SECURE PRACTICE

Sluttrapport fra sandkasseprosjektet med Secure Practice

Temaer: lovlighet, personvernkonsekvensvurdering (DPIA), rettferdighet, behandlingsansvarlig, tillit og transparens.

Februar 2022

Datatilsynet

Innhold

1. Sammendrag

Mål med sandkasseprosjektet

Å la seg profilere, kan gjøre livet enklere og mer interessant. Det er behagelig når strømmetjenester treffer med sine forslag. Og det er utvilsomt mer motiverende om et kurs er tilpasset akkurat din kunnskap og dine interesser. Det kan være store både personlige og samfunnsmessige gevinster i profilering. Men det er et tveegget sverd å søke mot et mer skreddersydd liv i det digitale samfunnet. Jo mer presis skreddersømmen er, jo mer presise personopplysninger finnes det om deg, med fare for misbruk.

Profilering på arbeidsplassen kan være spesielt utfordrende, siden relasjonen mellom arbeidstaker og arbeidsgiver har en iboende makt-ubalanse i seg. Arbeidstakere kan oppleve det som inngripende og krenkende. De kan føle seg overvåket og frykte misbruk av informasjonen.

Men kan det finnes en metode som utnytter fordelene ved profilering, samtidig som risikoen for ulempene blir redusert eller fjernet helt?

Det er utgangspunktet for dette sandkasseprosjektet, som tar for seg en ny tjeneste Secure Practice ønsker å tilby markedet for informasjonssikkerhet. Tjenesten skal bruke kunstig intelligens til å gi individuell og persontilpasset sikkerhetsopplæring til ansatte i kundenes virksomheter.

Folk er forskjellige, samtidig som sikkerhetsopplæring ofte blir for generell til å være effektiv. Men med kunstig intelligens kan Secure Practice tilby en skreddersydd og dermed mer pedagogisk opplæring. Både virksomheten og de ansatte vil dra nytte av bedre og mer interessant opplæring, og at folk unngår svindel og hacking.

I tillegg er det et formål med tjenesten, at virksomheten får oversikt på statistisk nivå om kunnskap og risiko, for å prioritere bedre tiltak. Baksiden av medaljen er en kartlegging som potensielt kan oppleves inngripende for den enkelte ansatte. Sandkasseprosjektet handler om hvordan en slik tjeneste kan gjøres personvernvennlig.

Konklusjoner

- **Behandlingsansvar:** Både arbeidsgiverne, Secure Practice og partene i fellesskap har ansvaret for å følge personvernreglene. Når KI-verktøyet er i bruk i virksomhetene, er arbeidsgiverne i utgangspunktet behandlingsansvarlig for behandlingen. Når Secure Practice tilbakeholder informasjon om hvilken arbeidstaker verktøyet profilerer, er arbeidsgiver og Secure Practice i fellesskap behandlingsansvarlige. Når KI-verktøyet blir utviklet videre i etterlæringsfasen, er bare Secure Practice behandlingsansvarlig.
- **Lovlighet:** Det er mulig å ta i bruk og videreutvikle tjenesten innenfor både generelle personvernregler i EU, og spesialregler for personvern i arbeidslivet i Norge.
- **Den registrertes rettigheter og friheter:** Med innovativ teknologi og formål som innebærer å forutsi personlige interesser og atferd besluttet prosjektet å gjennomføre en vurdering av personvernkonsekvensene (DPIA). Endelig viser vurderingene at tjenesten har lav risiko for diskriminering.
- **Åpenhet:** Prosjektet har vurdert om det er en rettslig forpliktelse å forklare den underliggende logikken i løsningen. Konklusjonen er at det ikke foreligger en rettslig plikt i dette konkrete tilfellet. Sandkassa anbefaler likevel mer åpenhet enn det som er rettslig påkrevd.

Veien videre

Denne sluttrapporten er viktig for å belyse forbudet mot overvåking i e-postforskriften[1]. Datatilsynet vil dele erfaringer fra prosjektet med Arbeids- og inkluderingsdepartementet, som er ansvarlig for forskriften, og for å gi departementet en mulighet til å klargjøre reglene.

[1] Forskrift om arbeidsgivers innsyn i e-postkasse og annet elektronisk lagret materiale, 2. juli nr. 1108. Forskriften er gitt av Arbeids- og inkluderingsdepartementet med hjemmel i arbeidsmiljøloven.

2. Om prosjektet

Secure Practice er en norsk teknologivirksomhet med fokus på den menneskelige delen av informasjonssikkerhetsarbeidet. Skytjenestene deres benyttes av over 500 virksomheter, med sluttbrukere i mer enn 30 land. En av tjenestene de i dag tilbyr, er MailRisk, som med kunstig intelligens hjelper ansatte med å finne ut om mistenkelig e-post er farlig eller trygg. Secure Practice utvikler og leverer også integrerte tjenester for e-læring og simulert phishing[2].

Nå vil Secure Practice bruke kunstig intelligens for å gi persontilpasset sikkerhetsopplæring til ansatte i kundenes virksomheter. Ved å ta utgangspunkt i hvilke interesser og kunnskaper den enkelte ansatte har om informasjonssikkerhet, skal opplæringen gjøres mer pedagogisk og målrettet, og dermed mer virkningsfull. Verktøyet skal i tillegg tilby virksomhetene rapporter med aggregert statistikk over ansattes kunnskaps- og interessenivå innenfor informasjonssikkerhet. Rapportene skal gjøre det mulig for arbeidsgiver å følge med på utviklingen over tid, og samtidig identifisere særskilte risikoområder og avdekke eventuelle behov for kollektive tiltak.

Profilering

Profilering er definert i personvernforordningen (GDPR) som:

«enhver form for automatisert behandling av personopplysninger som innebærer å bruke personopplysninger for å vurdere visse personlige aspekter knyttet til en fysisk person, særlig for å analysere eller forutsi aspekter som gjelder nevnte fysiske persons arbeidsprestasjoner, økonomiske situasjon, helse, personlige preferanser, interesser, pålitelighet, atferd, plassering eller bevegelser».

For å kunne gi persontilpasset opplæring, vil Secure Practice samle og sammenstille relevante data om de ansatte i kundens virksomhet. Profileringen plasserer hver enkelt sluttbruker i en av flere «risikokategorier», som blir utslagsgivende for hvilket opplæringstilbud han eller hun mottar i fortsettelsen. Re-kalkulering av risikoprofiler vil gjøres kontinuerlig og automatisk, slik at ansatte kan flyttes til ny kategori når de underliggende dataene tilsier dette.

Selve utviklingen av verktøyet bygger på et utvalg vitenskapelige studier relatert til menneskelig sikkerhetsatferd. Fra disse studiene har Secure Practice identifisert noen faktorer som det er ønskelig å kartlegge hos de ansatte. Det er fokusert på utvikling av en fleksibel statistisk modell og teknologisk løsning for behandling og kobling av ulike data i flere dimensjoner, inkludert tid. Med dette som utgangspunkt, kan selve vurderingen av risiko gjøres tilsvarende fleksibelt, ut fra hvilke hypoteser man til enhver tid legger til grunn i modellen. Dette er forutsetninger som i så måte er programmert inn i på forhånd, og «intelligensen» er i første omgang altså et produkt av kvaliteten på hypotesene.

Samtidig vil det være interessant å kunne benytte maskinlæring på data i historisk perspektiv. Såkalt etterlæring kan gjøres på bakgrunn av bruksdata for å identifisere mønstre for forbedring, eller eventuelt forverring.[3] Dette vil dermed kunne bidra til å forbedre hypotesene og utvikle enda mer treffsikre tiltak og anbefalinger i tjenesten i fremtiden.

Secure Practice har arbeidet med den nye tjenesten siden Innovasjon Norge innvilget et tilskudd til prosjektet i 2020. Forskningsrådet har også innvilget støtte til videreutvikling av teorier bak verktøyet, gjennom et forskningsprosjekt sammen med NTNU. Ved oppstart av sandkasseprosjektet hadde Secure Practice en teoretisk modell på plass, og teknisk implementasjon av den sentrale risikomodellen var innenfor rekkevidde. Og fordi den nye tjenesten integrerer mot eksisterende tjenesteplattform, er også mye «ferdig» i verktøyet. Samtidig har Secure Practice fortsatt hatt en rekke åpne spørsmål omkring både datainnsamling og brukergrensesnitt å ta stilling til.

