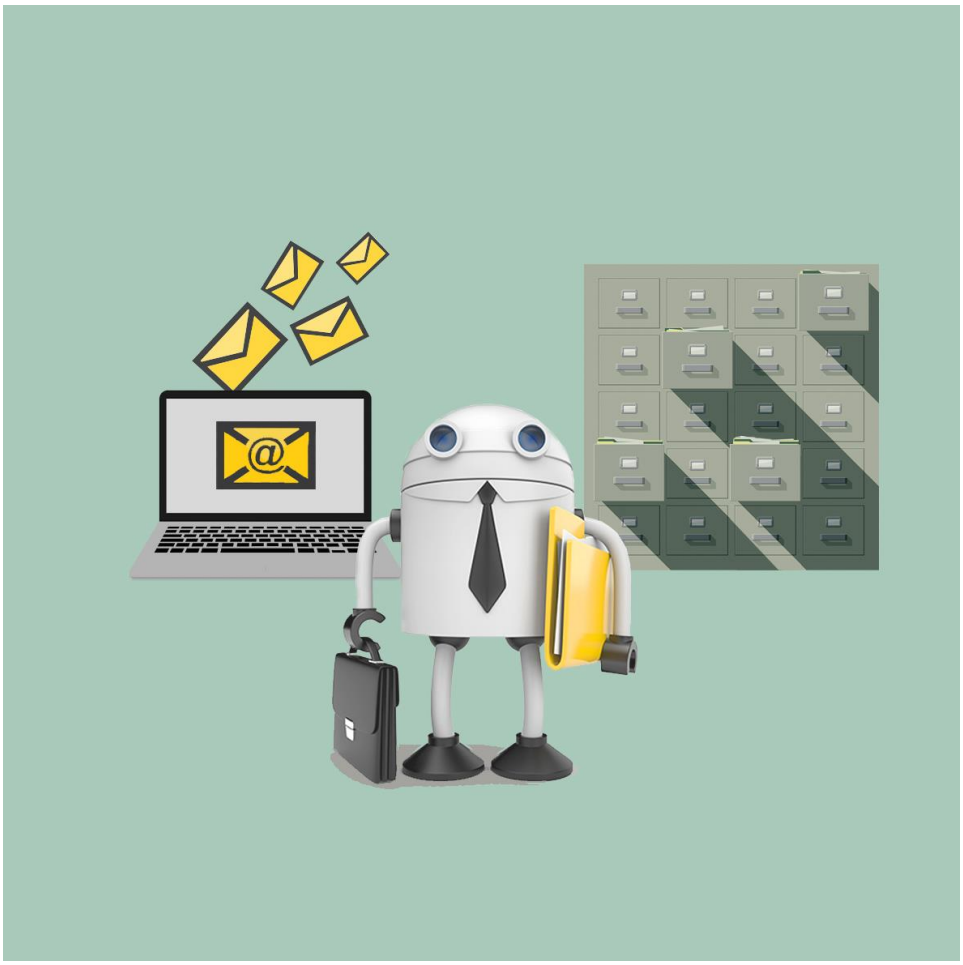


DIGITAL MEDARBEIDER

Sluttrapport fra sandkasseprosjektet med SIMPLIFAI og NVE

Temaer: lovlighet og innebygd personvern

Datatilsynet

Innhold

Hva er sandkassa?

I sandkassa utforsker deltakere sammen med Datatilsynet personvernrelaterte spørsmål, for å bidra til at tjenesten eller produktet deres etterlever regelverket og ivaretar personvernet på en god måte.

Datatilsynet tilbyr veiledning i dialog med deltakerne, og konklusjonene fra prosjektene er ikke vedtak eller forhåndsgodkjenning. Deltakerne står fritt i valget om å følge rådene de får.

Sandkassa er en verdifull metode for å utforske problemstillinger der jussen har få praktiske eksempler å vise til, og vi håper konklusjoner og vurderinger i rapporten kan være til hjelp for andre med liknende problemstillinger.

Sammendrag

Overgangen fra analog til digital postgang har utfordret de gamle systemene for journalføring og arkivering av post og viktige dokumenter. En rapport antyder at mer enn 25 prosent av all viktig informasjon i offentlig saksbehandling "går tapt" i sviktende arkiveringsrutiner.

Derfor har selskapet Simplifai utviklet en digital arkivmedarbeider, DAM, for å hjelpe saksbehandlere i offentlig sektor med å journalføre og arkivere dokumentasjon som blir sendt på e-post. DAM skal fungere som beslutningsstøtte for saksbehandlerne. Ønsket er å automatisere arbeidet så langt det lar seg gjøre på en forsvarlig måte, i håp om at det i større grad sikrer riktig og viktig arkivering.

I sandkassa har Simplifai og Datatilsynet sett på om personvernreglene åpner for at offentlige aktører kan ta i bruk en maskinlæringsløsning for å journalføre og arkivere e-post. Og sammen med NVE har de utforsket hvordan offentlige aktører kan gjøre informerte valg når de skal kjøpe intelligente løsninger, som for eksempel DAM.

Konklusjoner

- **Lovlighet.** Offentlige aktører har rettslig grunnlag for å bruke DAM som støtte ved beslutning om arkivering og journalføring. Det er riktignok usikkert om det rettslige grunnlaget åpner for å bruke personopplysninger til å videreutvikle modellen (etterlæring), med mindre personopplysningene er anonymiserte. Bruk av DAM er også lovlig innenfor nasjonale forskrifter til arbeidsmiljøloven. Datatilsynet anbefaler tekniske og organisatoriske tiltak, for eksempel instrukser som forbyr eller begrenser bruk av privat e-post.
- **Innebygd personvern.** Prosjektet har avdekket et stort behov for veiledning om hvordan det offentlige kan sikre innebygd personvern ved kjøp av intelligente løsninger. Prosjektet har gitt overordnede anbefalinger til hvilke steg en offentlig virksomhet kan ta: Skaff kunnskap, vurder om maskinlæring passer til det konkrete behovet og still krav.

Veien videre

Arbeidet med prosjektet har synliggjort et stort behov for veiledning og verktøy for å ivareta kravene til innebygd personvern ved kjøp av løsninger der den behandlingsansvarlige plikter å ivareta innebygd personvern.

Sandkasseprosjektet ser et behov for at flere aktører kommer sammen for å løfte kompetansen om innebygd personvern i anskaffelsesfasen i offentlig sektor. Det vil også være nyttig med praktiske eksempler på krav det offentlige kan stille i konkurranser om teknologi som bygger på maskinlæring.

