

Pierre Lison  
Norsk Regnesentral

Vår ref.: 2020/178

REKs ref.: 139585

Dato: 05.02.2021

## Maskinlæring til anonymisering av ustrukturerte persondata

Den nasjonale forskningsetiske komité for medisin og helsefag (NEM) viser til klage på REK sør-øst D sitt vedtak datert 24.11.2020. REK tok ikke klagen til følge, og klagen ble oversendt NEM for videre behandling.

NEM tok klagen til behandling i sitt møte den 19.01.2021. Klagen ble behandlet med hjemmel i forskningsetikkloven § 10 (3) og forvaltningsloven § 34.

### Kort om prosjektet

Dette prosjektet handler om å utvikle en algoritme for automatisk anonymisering av ustrukturerte persondata, herunder journaldata. Norsk Regnesentral er forskningsansvarlig.

Forkortet versjon av søkers beskrivelse av prosjektet:

*Offentlige etater, helseforetak og private bedrifter har utfordringer med personlig data som de behandler, samler inn eller produserer. Databaser med slike data er ofte svært verdifulle.*

*Data som kan inneholde personlig informasjon må behandles i henhold til gjeldende personvernregler. GDPR sier blant annet at persondata ikke kan utleveres til en tredjepart (eller bli brukt for andre formål enn det dataene ble samlet inn for) uten rettslig grunnlag, for eksempel gjennom å ha innhentet samtykke fra de berørte individene. En løsning er å benytte anonymiseringsteknikker for å beskytte personvernet til de registrerte individene. Dagens anonymiseringsteknikker fungerer ikke godt nok for ustrukturerte data. Dette er en stor utfordring, siden dataene i mange saksbehandlingssystemer i stor grad består av tekst (som for eksempel pasientjournaler). Manuell anonymisering er dessuten svært kostbart, følsomt for menneskelige feil og vil være inkonsistent.*

*CLEANUP-prosjektet skal tette dette teknologiske gapet og utvikle nye maskinlæringsmodeller for å automatisk anonymisere tekstdokumenter. CLEANUP vil også utvikle nye metoder for å evaluere kvaliteten til de nye tekstanonymiseringsteknikkene og samordne disse krav fra lover og reguleringer. Til sist vil CLEANUP undersøke hvordan de teknologiske løsningene kan integreres i de ulike organisasjonene, spesielt hvordan kvalitetskontroll bør gjennomføres og tilpasses behovene til den enkelte dataeier.*

*Konkret skal CLEANUP:*

- 1) samle tekstmaterialer fra ulike kilder, blant annet pasientjournaler fra avdøde pasienter*
- 2) annotere pasientjournalene for å markere deler av teksten som inneholder personopplysninger*
- 3) utvikle maskinlæringsmodeller (trent på den annoterte data nevnt ovenfor) som automatisk skal gjenkjenne personopplysninger i tekst, og redigere dokumentene for å maskere disse opplysningene.*
- 4) evaluere hvor godt disse anonymiseringsmodellene fungerer, blant annet i hvilke grad de bidrar til å redusere avsløringsrisikoen knyttet til ulike tekstdokumenter.*

Det skal brukes 3000 pasientjournaler fra personer som døde for ti år siden eller mer. Valg av pasientjournaler vil følge to kriterier:

- 1) pasientjournalene som skal hentes inn må enten allerede være digitalisert eller lette å lese via OCR (optisk tegngjenkjenning),
- 2) i tillegg vil «lengre» pasientjournaler prioriteres (som er mer utfordrende og interessante å anonymisere).