---

2 Phishing er en form for sosial manipulering, hvor en angriper forsøker å lure noen til å utføre en handling, for eksempel åpne et skadelig vedlegg i e-post. Simulert phishing er øvelser for å lære å gjenkjenne phishing-forsøk. Les mer på datatilsynet.no.

3 Lær mer om etterlæring i pkt. 4.1.1 og 4.1.2

3. Mål for sandkasseprosessen

Sandkasseprosjektet er delt opp i tre delmål, hver med sine egne leveranser. De tre delmålene har blitt organisert tematisk omkring de tre sentrale rollene som utspiller seg når tjenesten tas i bruk; arbeidsgiverne, Secure Practice og de ansatte.

1. **Behandlingsansvarlig. Hvem har ansvaret for å følge personvernreglene?**
   Er det Secure Practice som drifter tjenesten, virksomheten som innfører det på arbeidsplassen eller begge i fellesskap som er behandlingsansvarlig? Er svaret det samme i bruksfasen, som når verktøyet videreutvikles i etterlæringsfasen?

2. **Lovlighet. Kan verktøyet tas i bruk og videreutvikles lovlig?**
   Prosjektet vil avklare hvilket rettslig grunnlag de behandlingsansvarlige kan ha for å profilere ansatte med formål om å tilby individuelt tilpasset sikkerhetsopplæring i virksomheter og statistiske rapporter til virksomheter. Prosjektet vil også se på om en slik profilering rammes av forbudet mot overvåking i e-postforskriften.

3. **Den registrerte. Hvordan påvirker verktøyet arbeidstakerne?**
   Prosjektet vil avklare hvordan den registrerte påvirkes med tanke på hvilke opplysninger som legges til grunn for behandlingen, risiko ved slik behandling, rettferdighet, åpenhet og rutiner for utøvelse av den registrertes rettigheter.

4. Vurderinger og konklusjoner

4.1 Hvem har ansvaret for å følge personvernreglene?

Behandlingsansvar

Personvernforordningen benytter begrepene *behandlingsansvarlig* i artikkel 4 nr. 7, *databehandler* i artikkel 4 nr. 8 og *felles behandlingsansvarlige* i artikkel 26 for å plassere ansvaret for å følge reglene. Av ansvarlighetsprinsippet går det frem, at hovedansvaret for å sikre at behandlingen av personopplysninger er i tråd med personvernforordningen ligger hos den behandlingsansvarlige.

En behandlingsansvarlig er den som fastsetter formålene med og midlene for behandlingen av personopplysninger, mens en databehandler behandler personopplysninger på vegne av den behandlingsansvarlige. Felles behandlingsansvar inntrer hvor partene i fellesskap fastsetter formålene med og midlene for behandlingen av personopplysninger.

4.1.1 Vurderinger av de ulike rollene i bruksfasen

Et spørsmål som kom opp tidlig i sandkasseprosjektet var hvordan ansvarsfordelingen mellom Secure Practice og deres kunder skal være. Bakgrunnen for spørsmålet var at Secure Practice selv kom inn i prosjektet med et ønske om å unngå at kunden (altså arbeidsgiver) får innsyn i individuelle risikoprofiler for ansatte, som et personverntiltak for å ivareta arbeidstaker. Secure Practice var likevel i tvil om dette var mulig å garantere i praksis, selv om de kunne gjøre tekniske tiltak i løsningen for å unngå at kunden uten videre fikk innsyn i slike data.

I øvrige tjenester har Secure Practice sett på sin egen rolle som databehandler og utformet ordinære databehandleravtaler med kunden. Kunden er da behandlingsansvarlig og bestemmer hvordan personopplysningene skal brukes. I en rolle som databehandler har Secure Practice plikt til å utlevere alle personopplysninger kunden ønsker å få tilgang til. En iboende risiko ved en slik ansvarsfordeling i den nye tjenesten, er at Secure Practice ikke kan la være å utlevere personopplysninger om enkelte arbeidstakere til arbeidsgiverne, selv ved mistanke om at opplysningene skal brukes til andre formål (for eksempel for å påvirke lønnsnivå eller bonusutbetaling).

I dialog med Secure Practice utforsket Datatilsynet konsekvensene av ulike ansvarsfordelinger, og en alternativ løsning med felles behandlingsansvar ble foreslått. Secure Practice ønsker å holde igjen personopplysninger fra kundene deres. Ved å holde igjen personopplysninger ovenfor sine kunder har de en avgjørende innflytelse over behandlingen av disse personopplysningene som går lengre enn deres rolle som databehandler.[4] Det vil si at Secure Practice og kunden i fellesskap fastsetter formålene og midlene for behandlingen av arbeidstakers personopplysninger.

Personvernrådet viser i sin veileder til et skille mellom "essensielle" og "ikke-essensielle" midler. I avsnitt 40 i veilederen knyttes "essensielle" midler til de valg som er overlatt til den behandlingsansvarlige. Som det fremgår av veilederen er det en nær tilknytning mellom hva som er "essensielle midler" og til spørsmålet om hvorvidt behandlingen er lovlig, nødvendig og forholdsmessig. "Ikke-essensielle" midler knytter seg til den praktiske implementeringen, som eksempelvis sikkerhetstiltak.

Å minimere arbeidsgivers adgang til arbeidstakers personopplysninger, kan vurderes som et tiltak for å minske personvernulempene. I vurderingen av det rettslige grunnlaget, og dermed om verktøyet er lovlig eller ikke, er personvernulempene relevante, se pkt. 4.2 nedenfor.

Praksis fra EU-domstolen viser at parter kan bli felles behandlingsansvarlige, selv om behandlingen av personopplysninger ikke er jevnt fordelt mellom partene eller kunden ikke har tilgang til personopplysningene som behandles. Dette kan være tilfellet hvor tjenesten behandler personopplysningene for sine egne formål, og denne behandlingen bare kan utføres fordi kunden tilrettelegger for en slik behandling ved å velge tjenesten.

Slik tjenesten er skissert i sandkasseprosjektet vil Secure Practice' behandling av personopplysninger bare være mulig fordi kunden benytter seg av tjenesten. Secure Practice behandler disse personopplysningene ved å hindre at kunden får utlevert personopplysningene om hver enkelt ansatt. Ved å tilbakeholde personopplysninger fra kunden oppstår det dermed et felles behandlingsansvar mellom Secure Practice og deres kunder.[5]

Formålet ved å organisere behandlingen som et felles behandlingsansvar er å sikre at ansvarsfordelingen gjenspeiler den faktiske rollen som Secure Practice påtar seg i de enkelte behandlingssituasjonene. Det innebærer at det ikke er noen endring for de prosessene hvor Secure Practice rent faktisk opptrer som en databehandler og behandler personopplysninger på kundens vegne. Der behandlingen ikke utelukkende blir gjort på kundens vegne, stiller personvernforordningen imidlertid krav til at partene identifiserer dette.

Secure Practice og den enkelte kunde må kartlegge prosessene hvor de i fellesskap fastsetter formålene med og midlene for behandlingen, slik at de på en åpen måte fastsetter ansvaret seg imellom. En åpen fastsettelse av ansvar mellom Secure Practice og deres kunder er ment å forhindre ansvarspulverisering mellom virksomhetene når arbeidstakere ønsker å benytte seg av sine rettigheter etter personvernforordningen.

Fastsettelsen av ansvar kan gjennomføres i en kontrakt eller et annet dokument mellom kunden og Secure Practice. Uavhengig av hvordan dette blir gjort mellom Secure Practice og kunden må det kommuniseres utad, slik at arbeidstakere er kjent med hvor de kan be om innsyn i sine personopplysninger og benytte seg av sine rettigheter etter personvernforordningen.

---

[4] EDPB, Guidelines 07/2020 on the concepts of controller and processor in the GDPR, avsnitt 40.

