholder de statistiske egenskapene til det originale datasettet, uten å inneholde faktisk pasientinformasjon. Syntetiske data kan derfor øke mengden data for trening av ML-modeller. Fordelene med teknologien er:

- større og mer representative datagrunnlag,
- redusert risiko for re-identifisering av enkeltpersoner,
- forbedret datakvalitet.

Syntetiske data som er generert med en maskinlæringsmodell som er trent med et reelt datasett, kan fortsatt lekke informasjon fra det originale datasettet. I tillegg har bruk av syntetiske data til trening av ML-modeller andre utfordringer:

- evaluering av de syntetiske dataene med hensyn til mulige nye og/eller forsterkede skjevheter,
- mangel på standardiserte måter å evaluere statistisk likhet med det originale datasettet, nytte og personvern i syntetiske datasett,
- uklar kostnads- og nytteeffektivitet.

Bedre og mer automatiserte verktøy og metoder trengs for å redusere kostnader og tidsbruk ved generering og validering av syntetiske data. Det er også behov for standardiserte og objektive metoder

for evaluering om syntetiske data er egnet til formålet og har lav risiko for personvernbrudd. Syntetiske data kan med fordel kombineres med andre personvern fremmende teknologier for å redusere personvernrisiko.

Både føderert læring og syntetiske data kan bidra til personvernbevarende dataanalyse i helse- og omsorgstjenester og sikre forsvarlig og nyttig bruk av KI. Begge teknologiene har sine bruksområder, fordeler og utfordringer. Samtidig ser vi behov for mer forskning, videreutvikling av verktøy og metoder, samt nødvendigheten av praktiske utprøvinger for at mulighetene som ligger i teknologiene skal kunne realiseres fullt ut.

7 Referanseliste

1. Helsedirektoratet. Det nasjonale koordineringsprosjektet «Bedre bruk av kunstig intelligens» [Internett]. 2023 [sitert 25. august 2023]. Tilgjengelig på: <https://www.helsedirektoratet.no/tema/kunstig-intelligens/gi-innspill-og-bidra>
2. Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten [Internett]. Direktoratet for e-helse; 2022 okt [sitert 28. juni 2023]. Report No.: IE-1105. Tilgjengelig på: <https://www.ehelse.no/publikasjoner/tilgang-til-data-til-kunstig-intelligens-i-helse-og-omsorgstjenesten>
3. Datatilsynet. Hva er personopplysninger? [Internett]. 2023 [sitert 30. august 2023]. Tilgjengelig på: <https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/>
4. Datatilsynet. Digitale tjenester og forbrukeres personopplysninger [Internett]. 2023 [sitert 30. august 2023]. Tilgjengelig på: <https://www.datatilsynet.no/personvern-pa-ulike-omrader/kundehandtering-handel-og-medlemskap/digitale-tjenester-og-forbrukeres-personopplysninger/>
5. Datatilsynet. Spesielt om særlige kategorier av personopplysninger (sensitive personopplysninger) - forbud og unntak. [Internett]. 2023 [sitert 30. august 2023]. Tilgjengelig på: <https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/om-behandlingsgrunnlag/spesielt-om-sarlige-kategorier-av-personopplysninger/>
6. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [Internett]. 2016 [sitert 30. august 2023]. Tilgjengelig på: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>
7. Datatilsynet. Om personopplysningsloven med forordning og når den gjelder [Internett]. 2023 [sitert 30. august 2023]. Tilgjengelig på: <https://www.datatilsynet.no/regelverk-og-verktoy/lover-og-regler/om-personopplysningsloven-og-nar-den-gjelder/>
8. Datatilsynet. Grunnleggende personvernprinsipper [Internett]. 2019 [sitert 28. juni 2023]. Tilgjengelig på: <https://www.datatilsynet.no/rettigheter-og-plikter/personvernprinsippene/grunnleggende-personvernprinsipper/>
9. Datatilsynet. Behandlingsgrunnlag [Internett]. 2023 [sitert 30. august 2023]. Tilgjengelig på: <https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/om-behandlingsgrunnlag/>
10. ICT&health. The European Health Data Space proposal (EHDS) explained [Internett]. 2022 [sitert 14. mai 2023]. Tilgjengelig på: <https://ictandhealth.com/the-european-health-data-space-proposal-ehds-explained/news/>
11. Europakommisjonen. Forslag til dataloven [Internett]. 2022 [sitert 10. august 2023]. Tilgjengelig på: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
12. Europakommisjonen. Lov om datastyring [Internett]. 2022 [sitert 10. mai 2023]. Tilgjengelig på: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020PC0767>
13. Europakommisjonen. NIS2-direktivet om cybersikkerhet [Internett]. 2023 [sitert 10. mai 2023]. Tilgjengelig på: <https://digital-strategy.ec.europa.eu/en/policies/nis2-directive>