Og – for å gjøre det lettere å bruke maskinlæring til å løse utfordringene med å journalføre og arkivere dokumentasjon i tide, anbefaler sandkasseprosjektet at ny arkivlovgivning regulerer dette på en ansvarlig måte.

Januar 2023

Denne pdf-en tilsvarer den første versjonen av rapporten, slik den ble publisert på Datatilsynets sider januar 2023. Teknologien og jussen er stadig i utvikling, så det *kan* være behov for å justere eller presisere rapportene med tiden.

Dersom denne pdf-en skiller seg fra det som står på Datatilsynets nettsider, kan du ta utgangspunkt i at det er nettsidens tekst som er gjeldende råd.

Om prosjektet

Simplifai har utviklet en digital arkivmedarbeider, DAM, for å hjelpe saksbehandlere i offentlig sektor med å journalføre og arkivere dokumentasjon som blir sendt på e-post. Siden offentlig sektor forvalter viktige samfunnsoppgaver, har sektoren egne krav til å behandle dokumenter. Relevant korrespondanse skal journalføres, slik at allmenheten kan få innsyn i hva som skjer. Akrivverdig dokumentasjon som er brukt i saksbehandlingen skal tas vare på for ettertida, gjennom arkivering.

Simplifai ble opprettet i 2018, med en ide om å gjøre kunstig intelligens enkelt og tilgjengelig for alle. I 2022 består Simplifai av ca. 140 ansatte, og leverer ulike digitale medarbeidere verden over. Hovedsektorene for leveranse har vært offentlig sektor, i tillegg til bank, forsikring og finans. Felles for alle sektorene er strenge lovkrav. Derfor jobber Simplifai aktivt med å utvikle sikre løsninger for automatisering av arbeidsprosesser relatert til digital kommunikasjon og dokumenter med bruk av kunstig intelligens.

Prosjektet ønsket å invitere med en offentlig aktør, som kunne dele kompetanse og erfaring om arkivering og personvern i praksis. Ressurspersoner fra Norges vassdrags- og energidirektorat (NVE) ga nyttige perspektiv ut fra rollen som en potensiell kjøper av DAM.

Hvorfor trenger vi DAM?

Den tida da alle sendte brev via postbud, sørget dyktige arkivarer i offentlig sektor for å registrere brevene ut og inn på riktig sak. Saksbehandlerne trengte ikke selv å tenke på journalføring og arkivering. I dag går mye korrespondanse direkte til saksbehandlerne på e-post. I tillegg foregår mye intern kommunikasjon på andre plattformer. Mengden dokumentasjon har vokst og framstår som overveldende for mange av oss.

Med digitalisering har ansvaret for arkivering og journalføring i større grad flyttet seg fra arkivarene til saksbehandlerne. Flertallet av saksbehandlere mener de får arkivert mindre enn 75 prosent av all viktig informasjon, ifølge en undersøkelse fra Menon.

Les rapporten: Mangelfull arkivering koster Norge dyrt (arkivverket.no)

Og dette skjer på alle nivåer. I en rapport fra Riksrevisjonen om arkivering og åpenhet i statlig forvaltning, peker de på at manglende arkivering og mangelfull journalføring av dokumenter i viktige enkeltsaker svekker mulighetene for offentlig debatt og demokratisk innsyn og kontroll. De viser blant annet til koronakommisjonens konklusjoner om at en sentral del av samspillet mellom regjeringen og Stortinget under håndteringen av covid-19-pandemien ikke er dokumentert. «Den manglende dokumentasjonen og arkiveringen svekker mulighetene til å få kunnskap om og innsyn i sentrale vurderinger og beslutninger som ble tatt i forbindelse med håndteringen av pandemien, omtalt som den største nasjonale krisen siden andre verdenskrig.»

Les rapporten fra riksrevisjonen (riksrevisjonen.no)

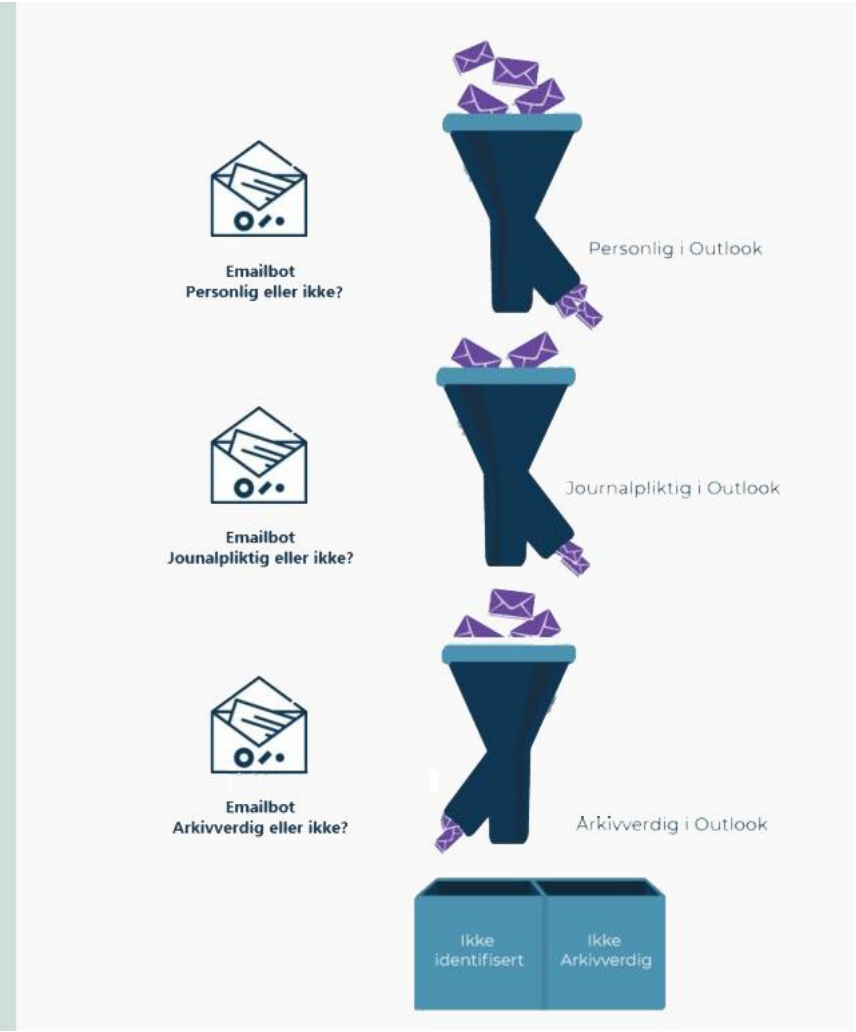
Hvordan fungerer den digitale medarbeideren (DAM)?