[5] EU-domstolen har argumentert med at sammenfallende interesser taler for et felles behandlingsansvar i dom av 29. juli 2019, *Fashion ID* C-40/17, avsnitt 80. "As to the purposes of those operations involving the processing of personal data, it appears that Fashion ID's embedding of the Facebook 'Like' button on its website allows it to optimise the publicity of its goods by making them more visible on the social network Facebook when a visitor to its website clicks on that button. The reason why Fashion ID seems to have consented, at least implicitly, to the collection and disclosure by transmission of the personal data of visitors to its website by embedding such a plugin on that website is in order to benefit from the commercial advantage consisting in increased publicity for its goods; those processing operations are performed in the economic interests of both Fashion ID and Facebook Ireland, for whom the fact that it can use those data for its own commercial purposes is the consideration for the benefit to Fashion ID." (Våre understrekinger).

Kunstig intelligens i tre faser:

Vi deler ofte opp kunstig intelligens i tre faser: *utvikling*, *bruk* og *etterlæring*. Personvernspørsmål oppstår i alle fasene.

Utviklingsfasen: Noen ganger brukes maskinlæring for å få innsikt i store datamengder, innsikt som i neste omgang kan brukes til å utvikle nye løsninger. Secure Practice har for eksempel gjort analyse av e-post for å finne hvilke egenskaper som er særskilt gjeldende for *mistenkelig* e-post. Slike funn kunne i neste omgang gjøres om til kunstig intelligens for deteksjon av skadelig e-post, gjennom konkrete mønstre som utviklerne legger inn i programvaren.

Bruksfasen: Maskinlæring kan også brukes til å oppdage nye sammenhenger i sanntid, uten at det må være mennesker involvert i prosessen. Her vil altså graden av automatikk være høy, i motsetning til utviklingsfasen, selv om det for øvrig kan være mye likt. Secure Practice bruker allerede maskinlæring i sanntid i MailRisk-tjenesten sin, uten at mennesker er involvert, for å identifisere og klassifisere helt nye svindelkampanjer og cyberangrep som ikke spamfiltre har klart å stoppe.

Etterlæringsfasen: Ligner mer på utviklingsfasen igjen, men gjerne med mer data tilgjengelig, og ikke minst data om resultatene etter at systemet fikk gjort jobben sin. Etterlæring kan være avgjørende for å verifisere eller forbedre korrektheten i en maskinlæringsmodell. Særlig gjelder dette forhold som forandrer seg over tid, slik som svindelmetoder brukt i skadelig e-post. Etterlæring er også aktuelt for komplekse forhold, slik som sikkerhetsatferd blant virksomhetens ansatte.

Secure Practice har ikke brukt personopplysninger i utviklingsfasen av tjenesten som prosjektet gjelder. Derfor har dette prosjektet valgt å fokusere på bruksfasen, som illustrerer de særskilte utfordringene på arbeidsplassen tydeligst. Vi har også kort omtalt etterlæring enkelte steder, for å illustrere at maskinlæring er en kontinuerlig prosess.

4.1.2 Vurderinger av de ulike rollene i etterlæringsfasen

Bruken av personopplysninger til etterlæring for å forbedre egne produkter er, i motsetning til den øvrige behandlingen, primært av interesse for Secure Practice, selv om også deres kunder kan være tjent med resultatet av en slik etterlæring.

Det er Secure Practice som fastsetter formålene med og midlene for denne behandlingen. Etterlæringen kunne imidlertid ikke blitt gjennomført uten personopplysningene som benyttes ellers i tjenesten.

For å sikre at Secure Practice alene har behandlingsansvar for etterlæringsfasen, bør det etableres kontroller i brukergrensesnittet, slik at kunden kan hindre bruken av personopplysninger til dette formålet. Et slikt skille sikrer at det ikke blir utydelig når Secure Practice opptrer som behandlingsansvarlig og selskapet opptrer som databehandler. Videre bør informasjon om behandlingen klart presenteres ovenfor kunder som ønsker å benytte seg av denne funksjonaliteten.

4.2 Kan KI-verktøyet tas i bruk og videreutvikles lovlig?

I dette kapitlet ser vi på rettslig grunnlag for løsningen, og på det nasjonale forbudet mot overvåking i e-postforskriften.

4.2.1 Hvilke regler gjelder?

Alle som vil samle inn, lagre eller på annen måte *behandle* personopplysninger må følge kravene i personvernforordningen (GDPR) . Vi skal her se nærmere på et grunnleggende krav som handler om at du må ha et rettslig grunnlag (samtykke, avtale, hjemmel i lov osv.) for å behandle personopplysninger.

Personvernforordningen skal gjelde på samme måte i 30 land i Europa. I tillegg har Norge spesialregler om personvern i arbeidslivet. Disse spesialreglene er gitt i forskrifter til arbeidsmiljøloven[6]. Den aktuelle forskriften for Secure Practice, er forskrift om arbeidsgivers innsyn i e-postkasse og annet elektronisk lagret materiale av 2. juli 2018 og blir populært kalt «e-postforskriften». E-postforskriften gjelder arbeidsgiveres innsyn i alt elektronisk utstyr arbeidstakere bruker i jobbsammenheng og er ment å beskytte arbeidstakere mot unødvendig inngripende overvåking eller kontroll. E-postforskriften § 2, andre ledd forbyr overvåkning av de ansattes bruk av elektronisk utstyr. I sandkasseprosjektet har vi vurdert hvordan dette påvirker verktøyet Secure Practice utvikler.

4.2.2 Bruksfasen

Før Secure Practice tilbyr verktøyet på markedet, er det viktig å få avklart om framtidige kunder vil ha lov til å ta verktøyet i bruk. I den første workshopen i sandkasseprosjektet startet vi med å se på om en arbeidsgiver som kjøper tjenesten har rettslig grunnlag for å behandle personopplysninger fra arbeidstakerne i dette verktøyet. Vi så også nærmere på hvordan tjenesten berøres av arbeidsgivers forbud mot å overvåke arbeidstakerne, slik det står i e-postforskriften.

Vi nevner kort at de prosessene hvor Secure Practice og arbeidsgiveren har et felles behandlingsansvar, må begge ha et rettslig grunnlag for behandlingen. I dette kapitlet om bruksfasen har vi valgt å fokusere på arbeidsgiveren som behandlingsansvarlig.

Hvilke personopplysninger var aktuelle i KI-verktøyet og ble vurdert i sandkasseprosjektet?

Før vi vurderte rettslig grunnlag, listet Secure Practice opp alle *mulige* opplysninger som kunne ha en verdi for å kartlegge ansattes interesser og kunnskap innen informasjonssikkerhet. Målet med øvelsen var å vurdere nærmere hvilke opplysninger som i det hele tatt er aktuelle, og ta stilling til hva som kan være greit å bruke fra et personvernperspektiv.

Opplysninger Secure Practice vurderte både som ønskelige og i den «riktige» enden av personvernskalaen, var data om gjennomføring av e-læring, phishing-øvelser og rapportering fra MailRisk-tjenesten, i tillegg til svar på spørreundersøkelser og quiz-er om kunnskap og vaner relatert til informasjonssikkerhet.

Deretter ble det listet opp en rekke opplysninger som kunne befinne seg i grenseland, både med tanke på personvern og nytteverdi. Disse var de viktigste å få diskutert, for eksempel demografiske data som *alder* og *ansiennitet* for de ansatte, og data fra andre sikkerhetssystemer som eksempelvis weblogger.

Til slutt var det opplysninger som egentlig aldri var aktuelt fra et personvernperspektiv å benytte, men som likevel ble tatt med på den opprinnelige listen på grunn av potensiell nytteverdi. Eksempel på denne kategorien er psykologiske tester eller opplysninger hentet fra profiler på sosiale nettverk.

---

[6] Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. av 17. juni 2005

Kort om det aktuelle rettslige grunnlaget – berettigede interesser

Secure Practice og Datatilsynet vurderte at "berettigede interesser", etter personvernforordningen artikkel 6 nr. 1 bokstav f, var det mest aktuelle rettslige grunnlaget i dette prosjektet.[7]

Bestemmelsen gir tre vilkår, som alle må være oppfylt for at en behandling skal være lovlig:

1. Formålet med behandlingen må være knyttet til en berettiget interesse
2. Behandlingen må være nødvendig for å oppnå formålet
3. Arbeidstakerens interesser, rettigheter og friheter må ikke overstige arbeidsgiverens interesser. Kort sagt kaller vi dette trinnet for «interesseavveiing».