14. Lov om behandling av personopplysninger (personopplysningsloven) [Internett]. [sitert 21. mars 2023]. Tilgjengelig på: <https://lovdata.no/dokument/NL/lov/2018-06-15-38/>
15. OECD. Emerging privacy-enhancing technologies. Current regulatory and policy approaches [Internett]. 2023 [sitert 30. august 2023] s. 51. Report No.: DSTI/CDEP/2022/10/FINAL. Tilgjengelig på: <https://www.oecd-ilibrary.org/docserver/bf121be4-en.pdf?expires=1693395320&id=id&ac-name=guest&checksum=BC55A4E69890E525EE803EFDB1AE4717>
16. McMahan B, Ramage D. Federated Learning: Collaborative Machine Learning without Centralized Training Data [Internett]. Google AI Blog. 2017 [sitert 18. juli 2022]. Tilgjengelig på: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>
17. Red Hat. What is edge machine learning? [Internett]. 2022 [sitert 12. oktober 2023]. Tilgjengelig på: <https://www.redhat.com/en/topics/edge-computing/what-is-edge-machine-learning>
18. Vaadata. What is Pseudonymisation? Techniques and Best Practices [Internett]. 2023 [sitert 12. oktober 2023]. Tilgjengelig på: <https://www.vaadata.com/blog/what-is-pseudonymisation-techniques-and-best-practices/>
19. Coutinho-Almeida J, Rodrigues PP, Cruz-Correia RJ. GANs for Tabular Healthcare Data Generation: A Review on Utility and Privacy. I: Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings [Internett]. Berlin, Heidelberg: Springer-Verlag; 2021 [sitert 7. juni 2023]. s. 282–91. Tilgjengelig på: https://doi.org/10.1007/978-3-030-88942-5_22
20. Wood A, Altman M, Bembenek A, Bun M, Gaboardi M, Honaker J, mfl. Differential Privacy: A Primer for a Non-Technical Audience. Vanderbilt J Entertain Technol Law [Internett]. 2018;21:209. Tilgjengelig på: <https://heinonline.org/HOL/Page?handle=hein.journals/vanep21&id=219&div=&collection=>
21. Wood A, Najarian K, Kahrobaei D. Homomorphic Encryption for Machine Learning in Medicine and Bioinformatics. ACM Comput Surv [Internett]. august 2020 [sitert 25. oktober 2022];53(4):70:1-70:35. Tilgjengelig på: <http://doi.org/10.1145/3394658>
22. Andersen A. Using secure multi-party computation when processing distributed health data. I: Proceedings of the 2013 International Conference on Security & Management [Internett]. CSREA Press; 2013. Tilgjengelig på: <http://worldcomp-proceedings.com/proc/p2013/SAM9777.pdf>
23. Ethereum. What are zero-knowledge proofs? [Internett]. 2023 [sitert 12. oktober 2023]. Tilgjengelig på: <https://ethereum.org/en/zero-knowledge-proofs/>
24. Schwarz, C.G., Kremers, W.K., Lowe, V.J., Savvides, M., Gunter, J.L., Senjem, M.L., Vemuri, P., Kantarci, K., Knopman, D.S., Petersen, R.C., Jack, C.R. Face recognition from research brain PET: An unexpected PET problem. NeuroImage. 2022;258.
25. Schwarz, C.G., Kremers, W.K., Therneau, T.M., Sharp, R.R., Gunter, J.L., Vemuri, P., Arani, A., Spsychalla, A.J., Kantarci, K., Knopman, D.S., Jack, C.R. Identification of anonymous mri research participants with face recognition software. N Engl J Med. 2019;
26. Schwarz, C.G., Kremers, W.K., Wiste, H.J., Gunter, J.L., Vemuri, P., Spsychalla, A.J., Kantarci, K., Schultz, A.P., Sperling, R.A., Knopman, D.S., Petersen, R.C., Jack, C.R. Changing the face of

