Risikokatalog over risici vedrørende databeskyttelse i AI-løsninger

Til brug for udarbejdelse af konsekvensanalyse vedrørende databeskyttelse

1. Formål og anvendelsesområde

Dette risikokatalog indeholder et overblik over en række typiske risici, der er forbundet med behandling af personoplysninger i systemer baseret på kunstig intelligens (”AI-løsninger”). Der er tale om risici for de registreredes rettigheder samt frihedsrettigheder og ikke for kommunen, f.eks. tab af omdømme eller bøder m.v. Risikokataloget indeholder også relevante forslag til tekniske og organisatoriske foranstaltninger, som kommunen kan implementere for at håndtere de pågældende risici. Risikokataloget kan anvendes til identifikation og mitigering af risici i hele AI-løsningens livscyklus, herunder udvikling og drift. Risikokataloget kan anvendes både ved egenudvikling og anskaffelse af standard-løsninger.

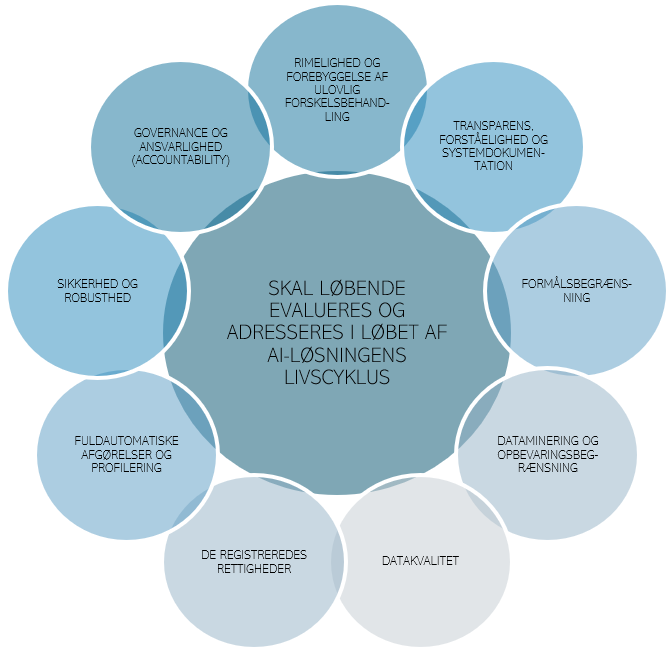
Formålet med risikokataloget er at give vejledning til kommunen, der skal lave en konsekvensanalyse vedrørende databeskyttelse i forbindelse med behandling af personoplysninger i en AI-løsning, jf. databeskyttelsesforordningens artikel 35. Kataloget kan bruges som inspiration til at identificere typiske risici vedrørende databeskyttelse i AI-løsninger og give forslag til tilhørende tekniske og organisatoriske afhjælpende foranstaltninger til at håndtere disse risici, så disse kan dokumenteres i konsekvensanalysen og der kan opnås et acceptabelt risikoniveau, inden AI-løsningen idriftsættes.

Risikokataloget har særligt fokus på typiske risici i AI-løsninger, der involverer beslutningstagen – enten i form af fuldautomatiske afgørelser eller beslutningsstøtte. Men kataloget kan også anvendes til andre former for brug af AI, f.eks. til modeller til svigbekæmpelse eller risikovurdering af borgere/kunder samt profilering til brug for målrettet vejledning af borger, kunder og markedsføring m.v.

Risiciene og de tilhørende afhjælpende foranstaltninger er ikke udtømmende, og valg af risici og foranstaltninger skal tilpasses til de konkrete omstændigheder. Ikke alle risici og afhjælpende foranstaltninger er relevante i alle projekter, systemer m.v., og valget heraf vil bero på den konkrete konsekvensanalyse samt den specifikke teknologi, der er tale om, samt organisationens kompetencer, modenhed m.v. Risikokataloget er således til inspiration og vejledning – ikke en facitliste.

Kataloget indeholder afslutningsvist en ikke-udtømmende oversigt over mulige databeskyttelsesretlige skadevirkninger, som de oplistede risici kan medføre for de registrerede. Det er således vigtigt, at kommunen i konsekvensanalysen vedrørende databeskyttelse for hver enkelt identificeret risiko vurderer, hvilken konkret skadevirkning det vil have for de registrerede, hvis den pågældende risiko realiseres.

1. Risikokatalogets risikotemaer

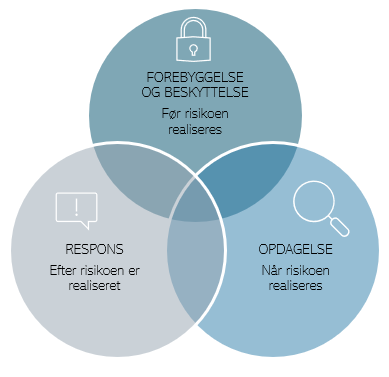


Figur 1. Risikokatalogets risikotemaer

En risiko defineres i dette risikokatalog som et scenarie, der beskriver en hændelse og konsekvenserne heraf, som vurderes i forhold til alvoren for de registrerede og sandsynlighed. Sådanne risici omfatter også risikoscenarier, der kan indebære, at der sker ulovlig behandling af personoplysninger. Ved sådanne risikoscenarier forstås mulige situationer, der kan opstå utilsigtet, og som indebærer en afvigelse af den tiltænkte, lovlige behandlingsaktivitet. Det gælder f.eks. utilsigtet indsamling af flere oplysninger end nødvendigt i lyset af formålet, utilsigtet videregivelse af personoplysninger, som kommunen ikke har hjemmel til at videregive, eller en utilsigtet manglende sletning af oplysninger, når kommunen ikke længere har behov for oplysningerne.

Risikokataloget er overordnet opdelt i følgende risikotemaer vedrørende databeskyttelse og AI:

1. Risikokatalogets forslag til afhjælpende foranstaltninger/kontroller



Figur 2. Fokus for afhjælpende foranstaltninger/kontroller

For hver enkelt risiko gives der i risikokataloget forslag til, hvordan man kan håndtere denne risiko, dvs. hvilke foranstaltninger eller kontroller, man med fordel kan overveje at implementere for at håndtere risikoen.

Der sondres overordnet mellem organisatoriske og tekniske foranstaltninger. De organisatoriske foranstaltninger omfatter f.eks. undervisning af medarbejdere og awarenessaktiviteter, udarbejdelse af kontroller, revision og audits, udarbejdelse af interne retningslinjer og procedurer vedrørende ansvarsfordeling og ledelsesgodkendelse m.v. De tekniske foranstaltninger vedrører f.eks. logning, pseudonymisering, tests og forskellige tekniske metoder til at gøre beslutninger baseret på AI transparente og forklarlige (såkaldt ”Explainable AI” eller ”XAI”).

Foranstaltningerne kan have et forskelligt fokus, alt efter om de tager sigte på (1) forebyggelse og beskyttelse af, at risikoen ikke realiseres, (2) opdagelse af, at risikoen er blevet realiseret eller (3) respons, dvs. handling når det er et konstateret, at risikoen er realiseret, f.eks. at man har opdaget en fejl i algoritmen, jf. figur 2.

Ved udarbejdelsen af dette risikokatalog er der bl.a. inddraget de risici og eksempler på foranstaltninger, som indgår i Datatilsynets skabelon til en konsekvensanalyse vedrørende databeskyttelse ved brug af AI, som blev offentliggjort af Datatilsynet den 22. maj 2024. Disse er markeret med blåt i tabellen nedenfor.