DAM er en løsning som støtter den ansatte med å velge hvilke e-poster som skal arkiveres og journalføres. Den hjelper også til med selve gjennomføringen av journalføring og lagring.

Løsningen er utviklet ved hjelp av maskinlæring. Metoden som er brukt i dette prosjektet kalles prosessering av naturlig språk (NLP), nærmere bestemt naturlig språkforståelse (NLU), som innebærer at løsningen kjenner igjen tekst i e-postene, inkludert tekst i eventuelle vedlegg. En nyutvikling for den digitale medarbeideren er at maskinlæring (NLU/NLP) er kombinert med optisk tegngjenkjenning (OCR).

Den digitale medarbeideren er bygd opp av tre moduler – eller «boter». Før den digitale medarbeideren foreslår hva som skal skje med e-posten videre, vurderer hver bot ett av disse spørsmålene videre:

1. Er e-posten av personlig karakter?
2. Er e-posten journalpliktig?
3. Er e-posten arkivpliktig?



Basert på vurderingen, vil en e-post bli merket med kategorier for om den er personlig/ikke-personlig, om den er journalpliktig og om den er arkivpliktig. Hvis den ansatte er enig i vurderingen, kan e-posten markeres med kategorien «Arkivering OK». Den ansatte godkjenner da selv arkivering av denne e-posten. Videre kan den ansatte selv arkivere e-posten, eller så kan e-post som er merket med «Arkivering OK» bli arkivert av den digitale arkiveringsassistenten påfølgende natt, basert på den aktive handlingen til den ansatte.

Den ansatte kan velge å legge til egne regler som unntar innhold i e-postboksen fra behandling av DAM. Dette kan f.eks. være å unnta enkelte avsendere, eller nøkkelord i e-postemnet eller e-postinnholdet.

Simplifais mål er at den digitale medarbeideren skal bidra til at

- flere e-poster blir journalført og arkivert
  - Partene som er i kontakt med det offentlige får større rettssikkerhet.
  - Det offentlige og media får en bedre kontrollfunksjon.
- de ansatte får frigjort tid som kan brukes til kjerneoppgavene

Mål for sandkasseprosessen

Før Simplifai ble tatt opp i Datatilsynets sandkasse, hadde selskapet deltatt i et StartOff-prosjekt sammen med Arkivverket, administrert av Direktoratet for forvaltning og økonomistyring (DFØ). Målet med prosjektet var å utforske hvordan kunstig intelligens kan brukes til å automatisere eller understøtte arbeidet med journalføring og arkivering av e-post. Det var i dette prosjektet DAM ble utviklet.

Se presentasjonen «Fullautomatisert epostarkivering for Arkivverket» (på YouTube.com)

Hovedfokus i StartOff-prosjektet var å finne en god balanse mellom hensynet til personvern og ønsket om effektiv ivaretakelse av plikter som følger av offentlighetsloven og arkivloven. I prosjektet ble det klart at de juridiske utfordringene ved bruk av kunstig intelligens til å behandle e-post ikke var knyttet til arkivbestemmelsene, men til personopplysningsloven og arbeidsmiljøloven. Prosjektet antok at den juridiske kompleksiteten var lavere ved å ta i bruk DAM på virksomhetens sentrale e-postkasse, slik som post@datatilsynet.no i stedet for på personlige e-postkasser. Personlige e-postkasser, som for eksempel kari.nordmann@datatilsynet.no, ble ikke vurdert i StartOff-prosjektet.

Det er likevel først når DAM kobles til de personlige e-postkassene, effekten av tjenesten virkelig slår inn, siden det er de personlige e-postkassene som brukes mest i saksbehandling. Samtidig reiser koblingen til personlige e-postkasser viktige spørsmål knyttet til personvern. Dette var bakgrunnen for søknaden til Datatilsynets sandkasse for ansvarlig kunstig intelligens.

Sandkasseprosessen har hatt **to hovedmål:**

1. Vurdere om DAM er lovlig å ta i bruk for en offentlig aktør

Å vurdere om en DAM er lovlig å ta i bruk og videreutvikle for en offentlig aktør, var sentralt for Simplifai før de ville gå videre med løsningen ut i markedet. Hovedmålet om lovlighet ble delt inn i tre problemstillinger:

1. Utforske hvilket rettslig grunnlag i personvernforordningen som er aktuelt når en offentlig aktør tar i bruk den planlagte løsningen.

2. Undersøke om bruken av særskilte kategorier personopplysninger er tillatt etter personvernforordningen

3. Avklare om den planlagte løsningen vil være i strid med e-postforskriftens forbud mot overvåking

Prosjektet besluttet å ikke vurdere spørsmål knyttet til å overføre personopplysninger til land utenfor EØS ved bruk av skyløsninger, av ressurshensyn. I spørsmål om lovlighet har prosjektet bygd på Simplifais vurdering av at selskapet er databehandler. Prosjektet har altså ikke selv vurdert om Simplifai er behandlingsansvarlig eller databehandler.

2. Gi anbefalinger om innebygd personvern til offentlige aktører som skal anskaffe en løsning som helt eller delvis er basert på maskinlæring (kunstig intelligens)

Som leverandør av intelligente arkivløsninger, ønsker Simplifai å utvilke en tjeneste som både ivaretar regelverkskrav og andre behov offentlige virksomheter har, når det kommer til personvern. Intelligente løsninger er på full fart inn i offentlig sektor, og bestillerkompetansen er veldig viktig for at det innføres på en god måte.

I dette prosjektet samarbeidet Datatilsynet, Simplifai og NVE med å lage anbefalinger til hvordan offentlige aktører kan stille krav til løsninger som bygger på maskinlæring i anskaffelsesprosesser, slik at det offentlige oppfyller plikten til innebygd personvern.

Er DAM lovlig å ta i bruk?

Lovlighet er et grunnleggende prinsipp i personvernforordningen (GDPR). Kjernen er at alle som vil samle inn, lagre eller på annen måte behandle personopplysninger må følge kravene i forordningen og annen lovgivning. Vi skal her se nærmere på kravet som handler om at du må ha et rettslig grunnlag for å behandle personopplysninger.