Vilkår nr. 1 - berettiget interesse

Som nevnt ovenfor, er formålene med å bruke verktøyet todelt:

1) Å gi de ansatte individuelt tilpasset opplæring i informasjonssikkerhet.
2) Å gi statistiske rapporter til virksomhetene, som på gruppenivå beskriver ansattes kunnskap- og interessenivå innenfor informasjonssikkerhet.

Begge formålene er knyttet til virksomhetens interesse i å bedre informasjonssikkerheten. Vi vurderte bedre informasjonssikkerhet til å være i interessene til både virksomheten selv, de ansatte, tredjeparter som kunder og samarbeidspartnere og samfunnet som helhet. Det var lett å slå fast at bedring av informasjonssikkerhet utgjør en berettiget interesse og at det første vilkåret dermed er oppfylt. Diskusjonene i prosjektet har derfor dreid seg om de to siste vilkårene.

Vilkår nr. 2 - nødvendighet

Nyttige spørsmål i vurderingen av hvilke behandlinger som er nødvendige er:

- Vil behandlingen av disse personopplysningene faktisk bidra til å oppnå formålene?
- Kan formålene oppnås uten å behandle disse personopplysningene eller ved å behandle færre personopplysninger?

De ansattes kunnskap og interesse for informasjonssikkerhet må kartlegges for å oppnå begge formålene – både individuelt tilpasset sikkerhetsopplæring og statistisk rapportering til virksomhetene. Diskusjonene i sandkassen dreide seg her om hvordan Secure Practice kan minimere bruk av personopplysninger og å sikre at de opplysningene som brukes faktisk bidrar til å oppnå formålene.

Vi dannet fokusgrupper av ansatte fra fagforeningen Negotia og en stor bedrift, for å få potensielle brukeres perspektiv i vurderingene. Der kom det interessante innspill til vurderingen om hvordan tilsynelatende hensiktsmessige opplysninger kan være det stikk motsatte.

En av metodene Secure Practice ønsker å bruke i kartleggingen er spørreundersøkelser. Der skal arbeidstakerne ta stilling til forskjellige utsagn, som for eksempel «jeg har bevisst brutt regler for informasjonssikkerhet på jobb». Fokusgruppen med representanter fra arbeidstakerorganisasjonen kommenterte at det er uklart hvilke konsekvenser det ville kunne få for den enkelte ansatte å svare på et slikt spørsmål, dersom de faktisk hadde brutt reglene. I den andre fokusgruppen ble det lagt vekt på betydningen av anonymitet overfor arbeidsgiver, dersom man skal gi et ærlig svar på dette.

Med tanke på formålet om individuelt tilpasset opplæring i informasjonssikkerhet vil ikke spørsmålet bidra til å oppnå formålet, dersom arbeidstakerne som har brutt sikkerhetsreglene svarer nei, fordi de er redde for konsekvensene av å svare ja.

Når det gjelder formålet om statistiske rapporter til virksomhetene, kunne spørsmålet isolert sett bidratt til å oppnå formålet, så lenge noen av de ansatte ville svart ja. En slik tilbakemelding kunne gitt virksomheten en pekepinn om at de

---

7 Les gjerne mer om bestemmelsen på Datatilsynets nettsider om behandlingsgrunnlag, og på det britiske datatilsynets nettsider under tittelen What is the 'legitimate interests' basis?

interne reglene for informasjonssikkerhet er dårlig utformet eller ikke passer inn i arbeidshverdagen til de ansatte. Eksemplet illustrerer at den behandlingsansvarlige må vurdere rettslig grunnlag separat for hvert formål behandlingen skal ivareta.

Som det går fram av denne rapportens punkt 4.4, fikk Secure Practice råd om å reformulere noen av spørsmålene for å lettere kunne få ærlige svar.

Når Secure Practice har identifisert personopplysningene som er nødvendige for at verktøyet skal kunne fungere som individuell opplæring og en overordnet informasjon til arbeidsgiverne, vil neste steg være en interesseavveiing.

Vilkår nr. 3 - interesseavveiing

Det tredje vilkåret handler om at arbeidsgiver ikke kan innføre tiltak hvis arbeidstakeres interesser, rettigheter og friheter veier tyngre enn arbeidsgiverens interesser. Så en interesseavveiing handler kort sagt om å finne en balanse mellom interessene på begge sider, slik at inngrepet i personvernet blir forholdsmessig. For å gjøre en slik avveiing, startet vi med å undersøke hvordan arbeidstakere berøres av verktøyet.

Ansatte er i et ujevnt maktforhold overfor arbeidsgiver. Derfor er det ekstra viktig for ansatte at personopplysningene om dem ikke blir misbrukt til nye formål. En ansatt som blir "lurt" i en phishing-øvelse, forventer at resultatet bare blir brukt til opplæringsformål, og ikke for å vurdere hvilke oppgaver eller goder han eller hun får. Men hvis arbeidsgiver får tilgang til denne informasjonen, finnes det en risiko for slikt misbruk.

Et særtrekk ved løsninger som bruker kunstig intelligens er at de ofte behandler svært mange personopplysninger.[8] Verktøyet vi diskuterer i dette prosjektet er ment å kartlegge både kunnskap og interesser, og er egnet til en detaljert kartlegging av de ansatte. Dette kan oppleves som inngripende. I tillegg kan løsningene innen kunstig intelligens gi et uventet resultat.

Vi vurderer hensynet til de ansatte å veie tungt, og dermed stilles det større krav til interessene knyttet til arbeidsgiverne. På den andre siden veier også interessen knyttet til bedre informasjonssikkerhet tungt. Som nevnt i punktet om berettiget interesse, er bedre informasjonssikkerhet en interesse som kommer den enkelte arbeidstaker til gode, ikke bare arbeidsgiverne.

I avveiingen mellom interessen knyttet til informasjonssikkerhet og hensynet til de ansattes personvern vurderte vi særlig disse momentene:

- Hvordan de ansatte kommer til å oppleve kartleggingen og hvilke konsekvenser det kan ha for dem. Positive konsekvenser kan være at de får skreddersydd hjelp og oppfølging til å styrke kompetansen sin på informasjonssikkerhet, og at de unngår å bli rammet av svindel og hacking med konsekvensene det måtte medføre. Potensielle negative konsekvenser kan være at ansatte føler usikkerhet om hvordan dataene om dem blir brukt, og om det kan komme negative konsekvenser dersom de avslører lite kunnskap eller interesse for informasjonssikkerhet.
- Hvordan arbeidsgiverne involverer arbeidstakerne før de innfører verktøyet.
- Hvilken informasjon arbeidsgiveren får tilgang til om den enkelte ansatte.
- Hvordan Secure Practice gir informasjon til de ansatte i verktøyet og i personvernerklæring.
- Hvilke tekniske garantier som er bygd inn for å hindre at utenforstående får tilgang til personopplysningene om de ansatte.

Når det gjelder kulepunktene om mulige konsekvenser for de ansatte og hvilken informasjon arbeidsgiveren får tilgang til om enkeltansatte, arbeidet Secure Practice ut fra en forutsetning om at arbeidsgiver ikke skulle ha tilgang til målinger knyttet til hver enkelt ansatt gjennom verktøyet. Sandkassas anbefaling er å innføre både *juridiske* og *tekniske* garantier for at opplysninger om de ansatte ikke skal komme på avveie.

---

[8] En eventuell diskriminering i verktøyet vil også ha stor virkning, siden den gjelder den enkeltes karriere. Når det gjelder mulig diskriminering i løsningen, viser vi til vurderingen fra Likestillingsombudet nedenfor om at risikoen for diskriminering er lav (se pkt. 4.5). Risiko for diskriminering er derfor ikke vektlagt i interesseavveiingen.

Med juridiske garantier mener vi, at Secure Practice regulerer i kontrakt at arbeidsgiverne ikke får tilgang til opplysninger om enkeltansatte – heller ikke på særskilt forespørsel. Med tekniske garantier viser vi til de tiltakene Secure Practice allerede har satt i gang for å hindre arbeidsgiveren eller andre i å få tilgang til disse opplysningene. Tjenesten bruker både pseudonymisering og kryptering for å beskytte personopplysningene. Navnet på den ansatte erstattes med en unik identifikator, og koblingen mellom navn og identifikator lagres i en separat database. Secure Practice har samtidig identifisert et behov for å lagre brukerens e-postadresse i tilknytning til brukerens øvrige opplysninger. En personlig e-postadresse etablerer en tydelig kobling til et enkeltindivid, og vil derfor utfordre den opprinnelige målsettingen om en pseudonymisert database.