1. Skema med risici fordelt på risikotemaer samt relevante afhjælpende foranstaltninger/kontroller

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nr. | Risiko | Forklaring af risikoen | Retsgrundlag | Relevante afhjælpende foranstaltninger/kontroller | |
| **Type** | **Beskrivelse og fokus for tiltag** |
| 1. Rimelighed og forebyggelse af ulovlig forskelsbehandling   AI-løsningens ukorrekte output eller afgørelser fører til ulovlig forskelsbehandling af de registrerede (se også punkt 5 vedrørende datakvalitet). | | | | | |
|  | Risiko for manglende iagttagelse af princippet om rimelighed i afdækningsfasen (forud for udvikling og idriftsættelse) og i udviklingsfasen | Hvis princippet om rimelighed ikke iagttages tilstrækkeligt i afdæknings- og udviklingsfaserne er der risiko for, at AI-løsningen kan resultere i urimelige udfald med manglende eller mangelfuld fastsættelse af passende og tilstrækkelige foranstaltninger. | Artikel 5, stk. 1, litra a, og præambelbetragtning nr. 39, 40 samt 58 | Organisatoriske og tekniske foranstaltninger | * Der skal foretages vurdering og dokumentation af de forskellige måder, hvorpå AI-løsningen kan resultere i urimelige udfald, samt hvilke passende tekniske og organisatoriske foranstaltninger, der implementeres med henblik på kontinuerligt at mitigere disse risici. * Der skal generelt foretages vurdering og dokumentation af, hvilke data der er nødvendige for at sikre et repræsentativt, pålideligt og relevant træningsdatasæt. * Datasæt, som skal bruges til udvikling, skal renses og justeres for at sikre, at data er repræsentative, pålidelige og relevante. * Der bør indregnes risiko for tidligere diskrimination. * Det bør overvejes, om der skal indsamles specifikke datapunkter til brug for bias-analyse. * Der skal generelt foretages vurdering og dokumentation af, hvilke datakilder eller features, som ikke bør behandles på grund af risikoen for direkte eller indirekte diskrimination. Samtidig bør det vurderes, om visse af disse data eller features desuagtet skal indsamles til brug for bias analyse. * Det bør overvejes, om der er behov for at behandle yderligere data for at gennemføre bias-analyse, og om der er behov for at lave labels for data, som allerede foreligger, eller om der skal indsamles yderligere data. * Det bør overvejes, om visse data skal indsamles eller features skal indgå som led i bias-analysen. * Hvis det er nødvendigt at behandle andre kategorier af persondata til brug for bias-analysen, end de typer, der allerede indgår i trænings- og testdata, skal der identificeres et behandlingsgrundlag for disse data. * Der bør fastsættes kriterier for mærkning, som er lette at forstå, omfatter beskrivelser af og eksempler for alle mærker. Kriterierne bør også omfatte "gråzone"-tilfælde. Generelt skal der foretages fastlæggelse og dokumentation af klare kriterier og procedurer for mærkning af træningsdata. * Der bør udarbejdes vejledninger for mærkning og annotering. * Det bør sikres, at medarbejderne med ansvaret for mærkning og annotering er tilstrækkeligt kvalificerede, og at sammensætningen af medarbejderne er tilstrækkeligt mangfoldig. * Relevante interessenter eller interesseorganisationer kan inddrages med henblik på at definere mærkningskriterierne. * Processen for mærkning og annotering bør indeholde procedurer for kvalitetssikring, herunder af subjektive og objektive mærker. Ved subjektive mærker forstås tilfælde, hvor der ikke findes et korrekt mærke, og mærket tildeles skønsmæssigt af medarbejderen. Ved objektive mærker forstås tilfælde, hvor der findes et korrekt mærke, men medarbejderen er ikke bekendt med, hvordan det anvendes. Processen bør også omfatte, at flere medarbejdere mærker samme datapunkter uafhængigt af hinanden med henblik på sammenligning. * Der kan udarbejdes statistikker for, hvor ofte medarbejderne, der forestår mærkningen, er enige. * Hvis der benyttes automatisk mærkning, bør der fastsætte procedurer for kvalitetssikring heraf. * Der skal foretages vurdering og dokumentation af, i hvilket omfang modellen er tilbøjelig til *overfitting* * Features, som sandsynligvis kan medføre *overfitting*, skal fjernes. * Nøglemetrikker for modellens ydeevne bør overvejes med henblik på at identificere mulige årsager til *overfitting*. * Der kan indsamles yderligere data for at sikre, at træningsdata er repræsentative for den population, som AI-løsningen skal bruges over for. Dette skal imidlertid afvejes over for de registreredes ret til ikke at blive udsat for overflødig, ulovlig eller urimelig behandling af personoplysninger. |
|  | Risiko for, at princippet om rimelighed ikke iagttages tilstrækkeligt i testfasen | Hvis princippet om rimelighed ikke iagttages tilstrækkeligt i testfasen er der risiko for, at AI-løsningens output eller afgørelser fører til ulovlig forskelsbehandling eller lignende af de registrerede. | Artikel 5, stk. 1, litra a, og præambelbetragtning nr. 39, 58 samt 60 | Organisatoriske og tekniske foranstaltninger | * Det bør vurderes og begrundes, om der skal foretages (i) indsamling af flere data for at reducere et disproportionalt antal statistiske fejl eller (ii) ikke foretages indsamling af sådanne data grundet risiciene, som det kan indebære for de registreredes øvrige rettigheder og frihedsrettigheder. * Det kan overvejes at bruge *feature engineering[[1]](#footnote-1)*-teknikker for at aggregere features, der bruges til træning, og for at undgå at bruge personoplysninger. * Generelt skal der udføres test af, om AI-løsningen genererer samme output for forskellige personer med samme karakteristika. * Der bør fastlægges procedurer for overvågning af AI-løsningen og dens output. * De forskellige typer af fejl (f.eks. falske positiver, falske negativer mv.) bør måles. * Det bør overvejes, om datasættet, der bruges til test, er tilstrækkeligt * AI-løsningen skal testes med forskellige datasæt for at sikre, at løsningen genererer ens udfald for forskellige registrerede med samme karakteristika. Hvis det ikke er tilfældet, skal løsningen gentrænes. * Det kan overvejes at dokumentere AI-løsningens begrænsninger i et *modelkort[[2]](#footnote-2).* |
|  | Risiko for, at princippet om rimelighed ikke iagttages tilstrækkeligt i drift- og monitoreringsfasen | Hvis princippet om rimelighed ikke iagttages tilstrækkeligt i drifts- og/eller monitoreringsfasen er der risiko for, at AI-løsningens output eller afgørelser fører til ulovlig forskelsbehandling eller lignende af de registrerede. | Artikel 5, stk. 1, litra a, og præambelbetragtning nr. 39, 58 samt 60 | Organisatoriske foranstaltninger | * Der skal fastlægges procedurer for overvågning og regelmæssig test af AI-løsningen. * Procedurerne for overvågning og test bør indeholde metrikker og tærskelværdier, som udløser review og test. * Tidligere versioner af modellen, som kan reverteres tilbage til i tilfælde af væsentlig modeldrift, bør gemmes. * Der kan antages en uafhængig tredjemand til at gennemføre regelmæssige algoritmetilsyn. |