I sandkasseprosjektet med Simplifai har vi utforsket handlingsrommet for bruk av personopplysninger i et beslutningsstøtteverktøy basert på maskinlæring.

Personvernforordningen skal gjelde på samme måte i 30 land i Europa. I tillegg har Norge spesialregler om personvern, blant annet i arbeidslivet. Disse spesialreglene er gitt i forskrifter til arbeidsmiljøloven og er ment å beskytte arbeidstakere mot unødvendig inngripende overvåking eller kontroll.

Hvordan gikk vi fram i vurderingen?

Sandkasseprosjektet drøftet i starten om vi skulle ta utgangspunkt i en full automatisering av DAM, eller DAM som beslutningsstøtte. Selv om automatisering vil gi størst effekt, er også risikoen for brudd på personvernreglene høyere, ved at innhold fra personlige e-postkasser blir sendt rett til offentlig journal. Prosjektet besluttet å fokusere på DAM som beslutningsstøtte.

Prosjektet gjennomførte to workshops med lovlighet som tema. Som forberedelse til samlingene, hadde Simplifai gjort rede for tekniske løsninger og formålene med de ulike modulene eller «botene» i løsningen. I tillegg ga NVE sitt syn på bl.a. formålene med de ulike behandlingene i modulene og hvilket rettslig grunnlag direktoratet så for seg for å ta løsningen i bruk. Ut fra dette grundige forarbeidet, skisserte Datatilsynet rammer for bruk, etterlæring og mulige tiltak for å gjøre DAM mer personvernvennlig.

Vi valgte å dele spørsmålet om rettslig grunnlag inn i faser: (1) bruksfasen, der algoritmemodellen brukes i saksbehandlingen og (2) etterlæringsfasen, der algoritmemodellen blir forbedret.

Simplifai kom til sandkassa med et utviklet verktøy. Prosjektet avgrenset mot spørsmålet om Simplifai hadde rettslig grunnlag for å utvikle DAM.

Rettslig grunnlag for bruksfasen

I denne fasen brukes DAM som beslutningsstøtteverktøy og skal benyttes som en veiledning for saksbehandlere i vurderingen av om e-poster er av personlig karakter, journalpliktig og arkivpliktig.

Behandling av personopplysninger i forbindelse med journalføringen og arkiveringen skjer i dag med hjemmel i personvernforordningen art. 6 nr. 1 bokstav c), «behandlingen er nødvendig for å oppfylle en rettslig forpliktelse som påhviler den behandlingsansvarlige.»

Den rettslige forpliktelsen til å ha arkiv finner vi i arkivloven § 6 og forskrift om offentlege arkiv §§ 9 og 10. For journalplikt finner vi i tillegg forpliktelsen i offentligetsloven § 10.

Sandkasseprosjektet foreslår at offentlige virksomheter bygger på artikkel 6 nr. 1 bokstav c også når de benytter digitale verktøy, som DAM, for å effektivisere og organisere arbeidet med arkiv.

Rettslig grunnlag for etterlæringsfasen

Det er et stort teknologisk framskritt, at en del kunstig intelligens-vertøy kan ha evnen til å lære. Når algoritmen lærer etter hvert som den blir brukt, kaller vi det etterlæring. Inferensen (altså resultatet av at algoritmen er benyttet på et nytt datagrunnlag) inngår i algoritmen, og slik blir verktøyet dynamisk justert for å bli mer treffsikkert.

Etterlæring kan være en utfordring for personvernjussen. Selv om det finnes rettslig grunnlag til å bruke kunstig intelligens til arkivering, for eksempel basert på behovet for å oppfylle en rettslig forpliktelse, er det ikke gitt at det finnes rettslig grunnlag for etterlæring av algoritmen.

I personvernforordningen artikkel 6 nr. 1 er det listet opp seks alternative vilkår som kan gi rettslig grunnlag for behandling av personopplysninger. De alternativene som det er mest aktuelt å vurdere i dette tilfellet, er vilkårene i bokstav c og bokstav e:

- Behandlingen er nødvendig for å oppfylle en rettslig forpliktelse (bokstav c).
- Behandlingen er nødvendig for å utføre en oppgave i allmenhetens interesse (bokstav e).

I begge tilfellene krever loven at det i tillegg finnes hjemmel for behandlingen i en annen lov jf. art. 6 nr. 3.

Dagens arkivlov sier ingenting om at e-post eller annet arkivverdig material kan brukes for å videreutvikle maskinlæringssystemer eller andre digitale verktøy. DAM hadde vært på tryggest grunn om arkivloven eksplisitt åpnet for at materialet behandlet i forbindelse med utførelse av arkiveringsplikten samtidig kan benyttes til forbedring eller videreutvikling av digitale verktøy. Det er likevel mulig at etterlæringen kan anses som et utslag av å bruke verktøyet, og på den måten har rettslig grunnlag i artikkel 6 nr. 1 bokstav c, på samme måte som for bruksfasen.

Det er et klart samfunnsmessig behov for å effektivisere og forbedre rettsikkerheten knyttet til arkivering. Hvordan tilgangen til arkivmateriale for videreutvikling av maskinlæring skal løses, er en oppgave for lovgiveren. Datatilsynet har sett eksempler på at enkelte virksomheter får hjemmel til å benytte innsamlede data til utvikling av it-systemer, se lov om statens pensjonskasse § 45 b.

En måte å løse denne utfordringen på, er å bygge algoritmen på en slik måte at inferensen uansett ikke vil inneholde personopplysninger. Da gjelder ikke personvernforordningen, og det er fritt frem for å trene algoritmen. En annen variant, dersom inferensen inneholder personopplysninger, er å anonymisere dataene før algoritmen etterlæres. Da er det ikke lenger personopplysninger, og modellen kan trenes uten krav om rettslig grunnlag.

Et alternativ kan også være om den enkelte bruker av DAM selv kan skru av eller på etterlæring av algoritmen, enten generelt eller med mulighet for å unnta den enkelte e-post fra inferensen. Dette vil dessuten være et positivt tiltak med tanke på åpenhet om hvordan verktøyet behandler personopplysninger, og det kan sikre at f.eks. privat e-post ikke inngår i treningen. En bivirkning av en slik løsning kan riktignok bli, at datagrunnlaget kan sementere en feilaktig arkiveringspraksis.

Er det lov å behandle særlige kategorier av personopplysninger?