For å imøtekomme denne utfordringen har Secure Practice valgt å kryptere både e-postadresser, navn og andre direkte identifikatorer i selve databasen, slik at disse ikke lenger skal være tilgjengelige i klartekst. Nøklene som brukes for å kryptere og dekryptere slike identifikatorer lagres adskilt fra databasen. På denne måten vil en person som får tilgang til brukerdatabasen, ikke få tilgang til informasjon som er nødvendig for å koble personopplysningene til enkeltindivider.

Når både juridiske og tekniske garantier kommer på plass, vurderer sandkassa risikoen som liten for at arbeidsgiver eller andre vil kunne få tilgang til og eventuelt misbruke personopplysningene som verktøyet samler inn. Dette bidrar til at interessene samlet sett veier tyngst på den behandlingsansvarliges side. Det vil derfor være mulig å bruke berettiget interesse som rettslig grunnlag i bruksfasen.

Retten til å protestere

Når berettiget interesse brukes som rettslig grunnlag, har arbeidstakerne en rett til å protestere mot behandlingen av personopplysningene etter artikkel 21 i personvernforordningen. Dersom en ansatt protesterer mot å bruke hennes personopplysninger i verktøyet, må arbeidsgiverne ta hensyn til de særlige forholdene den ansatte har pekt på i protesten. Overfor den som har protestert skal arbeidsgiveren fortsatt gjøre en interesseavveiing, men må vise at det er «*tvingende berettigede grunner*» til å bruke personopplysningene i verktøyet.[9]

Vurderingen må altså gjøres individuelt, ut fra begrunnelsen den ansatte som protesterer har gitt. En behandlingsansvarlig som skal vurdere en protest må i større grad vurdere alternativer til spesialtilpasset opplæring og rapportering på statistisk nivå. Secure Practice ser for seg at de som protesterer, enten vil få protesten behandlet av arbeidsgiver, eller få den innvilget gjennom verktøyet uten videre manuell behandling, da det kan være utfordrende for Secure Practice å vurdere grunnlaget for protesten.

Dersom mange ansatte protesterer og ikke bruker verktøyet, blir naturlig nok en mindre andel av virksomheten kartlagt i verktøyet. De behandlingsansvarlige må være oppmerksom på at dette kan gjøre det lettere å identifisere arbeidstakerne i rapporteringen på statistisk nivå.

4.2.3 Etterlæring

Bruk av personopplysninger til etterlæring for å forbedre verktøyet, krever også et rettslig grunnlag. [10] Secure Practice er behandlingsansvarlig alene for denne fasen. På samme måte som for bruksfasen, er formålene med å forbedre tjenesten knyttet til en berettiget interesse. Etterlæringen er forventet å gjøre tjenesten mer treffsikker, etter hvert som personopplysninger fra flere ansatte blir lagt inn i verktøyet. Det vil trolig øke informasjonssikkerheten i kundenes virksomheter.

I motsetning til i bruksfasen, er likevel formålene tydeligere knyttet til kommersielle interesser, siden en forventet økt kvalitet kan gi økt salg. Vi nevner derfor kort at også kommersielle interesser er berettigede interesser, slik at det første vilkåret er oppfylt.

Når det gjelder det andre vilkåret, må Secure Practice ta aktivt stilling til hvilke personopplysninger som er nødvendige for å oppnå hvert enkelt formål. Vi går ut fra samme type personopplysninger som i bruksfasen er aktuelle for å gjøre tjenesten mer treffsikker. Dersom Secure Practice skal teste verktøyet mot mulig diskriminering, krever det en nærmere

---

[9] Se artikkel 21 i personvernforordningen.

[10] Se pkt. 4.1 for en nærmere forklaring av etterlæring. I sandkasseprosjektet ble det avgrenset mot problemstillinger knyttet til gjenbruk av personopplysninger og begrensninger som følger av formålsbegrensningsprinsippet i artikkel 5 nr. 1 bokstav b og artikkel 6 nr. 4. Det franske datatilsynet har publisert veiledning om bruk av personopplysninger til etterlæring på sine nettsider (cnil.fr) under tittelen «Sous-traitants: la réutilisation de données confiées par un responsable de traitement».

vurdering av hvilke opplysninger som er nødvendige. For dette formålet kan for eksempel tilgang på demografiske data som alder og kjønn være av større nytteverdi. Denne vurderingen har vi likevel ikke gått inn på i dette prosjektet.

I avveiingen mellom hensynet til Secure Practice sine interesser og de ansattes personvern, kan Secure Practice legge vekt på at økt treffsikkerhet vil komme virksomhetene og de ansatte til gode. Hvis løsningen ikke blir oppdatert med etterlæring, kan verktøyet bli utdatert og potensielt ikke fungere etter hensikten. Mulige personvernulemper for de ansatte kan være større i denne fasen, fordi personopplysningene blir brukt til å videreutvikle verktøyet til *nye* virksomheter. Med juridiske og tekniske garantier på plass, som vi gjorde rede for i bruksfasen, er det vår oppfatning at det vil være mulig å bruke berettiget interesse som rettslig grunnlag, også i etterlæringsfasen.

Også i denne fasen kan ansatte protestere mot behandlingen.

4.2.4 Det nasjonale forbudet mot overvåking i e-postforskriften

Så langt har vi forholdt oss til personvernforordningen. Men det er også relevant å vurdere tjenesten opp mot e-postforskriften med sitt forbud mot «å overvåke arbeidstakers bruk av elektronisk utstyr, herunder bruk av Internett». I motsetning til personvernforordningen, må det altså noe mer til enn å behandle personopplysninger for at disse spesialreglene begynner å gjelde.

En annen viktig forskjell fra personvernforordningen, er når en eventuell overvåking er lovlig. Det finnes bare to lovlige tilfeller. Enten for å «administrere virksomhetens datanettverk» eller for å «avdekke eller oppklare sikkerhetsbrudd i nettverket».

En arbeidsgiver vil altså ha lov til å bruke verktøyet så lenge bruken ikke innebærer overvåking av arbeidstakerne eller ett av de to tilfellene nevnt ovenfor er oppfylt. Men kan bruk av KI-verktøyet 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.[11] 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.

KI-verktøyet som Secure Practice utvikler er satt sammen av mange metoder for å kartlegge arbeidstakerne. Metodene spenner fra spørreundersøkelser og quiz, til å registrere hvordan arbeidstakerne reagerer på simulerte phishing-øvelser og aktivitet i læringsplattformen. Isolert sett vil ikke dette regnes som å overvåke bruk av elektronisk utstyr. Men spørsmålet er om den samlede kartleggingen rammes av forbudet mot å overvåke.

Praksis fra Datatilsynet er ikke entydig, med tanke på om det er et krav at arbeidsgiver faktisk skal se opplysningene eller metadata for at det skal regnes som overvåking. Overvåkingsbegrepet favner vidt, og kan tyde på 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.

Etter diskusjonene i sandkassen var det enighet om at kartleggingen i verktøyet ikke vil omfattes av forbudet mot overvåking. Vi har særlig lagt vekt på at de tekniske og juridiske tiltakene som er satt i verk for at arbeidsgiveren ikke skal få tilgang til opplysningene som blir samlet inn om hver ansatt.

De statistiske rapportene til arbeidsgiverne om nivået på informasjonssikkerhet blant bedriftens ansatte, vil lettere kunne oppleves som overvåking. Rapporteringen skal skje på gruppenivå. Hvor mange ansatte bedriften har og hvor store grupper personalet deles inn i, vil trolig påvirke hvordan de ansatte opplever verktøyet. Vi har gått ut fra at data vil

---

[11] Fornyings- og administrasjonsdepartementets høringsnotat: Forslag til regler om arbeidsgivers tilgang til ansattes e-post mv, 17.10.2006, s. 15: «Overvåke» innebærer at tiltaket har en viss varighet eller at det skjer gjentatte ganger. Motsatsen er det situasjonsbetingete innsyn som er tillatt etter forskriften her. Bestemmelsen omfatter både automatisk og manuell overvåkning. Bestemmelsen må sees i sammenheng med § 7-11 siste ledd, som forbyr bruk av aktivitetslogger for kontroll og overvåkning av enkeltpersoner.» (Sitert fra Signhild Blekastad og Marion Holthe Hirst: Personvern og kontroll i arbeidslivet s. 310.)

gis på en måte som ikke setter arbeidsgiveren i stand til å identifisere enkeltarbeidstakere. Hvilken informasjon arbeidstakerne får, vil også spille inn på hvordan de opplever tjenesten.