- neuroimaging research: comparing a new MRI de-facing technique with popular alternatives. *Neuroimage*. 2021;231.
27. NVIDIA. What is federated learning? [Internett]. 2019 [sitert 28. juni 2023]. Tilgjengelig på: <https://blogs.nvidia.com/blog/2019/10/13/what-is-federated-learning/>
 28. World Economic Forum. Sharing Sensitive Health Data in a Federated Data Consortium Model. An Eight-Step Guide [Internett]. 2020 [sitert 28. juni 2023]. Tilgjengelig på: https://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf
 29. Sheller, M. J., Reina, G. A., Edwards, B., Martin, J. & Bakas, S. Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. I: International MICCAI Brain lesion Workshop. Springer; 2018. s. 92–104.
 30. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, mfl. The future of digital health with federated learning. *Npj Digit Med* [Internett]. september 2020 [sitert 29. juni 2022];3(1):1–7. Tilgjengelig på: <https://www.nature.com/articles/s41746-020-00323-1>
 31. Tian, L., Kumar Sahu, A., Talwalkar, A. S., Smith, V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process Mag*. 2020;37.
 32. Open Data Institute. Federated learning: an introduction. Considerations and practical guidance for prospective adopters [Internett]. 2023 [sitert 28. juni 2023]. Tilgjengelig på: <https://www.theodi.org/article/federated-learning-an-introduction-report/>
 33. Anderson, M.L. & Oates, T. A review of recent research in metareasoning and metalearning. *AI Mag*. 2007;
 34. Caruana, R. Multitask Learning. I: *Machine Learning*. Springer; 1997. s. 41–75.
 35. Kim, H. Recent Improvements of MPI Communication for DDLS [Internett]. 2021 [sitert 28. juni 2023]. Tilgjengelig på: <https://hk3342.medium.com/recent-improvements-of-mpi-communication-74e3c4a1ccb4>
 36. Qiu, X., Parcollet, T., Beutel, D. J., Topal, T., Mathur, A., Lane, N. D. Can Federated Learning Save The Planet? I: Tackling Climate Change with Machine Learning workshop at NeurIPS [Internett]. 2020 [sitert 28. juni 2023]. Tilgjengelig på: <https://arxiv.org/abs/2010.06537>
 37. Benmalek, M., Benrekia, M.A., Challal, Y. Security of Federated Learning: Attacks, Defensive Mechanisms, and Challenges. *Rev Sci Technol L'Information – Sér RIA Rev D'Intelligence Artificielle*. 2022;36 (1):49–59.
 38. Chen, Y., Su, L., Xu, J. Distributed statistical machine learning in adversarial settings: Byzantine gradient descent. *ACM Meas Anal Comput Syst*. 2017;1(2):1–25.
 39. Ang, F., Chen, L., Zhao, N., Chen, Y., Wang, W., Yu,, F.R. Robust federated learning with noisy communication. *IEEE Trans Commun* [Internett]. 2020 [sitert 10. august 2023];68(6):3452–64. Tilgjengelig på: <https://ieeexplore.ieee.org/document/9026922>
 40. Zhu, Z., Hong, J., Zhou, J. Data-Free Knowledge Distillation for Heterogeneous Federated Learning. I: The 38th International Conference on Machine Learning. 2021. s. 12878–89.
 41. evervault. What is a Trusted Execution Environment (TEE)? [Internett]. 2023 [sitert 12. oktober 2023]. Tilgjengelig på: <https://evervault.com/blog/what-is-a-trusted-execution-environment-tee>