I e-postkassa til en saksbehandler i NVE kan DAM finne informasjon fra den lokale fagforeninga, sammen med e-post til sjefen om sykdom. Noen bruker kanskje også virksomhetens e-post til privat kommunikasjon. Der kan DAM også finne informasjon både om saksbehandlerens religion og seksuelle orientering. Alle disse eksemplene regnes som særskilte kategorier, jf. personvernforordningen artikkel 9. Det er bare lov å behandle disse opplysningene dersom vilkårene i artikkel 9 nr. 2 er oppfylt.

Sandkassa antar at NVE kan bygge på vilkåret om «at behandlingen er nødvendig av hensyn til viktige allmenne interesser», i bokstav g for å bruke DAM. Det er illustrerende i forarbeidene til Nav-loven, der departementet legger til grunn at effektiv saksbehandling i Arbeids- og velferdsetaten og Statens pensjonskasse omfattes av «viktige allmenne interesser» i personvernforordningen artikkel 9 nr. 2 bokstav g.

En annen viktig del av vilkåret er at behandlingen skal «stå i et rimelig forhold» til målet som søkes oppnådd. I denne forholdsmessighetsvurderingen har sandkassa lagt vekt på at forslaget bare skal gå til saksbehandleren, og at inngrepet i personvernet derfor er begrenset. Vi anbefaler dessuten instrukser som forbyr eller begrenser bruk av virksomhetens e-post til private gjøremål for å begrense omfanget av privat e-post som blir behandlet i løsningen.

Forbudet mot overvåking i e-postforskriften

Så langt har vi forholdt oss til personvernforordningen. Men det er også relevant å vurdere DAM opp mot e-postforskriften med sitt forbud mot «å overvåke arbeidstakers bruk av elektronisk utstyr, herunder bruk av Internett» (§ 2, andre ledd).

En arbeidsgiver vil altså ha lov til innføre DAM, så lenge bruken ikke innebærer overvåking av arbeidstakerne. Kan innføring av DAM sees på som overvåking av de ansattes bruk av elektronisk utstyr?

Hva som ligger i «å overvåke» er ikke nærmere definert i forskriften. I forarbeidene til tilsvarende regler i den gamle loven er det framhevet at tiltaket skal ha en viss varighet eller skje gjentatte ganger. Overvåking står i motsetning til enkeltstående innsyn, som er tillatt i flere situasjoner. I forarbeidene er det også understreket at det ikke bare er et spørsmål om formålet er å overvåke. Arbeidsgiveren skal også legge vekt på om de ansatte kan oppleve situasjonen som overvåkning.

Praksis fra Datatilsynet er ikke entydig, med tanke på om arbeidsgiver faktisk skal se personopplysningene for at det skal regnes som overvåking. Overvåkingsbegrepet favner vidt, og en naturlig språklig forståelse av begrepet kan tale for at også innsamling og systematisering rammes av forbudet. At bestemmelsen retter seg mot *arbeidsgivers* overvåking trekker i retning av at arbeidsgiveren i det minste må kunne ha adgang til opplysningene om arbeidstakerne for å rammes av forbudet. Dette var også sandkassas standpunkt i prosjektet med Secure Practice.

Etter diskusjonene i sandkasseprosjektet var det enighet om at DAM som beslutningsstøtte nok ikke vil omfattes av forbudet mot overvåking. Vi har lagt vekt på at løsningen bare lager et forslag i saksbehandlerens e-postkasse og at informasjonen om kategoriene ikke blir sendt videre.

I tillegg anbefaler sandkasseprosjektet tekniske og juridiske tiltak for at arbeidsgiveren ikke skal få tilgang til opplysningene som blir samlet inn om hver ansatt i løsningen. Et eksempel på tiltak er å innføre instrukser som forbyr eller begrenser bruk av virksomhetens e-post til private gjøremål, slik som nevnt under avsnittet om særskilte kategorier.

Innebygd personvern ved kjøp av intelligente løsninger

Prinsippet om innebygd personvern har eksistert lenge, men ble ikke lovfestet før personvernforordningen kom i 2018. Kravet bygger på tanken om, at et system må ha personvernmekanismer bygget inn i løsningen fra start for at den som bruker systemene skal kunne etterleve personvernregelverket som helhet. Å slenge på tiltak for å «fikse» personvernproblemstillinger i ettertid, vil i lengden gi mindre personvernvennlige løsninger enn de som har personvern bygget inn fra start.

Artikkel 25: Innebygd personvern og personvern som standardinnstilling

1. Idet det tas hensyn til den tekniske utviklingen, gjennomføringskostnadene, behandlingens art, omfang, formål og sammenhengen den utføres i, samt risikoene av varierende sannsynlighets- og alvorlighetsgrad for fysiske personers rettigheter og friheter som behandlingen medfører, skal den behandlingsansvarlige, både på tidspunktet for fastsettelse av midlene som skal brukes i forbindelse med behandlingen, og på tidspunktet for selve behandlingen, gjennomføre egnede tekniske og organisatoriske tiltak, f.eks. pseudonymisering, utformet med sikte på en effektiv gjennomføring av prinsippene for vern av personopplysninger, f.eks. dataminimering, og for å integrere de nødvendige garantier i behandlingen for å oppfylle kravene i denne forordning og verne de registrertes rettigheter.
2. Den behandlingsansvarlige skal gjennomføre egnede tekniske og organisatoriske tiltak for å sikre at det som standard bare er personopplysninger som er nødvendige for hvert spesifikke formål med behandlingen, som behandles. Nevnte forpliktelse får anvendelse på den mengden personopplysninger som samles inn, omfanget av behandlingen av opplysningene, hvor lenge de lagres og deres tilgjengelighet. Nevnte tiltak skal særlig sikre at personopplysninger som standard ikke gjøres tilgjengelige for et ubegrenset antall fysiske personer uten den berørte personens medvirkning.
3. En godkjent sertifiseringsmekanisme i henhold til artikkel 42 kan brukes som en faktor for å påvise at kravene fastsatt i nr. 1 og 2 i denne artikkel overholdes.

Hvem må sørge for innebygd personvern?

For å etterleve artikkel 25, er en avhengig av at programmerere og utviklere implementerer egnede tekniske løsninger. Det er bare de som utvikler selve løsningen, som kan bygge inn gode personvernmekanismer fra start. Likevel ligger ikke ansvaret for etterlevelse av kravet hos produsenten av programvare eller hos databehandler, men hos den behandlingsansvarlige.