Sandkassa har kommet til at kartleggingen, som bare formidles til arbeidstakerne, ikke rammes av forbudet mot overvåking. Når det gjelder måling av sikkerhetsnivået for de ansatte på gruppenivå, må arbeidsgiveren vurdere utformingen nærmere. Vi anbefaler at arbeidsgiveren drøfter hvordan rapporteringen kan skje med de ansatte på forhånd.

Siden bruk av verktøyet ikke regnes som overvåking i dette tilfellet, trenger vi ikke å vurdere resten av bestemmelsen.

4.2.5 Hvilke (tilgrensende) regler har ikke blitt vurdert i sandkasseprosjektet?

Vi har ikke vurdert om arbeidsgiverens bruk av KI-verktøy går inn under reglene om kontrolltiltak i arbeidsmiljøloven kap. IX. Arbeidstilsynet håndhever disse reglene, og du kan lese mer på deres hjemmesider.

Vi har heller ikke gått inn på reglene som gjelder *tilgang* til informasjon som er lagret på en brukers terminalutstyr, datamaskin, telefon osv., som er regulert i e-komloven.

4.3 Personvernkonsekvensvurdering – DPIA

Dersom det er sannsynlig at en type behandling av personopplysninger vil medføre høy risiko for folks rettigheter og friheter, skal den behandlingsansvarlige vurdere hvilke konsekvenser den planlagte behandlingen vil ha for personvernet. Dette gjelder særlig ved bruk av ny teknologi.

Det kan være en utfordring å fastslå med sikkerhet når det foreligger høy risiko. Når dette er usikkert, anbefaler Datatilsynet at man gjennomfører en personvernkonsekvensvurdering, også kalt DPIA (Data Protection Impact Assessment). Dette kan være et godt verktøy for å sikre at også de øvrige kravene i personvernforordningen følges.

Datatilsynet har utarbeidet en liste over behandlingsaktiviteter som alltid krever at det gjennomføres en vurdering av personvernkonsekvenser.[12] Fra denne listen er følgende elementer relevante for verktøyet Secure Practice utvikler:

- Behandling av personopplysninger med kunstig intelligens, som er innovativ teknologi.
- Behandling av personopplysninger for systematisk monitorering av ansatte.
- Behandling av personopplysninger, der formålet er å tilby en tjeneste eller utvikle produkter for kommersiell bruk, som involverer å forutsi jobbprestasjoner, økonomi, helse, personlige preferanser eller interesser, pålitelighet, adferd, lokasjon eller bevegelsesmønster. (Særlige kategorier av personopplysninger eller svært personlige opplysninger og evaluering/poengsetting).

Sandkassa konkluderte derfor med at bruk av verktøyet til Secure Practice krever en vurdering av personvernkonsekvenser. Det er den behandlingsansvarliges ansvar at en DPIA gjennomføres. I praksis betyr dette at virksomheter som kjøper den nye tjenesten av Secure Practice, også må gjøre en DPIA.

For mange små og mellomstore virksomheter kan det være krevende å gjennomføre en tilstrekkelig vurdering rundt personvernkonsekvensene av et verktøy som baserer seg på kunstig intelligens. Dette forutsetter blant annet kjennskap til personvernregelverket og andre grunnleggende rettigheter, kunstig intelligens og kjennskap til systemets logikk i tillegg til særegne forhold på den enkelte arbeidsplass.

[12] Finnes på datatilsynet.no under tittelen «Når må man gjennomføre en vurdering av personvernkonsekvenser». Listen er bygd på en veileder fra Artikkel 29-gruppen, Guidelines on Data Protection Impact Assessment (DPIA) (wp248). Veilederen er godkjent av Personvernrådet.

Det asymmetriske forholdet mellom kunde og tjenestetilbyder gir seg ofte til kjenne i det digitale samfunnet. Satt på spissen, kan det av og til virke som om leverandøren setter krav til kunden fremfor omvendt. En tilsvarende dynamikk kan også finnes mellom en tilbyder av KI-tjenester og kunde. I denne situasjonen vil tilbyder også inneha rollen som fagekspert, som kan belyse både fordeler og ulemper med teknologien de omsetter.

Datatilsynet og Secure Practice var enige om at ansvarlig bruk av KI forutsetter en behandlingsansvarlig med et godt informasjonsgrunnlag, som setter den behandlingsansvarlige i stand til å gjennomføre de riktige vurderingene. På bakgrunn av dette besluttet sandkassa å inkludere vurdering av personvernkonsekvenser i prosjektet.

Det er viktig å presisere at det ikke er tilstrekkelig å vurdere personvernkonsekvensene av et verktøy i seg selv. Vurderingen skal også ta høyde for hvilken sammenheng verktøyet blir brukt i. Denne sammenhengen vil ofte variere fra kunde til kunde, og hvilken del av landet (eller verden) kunden holder til i. Det betyr at Secure Practice kan gjøre en del av utredningsarbeidet for kunden på forhånd, men vurderingene for hvert konkrete forhold må kunden gjøre selv.

For Datatilsynets del var det viktig å kunne bidra med effektiv veiledning i prosessen for vurdering av personvernkonsekvenser, uten at rådene etterlot så lite manøvreringsrom at de utfordret Secure Practice' eierskap til prosessen. Dette var spesielt viktig ettersom tjenestens utvikling var i en tidlig fase, da Datatilsynet ga tilbakemeldinger. Secure Practice arrangerte selv en workshop for vurdering av personvernkonsekvensene sammen med Datatilsynet, og gjorde deretter jobben med å dokumentere resultatene.

Datatilsynet gav tilbakemelding om ulike problemstillinger knyttet til potensielle personvernkonsekvenser og utformingen av selve vurderingen:

- Å publisere vurderingen av personvernkonsekvensene på nett kan være ett av flere tiltak for å legge til rette for en åpen behandling av personopplysninger. Å publisere vurderingen er i seg selv ikke godt nok til å vareta informasjonsplikten. Den registrerte skal informeres på en kortfattet, åpen, forståelig og lett tilgjengelig måte.
- Det er viktig å sikre rutiner slik at personvernerklæringen og vurderingen av personvernkonsekvenser oppdateres parallelt. Det vil si at endringer og nye løsninger som implementeres i produksjonsløsningen må gjenspeiles og håndteres i vurderingen av personvernkonsekvenser og i personvernerklæringen når det er relevant.
- Det er viktig å unngå for vage og brede formuleringer. Beskrivelsene bør være så presise som mulig, slik at det er mulig å se hva som er vurdert. Det må fremstå tydelig hva som er gjenstand for vurdering.
- Budskap som er rettet mot kunden (virksomhet) må unngå formuleringer som kan forveksles med det rettslige grunnlaget for behandlingen av arbeidstakers personopplysninger. Det var særlig aktuelt der kundens avtale med Secure Practice ble nevnt, og dette kunne forveksles med henvisning til personvernforordningen artikkel 6 nr. 1 bokstav b («behandlingen er nødvendig for å oppfylle en avtale som den registrerte er part i […]»).
- Terskelverdier for konsekvens og sannsynlighet må ta særlig hensyn til risikoen for den enkelte registrerte, og at en vurdering av personvernkonsekvenser ikke utelukkende dreier seg om hvor mange som blir berørt, men også konsekvensene for den enkelte registrerte.
- Datatilsynet anbefalte at risikoen for den registrertes friheter og rettigheter utredes nærmere med hensyn til den tiltenkte ansvarsfordelingen mellom Secure Practice og virksomheten. Dette kan tydeliggjøre både rolle- og ansvarsfordeling for den registrerte.