|  | Risiko for, at anvendelsen af systemet vil medføre ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias). | Hvis systemets datasæt (både udvikling og drift) er præget af historisk (utilsigtet) skævhed, vil systemets beslutningsprocesser med høj sandsynlighed også medføre utilsigtet direkte eller indirekte ulovlig forskelsbehandling. Selvom rådata (inputdata) registreres korrekt, kan datasættet muligvis ikke være fuldt repræsentativt, eller analysen af dataene kan indeholde skjulte skævheder. Ulovlig forskelsbehandling kan også opstå som følge af en forkert udvikling og programmering af algoritmer. | Artikel 5, stk. 1, litra a | Organisatoriske foranstaltninger | * Gennemgående træning af AI-systemet med forskellige træningsdata, inden systemet anvendes i praksis. (Forebyggelse og beskyttelse) * Sikring af repræsentative træningsdata. Test for brug af modellen på specifikke populationer eller udsatte grupper. (Forebyggelse og beskyttelse og Opdagelse) * Uddannelse af AI-udviklere og nøglepersoner, så de kan identificere og håndtere ulovlig forskelsbehandling i AI-systemer. (Forebyggelse og beskyttelse) * Overvejelse og beskrivelse af, hvordan ulovlig forskelsbehandling kan ske på det pågældende område, eventuelt med en definition af ”skævhed” og ”retfærdighed”, som anvendes under udvikling og anvendelse af systemet. Etablering af klare og målbare kriterier til at måle og teste de anvendte definitioner. (Forebyggelse og beskyttelse) * Regelmæssige kvalitetssikringskontroller og tests af AI-systemet for at sikre, at de registrerede behandles rimeligt og ikke forskelsbehandles, enten på grundlag af følsomme personoplysninger eller andet grundlag. (Forebyggelse og beskyttelse) * Indførelse af test og procedurer for overvågning af potentielle skævheder under udvikling og anvendelse af systemet. (Forebyggelse og beskyttelse og Opdagelse) * Hvis træningsdataene afspejler tidligere ulovlig forskelsbehandling, kan man udskifte træningsdataene og træne modellen på ny. (Respons) |
|  | Risiko for manglende hensyntagen til målgruppen ved behandling af personoplysninger om sårbare registrerede, herunder børn og handikappede m.v. | Når kommunen behandler personoplysninger om sårbare registrerede, herunder børn og handikappede m.v., vil der på grund af den øgede skævhed i magtfordelingen mellem den registrerede og kommunen være øget risiko for, at enkeltpersoner kan være ude af stand til på en nem måde at overskue eller modsætte sig behandlingen af deres oplysninger eller udøve deres rettigheder. | Artikel 5 og artikel 12-14 | Organisatoriske foranstaltninger | * I forbindelse med opfyldelse af oplysningspligten eller ved anden form for kommunikation har kommunen særligt fokus på, at oplysninger om behandlingen til sårbare registrerede skal være fyldestgørende, letforståelige og lettilgængelige under hensyntagen til denne gruppe. (Forebyggelse og beskyttelse) |
| 1. Transparens, forståelighed og systemtransparens   Det er ikke muligt fuldt ud at forklare AI-systemets output, dvs. hvordan systemet er nået frem til en afgørelse eller forslag hertil – beslutningsprocessen eller logikken i systemet er en ”black box”. | | | | | |
|  | Risiko for manglende opfyldelse af princippet om gennemsigtighed i udviklingsfasen | Risiko for manglende iagttagelse af princippet om gennemsigtighed kan medføre negative konsekvenser for de registrerede. | Artikel 5, stk. 1, litra a, og 12-15 samt præambelbetragtning nr. 60-62 | Organisatoriske foranstaltninger | * Der skal generelt foretages vurdering og dokumentation af, i hvilket omfang den information, der gives til de registrerede, er forståelig og lettilgængelig. * De registrerede skal oplyses om behandlingen på en måde, som er lettilgængelig og letforståelig. Der skal bruges klart og enkelt sprog. * Det bør undersøges, i hvilket omfang den information, som de registrerede skal modtage, gives på en lettilgængelig og letforståelig måde. Det kan f.eks. ske gennem målgruppeundersøgelser eller lignende. |
|  | Risiko for manglende opfyldelse af princippet om gennemsigtighed i testfasen | Risiko for manglende iagttagelse af princippet om gennemsigtighed kan medføre negative konsekvenser for de registrerede. | Artikel 5, stk. 1, litra a, og 12-15 samt præambelbetragtning nr. 60-62 | Organisatoriske foranstaltninger | * Der skal generelt foretages vurdering og dokumentation af, hvor forklarlig AI-løsningen er, samt overvejes. hvilke supplerende foranstaltninger eller værktøjer der kan bruges til at understøtte forklaring af beslutninger truffet af AI-løsningen til de personer, der bliver påvirket af beslutningen. Der skal endvidere defineres og dokumenteres såvel en klageadgang som en adgang til at give feedback, der giver berørte personer mulighed for at give feedback om de forklaringer, som de modtager. * Det bør testes, hvor effektive forklaringerne er, ved at måle hvor godt de registrerede forstår, hvorfor modellen kom frem til den pågældende beslutning eller afgørelse, eller hvordan modellens output bidrog til beslutningen. * Der bør designes forklaringer, som passer til behovene hos dem, der interagerer med AI-løsningen på forskellige tidspunkter. Det kan f.eks. være, hvor nogen interagerer med løsningen for første gange i modsætning til nogen, som er regelmæssige brugere. * Det kan overvejes, hvornår der er behov for at forklare automatiske resultater til de forskellige grupper, som interagerer med AI-løsningen, og hvorvidt der findes en let måde at anfægte disse resultater og opnå menneskelig indgriben. * Uddybende forklaringer, visualiseringer, grafiske fremstillinger eller kombinationer heraf bør overvejes som del af forklaringen. * Der kan proaktivt rækkes ud til de registrerede for at vurdere, hvordan de forklaringer, som afgives, kan forbedres. |
|  | Risiko for manglende forklarlighed og transparens i forhold til systemets beslutningsprocesser og datagrundlag. | Det er et krav, at kommunen skal kunne forklare, hvordan systemet når frem til en beslutning (output), dvs. mellemregningerne eller logikken i systemet samt da-tagrundlaget. Både de registrerede, der er adressat for en beslutning, medarbejdere hos kommunen selv samt tilsynsmyndigheder m.v. skal efter omstændighederne kunne få indsigt heri.  Hvis kommunen ikke er i stand til at forklare, hvordan systemet virker og kvaliteten af systemets output samt rationalet bag dette, skaber det risiko for, at fejl ved systemet ikke opdages eller at outputtet ikke kan kommunikeres til de registrerede og andre interessenter. | Artikel 5, stk. 1, litra a, artikel 13-15 og artikel 22 | Organisatoriske foranstaltninger | * Sikring af, at modellen kan forklares, herunder hvordan den når frem til (forslag til) beslutninger, logikken og de kriterier/data den lægger vægt på. Det skal sikres, at et menneske (1) kan forudsige, hvordan systemets output vil ændre sig, hvis input ændrer sig, (2) kan identificere de mest vigtige inputvariable og deres vægt og betydning for afgørelsens resultat, (3) identificere, hvornår outputtet sandsynligvis er forkert. (Forebyggelse og beskyttelse) * Dokumentation af datasættene og de processer, der fører til AI-systemets beslutninger, herunder træningsdata, testdata og algoritmer. Hvis der er tale om et regelbaseret system, bør programmeringsmetoden eller den måde, systemet er bygget på, dokumenteres. Hvis der er tale om et læringsbaseret AI-system, bør der ske dokumentation af metoden til oplæring af algoritmen, herunder hvilke inputdata der er indsamlet og udvalgt, og hvordan dette er sket. (Forebyggelse og beskyttelse) * Dokumentation af testmetoder og resultater af disse. (Forebyggelse og beskyttelse) * Sikring af, at enkeltpersoner, der er adressat for en fuldautomatisk afgørelse, kan få en individuel forklaring på grundlaget for og logikken bag afgørelsen og de forventede konsekvenser af en sådan behandling. (Forebyggelse og beskyttelse) |
| Tekniske foranstaltninger | * AI-systemet designes, så det understøtter transparens og ikke udgør en uforklarlig "black box”. Forklarlighed udgør et grundprincip for valg af systemets matematiske model (algoritme), f.eks. ved brug af beslutningstræer i stedet for et neuralt netværk etc. (Forebyggelse og beskyttelse) * Brug af værktøjer og teknologier i relevant omfang til at understøtte forklarlighed (Explainable AI). (Forebyggelse og beskyttelse) |
|  | Risiko for, at registrerede ikke kender til AI-systemets karakter, kompetencer og begrænsninger. | Det udgør en risiko for brugerne af systemet, hvis de ikke har den information, det kræves for at forstå og interagere med AI-systemet i en tilfredsstillende grad, herunder vurdere systemets begrænsninger. | Artikel 5, stk. 1, litra a | Tekniske foranstaltninger | * Klar kommunikation af AI-systemets kapaciteter og begrænsninger til slutbrugerne. Dette kan omfatte en beskrivelse af AI-systemets nøjagtighed og dets begrænsninger. (Forebyggelse og beskyttelse) |