Se Datatilsynets veileder om innebygd personvern.

Det er sånn forordningen er lagt opp. Den behandlingsansvarlige har ansvar for behandlingen av personopplysninger jf. artikkel 5 nr. 2. Den bevisste behandlingsansvarlige vil etterspørre gode personvernmekanismer når de kjøper produkter. På den måten vil produsenter som kan gi garantier for innebygde personvernløsninger ha en konkurransefordel. Dette forutsetter en kompetanse hos den behandlingsansvarlige, som skal kjøpe inn maskinlæringsløsningen, både om

- innebygd personvern
- maskinlæring
- kontrakter
- anskaffelsesreglene (for behandlingsansvarlige som må følge lov om offentlige anskaffelser)

Hva skal bygges inn?

Det et system skal ha innebygd er:

- grunnleggende personvernprinsipper
- den registrertes rettigheter og friheter

Dette betyr alt fra abstrakte krav som rettferdighet til mer praktiske krav som rutiner for sletting. Så, når har en virksomhet klart å bygge inn personvern? Hva som er å regne som egnede tekniske og organisatoriske tiltak, er ikke direkte definert i lovteksten. Datatilsynets veileder om innebygd personvern understreker at tiltakene må være egnet til å fremme personvernet på en effektiv måte, og gir eksempler på nøkkelelementer for de ulike personvernprinsippene. Hvilke tiltak som kreves og nivået på tiltakene, skal den behandlingsansvarlige bestemme ved å vurdere følgende momenter:

- Den tekniske utviklingen
- Gjennomføringskostnader
- Hvilken behandling som gjøres basert på art, omfang, formål og sammenheng
- Risikoen for de registrertes rettigheter og friheter – sannsynlighet og alvorlighetsgrad

Med andre ord legger artikkel 25 opp til at konteksten behandlingen foregår i skal være bestemmende for hvordan og til hvilken grad personvern bygges inn i løsningene. Alle forordningens bestemmelser gjelder uansett, men hvordan en spesifikk sletterutine bygges inn i løsningen, må bestemmes ut fra hva som er mest egnet.

Innebygd personvern og maskinlæring

Maskinlæring kan gjøre det vanskeligere å bestemme hvilke tiltak som vil være mest egnet for å sørge for innebygd personvern. Årsakene er at maskinlæring utfordrer personvernet på følgende måter:

- De fleste kjøper algoritmer som er ferdigutviklet (hyllevare) og gjenbrukes til ulike formål
- Trening av maskinlæringsløsninger krever store mengder data
- Vi vet fortsatt lite om hvordan maskinlæring kan gå ut over den enkeltes rettigheter og friheter og gi uventede effekter, for eksempel
    - diskriminering
    - mangel på åpenhet

Maskinlæringsalgoritmer kan være lite gjennomsiktige. Det kan være vanskelig å forklare logikken en maskinlæringsløsning baserer resultatene sine på, ofte kalt svart boks-problematikk.

Dersom risikoen for den enkeltes rettigheter og friheter for eksempel knytter seg til diskriminering, må tiltaket redusere risikoen for diskriminering. Dermed må algoritmene bygges slik at diskriminering ikke forekommer.

Maskinlæringsløsninger anskaffes ofte som såkalt «hyllevare». Det vil si at løsningen er utviklet av noen andre enn de som faktisk har behandlingsansvaret ved bruken av verktøyet. Kjøperen vil stå med ansvaret for å anskaffe en løsning som har innebygd personvern. Kjøperen (den behandlingsansvarlige) kan velge en databehandler som garanterer for etterlevelsen, men for å velge korrekte tiltak på korrekt nivå, må behandlingsansvarlig også forstå hvilken effekt løsningen har på de registrertes personvern. Fordi maskinlæringsløsninger ofte er komplekse og

vanskelige å forklare, vil dette føre til en større utfordring for de behandlingsansvarlige enn for teknologi som ikke inneholder maskinlæring.

I sandkasseprosjektet ønsket vi å belyse hvordan innebygd personvern kan ivaretas i anskaffelse av intelligente løsninger. For å forstå hva offentlige aktører trenger hjelp til, valgte vi å gjennomføre et case-studie for å kartlegge landskapet og utforme noen enkle anbefalinger.

Hvordan gikk vi fram for å lage anbefalingene?

For å avklare hvilken informasjon offentlige aktører trenger for å gjøre gode valg i innkjøp av maskinlæringsløsninger har prosjektgruppen valgt å gjennomføre intervju med en offentlig aktør (NVE). Formålet med intervjuet var å danne et bilde av hvilke behov offentlige virksomheter har til informasjon om innebygd personvern. Basert på intervjuet valgte vi å lage en anbefaling med punkter offentlige aktører kan følge opp. NVE ble også invitert til å gi innspill til et tidlig utkast til anbefalingene, for å se om de møtte behovet. I tillegg til NVE ga representanter fra politiet og Simplifai innspill til utkastet.

NVE ble aktuelle som case, fordi de allerede hadde vært i kontakt med Simplifai angående arkivsystemet tidligere. Som en offentlig aktør er NVE et interessant casevalg, fordi virksomheten i relativt liten grad behandler personopplysninger, sammenlignet for eksempel med en kommune. Likevel vil NVE være nødt til å behandle personopplysninger om sine egne ansatte, som DAM berører. NVE er også en offentlig aktør med erfaring i å utvikle løsninger selv, noe som skaper et interessant perspektiv i en bestillersituasjon som denne.

Tilbakemeldingen fra NVE etter intervjuet, er at det er stort behov for veiledning i å ivareta personvern ved anskaffelse av intelligente løsninger. Også de øvrige arbeidsmøtene om innebygd personvern i sandkasseprosjektet bekrefter behovet for veiledning.

Behov for informasjon om innebygd personvern støttes også av arbeidet Datatilsynet gjør utenom sandkassa, i det daglige arbeidet med kontroll og etterlevelse av personvernforordningen. Vi ser det spesielt er to momenter som skaper det omfattende informasjonsbehovet.

Hvorfor er informasjonsbehovet så stort? Våre funn

Et interessant funn er at det virker som om innebygd personvern forbindes utelukkende med informasjonssikkerhet. Informasjonssikkerhet er et bedre utredet felt enn personvern, så det kan være en årsak. Vi ser et behov for råd om hvordan virksomhetene kan etterspørre informasjon og stille krav utover informasjonssikkerhet, når de skal sjekke det innebygde personvernet i løsningen. Disse erfaringene deles også av Personvernkommisjonen.