Tilbakemeldingene på personvernkonsekvensvurderingen over er fokusert på bruksfasen. Vi nevner til slutt at Secure Practice også må vurdere personvernkonsekvensene for etterlæringsfasen, der de har selvstendig behandlingsansvar.

4.4 Hvordan forklare bruken av kunstig intelligens?

Å behandle personopplysninger på en åpen måte er et grunnleggende prinsipp i personopplysningsloven. Åpenhet setter den registrerte i stand til å bruke sine rettigheter og ivareta sine interesser. I sandkasseprosjektet diskuterte vi hvilke krav som stilles til å informere de registrerte om hvordan personopplysningene behandles. I tillegg drøftet vi konkrete problemstillinger knyttet til brukergrensesnittet.

4.4.1 Hvilke krav stilles til åpenhet?

Kravet om å informere den registrerte finner vi i personvernforordningens artikkel 13 til 15. Artikkel 13 regulerer hvilken informasjon som skal gis ved innsamling av personopplysninger fra den registrerte. Artikkel 14 regulerer hvilken informasjon, og når denne informasjonen skal gis, dersom personopplysninger ikke er samlet inn fra den registrerte selv. Artikkel 15 regulerer den registrertes rett til innsyn i personopplysningene som behandles om dem. I tillegg gir artikkel 12 en generell plikt til å gi informasjon på en kortfattet, åpen, forståelig og lett tilgjengelig måte og på et klart og enkelt språk.

Uavhengig om du bruker kunstig intelligens eller ikke er det visse krav til åpenhet dersom du behandler personopplysninger. Kort oppsummert er disse:

- De registrerte må få informasjon om hvordan opplysningene brukes, enten opplysningene hentes inn fra *den registrerte* selv eller fra andre.
- Informasjonen må være lett tilgjengelig, for eksempel på en hjemmeside, og være skrevet i et klart og forståelig språk.
- Den registrerte har rett til å få vite om det behandles opplysninger om henne og eventuelt innsyn i egne opplysninger.
- Det er et grunnleggende krav at all behandling av personopplysninger skal gjøres på en åpen måte. Det betyr at det er krav om å vurdere hvilke åpenhetstiltak som må til for at den registrerte skal kunne ivareta egne rettigheter.

I det første kulepunktet er det krav om å gi informasjon om hvordan opplysningene brukes. Det inkluderer blant annet kontaktinformasjon til den behandlingsansvarlige, formålet med behandlingen og hvilke kategorier personopplysninger som blir behandlet. Dette er informasjon som typisk formidles i personvernerklæringen.

Disse pliktene retter seg mot den behandlingsansvarlige. Når Secure Practice og arbeidsgiver har felles behandlingsansvar, må de fastsette hvilket ansvar hver av dem har for å oppfylle kravene i personvernforordningen.[13] Plikten til å fordele ansvar følger av artikkel 26 i personvernforordningen. Det innebærer blant annet informasjon om hvordan de registrerte kan utøve sine rettigheter og hvilke personopplysninger som blir behandlet om dem i verktøyet.

4.4.2 Har arbeidstakerne krav på å få informasjon om logikken til algoritmen?

For automatiserte avgjørelser som har rettsvirkning eller i betydelig grad påvirker en person, gjelder det særlige krav til informasjon. Det fremgår av artikkel 13 nr. 2 bokstav f at den behandlingsansvarlige i disse tilfellene skal opplyse om den underliggende logikken til algoritmen. Det samme gjelder etter artikkel 14 nr. 2 bokstav g når personopplysningene ikke er innhentet direkte fra den registrerte.

Retningslinjer fra Artikkel 29-gruppen

«Articles 13(2) (f) and 14(2) (g) require controllers to provide specific, easily accessible information about automated decision-making, based solely on automated processing, including profiling, that produces legal or simliarly signifant effects.

If the controller is making automated decisions as described in Article 22 (1), they must:

- tell the data subject that they are engaging in this type of activity;
- provide meaningful information about the logic involved; and
- explain the significance and envisaged consequences of the processing»

(Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, side 25.)

[13] Se nærmere om felles behandlingsansvar i pkt. 4.1. og 4.2.

Verktøyet i dette sandkasseprosjektet fører verken til rettsvirkninger for arbeidstakerne eller påvirker dem i betydelig grad. Behandlingen faller dermed utenfor artikkel 22 i forordningen. Informasjonsplikten etter artikkel 13 og 14 retter seg mot behandling som er omfattet av artikkel 22. Derfor følger det ikke noen plikt til å informere om hvordan algoritmen fungerer, direkte av denne bestemmelsen.

Prosjektet vurderte om åpenhetsprinsippet lest i lys av fortalen kunne tilsi en rettslig plikt til å informere om hvordan algoritmen fungerer.

Etter personvernforordningens artikkel 5 nr. 1 bokstav a skal den behandlingsansvarlige sikre at behandling av personopplysninger gjøres på et åpent og rettferdig vis. Fortalepunkt 60 fremhever i tilknytning til åpenhetsprinsippet at den registrerte bør få informasjon når det skjer profilering og hvilke konsekvenser profileringen har. Fortalepunktet viser til profilering i alminnelighet, og det fremstår dermed som noe videre enn artikkel 13 nr. 2 bokstav f og artikkel 14 nr. 2 bokstav g, som peker mot automatiserte avgjørelser med rettslige eller andre betydelige konsekvenser.

Punkt 60 i fortalen til personvernforordningen

«Prinsippene om rettferdig og åpen behandling krever at den registrerte informeres om at behandlingen skjer, samt om formålet med den. Den behandlingsansvarlige bør gi den registrerte eventuell ytterligere informasjon som er nødvendig for å sikre en rettferdig og åpen behandling, idet det tas hensyn til de særlige omstendighetene rundt behandlingen av personopplysningene og sammenhengen den skjer i. Den registrerte bør dessuten informeres om forekomsten av profilering og konsekvensene av dette. Dersom personopplysningene samles inn fra den registrerte, bør den registrerte også informeres om hvorvidt vedkommende har plikt til å gi personopplysningene, og om konsekvensene dersom de ikke gis. Nevnte informasjon kan gis sammen med standardiserte ikoner, slik at det gis en oversikt over den tiltenkte behandlingen på en lett synlig, forståelig og lettlest måte. Dersom ikonene presenteres elektronisk, bør de være maskinlesbare.»

Artikkel 29-gruppen har uttalt seg om åpenhet i behandlingssituasjoner som faller utenfor artiklene 13, 14 og 22. I retningslinjene om åpenhet fremheves særlig viktigheten av å informere om konsekvensene av at personopplysninger behandles og at behandlingen av personopplysninger ikke skal komme som en overraskelse på de som får personopplysningene sine behandlet.[14] At pliktene til å informere om den underliggende logikken etter artikkel 13 og 14 går lenger enn det generelle åpenhetsprinsippet, som omtalt i fortalepunkt 60 støttes også av veilederen om profilering og automatiserte avgjørelser.[15] Oppsummeringsvis er det vanskelig å se at det kan utledes en rettslig plikt av forordningen til å forklare den underliggende logikken som tilsvarer kravene som følger av artikkel 13 og 14 for verktøyet i dette prosjektet. Artikkel 29-gruppen uttaler uansett i den nevnte veilederen at det er god praksis å forklare den underliggende algoritmen, selv om den behandlingsansvarlige ikke har en plikt til det.

Sandkassa anbefaler dessuten informasjon om hvordan verktøyet fra Secure Practice fungerer, fordi det kan bidra til å skape tillit til KI-verktøyet. I avsnittet nedenfor viser vi et eksempel fra en fokusgruppe med ansatte, som understreket betydningen av klar og tydelig informasjon som forutsetning for å gi korrekte personopplysninger.

---

[14] Article 29 Working Party, Guidelines on transparency under Regulation 2016/679, WP260 rev. 1, avsnitt 41. Retningslinjene er godkjent av Personvernrådet.

[15] Se videre på s. 25 i Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev0.1) Retningslinjene er godkjent av Personvernrådet.

4.4.3 Hvordan og når er det best å gi informasjon til brukerne?

Personvernforordningen regulerer ikke i detalj hvordan brukergrensesnitt skal utformes. Men i forlengelsen av pliktspørsmålet ble det også drøftet *hvordan* og *når* løsningen skal informere brukeren.