| 1. Formålsbegrænsning   Personoplysninger viderebehandles til modeludvikling eller anvendelse af systemet på en måde, der er uforenelig med de formål, personoplysningerne blev indsamlet til, i strid med princippet om formålsbegrænsning | | | | | |
|  | Risiko for, at personoplysninger viderebehandles til modeludvikling eller anvendelse af systemet i strid med principperne om formålsbegrænsning og lovlighed. | Der er typisk videre rammer for at bruge personoplysninger indsamlet til forskellige formål til udvikling af en AI-model, når grundlaget for behandlingen er statistik. Omvendt er der snævrere rammer for brug af personoplysninger i nye sammenhænge, når personoplysningerne skal anvendes til brug for at træffe afgørelser overfor de registrerede. Der er derfor risiko for, at man i strid med formålsbestemthedsprincippet behandler personoplysninger ved anvendelsen af den udviklede AI-model. | Artikel 5, stk. 1, litra b, artikel 6, stk. 1 og 4, og 30 | Organisatoriske foranstaltninger | * Klar opdeling af behandling af personoplysninger i *udvikling og anvendelse* af modellen/systemet. (Forebyggelse og beskyttelse) * Dokumenterbare retningslinjer og procedurer for overholdelse af formålsbegrænsningsprincippet. (Forebyggelse og beskyttelse) * Udarbejdelse af dokumentation af (i) hvorfra data hidrører, (ii) hvert formål med behandling af persondata på hvert trin af AI-livscyklussen, (iii) vurdering om formålene er forenelige med det oprindelige formål og (iv) fastsættelse af intervaller for review af, om formålene forbliver forenelige. * Etablering af procedurer for data proveniens og data governance. Som led heri bør der gennemføres en analyse af datastrømme, og det skal afdækkes, hvilke oplysninger der tilgår, bliver behandlet i, og udgår fra AI-løsningen, med henblik på at sikre, at behandlingen er lovlig og forenelig med det oprindelige formål. |
|  | Risiko for viderebehandling af personoplysninger til andre formål end forskning og statistik (utilstrækkelig formålsbestemthed). | Når personoplysninger bliver indsamlet udelukkende med henblik på forskning og statistik efter de særlige bestemmelser herom i databeskyttelseslovens § 10, må oplysningerne ikke efterfølgende anvendes – herunder videregives – til andre formål end forskning og statistik, medmindre resultaterne er anonymiserede i sådan en grad, at de ikke kan føres tilbage til enkelte individer. | Artikel 5, stk. 1, litra b, og data-beskyttelseslovens § 10, stk. 2 | Organisatoriske foranstaltninger | * Implementering af effektiv data governance og skarp opdeling af data i administrative og statistiske/forskningsmæssige formål. (Forebyggelse og beskyttelse) * Udarbejdelse af retningslinjer og procedurer for adgangskontrol og at personoplysninger alene må anvendes til forskningsmæssige og statistiske formål, og at alle resultater skal anonymiseres effektivt. (Forebyggelse og beskyttelse) * Overholdelse af kravene til videregivelse i bekendtgørelse nr. 1509 af 18. december 2019 om videregivelse af personoplysninger omfattet af databeskyttelseslovens § 10, stk. 1, og 2. (Forebyggelse og beskyttelse) |
|  | Risiko for uberettiget videregivelse til tredjemand til brug for denne tredjemands egne formål. | Anvendelse af AI-løsninger som Software as a service (SaaS) kan medføre, at der videregives personoplysninger til leverandøren af løsningen til leverandørens egne formål – f.eks. markedsføring, profilering eller træning/forbedring af løsningen – i strid med formålsbegrænsningsprincippet og uden hjemmel hertil. | Artikel 5, stk. 1, litra b, samt artikel 6-10. | Organisatoriske foranstaltninger | * Grundigt review af leverandørens aftalevilkår med henblik på at konstatere, om leverandøren forbeholder sig ret til at anvende personoplysninger til egne formål. * Overvejelser om konfiguration af løsningen, så sådanne behandlinger til leverandørens egne formål undgås eller begrænses til det lovlige. |
| 1. Dataminimering og opbevaringsbegrænsning   AI-systemet udvikles, uden at der sker en tilstrækkelig vurdering af, om behandlingen af personoplysningerne til udvikling eller anvendelse af systemet er begrænset til det nødvendige, relevante og tilstrækkelige. Personoplysningerne er desuden opbevaret i et længere tidsrum end nødvendigt i lyset af behandlingens formål. Dette medfører risiko for overtrædelse af princippet om dataminimering og princippet om opbevaringsbegrænsning | | | | | |
|  | Risiko for manglende iagttagelse af princippet om dataminimering i afdækningsfasen (forud for udvikling og idriftsættelse) | Hvis princippet om manglende dataminimering ikke iagttages tilstrækkeligt i afdækningsfasen er der risiko for, at den i behandlingsaktiviteten omfattede data ikke er rigtig, tilstrækkelig, relevant og/eller begrænset til det nødvendige. | Artikel 5, stk. 1, litra c, samt betragtning nr. 39 | Organisatoriske foranstaltninger | * Det bør vurderes, hvilke privatlivsfremmende teknologier, der kan være passende for den pågældende AI-løsning for at sikre dataminimering. * Domæneeksperter bør inddrages for at sikre, at persondata, som forventes at skulle indsamles, er passende og tilstrækkelige. |
|  | Risiko for behandling af et unødvendigt omfang af personoplysninger til udvikling og anvendelse af systemet i strid med dataminimeringsprincippet. | Hvis AI-systemet udvikles, uden at der sker en tilstrækkelig vurdering af, om behandlingen af personoplysningerne til udvikling eller anvendelse af systemet er begrænset til det nødvendige, relevante og tilstrækkelige, medfører dette en risiko for overtrædelse af princippet om dataminimering. | Artikel 5, stk. 1, litra c | Organisatoriske foranstaltninger | * Overvejelse og dokumentation af, om brug af fiktive, anonymiserede og aggregerede oplysninger er tilstrækkelig til oplæring, modeludvikling og test. (Forebyggelse og beskyttelse) * Overvejelse og dokumentation af, om hver enkelt oplysningstype er relevant i hver fase af modeludviklingen, herunder ved inddragelse af personer med domænekendskab. (Forebyggelse og beskyttelse) * Brug af læringskurver til at dokumentere, hvornår yderligere kriterier ikke forbedrer modellens prædiktionsevne og præcision. (Forebyggelse og beskyttelse) * Brug af pseudonymisering af personoplysninger, herunder klare retningslinjer for, hvornår pseudonymisering er et krav. (Forebyggelse og beskyttelse) * Udarbejdelse og implementering af dokumenterede retningslinjer og procedurer for regelmæssig periodisk evaluering af proportionaliteten i databehandlingen med henblik på at eliminere datakilder, der ikke længere er nødvendige i lyset af formålet. (Forebyggelse og beskyttelse) * Risikoen forsøges reduceret ved at udarbejde og implementere retningslinjer i projektet for regelmæssig periodisk evaluering af proportionaliteten i databehandlingen med henblik på at eliminere datakilder, der ikke længere er nødvendige i lyset af formålet. * Brug af en uafhængig audit af modellen med henblik på at konstatere, om kriterierne (fortsat) er relevante. Brug af test med andre systemer med færre kriterier for at konstatere, om samme resultater kan opnås med brug af færre kriterier. (Opdagelse) * Sletning af kriterier, der ikke er relevante. (Respons) * Sletning af træningsdata, der ikke længere er relevante. (Respons) * Brug af en model med færre kriterier og grundig dokumentation af begrundelsen herfor. (Respons) * Det skal overvejes, om data er blevet duplikeret eller kopieret under trænings- og testfasen. * Trade-off mellem dataminimering og statistisk nøjagtighed bør overvejes, samt hvorvidt der kan fjernes datapunkter, uden at det væsentligt påvirker nøjagtigheden af modellen. |