Les mer om det i Personvernkommisjonens NOU 2022: 11 (på regjeringen.no)

Maskinlæring og kunstig intelligens er såpass komplekst for de fleste virksomheter uten større teknisk kompetanse, at de de mister oversikt over hvilke tiltak som er «egnet». Forskjellen i teknologisk kunnskap mellom utvikleren av verktøyet og den behandlingsansvarlige er enda større innenfor maskinlæring enn ellers. Det gir produsentene en sentral rolle i å gi informasjon om produktet. Vår oppfatning er at kundene til nå i liten grad har etterspurt innebygd personvern for hyllevare som bygger på maskinlæring.

Bakgrunnen for valgene

Fordi både innebygd personvern og maskinlæring er kompliserte og til en viss grad nye tema, blir anbefalingen vi lager i denne omgang noe overordnet. Eksemplene på type krav som kan stilles til leverandører av maskinlæringsverktøy er likevel valgt fordi de er særlig aktuelle for maskinlæringsløsninger.

Anbefalinger om innebygd personvern ved anskaffelse av løsninger som bygger på maskinlæring

Øk kompetansen

Opplæring av ansatte er et sentralt tiltak for å sørge for innebygd personvern i praksis. Øk kompetansen om innebygd personvern, maskinlæring, kontrakter og anskaffelser.

Både sikkerhetsansvarlig og virksomhetens personvernombud kan være gode ressurser ved anskaffelse av maskinlæringsløsninger og bør involveres på et tidlig tidspunkt.

Her er en liste over nyttige kilder:

Innebygd personvern

- Innebygd personvern og personvern som standard | Datatilsynet

Veilederen går gjennom alle personvernprinsippene og gir enkle råd for hvordan de kan bygges inn. Den gir en god beskrivelse av innebygd personvern som krav.

- Programvareutvikling med innebygd personvern | Datatilsynet

Veilederen går gjennom de forskjellige fasene i en utviklingsprosess og beskriver hvordan personvern kan bygges inn. Den er designet for teknologer, men enhver kan dra nytte av de praktiske rådene for hvordan personvern bør bygges inn.

Maskinlæring

- Kunstig intelligens og personvern | Datatilsynet

Lettlest rapport som beskriver kunstig intelligens og sammenhengen med personvern.

- Et gratis introduksjonskurs om kunstig intelligens for alle (elementsofai.com)

Et dekkende og godt designet online kurs om kunstig intelligens for de som vil vite mer om mekanismene bak.

- Teknologirådets rapport om kunstig intelligens og maskinlæring (pdf)

Omfattende rapport om kunstig intelligens fra et teknisk perspektiv. Rapporten går også gjennom mulige utfordringer maskinlæring kan skape.

Hvordan stille kravene og evaluere i en anskaffelsesprosess

- Anskaffelser.no

Generell informasjon om anskaffelser, særlig hvordan en offentlig oppdragsgiver kan gå fram for å få informasjon om produkter og stille krav.

Vurder: Er maskinlæring mest hensiktsmessig?

Hvilket behov skal maskinlæringsverktøyet løse? Vår anbefaling er å vurdere hvilken kontekst dere skal behandle personopplysningene i og vurdere om et regelstyrt verktøy kan føre til en mer personvernvennlig løsning enn et maskinlæringsverktøy.Skal dere for eksempel bruke verktøyet som beslutningsstøtte for å fatte vedtak overfor innbyggerne, må dere stille andre krav til verktøyet enn om det skal brukes til å effektivisere interne beslutningsprosesser.

> ### Regelstyrte verktøy
>
> Regelstyrte verktøy er verktøy der algoritmen er statisk og baserer seg på faste regler, i motseting til algoritmer som er dynamiske og gjør prediksjoner basert på mønstre i datagrunnlaget.

Til hjelp ved sammmenlikningen, vil vi vise til noen utfordringer ved maskinlæringsverktøy, som Personvernkommisjonen har løftet fram i sin rapport 26. september 2022.

For det første kan maskinlæringsalgoritmer kreve store mengder data for å opparbeide en nøyaktig modell.[1] Et gjennomgående krav i personvernforordningen er at en behandling bare skal omfatte personopplysninger som er nødvendige for å oppfylle behandlingens formål (prinsipp om dataminimering) og at personopplysninger ikke skal benyttes til annet formål enn de opprinnelig ble samlet inn for (formålsbegrensning). Videre er det i bestemmelsen om innebygd personvern spesifisert at den behandlingsansvarlige skal sikre at det «som standard» bare er personopplysninger som er nødvendige for behandlingens formål, som blir behandlet, se artikkel 25 nr. 2.

For det andre trekker kommisjonen frem at datagrunnlaget som samles inn og brukes til å trene en maskinlæringsalgoritme kan inneholde feil og mangler. Forekommer det slike skjevheter i datagrunnlaget vil resultatet, altså maskinlæringsalgoritmens prediksjoner, også være preget av skjevheter. Virksomheten må med dette være obs på hvilke data de benytter når de trener en maskinlæringsalgoritme. Avhengig av behandlingens kontekst kan også datagrunnlaget bli utdatert, og dermed misvisende. For å ivareta like godt personvern gjennom løsningens livsløp vil det dermed være nødvendig å gjøre justeringer tilstrekkelig ofte for å opprettholde prediksjonenes nøyaktighet.

Videre vil maskinlæringsalgoritmer sjeldent gi tilstrekkelig mulighet for åpenhet og forutberegnelighet. Maskinlæringsløsninger er lite transparente, sier personvernkommisjonen i sin rapport. Maskinlæringsalgoritmer vil også i mange tilfeller være dynamiske. Det vil si at logikken kan endre seg, også etter at algoritmen er tatt i bruk.

Til slutt peker kommisjonen på problemstillingen med at prediksjoner fra maskinlæringsløsninger kan benyttes ukritisk. Denne problemstillingen er også trukket frem i sluttrapporten for NAV-sandkasseprosjektet. Det vil i praksis kunne føre til at det som er ment som et beslutningsstøttesystem i realiteten blir et automatisert beslutningssystem.

Datatilsynet vil også trekke frem at det er viktig at virksomheten tenker over hvilke rettigheter den behandlingsansvarlige fremdeles skal sørge for å ivareta for de registrerte, og hvorvidt man som behandlingsansvarlig fremdeles egner å etterleve sine plikter dersom man velger å benytte maskinlæringsløsninger fra en ekstern utvikler.