### **Kort sakshistorikk**

Saken ble først behandlet hos REK sør-øst D i møte den 10.06.2020. I sitt vedtak av 30.06.2020 hadde REK en rekke merknader, og kom fram til at det ikke forelå tilstrekkelig informasjon til å treffe et endelig vedtak. Ettersom flere aktører var involvert, ønsket REK en redegjørelse for hvilken rolle disse aktørene skulle spille i gjennomføringen av prosjektet. REK ba også om en avklaring på hvor mange og hvem som skal ha tilgang til journalopplysningene. Videre ble det stilt spørsmål ved hvilke kriterier som skal legges til grunn i vurderingen av hvorvidt algoritmen fungerer godt nok. Det var også usikkerhet rundt hvorfor det skulle benyttes reelle journaler, og de ba om en vurdering av om prosjektet kunne gjennomføres med fiktive pasientjournaler. REK ba videre om tilbakemelding på prosjektets forventede samfunnsnytte.

Tilbakemeldingen fra prosjektleder ble mottatt 06.08.2020, og behandlet i møte den 19.08.2020. I sitt vedtak av 09.09.2020 kom REK fram til at spørsmålet om hvilke kriterier som skal legges til grunn i vurderingen av hvorvidt algoritmen fungerer godt nok fortsatt var uklart. REK mente også at spørsmålet om samfunnsnyttien ikke ble besvart godt nok. Det ble ikke lagt frem tilstrekkelig gode indikasjoner på at prosjektet ville gi resultater som senere vil kunne være samfunnsnyttig. Søknaden fikk avslag.

Den 18.09.2020 sendte søker kommentarer til klagen. Søker mente avslaget var basert på en misforståelse. I sitt vedtak av 24.11.2020 opprettholdt REK sitt vedtak av 09.09.2020. REK mente at prosjektet fremdeles burde gjennomføres basert på samtykke fra nålevende og samtykkekompetente pasienter. Saken ble oversendt til NEM for videre behandling.

Saken ble behandlet i NEM den 19.01.2021.

### **NEMs vurdering**

Ingen av de tilstedeværende medlemmene av komiteen var inhabile, og klagen ble behandlet i full komité.

Klagen gjaldt spørsmålet om det kan gis dispensasjon fra taushetsplikt slik at 3000 journaler fra Norsk Helsearkiv kan utleveres til bruk i forskningsprosjektet.

Taushetsplikten er en plikt som etter norsk rett står sterkt. Den verner om den enkeltes private interesser, og sikrer allmennhetens tillit til helsepersonell slik at man ikke unnlater å oppsøke hjelp av frykt for at opplysninger skal komme på avveie.

For deltakelse i forskningsprosjekter er hovedregelen at det skal være gitt et informert samtykke. Samtykke sikrer deltakernes frihet og selvbestemmelse. REK er gitt myndighet til å kunne gi dispensasjon fra taushetsplikten for tilgang til taushetsbelagte helseopplysninger fra helsepersonell eller helsetjenesten for annen forskning, jf. helsepersonelloven § 29. For å kunne gi fritak fra hovedregelen om samtykke må forskningen være av vesentlig interesse for samfunnet veid opp mot hensynet til deltakernes velferd og integritet må være ivaretatt og samtykke må være vanskelig å innhente. Vurderingen er en helhetsvurdering hvor blant annet opplysningenes sensitivitet, form, samt hvor og hvor lenge de skal oppbevares tillegges vekt.

NEM mener prosjektet har en samfunnsnytte. Å utvikle en algoritme som skal sørge for anonymisering av journalopplysninger er et relativt nytt felt med et stort potensial for samfunnsnytte. Maskinlæring er viktig innen medisinsk forskning. NEM mener det er lagt frem tilstrekkelig gode indikasjoner på at prosjektet vil være samfunnsnyttig. Dette støttes også av utkastet til DPIA fra NSD som er vedlagt i klagen fra søker. Anonymiserte helsedata vil gi mer forskning og sikre bedre utnyttelse av denne type datasett til beste for individ og samfunn. Med gode verktøy for anonymisering kan man dele mer, og tilrettelegge for mer og bedre forskning. Bedre anonymiseringsteknikker vil gi en bedre beskyttelse for pasientens helseopplysninger når opplysningene deles eller behandles av datasystemer. På sikt vil dette også kunne føre til økt kunnskap og bedre behandlingsformer.