|  | Risiko for, at personoplysninger opbevares i længere tid end formålet tilsiger (dataophobning). | Hvis kommunen indsamler personoplysninger til et bestemt formål, men fortsætter med at opbevare oplysningerne efter formålets ophør, udgør dette en unødvendig risiko for krænkelse af den registrerede. Unødig dataophobning kan også forekomme pga. generering af nye, særskilte trænings-, validerings- eller testdatasæt. | Artikel 5, stk. 1, litra c og e | Organisatoriske foranstaltninger | * Udarbejdelse af slettepolitik for personoplysninger og dokumentation for overholdelse heraf. (Forebyggelse og beskyttelse) * Sletning af trænings- og testdata, der ikke længere er relevante. (Respons) * Gennemførelse af opfølgning på, at sletning faktisk sker i overensstemmelse med de fastsatte slettefrister. (Forebyggelse og beskyttelse og Opdagelse) |
| Tekniske foranstaltninger | * Implementering af automatiske sletningsprocedurer, der aktiveres ved formålets ophør. (Forebyggelse og beskyttelse og Respons) * Oprettelse af ”tombstones” eller sletteflag, så oplysninger slettes ved batchkørsler på foruddefinerede tidspunkter (hvis muligt). (Respons) |
|  | Risiko for uforholdsmæssig behandling i forbindelse med samkøring i kontroløjemed eller profilering. | Da samkøring i kontroløjemed per definition udgør en indgribende behandlingssituation, skærper en sådan behandling opmærksomheden på kravene i bl.a. andet forordningens artikel 5 om principper for behandling af personoplysninger, herunder dataminimerings-princippet. Tilsvarende gælder efter omstændighederne profilering. Det skal sikres, at der ikke samkøres eller behandles flere oplysninger end nødvendigt i lyset af formålet. | Artikel 5, stk. 1, litra c | Organisatoriske foranstaltninger | * Fokus på og dokumentation af proportionalitetsafvejning, herunder navnlig i forhold til (1) om der skal foretages samkøring i kontroløjemed/profilering, (2) hvor mange oplysningstyper, der skal indgå i samkøringen/profileringen, samt (3) i hvilket omfang samkøringen/profileringen er forventelig for de registrerede. (Forebyggelse og beskyttelse) * Se om profilering tillige nedenfor i punkt 7. |
| 1. Datakvalitet   Fejl i modellen eller i inputdata fører til forkerte beslutninger (output) (se også punkt 1 om rimelighed og forebyggelse af ulovlig forskelsbehandling). | | | | | |
|  | Risiko for fejl i modellen (algoritmen) i AI-løsningen fører til forkerte afgørelser m.v. | Hvis den udviklede AI-model er fejlbehæftet, vil resultaterne fra modellen (output) ligeledes være fejlbehæftet. Dette medfører, at sagsbehandlernes beslutningsgrundlag er ukorrekt, og der vil således være en risiko for, at den registrerede får en forkert afgørelse, rådgivning, service eller ydelse m.v.  Hvis AI-modellen f.eks. trænes til at forudsige et bestemt resultat, kan der være risiko for, at modellen ikke løbende trænes eller opdateres til at tage hensyn til kriterier med høj signifikans, som endnu ikke er kortlagt. Resultatet vil være, at modellen ikke har lige så stor forklaringskraft og prædiktionsevne, som den kunne have haft. | Artikel 5, stk. 1, litra a og d, og artikel 22, stk. 1 | Organisatoriske foranstaltninger | * Udarbejdelse af klar dokumentation og begrundelse for formålet med modellen, valg af model og de kriterier og deres vægtning, som indgår i modellen. (Forebyggelse og beskyttelse) * Beskrivelse af modellens begrænsninger i lyset af formålet, så alle relevante interessenter er opmærksom på, under hvilke omstændigheder den bør bruges. (Forebyggelse og beskyttelse) * Sikring af, at denne modelspecifikation godkendes af den relevante ledelse. (Forebyggelse og beskyttelse) * Fastsættelse af målbare kriterier og terminologi for modellens performance og nøjagtighed, som den skal leve op til og evaluering på, om det sker tilfredsstillende. Det skal fremgå, hvordan nøjagtighed måles og kontrolleres. (Forebyggelse og beskyttelse) * Udarbejdelse af beskrivelse og vurdering af mulige fejl og deres konsekvenser for de registrerede. (Forebyggelse og beskyttelse) * Test og revision af modellen og dens præcision og data før go live samt med et passende interval efter go live eller ved ændringer. Gennemførte tests, resultater heraf samt foranstaltninger truffet i medfør heraf skal kunne dokumenteres. (Forebyggelse og beskyttelse) * Gentræning/-kalibrering af modellen eller valg af en anden og bedre egnet model og grundig dokumentation for ændringen. (Respons) * Hvis der skal foretages centrale afvejninger – f.eks. mellem datakvalitet og dataminimering – skal ledelsen inddrages. (Forebyggelse og beskyttelse) * Træning og gentræning af AI-udviklere for at sikre, at de er opdateret på seneste viden. (Forebyggelse og beskyttelse og Respons) * Overvejelse af, om konstaterede fejl giver anledning til genoptagelse af sager. (Respons) * Der bør gennemføres tests ved brug af nye datasæt for at sikre, at AI-løsningen fortsat er retvisende. * Dokumenterbare retningslinjer og processer for kravstillelse og due diligence for eksterne leverandører af AI-løsninger. Krav om, at løsningen skal overholde på forhånd fastsatte krav om statistisk sikkerhed og krav om regelmæssig testning. (Forebyggelse og beskyttelse) |
|  | Risiko for behandling af ukorrekte, utilstrækkelige og ikke ajourførte data fører til forkerte afgørelser (utilstrækkelig rigtighed). | Ukorrekte eller ikke ajourførte inputdata kan føre til, at AI-løsningens resultater er ukorrekte og at den registrerede dermed får en forkert rådgivning, service eller ydelse m.v. (”Garbage in – garbage out”). Det gælder både ved træning og anvendelse af modellen. Tilsvarende gælder ukorrekte, utilstrækkelige eller ikke ajourførte udledte data. | Artikel 5, stk. 1, litra d | Organisatoriske foranstaltninger | * Etablering af en beskrivelse af håndtering af forskellige datatyper, herunder en beskrivelse af, hvordan alle personoplysninger, der anvendes til træning, test og anvendelse af AI-systemet, er så korrekte, relevante, repræsentative, komplette og ajourførte som muligt. (Forebyggelse og beskyttelse). * Indførelse af foranstaltninger med det formål at sikre, at de anvendte data er fuldstændige, ajourførte og korrekte, samt om der er behov for yderligere data, f.eks. for at forbedre nøjagtigheden eller eliminere fejl. (Forebyggelse og beskyttelse) * Klar information til de registrerede om de personoplysninger, der behandles, og hvordan AI-systemet fungerer, således at de registrerede kan rette eventuelle unøjagtige oplysninger og forbedre datakvaliteten. (Forebyggelse og beskyttelse, Opdagelse og Respons) |
| 1. De registreredes rettigheder   AI-systemet understøtter ikke i tilstrækkeligt omfang de registreredes rettigheder. | | | | | |
|  | Risiko for manglende muligheder for effektiv rettigheds-udøvelse for de registrerede. | Personoplysninger kan indgå i (1) træningsdata ved udvikling af systemet, (2) testdata ved test af systemet, (3) data i selve modellen, (4) inputdata ved anvendelse af systemet samt (5) outputdata ved anvendelse af systemet, f.eks. en afgørelse eller forudsigelse. De registreredes rettigheder skal kunne håndteres i forhold til alle disse anvendelser af personoplysninger. | Artikel 15-22 og præambelbetragtning nr. 60-71 | Organisatoriske foranstaltninger | * Der skal generelt udarbejdes dokumentation af, hvordan udøvelsen af de registreredes rettigheder vil blive faciliteret gennem hele AI-løsningens livscyklus. * Implementering af dokumenterbare retningslinjer og procedurer for håndtering af de registreredes rettigheder ved modeludvikling og -anvendelse. (Forebyggelse og beskyttelse) * Træning af medarbejdere i identifikation og håndtering af rettighedsanmodninger. (Forebyggelse og beskyttelse) * Der bør organiseres workshops i organisationen for at sikre, at udviklerne er klar over, at der skal stilles funktionalitet til rådighed som tillader organisationen at håndtere anmodninger fra de registrerede om udøvelsen af deres rettigheder. * Der kan gennemføres brugertests for at indsamle feedback om, hvor lettilgængeligt og letforståeligt de informationer, som der skal gives de registrerede, er. * Det skal sikres, at der er adgang til meningsfyldt menneskeligt review, når AI-løsningen træffer fuldautomatiske afgørelser, som har retsvirkning eller på tilsvarende vis betydeligt påvirker de registrerede. |