42. Qammar A, Karim A, Ning H, Ding J. Securing federated learning with blockchain: a systematic literature review. *Artif Intell Rev* [Internett]. september 2022 [sitert 26. oktober 2022]; Tilgjengelig på: <https://doi.org/10.1007/s10462-022-10271-9>
43. Truong, N., Sun, K., Wang, S., Guitton, F., Guo, Y. Privacy preservation in federated learning: an insightful survey from the GDPR perspective. *Comput Secur.* 2021;110:102402.
44. Kurupathi, S., Maass, W. Survey on federated learning towards privacy preserving AI. *Comput Sci Inf Technol.* 2020;10(11):235.
45. Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M., Bakhtiari, M., Völker, U., Baumbach, J., Baumbach, L., Buchholtz, G. Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review. *J Med Internet Res* [Internett]. 2023 [sitert 28. juni 2023];25:e41588. Tilgjengelig på: <https://www.jmir.org/2023/1/e41588>
46. Lee J, Sun J, Wang F, Wang S, Jun CH, Jiang X. Privacy-Preserving Patient Similarity Learning in a Federated Environment: Development and Analysis. *JMIR Med Inform* [Internett]. april 2018 [sitert 26. oktober 2022];6(2):e20. Tilgjengelig på: <http://medinform.jmir.org/2018/2/e20/>
47. Universitetet i Bergen. Forskningsnettverk i primærhelsetjenesten. *PraksisNett* [Internett]. 2023 [sitert 28. juni 2023]. Tilgjengelig på: <https://www.uib.no/praksisnett>
48. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated Electronic Health Records. *Int J Med Inf.* 2018;112:59–67.
49. Huang, L. et al. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *J Biomed Inf.* 2019;99:103291.
50. Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., Feng, A. Privacy-preserving federated brain tumour segmentation. I: International Workshop on Machine Learning in Medical Imaging. Springer; 2019. s. 133–41.
51. HTN. Owkin, NVIDIA and King's College London partner to deliver AI [Internett]. 2019 [sitert 28. juni 2023]. Tilgjengelig på: <https://htn.co.uk/2019/12/04/owkin-nvidia-and-kings-college-london-partner-to-deliver-ai/>
52. Bigpicture [Internett]. 2023 [sitert 10. august 2023]. Tilgjengelig på: <https://bigpicture.eu/>
53. Labelia Labs. HealthChain [Internett]. 2023 [sitert 19. mai 2023]. Tilgjengelig på: <https://www.labelia.org/en/healthchain-project>
54. University of Pennsylvania. Perelman School of Medicine. Center for Biomedical Image Computing & Analytics. Federated Tumor Segmentation (FeTS) initiative [Internett]. 2023 [sitert 10. juni 2023]. Tilgjengelig på: <https://www.med.upenn.edu/cbica/fets/>
55. Trustworthy Federated Data Analytics project (TFDA) [Internett]. 2022 [sitert 6. juni 2023]. Tilgjengelig på: <https://tfda.hmsp.center/>
56. DKTK. German Cancer Consortium. Joint Imaging Platform [Internett]. 2023 [sitert 10. mai 2023]. Tilgjengelig på: <https://jip.dtkk.dkfz.de/jiphomepage/>
57. OPTIMA [Internett]. 2020 [sitert 10. mai 2023]. Tilgjengelig på: <https://www.optima-oncology.eu/>