I prosjektet diskuterte vi blant annet konkrete problemstillinger knyttet til utformingen av brukergrensesnittet. Et viktig punkt var om du som ansatt bør få en forklaring på hvorfor KI-verktøyet serverer deg akkurat dette forslaget, enten du blir oppfordret til å gjennomføre en spesifikk opplæringsmodul eller ta en spesifikk quiz, og hvordan det eventuelt bør gjøres.

Et konkret eksempel kunne være at en ansatt fikk forslag om å gjennomføre en bestemt type opplæring, fordi de hadde blitt lurt av en phishing-øvelse. Slik informasjon sier noe om den underliggende logikken i algoritmen. Det ble særlig drøftet om en slik detaljert informasjon kunne gi brukeren en følelse av overvåking, som igjen kunne lede til mindre tillit. Argumentene som talte for å gi denne typen informasjon, var at de registrerte trenger den for å forstå hvordan opplysningene brukes og at forståelsen kan bygge tillit til løsningene.

I den første fokusgruppen var det stor villighet til å ta i bruk løsningen og dele data, hvis det er et konstruktivt bidrag til å oppnå formålene om bedre informasjonssikkerhet i virksomheten. Deltakerne poengterte, at det er viktig at kommunikasjonen med de ansatte er tydelig og klar. Det er viktig at det tidlig i prosessen avklares hvordan dataene skal lagres og benyttes i virksomhetens arbeide. Usikkerhet rundt hvordan dataene brukes, øker faren for at de ansatte tilpasser svarene sine til det de tror er "riktig", eller at de ikke er villige til å dele data. Dette er et interessant funn, fordi algoritmen blir mindre treffsikker hvis dataene den baserer seg på er unøyaktig og ikke representerer den reelle situasjonen brukeren er i.

I fokusgruppen med arbeidstakerorganisasjonen Negotia var det større fokus på åpenhet generelt, som en forutsetning for at arbeidstakere skal kunne stole på løsningen. Punktene som ble fremhevet var blant annet knyttet til hva arbeidsgiver har tilgang til av informasjon, hvordan kontrakten med virksomheten er utformet, viktigheten av å involvere de ansatte eller tillitsvalgte fra tidlig i prosessen, og at en slik løsning kan oppleves ulikt av arbeidstakere avhengig av situasjonen, for eksempel om de har høy eller lav tillit til arbeidsgiver. Risikoen knyttet til at svarene kunne spores tilbake til den enkelte arbeidstaker, ble også fremhevet i fokusgruppen. Denne fokusgruppen advarte mot å utforme spørsmål på en slik måte at svarene kunne skade arbeidstakerne, dersom de ble kjent for arbeidsgiveren.

4.5 Rettferdighet og forskjellsbehandling

Personvernforordningen plasserer rettferdighet blant grunnprinsippene for behandlingen av personopplysninger i artikkel 5, og blir nevnt samtidig som åpenhet med hensyn til den registrerte, og behandlingens lovlighet. Et lignende skille blir benyttet i EUs ekspertgruppes etiske retningslinjer for pålitelig kunstig intelligens[16] og gjenspeiles i Nasjonal strategi for kunstig intelligens, med hovedprinsippene *lovlig*, *etisk* og *sikker*.

Det er fortsatt lite praksis som kan belyse det konkrete innholdet i rettferdighetsprinsippet nærmere, men det foreligger noe veiledning i personvernforordningens fortalepunkter. Det er også verdt å nevne at prinsippet om en rettferdig behandling av personopplysninger er ment å være en fleksibel rettslig standard, som kan tilpasses den konkrete behandlingssituasjonen.

Den konkrete situasjonen til den registrerte må tas i betraktning, når man vurderer om behandlingen er forenelig med forordningens krav. Man må ta hensyn til hvilke rimelige forventninger den registrerte har til beskyttelsen av sine personopplysninger, samt eventuell makt-ubalanse mellom den registrerte og den behandlingsansvarlige.

[16] Du finner "Ethics guidelines for trustworthy AI" på europa.eu.

Personvernrådet fremhever i sin veiledning 4/2019[17] om innebygd personvern flere momenter som inngår i rettferdighetsprinsippet, blant annet ikke-diskriminering, den registrertes forventninger, behandlingens bredere etiske problemstillinger, og respekt for rettigheter og friheter. For å sikre at løsningen ble vurdert også i et videre rettferdighetsperspektiv, involverte derfor prosjektet Likestilling- og diskrimineringsombudet (LDO).

LDO presenterte i en egen workshop sammenhenger mellom personvernreglene og likestillings- og diskrimineringsloven. Ombudet ga også en innføring i hvordan en aktør kan vurdere om det skjer ulovlig diskriminering, etter likestillings- og diskrimineringsloven § 6- §9.

LDO framhevet følgende om Secure Practice' tjeneste:

I en tidlig fase av prosjektet lanserte Secure Practice et alternativ, der de med best score i verktøyet kunne fungere som ambassadører i virksomhetene. LDO pekte da på risikoen for at ambassadørene ville få en mer positiv karriereutvikling enn andre, dersom arbeidsgiver hadde tilgang til hver enkelt score for arbeidstakerne. LDO viste også til at dersom KI-verktøyet premierte egenskaper og interesser som for eksempel er mest vanlig hos enkelte grupper, ville det foreligge en risiko for indirekte diskriminering i modellen. Indirekte diskriminering er som hovedregel ulovlig, etter likestillings- og diskrimineringsloven § 8.

Når Secure Practice nå gjør ytterligere tiltak for å sikre at arbeidsgiverne ikke får tilgang til den enkeltes score, mener LDO det er et godt grep, som reduserer risikoen for at arbeidsgiver skal kunne benytte informasjon om arbeidstakernes kunnskap om cybersikkerhet til andre formål enn tiltenkt. Samtidig mener LDO det er viktig at Secure Practice på sikt spesifiserer hvilke demografiske data som skal samles inn, hvordan disse dataene vil bli brukt, samt begrunnelsen for og sakligheten av den aktuelle databruken.

LDO oppfordrer Secure Practice til å påse at opplæringstilbudet og kurstilpasningen fungerer like godt for alle grupper; kvinner og menn, yngre og eldre, og personer med funksjonsnedsettelser osv. For å oppnå dette bør Secure Practice unngå å spille på stereotype forestillinger om ulike grupper, når opplæringen skal tilpasses de ulike arbeidstakerne. Å spille på stereotypier er ikke nødvendigvis diskriminering, men kan bidra til å bygge opp under tradisjonelle forestillinger om personer tilhørende bestemte grupper. Slike forestillinger kan være lite treffsikre, noe som vil redusere verdien av opplæringstilpasningen. Av øvrige tiltak oppfordres Secure Practice til å periodevis teste og eventuelt justere egen tjeneste, for å påse at tilpasningsfunksjonene fungerer like godt for alle brukere.

Under disse forutsetningene mener LDO at den forskjellsbehandlingen de ansatte utsettes for ved at opplæringen tilpasses den enkeltes kompetansenivå, er en saklig form for forskjellsbehandling, jf. likestillings- og diskrimineringsloven §§ 6 og 9.

17 European Data Protection Board Guidelines 4/2019 on Article 25 Data Protection by Design and by Default

5 Veien videre

I sandkassa har Secure Practice fått utforsket spørsmål de har hatt i en produktutviklingsprosess, hvor innebygd personvern er avgjørende. Ved å ta initiativ til dette prosjektet, og være i forkant, kan resultatene brukes av Secure Practice til å sikre at den nye tjenesten de lanserer gir godt personvern fra starten.

Denne sluttrapporten er viktig for å belyse forbudet mot overvåking i e-postforskriften. Datatilsynet vil dele erfaringer fra prosjektet med Arbeids- og inkluderingsdepartementet, som er ansvarlig for forskriften, og for å gi departementet en mulighet til å klargjøre reglene.

Datatilsynet arbeider også med å oppdatere informasjonen om personvern i arbeidslivet på nettsidene. Mer enn hver fjerde henvendelse til Datatilsynets veiledningstjeneste og et stort antall klager gjelder dette temaet. Bare i 2021 kom det 1677 spørsmål om personvern i arbeidslivet, så dette er et viktig tema å spre kunnskap om.

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