|  | Risiko for manglende eller utilstrækkelig opfyldelse af oplysningspligten. | De registrerede skal som udgangspunkt underrettes om kommunens behandling af deres personoplysninger. Det gælder som udgangspunkt også ved træning/udvikling og anvendelse af en AI-model. Oplysningspligten gælder som udgangspunkt også, hvis kommunen ønsker at viderebehandle personoplysningerne til et andet formål end det, hvortil de er indsamlet. | Artikel 12-14 | Organisatoriske foranstaltninger | * Udvælgelse og implementering af metode og retningslinjer til opfyldelse af oplysningspligten – både ved indsamling og hvis formålet med behandlingen skifter. (Forebyggelse og beskyttelse) * Implementering af arbejdsgange for dokumentation af oplysningspligtens overholdelse. (Forebyggelse og beskyttelse). Der børføres log over indsigelser, som er modtaget vedrørende AI-løsningen for at følge op på indsigelsen og håndteringen heraf. Endvidere kan indsigelser analyseres for at fastlægge trends, problemstillinger og risici. |
|  | Risiko for manglende rettidig opfyldelse af retten til at gøre indsigelse med behandlingen. | Den registrerede kan efter omstændighederne have ret til at gøre indsigelse mod behandling af sine personoplysninger, herunder ved profilering baseret på disse bestemmelser. Senest på tidspunktet for den første kommunikation med den registrerede skal denne udtrykkeligt gøres opmærksom på indsigelsesretten, og oplysninger herom skal meddeles klart og adskilt fra alle andre oplysninger. | Artikel 21 | Organisatoriske foranstaltninger | * Samme som for punkt 6.2. |
|  | Risiko for manglende evne til at forklare logikken bag fuldautomatiske afgørelser. | Hvis AI-løsningen indebærer fuldautomatiske afgørelser, herunder ved brug af profilering, skal den registreredes oplyses herom samt modtage meningsfulde oplysninger om logikken heri samt betydningen og de forventede konsekvenser af en sådan behandling for den registrerede. Den registrerede har også ret til indsigt i disse oplysninger. | Artikel 13-15 og artikel 22 | Organisatoriske foranstaltninger | * Samme som punkt 6.2 ovenfor. * Se også foranstaltninger under punkt 2. Transparens, forståelighed og systemtransparens. |
| 1. Fuldautomatiske afgørelser og profilering   Risiko for manglende menneskelig kontrol og overvågning af systemets afgørelser og risiko for de facto fuldautomatiske afgørelser i beslutningsstøttesystemer. | | | | | |
|  | Risiko for manglende menneskelig kontrol og overvågning af AI-systemets afgørelser (output). | AI-systemer, der udfører fuldautomatiske afgørelser, har særlige risici for manglende menneskelig kontrol og overvågning af, om systemets afgørelser (output) er korrekte. Det gælder i særlig grad ved dynamiske modeller, der ændrer sig i takt med, at de modtager ny erfaring (data) fra sager. | Artikel 5 og artikel 22 | Organisatoriske foranstaltninger | * Løbende overvågning og mulighed for menneskelig overvågning og indgriben før/efter AI-løsningen har truffet en afgørelse. (Forebyggelse og beskyttelse og Respons) * Ved den manuelle kontrol med outputtet af systemet er der adgang til relevant viden om, hvorvidt den forudsatte kvalitet af outputtet er tilstrækkelig og i overensstemmelse med fastsatte kvalitetskrav, f.eks. at der er en kvalitetsscore af outputtet tilgængelig for brugeren, der kan forklares. * Udførelse af stikprøver af de automatiske afgørelser, således at et menneske kan vurdere, om afgørelserne er korrekte. (Opdagelse) * Overvejelse af genoptagelse af sager, hvis afgørelserne ikke er korrekte. (Respons) |
|  | Risiko for manglende kontinuitet, hvis AI-systemet skal stoppes. | Der kan være risiko for, at den forretningsmæssige opgave ikke kan løses, hvis AI-systemet af en eller anden grund – typisk på grund af en fejl eller en sikkerhedshændelse – skal stoppes. Det gælder særligt i fuldautomatiske afgørelsessystemer. | Artikel 5, artikel 22 og artikel 32 | Organisatoriske foranstaltninger | * Implementering af en fallback-plan, så sagsbehandlere kan overtage beslutningstagen og sagsbehandlingen fortsættes, hvis systemet af forskellige årsager må stoppes. (Forebyggelse og beskyttelse) |
|  | Risiko for de facto fuldautomatiske afgørelser i et beslutningsstøttesystem. | Der kan være risiko for, at den menneskelige sagsbehandler ikke foretager en reel vurdering og efterprøvelse af AI-systemets forslag til en afgørelse, dvs. at sagsbehandleren ukritisk lægger systemets forslag til afgørelse til grund (”rubber stamping”). Dette kan have flere årsager, f.eks. at sagsbehandleren ikke kan forstå, hvordan AI-systemet er nået frem til beslutningen (manglende transparens). Den manglende efterprøvelse kan også skyldes, at sagsbehandleren ”stoler på” AI-systemet og dermed stopper med at bruge sin egen kritiske sans og undlader at overveje, om beslutningen kan være forkert (”automation bias”). | Artikel 5 og artikel 22 samt Artikel 13, stk. 2, litra f, artikel 14, stk. 2, litra g, og artikel 15, stk. 1, litra h | Organisatoriske foranstaltninger | * Der skal designes og implementeres passende træning og uddannelse af de sagsbehandlere, der skal foretage den reelle vurdering og efterprøvelse af AI-systemets forslag (”menneskelige reviewers”). * AI-løsningen skal designes på en måde, som sikrer, at menneskelige reviewers har meningsfuld indflydelse på beslutningen, herunder beføjelser og kompetence til at gå imod anbefalingen eller beslutningen samt inddrage relevante faktorer, som ikke var en del af inputdata. * Der kan fastlægges procedurer for, hvornår beslutninger, der er truffet af AI-løsningen, skal genoptages. * Der kan fastlægges procedurer for, hvornår beslutninger, der er truffet af AI-løsningen, skal genoptages. * Klar og dokumenteret grænsedragning mellem fuldautomatiske afgørelser og beslutningsstøtte i de situationer, hvor der anvendes en kombination af de to. Dokumentation af overvejelserne af den passende fordeling mellem systemets og menneskets arbejdsopgaver. (Forebyggelse og beskyttelse) * Uddannelse af medarbejdere med henblik på at sikre, at de forstår, hvordan AI-systemet fungerer og kommer frem til beslutninger og systemets begrænsninger, samt give medarbejderne klare kriterier for, hvornår medarbejderne forventes at ændre systemets beslutninger. (Forebyggelse og beskyttelse) * Implementering af dokumenterede retningslinjer og procedurer, der sikrer, at systemets forslag til afgørelser efterprøves af personer med den fornødne kompetence og mulighed for at ændre afgørelsen. (Forebyggelse og beskyttelse) * Understøttelse af, at beslutningsprocessen og ”logikken” i systemet er forklarlig og transparent. (Forebyggelse og beskyttelse) * Løbende dokumenteret revision og statistik på, hvor mange forslag til afgørelser der omgøres af sagsbehandlerne – en meget lav omgørelsesprocent kan være tegn på, at medarbejderne ikke foretager en reel vurdering af systemets forslag til afgørelser. (Opdagelse) * Udførelse af stikprøver af afgørelser med henblik på at efterprøve, om den menneskelige sagsbehandler har foretaget en reel vurdering og korrekte afgørelser. (Opdagelse) * Re-design af AI-systemet med henblik på at gøre forslagene til beslutninger mere transparente og forklarlige. (Respons) * Overvej genoptagelse af sager, hvis afgørelserne ikke er korrekte. (Respons) |