58. Vantage. Comparing the quality of breast cancer care in Norway and the Netherlands using Vantage6 [Internett]. 2023 [sitert 10. juni 2023]. Tilgjengelig på: <https://distributedlearning.ai/news/comparing-quality-breast-cancer-care-norway-and-netherlands-using-vantage6/>
59. Kreftregisteret. FLORENCE: føderert læring på kreftdata [Internett]. [sitert 29. juni 2023]. Tilgjengelig på: <https://www.kreftregisteret.no/Forskning/Prosjekter/omop-prosjekter/florence/>
60. Siegel, M. Federated ML Model Evaluation. [Internett]. 2022 [sitert 28. juni 2023]. Tilgjengelig på: <https://www.bitfount.com/pets-explained/federated-ml-model-evaluation>
61. Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P., Duncan, J.S. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. *Med Image Anal.* 2020;65:101765.
62. Data.org. Epiverse. The Global Epidemic Response of the Future [Internett]. 2023 [sitert 25. august 2023]. Tilgjengelig på: <https://data.org/initiatives/epiverse/>
63. HealthData@EU Pilot [Internett]. 2022 [sitert 22. mai 2023]. Tilgjengelig på: <https://ehds2pilot.eu/>
64. FederatedHealth: A Nordic Federated Health Data Network. Nordic Innovation. [Internett]. 2023 [sitert 23. mai 2023]. Tilgjengelig på: <https://www.nordicinnovation.org/programs/federated-health-nordic-federated-health-data-network>
65. Elixir. Federated Human Data Community [Internett]. 2020 [sitert 6. juni 2023]. Tilgjengelig på: <https://elixir-europe.org/communities/human-data>
66. Dou, Q., So, T.Y., Jiang, M. et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *Npj Digit Med.* 2021;4:60.
67. Sankar, L., Zhao, M., Trieu, N., Berisha, V. FACT: Federated Analytics based Contact Tracing for COVID-19 [Internett]. 2020 [sitert 28. juni 2023]. Tilgjengelig på: <https://sankar.engineering.asu.edu/fact-federated-analytics-based-contact-tracing-for-covid-19/>
68. MELLODDY. Machine Learning Ledger Orchestration for Drug Discovery [Internett]. 2023 [sitert 30. mai 2023]. Tilgjengelig på: <https://www.melloddy.eu/>
69. Kings College London. Launch of the London Medical Imaging & Artificial Intelligence Centre for Value-Based Healthcare [Internett]. 2019 [sitert 21. mai 2023]. Tilgjengelig på: <https://www.kcl.ac.uk/news/launch-of-the-london-medical-imaging-artificial-intelligence-centre-for-value-based-healthcare>
70. Kings College London [Internett]. 2023 [sitert 19. mai 2023]. Tilgjengelig på: <https://www.kcl.ac.uk/>
71. NVIDIA. Health and Life Sciences [Internett]. 2023 [sitert 28. juni 2023]. Tilgjengelig på: <https://www.nvidia.com/en-us/industries/healthcare-life-sciences/>
72. OWKIN [Internett]. 2023 [sitert 19. mai 2023]. Tilgjengelig på: <https://owkin.com/>
73. Bitfount [Internett]. 2023 [sitert 30. mai 2023]. Tilgjengelig på: <https://www.bitfount.com/>
74. Moorfields øyesykehus [Internett]. 2023 [sitert 19. mai 2023]. Tilgjengelig på: <https://www.moorfields.nhs.uk/>