Dersom dere kommer til at en maskinlæringsverktøyet er mest hensiktsmessig til å løse behovet deres, er neste spørsmål hvordan dere kan gjøre tiltak ved å stille krav til produktet.

---

[1] Se også: DT KI rapport s. 17.

Spør, grav og still krav!

Vi har noen forslag til hva dere som behandlingsansvarlige kan spørre etter av dokumentasjon for å kunne vurdere innebygd personvern i de forskjellige løsningene som blir tilbudt.

Få løsningen forklart på en måte dere forstår

Maskinlæringsløsninger kan være svært kompliserte, og det er viktig at alle som skal bruke løsningen forstår hvordan den virker. For å sikre krav til innebygd personvern i anskaffelsesfasen er det viktig at de som er med og vurderer de ulike tilbudene i en konkurranse, forstår hvordan personvern kan ivaretas ved bruk.

Vi anbefaler, at dere som skal skaffe en maskinlæringsløsning ber om en lettforståelig og tilstrekkelig beskrivelse av hva løsningen faktisk gjør.

Be om å få se dataflyt og behandlingsprotokoll

En leverandør som opptrer som databehandler må gjøre rede for hvilke behandlingsaktiviteter de utfører på vegne av den behandlingsansvarlige, jf. artikkel 30 nr. 2. Dere som er behandlingsansvarlige kan be om å få se en slik behandlingsprotokoll, før dere inngår kontrakt om anskaffelse av en maskinlæringsløsning. Også leverandører som ikke er databehandlere bør kunne gjøre rede for hvilke personopplysninger som vil bli behandlet i løsningen. Behandlingsprotokollen er sentral for å få oversikt over hvordan behandlingen faktisk foregår. Behandlingsprotokollen vi vil være nyttig å se i sammenheng med den enkel beskrivelsen nevnt i punktet over.

Det kan også være relevant å få en beskrivelse av hvordan innsamlet data beveger seg i løsningen, altså dataflyten.

Spør om hvordan krav til åpenhet er ivaretatt

En større problemstilling knyttet til bruk av maskinlæringsløsninger er hvordan man skal sørge for å kunne forklare avgjørelser der behandlingsstøtte fra en maskinlæringsløsning er benyttet.

Som beskrevet ovenfor er manglende åpenhet, eller transparens, en gjennomgående problemstilling ved maskinlæringsløsninger. Likevel må virksomheten etterleve enkelte krav til å informere den registrerte. Denne informasjonsplikten gjelder også for den underliggende logikken, i visse typer automatiserte avgjørelser.

Det er derfor viktig å få avklart før kontrakt inngås om det for eksempel finnes måter å fremstille hvordan algoritmen vekter variabler og hvor nøyaktig algoritmen er. Sistnevnte kan for eksempel løses ved at løsningen som benyttes viser hvor stor sannsynlighet det er for at prediksjonen er korrekt.

Spør etter mekanismer for å fange opp og bøte med algoritmeskjevhet

Maskinlæring skaper en del nye problemstillinger knyttet til etikk i systemer. Mulig skjevhet i maskinlæringsalgoritmene kan utfordre prinsippet om rettferdighet i personvernforordningen artikkel 5 nr. 1 bokstav a.

Se også Datatilsynets KI-rapport på s. 15 (pdf).

Personvernkommisjonen peker på at slike skjevheter blant annet kan oppstå når det er manglende transparens i løsningen. Videre vil skjevheter forsterkes dersom løsningen benyttes ukritisk, eller mates med feil data.

I anskaffelse av maskinlæringsløsnigner kan det være lurt å undersøke om det finnes mekanismer i løsningen for å fange opp mulige skjevheter, hvor ofte algoritmen bør justeres og hvordan. Dersom det er mulig å finne ut hvilke situasjoner en algoritme er mindre nøyaktig i, vil dere lettere kunne iverksette egnede tiltak for å minske konsekvensene av denne skjevheten. Et annet alternativ vil være å retrene algoritmen så snart nøyaktigheten faller under en forhåndsbestemt tålegrense.

Veien videre

I dette prosjektet lærte vi at de største utfordringene knyttet til å ha rettslig grunnlag for å behandle personopplysninger med maskinlæring oppstår i etterlæringsfasen. Dersom det er et ønske om å kunne bruke maskinlæring for å løse utfordringene med at ikke all relevant dokumentasjon blir arkivert og journalført i tide, anbefaler sandkasseprosjektet at ny arkivlovgivning regulerer dette på en ansvarlig måte.

Deltakerne og Datatilsynet fått mulighet til å diskutere praktiske utfordringer og mulige tiltak for å fremme innebygd personvern i anskaffelse av løsninger som baserer seg på maskinlæring.

Ved å dypdykke i reglene for innebygd personvern har Simplifai fått større forståelse for hvilke krav de kan møte fra mulige kunder som vil kjøpe DAM. NVE har, på sin side, fått større bevissthet om ansvaret for løsningene som behandlingsansvarlig og hvordan direktoratet kan stille krav til løsningene i anskaffelsesfasen. Datatilsynet har lært mer om hvilke behov aktørene har for veiledning og hvordan Datatilsynet kan jobbe videre med å gi konkrete råd og eksempler.

Sandkasseprosjektet ser et behov for at flere aktører kommer sammen for å løfte kompetansen i offentlig sektor og lage praktiske eksempler på krav det offentlige kan stille i konkurranser om tekniske løsninger. På samme måte som det nå finnes maler for krav til leverandørenes kvalifikasjoner på anskaffelser.no, ønsker vi oss maler for krav til innebygd personvern for digitale verktøy. Slike eksempler i malverk vil også kunne gi mer forutsigbarhet for leverandørene.

Dersom EUs forslag til forordning om kunstig intelligens[2], blir vedtatt, blir behovet for å bruke kontrakter for å stille krav til produktene mindre. Forslaget stiller krav til utvalgte produkter som bruker maskinlæringsteknologi, ikke bare til de som bruker produktene.

[2] EUR-Lex - 52021PC0206 - EN - EUR-Lex (europa.eu)

Datatilsynets regulatoriske sandkasse for ansvarlig kunstig intelligens

Besøksadresse:
Trelastgata 3, Oslo

Postadresse:
Postboks 458 Sentrum
0105 Oslo

sandkasse@datatilsynet.no
Telefon: +47 22 39 69 00

datatilsynet.no/sandkasse
personvernbloggen.no
twitter.com/datatilsynet