|  | Risiko for manglende overblik over udledte personoplysninger ved profilering. | Der kan være risiko for, at kommunen ikke er klar over, at der udledes følsomme personoplysninger om de registrerede. Ved profilering kan der potentielt udledes følsomme personoplysninger ud fra ikke følsomme personoplysninger. Det kan f.eks. være muligt at udlede en persons sundhedstilstand ud fra oplysninger om fødevareindkøb. | Artikel 5, 9, 13-14 og 22 | Organisatoriske foranstaltninger | * Hvis der udledes følsomme personoplysninger ved profileringen, skal kommunen sikre, at behandlingen ikke er uforenelig med det oprindelige formål, at kommunen har identificeret et retligt grundlag for behandlingen af de følsomme personoplysninger samt at den registrerede oplyses om behandlingen, medmindre en af undtagelserne til oplysningspligten finder anvendelse. (Forebyggelse og beskyttelse) |
|  | Risiko for utilstrækkelig beskyttelse eller uproportional behandling ved brug af fuldautomatiske afgørelser over for børn. | Databeskyttelsesforordningen pålægger dataansvarlige yderligere forpligtelser i forbindelse med behandling af børns personoplysninger. Databeskyttelsesforordningen rummer ikke et forbud mod brug af fuldautomatiske afgørelser over for børn, men det er hovedreglen, at det ikke skal ske. Der kan dog være visse situationer, hvor dataansvarlige er nødt til at træffe afgørelser, der alene er baseret på automatisk behandling, herunder profilering, som har retsvirkning for eller på tilsvarende vis betydeligt påvirker børn, f.eks. for at beskytte deres velfærd. I så fald kan behandlingen foretages på grundlag af undtagelserne i artikel 22, stk. 2, litra a), b) og c), alt efter hvad der er relevant. | Artikel 22 | Organisatoriske foranstaltninger | * Udarbejdelse af interne retningslinjer for, hvordan behandlingen foregår og børnenes rettigheder sikres. * Udarbejdelse af adfærdskodeks for behandlingen, der omfatter beskyttelsesforanstaltninger for børn. |
| 1. Sikkerhed og robusthed   Utilstrækkelig sikkerhed i AI-systemet kan medføre brud på persondatasikkerheden. Manglende robusthed for systemet kan medføre, at systemet ikke kan performe i overensstemmelse med målsætningerne. | | | | | |
|  | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i afdækningsfasen (forud for udvikling og idriftsættelse) | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i afdækningsfasen kan medføre risiko for, at data ikke behandles tilstrækkeligt sikkert. | Artikel 5, stk. 1, litra f, og artikel 32-34 samt præambelbetragtning nr. 83 og 85-88 | Organisatoriske foranstaltninger | * Der skal foretages en forhåndsundersøgelse af leverandørerne og udviklerne af de tredjepartsmodeller, som skal anvendes til udvikling af løsningen, herunder vilkår og eventuelle privatlivspolitikker. Der må alene anvendes pålidelige leverandører. * Det bør sikres, at tredjepartsmodellerne er testet, og at sikkerheden er auditeret. * Der bør løbende ske overvågning af leverandørens eventuelle vilkår og sikkerhedsaudits for at sikre, at tredjepartsmodellen fortsat er robust, og at modellen ikke har ændret sig væsentligt fra udgangspunktet. * Der bør implementeres en passende politik for opdatering og patching af AI-løsningen, der er baseret på tredjepartsmodellen, for at sikre, at den nyeste udgave af modellen eller API'et benyttes. |
|  | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i udviklingsfasen | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i udviklingsfasen kan medføre risiko for, at data ikke behandles tilstrækkeligt sikkert. | Artikel 5, stk. 1, litra f, og artikel 32-34 samt præambelbetragtning nr. 83 og 85-88 | Organisatoriske foranstaltninger | * Der skal generelt etableres passende procedurer for ændringsstyring samt foretages implementering og dokumentation af streng adgangskontrol i forhold til, hvilke personer, der har adgang til træningsdata, kildekode og produktionsdata. * Der skal etableres passende procedurer for ændrings- og versionsstyring (change management). * Der skal etableres adskilte tekniske miljøer til henholdsvis udvikling, test og produktion, samt sikres, at de respektive datasæt, der skal bruges i de forskellige faser, behandles i de respektive miljøer. * Integriteten af datasæt, som skal bruges til henholdsvis udvikling, test og produktion, skal verificeres. Der kan tages udgangspunkt i eventuelle procedurer for data governance og data proviniens. * Medarbejderes adgangsrettigheder skal begrænses efter behov. Medarbejdere skal alene have de adgangsrettigheder, der er nødvendige for, at vedkommende kan udføre sine opgaver. Der skal ske regelmæssig kontrol af de tildelte adgangsrettigheder og inddragelse af rettighederne, når de ikke længere er nødvendige. Der skal føres logs over, hvem der har hvilke rettigheder, og hvordan disse rettigheder bliver brugt. * Der bør etableres klare revisionsspor for, hvordan datasæt flyttes og opbevares fra en placering til en anden under trænings- og testfasen. * Der bør fastlægges en beredskabsplan, som gøres tilgængelig for alle, der er involveret i udvikling og drift af AI-løsninger. Planen bør omfatte processer for indrapportering af hændelser, og hvem der er ansvarlig for at håndtere dem. Planen bør også omfatte vejledning i, hvordan hændelser hurtigt adresseres, hvem der håndterer hændelsen, og hvordan hændelsen kommunikeres til de øvrige dele af organisationen. * Der bør anvendes teknikker og værktøjer til at opdage anormaliteter i f.eks. træningsdata. * Der bør generelt foretages review og styring af anvendte kodebiblioteker, plugins mv. * Der bør anvendes pålidelige repositories og foretages validering af signaturen af pakker og kodebiblioteker, der anvendes, for at validere disses integritet. * Der kan foretages regelmæssig kodereview af de anvendte kodebiblioteker for at sikre, at der ikke er introduceret ondsindet kode. * Udviklerne kan modtage uddannelse for at gøres opmærksomme på risiciene ved forsyningskædeangreb og vigtigheden af at verificere de anvendte kodebiblioteker mv. |
|  | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i testfasen | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i testfasen kan medføre risiko for, at data ikke behandles tilstrækkeligt sikkert. | Artikel 5, stk. 1, litra f, og artikel 32-34 samt præambelbetragtning nr. 83 og 85-88 | Organisatoriske foranstaltninger | * Der skal udføres regelmæssige sikkerhedstests af AI-løsningen. * Der skal fastsættes passende procedurer og processer for patching og opdatering. Der kan som led heri med jævne mellemrum indhentes advarsler om sårbarheder fra relevante kilder for at sikre, at sårbarhederne ikke udnyttes. |
|  | Risiko for manglende iagttagelse af principperne om rimelighed og sikkerhed i drifts- og monitoreringsfasen | Risikoen indebærer, at medarbejdere ifm. drift eller monitorering opnår unødvendig og dermed ulovlig adgang til data på grund af manglende sikring af adgangsbegrænsning. | Artikel 5, stk. 1, litra a og f, samt betragtning nr. 39, 58 samt 60 og 83 | Tekniske foranstaltninger | * Generelt skal princippet om *least privilege[[3]](#footnote-3)* implementeres, og der skal ske uddannelse af medarbejdere i AI-løsningens anvendelsesområde * Medarbejderes adgangsrettigheder skal begrænses efter behov. Medarbejdere skal alene have de adgangsrettigheder, der er nødvendige for, at vedkommende kan udføre sine opgaver. De tildelte adgangsrettigheder skal regelmæssigt kontrolleres, og rettighederne skal inddrages, når de ikke længere er nødvendige. Der skal føres logs over, hvem der har hvilke rettigheder, og hvordan disse rettigheder bliver brugt. * Medarbejderne, der bruger AI-løsningen, skal uddannes i løsningens muligheder og begrænsninger. |