75. Data.org. Rockefeller [Internett]. 2023 [sitert 28. juni 2023]. Tilgjengelig på: <https://data.org/organizations/the-rockefeller-foundation/>
76. Data.org. Wellcome [Internett]. 2023 [sitert 28. juni 2023]. Tilgjengelig på: <https://data.org/organizations/wellcome/>
77. Sciensano. EHDS2 PILOT – European health data space pilot for secondary use of health data [Internett]. 2022 [sitert 10. juni 2023]. Tilgjengelig på: <https://www.sciensano.be/en/projects/european-health-data-space-pilot-secondary-use-health-data>
78. Federated European Genome-phenom Archive. Federated EGA Norway node. [Internett]. 2023 [sitert 20. mai 2023]. Tilgjengelig på: <https://ega.elixir.no/>
79. Kreftregisteret. OMOP-prosjekter [Internett]. 2023 [sitert 29. juni 2023]. Tilgjengelig på: <https://www.kreftregisteret.no/Forskning/Prosjekter/omop-prosjekter/>
80. Kristoffersen, E.S., Bjorvatn, B., Halvorsen, P.A., Nilsen, S., Fossum, G.H., Fors, E.A., Jørgensen, P., Øxnevad-Gundersen, B., Gjelstad, S., Bellika, J.G., Straand, J., Rørtveit, G. The Norwegian Praksis-Nett: a nationwide practice-based research network with a novel IT infrastructure. *Scand J Prim Health Care* [Internett]. 2022;40(2):217–26. Tilgjengelig på: <https://www.tandfonline.com/doi/full/10.1080/02813432.2022.2073966>
81. Behov for data til kunstig intelligens i helsetjenesten [Internett]. Direktoratet for e-helse; 2022 feb [sitert 28. juni 2023]. Report No.: IE-1096. Tilgjengelig på: <https://www.ehelse.no/publikasjoner/behov-for-data-til-kunstig-intelligens-i-helsetjenesten>
82. Introducing ChatGPT [Internett]. [sitert 17. september 2023]. Tilgjengelig på: <https://openai.com/blog/chatgpt>
83. Pichai S. An important next step on our AI journey [Internett]. Google. 2023 [sitert 17. september 2023]. Tilgjengelig på: <https://blog.google/technology/ai/bard-google-ai-search-updates/>
84. Azizi Z, Zheng C, Mosquera L, Pilote L, El Emam K. Can synthetic data be a proxy for real clinical trial data? A validation study. *BMJ Open* [Internett]. april 2021 [sitert 24. mai 2022];11(4):e043497. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8055130/>
85. Jordon J, Szpruch L, Houssiau F, Bottarelli M, Cherubin G, Maple C, mfl. Synthetic Data – what, why and how? [Internett]. arXiv; 2022 [sitert 24. januar 2023]. Tilgjengelig på: <http://arxiv.org/abs/2205.03257>
86. Lu Y, Wang H, Wei W. Machine Learning for Synthetic Data Generation: A Review [Internett]. arXiv; 2023 [sitert 7. juni 2023]. Tilgjengelig på: <http://arxiv.org/abs/2302.04062>
87. Rankin, D., Black, M., Bond, R., Wallace, J., Mulvenna, M., Epelde, G. Reliability of Supervised Machine Learning Using Synthetic Data in Healthcare: A Model to Preserve Privacy for Data Sharing. *JMIR Med Inform.* 8(7):e18910.
88. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, mfl. MIMIC-III, a freely accessible critical care database. *Sci Data.* mai 2016;3:160035.
89. Yale A, Dash S, Dutta R, Guyon I, Pavao A, Bennett KP. Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing* [Internett]. november 2020 [sitert 7. juni