Hensynet til prosjektdeltakernes velferd og integritet skal gå foran samfunnsnyttien. NEM er enig med REK om at kravene til å ivareta deltagernes integritet og velferd bør være strenge. Utlevering av opplysninger fra pasientjournal innebærer tilgang til sensitiv informasjon. Prosjektet skal bruke pasientjournaler avlevert til Norsk helsearkiv/Helsearkivregisteret som kun inneholder opplysninger om avdøde pasienter. Det skal brukes 3000 pasientjournaler fra personer som døde for ti år siden eller mer. REK mener at det etisk sett er mer problematisk å bruke taushetsbelagte opplysninger fra døde pasienters pasientjournaler, enn det ville ha vært å bruke opplysninger fra levende pasienters journaler basert på samtykke. Det er riktig at helsepersonellovens taushetsplikt også gjelder for døde pasienter, og at lovgrunnlaget for fritak for samtykke sånn sett er det samme for levende som for døde. Samtidig virker det rimelig å hevde at det er vanskeligere å se for seg hvordan en død persons integritet kan bli krenket enn en levende. Mens levende pasienter kan oppleve det som et integritets- og tillitsbrudd at forskere har fått tilgang på sensitive opplysninger i deres journal, uten at de har samtykket til det eller blitt informert om det, så kan ikke døde pasienter ha noen slik opplevelse. Det er dermed ikke helt åpenbart hva vi mener når vi snakker om integritetskrenkelse overfor døde.

Pasienter har rettsvern etter at de er døde, og i allmenmoralen er det også en anerkjennelse av en viss hensyntaken til døde menneskers tidligere interesser og ønsker. Eksempelvis vil man tenke at det er riktig at man oppfylder en avdød persons ønske om å donere sin kropp til forskning. Samtidig er det også en utbredt tanke at slik hensyntaken til ønsker og interesser, er sterkest

umiddelbart etter død. Man snakker ofte om et menneskes «siste vilje». Over tid blir det mindre relevant å tilskrive døde personer interesse og ønsker, eller tenke at det er relevant å finne ut av deres egentlige mening (da de levde) om bruk av helseopplysninger til forskning mange år etter deres død. I dette tilfelle er det 10 år eller mer siden pasientene døde.

NEM ser at REK har gode og sterke innvendinger i denne saken. NEM er imidlertid ikke overbevist om at det alltid er etisk mer forsvarlig å forske på levende som gir samtykke til deltakelse, enn døde hvor det gis dispensasjon fra taushetsplikt. En pasient som samtykker til at forskere kan lese dennes pasientjournal, kan likevel oppleve slik forskning som belastende selv om vedkommende ender opp med å si ja. På samme måte vil et risikofylt forskningsprosjekt være akkurat like risikofylt selv om deltakerne har samtykket. Samtykket «opphever» ikke verken risiko, ulempe eller belastning, selv om det riktignok forebygger et mulig opplevd integritetsbrudd som følge av å forske uten samtykke. Derfor sier NEM seg delvis enig i argumentasjonen fra prosjektleder som hevder at det totalt sett er mindre belastende å forske på døde enn levende pasienter.

Etter en helhetsvurdering er NEM kommet til at det kan gis dispensasjon fra taushetsplikt for utlevering av journaler fra Norsk helsearkiv.

Forskningsprosjektets formål er slik NEM vurderer det, ikke kontroversielt. Pasientene har vært døde i mer enn ti år, og forskningen har ingen følger for familie og pårørende. NEM stiller derfor ikke vilkår om at pårørende informeres om prosjektet.

## **Vedtak**

Klagen tas til følge. REK sitt vedtak datert 24.11.2020 omgjøres.

Vedtaket er endelig, og kan ikke påklages.

Med vennlig hilsen

Grete Dyb  
Professor dr.med.  
Komitéleder NEM

Camilla Bø Iversen  
Sekretariatsleder NEM

Kopi:

REK

Norsk Regnesentral