- 2023];416:244–55. Tilgjengelig på: <https://www.sciencedirect.com/science/article/pii/S0925231220305117>
90. Bellovin SM, Dutta PK, Reitinger N. Privacy and Synthetic Datasets. *Stanf Technol Law Rev* [Internett]. 2019 [sitert 19. april 2023];22(1). Tilgjengelig på: <https://papers.ssrn.com/abstract=3255766>
 91. van Breugel B, Kyono T, Berrevoets J, van der Schaar M. DECAF: Generating Fair Synthetic Data Using Causally-Aware Generative Networks. I: *Advances in Neural Information Processing Systems* [Internett]. Curran Associates, Inc.; 2021 [sitert 8. mars 2023]. s. 22221–33. Tilgjengelig på: <https://proceedings.neurips.cc/paper/2021/hash/ba9fab001f67381e56e410575874d967-Abstract.html>
 92. Tiwald P, Ebert A, Soukup DT. Representative & Fair Synthetic Data [Internett]. *arXiv*; 2021 [sitert 8. juni 2023]. Tilgjengelig på: <http://arxiv.org/abs/2104.03007>
 93. El Emam K, Mosquera L, Hoptroff R. *Practical Synthetic Data Generation. Balancing Privacy and the Broad Availability of Data*. O'Reilly Media, Inc.; 2020.
 94. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, mfl. Generative adversarial networks. *Commun ACM* [Internett]. oktober 2020 [sitert 21. februar 2023];63(11):139–44. Tilgjengelig på: <https://doi.org/10.1145/3422622>
 95. Figueira A, Vaz B. Survey on Synthetic Data Generation, Evaluation Methods and GANs. *Mathematics* [Internett]. januar 2022 [sitert 20. desember 2022];10(15):2733. Tilgjengelig på: <https://www.mdpi.com/2227-7390/10/15/2733>
 96. Kingma DP, Welling M. Auto-Encoding Variational Bayes [Internett]. *arXiv*; 2022 [sitert 21. februar 2023]. Tilgjengelig på: <http://arxiv.org/abs/1312.6114>
 97. Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: A systematic review. *Neurocomputing* [Internett]. juli 2022 [sitert 27. oktober 2022];493:28–45. Tilgjengelig på: <https://www.sciencedirect.com/science/article/pii/S0925231222004349>
 98. Ghosheh G, Li J, Zhu T. A review of Generative Adversarial Networks for Electronic Health Records: applications, evaluation measures and data sources [Internett]. *arXiv*; 2022 [sitert 7. juni 2023]. Tilgjengelig på: <http://arxiv.org/abs/2203.07018>
 99. Nik AHZ, Riegler MA, Halvorsen P, Storås AM. Generation of Synthetic Tabular Healthcare Data Using Generative Adversarial Networks. I: Dang-Nguyen DT, Gurrin C, Larson M, Smeaton AF, Rudinac S, Dao MS, mfl., redaktører. *MultiMedia Modeling*. Cham: Springer International Publishing; 2023. s. 434–46. (Lecture Notes in Computer Science).
 100. Yan C, Yan Y, Wan Z, Zhang Z, Omberg L, Guinney J, mfl. A Multifaceted benchmarking of synthetic electronic health record generation models. *Nat Commun* [Internett]. desember 2022 [sitert 13. juni 2023];13(1):7609. Tilgjengelig på: <https://www.nature.com/articles/s41467-022-35295-1>
 101. Nowok B, Raab GM, Dibben C. synthpop: Bespoke Creation of Synthetic Data in R. *J Stat Softw* [Internett]. oktober 2016 [sitert 1. september 2022];74:1–26. Tilgjengelig på: <https://doi.org/10.18637/jss.v074.i11>

102. Murtaza H, Ahmed M, Khan NF, Murtaza G, Zafar S, Bano A. Synthetic data generation: State of the art in health care domain. *Comput Sci Rev [Internett]*. mai 2023 [sitert 2. mai 2023];48:100546. Tilgjengelig på: <https://www.sciencedirect.com/science/article/pii/S1574013723000138>
103. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, mfl. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc [Internett]*. mars 2018 [sitert 16. mai 2022];25(3):230–8. Tilgjengelig på: <https://doi.org/10.1093/jamia/ocx079>
104. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, mfl. Attention is All you Need. I: *Advances in Neural Information Processing Systems [Internett]*. Long Beach, CA, USA: Curran Associates, Inc.; 2017 [sitert 3. juli 2023]. Tilgjengelig på: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
105. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare [Internett]*. mars 2023 [sitert 22. august 2023];11(6):887. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10048148/>
106. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. juli 2023;6(1):120.
107. Shoja MM, Van de Ridder JMM, Rajput V. The Emerging Role of Generative Artificial Intelligence in Medical Education, Research, and Practice. *Cureus [Internett]*. [sitert 22. august 2023];15(6):e40883. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10363933/>
108. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell [Internett]*. mai 2023 [sitert 22. august 2023];6:1169595. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10192861/>
109. Esteban C, Hyland SL, Rättsch G. Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs [Internett]. *arXiv; 2017 [sitert 15. mars 2023]*. Tilgjengelig på: <http://arxiv.org/abs/1706.02633>
110. Patki N, Wedge R, Veeramachaneni K. The Synthetic Data Vault. I: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 2016. s. 399–410.
111. Arora A, Arora A. Machine learning models trained on synthetic datasets of multiple sample sizes for the use of predicting blood pressure from clinical data in a national dataset. *PLOS ONE [Internett]*. mars 2023 [sitert 20. september 2023];18(3):e0283094. Tilgjengelig på: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0283094>
112. Benaim AR, Almog R, Gorelik Y, Hochberg I, Nassar L, Mashiach T, mfl. Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies. *JMIR Med Inform [Internett]*. februar 2020 [sitert 26. juli 2022];8(2):e16492. Tilgjengelig på: <https://medinform.jmir.org/2020/2/e16492>
113. Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. *BMC Med Res Methodol [Internett]*. mai 2020 [sitert 4. mai 2022];20(1):108. Tilgjengelig på: <https://doi.org/10.1186/s12874-020-00977-1>

114. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. I 2017.
115. Rabin MRI, Hussain A, Alipour MA, Hellendoorn VJ. Memorization and generalization in neural code intelligence models. *Inf Softw Technol* [Internett]. januar 2023 [sitert 8. februar 2023];153:107066. Tilgjengelig på: <https://www.sciencedirect.com/science/article/pii/S0950584922001756>
116. Arpit D, Jastrzębski S, Ballas N, Krueger D, Bengio E, Kanwal MS, mfl. A closer look at memorization in deep networks. I: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. Sydney, NSW, Australia: JMLR.org; 2017. s. 233–42. (ICML'17).
117. Emam KE, Mosquera L, Bass J. Evaluating Identity Disclosure Risk in Fully Synthetic Health Data: Model Development and Validation. *J Med Internet Res* [Internett]. november 2020 [sitert 4. august 2022];22(11):e23139. Tilgjengelig på: <https://www.jmir.org/2020/11/e23139>
118. Dwork C, McSherry F, Nissim K, Smith A. Calibrating Noise to Sensitivity in Private Data Analysis. I: Halevi S, Rabin T, redaktører. *Theory of Cryptography*. Berlin, Heidelberg: Springer; 2006. s. 265–84. (Lecture Notes in Computer Science; bd. 3876).
119. Jordon J, Yoon J, Schaar M van der. PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees. I 2022 [sitert 3. januar 2023]. Tilgjengelig på: <https://openreview.net/forum?id=S1zk9iRqF7>
120. Ficek J, Wang W, Chen H, Dagne G, Daley E. Differential privacy in health research: A scoping review. *J Am Med Inform Assoc JAMIA* [Internett]. august 2021 [sitert 9. mars 2023];28(10):2269–76. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8449619/>
121. Arora A, Arora A. Synthetic patient data in health care: a widening legal loophole. *The Lancet* [Internett]. april 2022 [sitert 10. mai 2022];399(10335):1601–2. Tilgjengelig på: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(22\)00232-X/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)00232-X/fulltext)
122. Stadler T, Oprisanu B, Troncoso C. Synthetic Data - Anonymisation Groundhog Day [Internett]. arXiv; 2022 [sitert 2. mars 2023]. Tilgjengelig på: <http://arxiv.org/abs/2011.07018>
123. van Breugel B, Sun H, Qian Z, van der Schaar M. Membership Inference Attacks against Synthetic Data through Overfitting Detection [Internett]. arXiv; 2023 [sitert 18. april 2023]. Tilgjengelig på: <http://arxiv.org/abs/2302.12580>
124. Chen D, Yu N, Zhang Y, Fritz M. GAN-Leaks: A Taxonomy of Membership Inference Attacks against Generative Models. I: *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security* [Internett]. New York, NY, USA: Association for Computing Machinery; 2020 [sitert 21. juni 2023]. s. 343–62. (CCS '20). Tilgjengelig på: <https://dl.acm.org/doi/10.1145/3372297.3417238>
125. Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical Black-Box Attacks against Machine Learning [Internett]. arXiv; 2017 [sitert 4. mai 2023]. Tilgjengelig på: <http://arxiv.org/abs/1602.02697>
126. Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* [Internett]. juni 2021 [sitert 16. februar 2023];5(6):493–7. Tilgjengelig på: <https://www.nature.com/articles/s41551-021-00751-8>