|  | Risiko for manglende iagttagelse af principperne om integritet og fortrolighed i drifts- og monitoreringsfasen | Risiko for manglende iagttagelse af princippet om integritet og fortrolighed i drifts- og/eller monitoreringsfasen kan medføre risiko for, at data ikke behandles tilstrækkeligt sikkert. | Artikel 5, stk. 1, litra f, og artikel 32-34 samt præambelbetragtning nr. 83 og 85-88 | Tekniske og organisatoriske foranstaltninger | * Generelt skal der ske fastsættelse og dokumentation af tekniske og organisatoriske foranstaltninger, som vil reducere behandlingssikkerhedsrisici. * AI-løsningen bør overvåges, når den er i drift, og enhver uregelmæssighed bør undersøges. * Der kan indføres foranstaltninger, som kan forebygge uregelmæssigheder, f.eks. at antallet af forespørgsler, der kan udføres af en given bruger inden for et tidsinterval, begrænses. * Anonym brug af AI-løsningen kan afskæres ved at implementere foranstaltninger, som kræver validering af brugerens identitet. * Der kan foretages regelmæssig fejlsøgning af AI-løsningen. * Der skal generelt fastsættes passende foranstaltninger for at sikre imod manipulation af input. * Der kan fastsættes begrænsninger for typen af inputdata. Begrænsningerne kan fastsættes ud fra logiske betragtninger om AI-løsningens anvendelsesområde, hvor input undersøges for uventede værdier eller mønstre. * Der skal generelt fastsættes passende foranstaltninger for at sikre imod nedbrud. * Medarbejderes adgangsrettigheder skal begrænses efter behov. Medarbejdere skal alene have de adgangsrettigheder, der er nødvendige for, at vedkommende kan udføre sine opgaver. Der skal regelmæssigt udføres kontrol over de tildelte adgangsrettigheder og rettigheder skal inddrages, når de ikke længere er nødvendige. Der skal føres logs over, hvem der har hvilke rettigheder, og hvordan disse rettigheder bliver brugt. * Der skal etableres backup af AI-løsningen, som gør det muligt at reetablere efter eventuelle nedbrud. Backup bør vedligeholdes og testes regelmæssigt. * Der bør fastlægges en beredskabsplan, som gøres tilgængelig for alle, der er involveret i udvikling og drift af AI-løsninger. Planen bør omfatte processer for indrapportering af hændelser, og hvem der er ansvarlig for at håndtere dem. Planen bør også omfatte vejledning i, hvordan hændelser hurtigt adresseres, hvem der håndterer hændelsen, og hvordan hændelsen kommunikeres til de øvrige dele af organisationen. * Der skal generelt fastsættes passende foranstaltninger for at sikre imod overbelastning af AI-løsningen. * Der kan fastsættes begrænsning for mængden af inputdata. Det kan f.eks. omfatte antallet input pr. bruger eller IP-adresse eller tilsvarende. * Der kan fastsætte begrænsninger på antallet af tilgængelige ressourcer, der bruges pr. input. * Der kan implementeres løbende overvågning af inputdata for at identificere unormale spidsbelastninger, der kan indikere et DoS-angreb. * Der skal generelt fastsættes passende foranstaltninger for at sikre, at AI-løsningen output ikke bidrager til eller forårsager sikkerhedsbrud. * Det kan sikres, at output fra AI-løsningen behandles på linje med almindelige brugere og der kan adopteres en *zero-trust[[4]](#footnote-4)*-tilgang, der sikrer, at alt output valideres. |
| Brud på persondatasikkerheden kan vedrøre dataene (trænings-, validerings-, test-, input- og outputdata) samt modellen herunder såkaldte model inversion attacks og membership inference attacks. Et brud kan også vedrøre den underliggende it-infrastruktur, dvs. både software og hardware. Hvis et AI-system angribes, kan både data og systemet ændres, således at systemet træffer anderledes, forkerte beslutninger, eller så det helt lukkes ned. For identifikation af risici og tilhørende afhjælpende foranstaltninger vedrørende temaet ”Robusthed og sikkerhed” kan du med fordel anvende følgende værktøjer og retningslinjer:   * Digitaliseringsstyrelsen, Erhvervsstyrelsen og Center for Cybersikkerhed: Vejledning: Tiltag til at sikre brugen af kunstig intelligens, januar 2020 * ENISA: Multilayer framework for good cybersecurity practices for AI, juni 2023 * Federal Office for Information Security, AI Security Concerns in a Nutshell, version 1.0 af 9. marts 2023 * OWASP: LLM AI Cybersecurity & Governance Checklist – From the OWASP Top 10 for LLM Applications Team, version 1.1 af 11. april 2024 * ISO/IEC 27001:2023, Anneks A, samt * ISO/IEC 29134:2023 og ISO/IEC 29151:2022, der i katalogform beskriver flere af de mulige sikkerhedsforanstaltninger, der kan benyttes. | | | | | |
| 1. Governance og ansvarlighed (accountability)   Manglende eller utilstrækkelig afklaring af ansvar for og ledelsesmæssig godkendelse af udvikling og brugen af AI-systemet kan medføre risiko for overtrædelse af reglerne. Tilsvarende gælder manglende eller utilstrækkelig dokumentation for centrale valg og fravalg samt retningslinjer og procedurer eller uddannelse af centrale medarbejdere. | | | | | |
|  | Risiko for, at der forud for anvendelse systemet ikke er dannet tilstrækkeligt overblik over risici | Hvis der ikke er dannet tilstrækkeligt overblik over risici, er der risiko for, at der ikke træffes korrekte og tilstrækkelige foranstaltninger til at mitigere risici, herunder høje risici ved behandlingen, som kan medføre negative konsekvenser for de registrerede. | Artikel 5, stk. 2, 35 og 36 samt betragtning nr. 74-77, 84, 89-92, 94 og 95. | Organisatoriske foranstaltninger | * Gennemføre konsekvensanalyse (DPIA), hvis behandlingsaktiviteten sandsynligvis vil indebære høj risiko (dette er normalt tilfældet for udvikling og brug af AI-systemer) * Hvis relevant skal de registreredes synspunkter vedrørende den påtænkte behandlingsaktivitet indhentes * Overvej om Datatilsynet skal høres inden påbegyndelse af behandlingsaktiviteten * Som led i konsekvensanalysen bør enheder i organisationen, som vil være involveret i projektet, konsulteres, mhp. at identificere relevante risici. * Domæneeksperter bør rådføres med henblik på at bistå med at identificere og mitigere relevante risici. |
|  | Risiko for manglende eller utilstrækkelig afklaring af ansvar for og ledelsesmæssig godkendelse af udvikling og brug af AI-systemet. | Hvis der ikke er en klar og kommunikeret fordeling af ansvar for og ledelsesmæssig godkendelse af udvikling og brug af AI-systemet, kan det medføre, at relevante beslutninger om AI-systemets udvikling, drift eller berostillelse er forkerte eller slet ikke tages. Tilsvarende gælder beslutninger vedrørende implementering og opfølgning på aftalte foranstaltninger. | Artikel 24-25 og artikel 5, stk. 2 | Organisatoriske foranstaltninger | Implementering af dokumenterede retningslinjer, beslutningsgange og procedurer, der sikrer ansvaret og ansvarligheden for AI-systemet og dets resultater i hele systemets livscyklus. Disse mekanismer skal klart regulere, hvornår ledelsen skal godkende centrale milepæle, herunder beslutning om opstart, udvikling af modellen, idriftsættelse og revision samt centrale performancekrav. (Forebyggelse og beskyttelse)   * Der bør udpeges en proces- eller systemejer for at sikre, at ansvaret for AI-løsningen forankres i organisationen. (Forebyggelse og beskyttelse) * Dokumenterbare retningslinjer og procedurer for adskillelse af udviklings- og produktionsmiljø og dokumentation af, at adskillelsen er sket. (Forebyggelse og beskyttelse) * Beskrivelse af modellens begrænsninger i lyset af formålet, så alle relevante interessenter er opmærksom på, under hvilke omstændigheder den bør bruges. (Forebyggelse og beskyttelse) * Dokumentation af valg og ændringer af AI-systemet, således at kommunen kan dokumentere valg og fravalg under udvikling, drift og opdatering af systemet og begrundelserne herfor. (Respons) * Dokumentation af ledelsens godkendelse af udvikling, test og anvendelse af AI-systemet. (Forebyggelse og beskyttelse) * Tilslutning til adfærdskodekser for revisionsprocesser vedrørende maskinlæring eller certificering/standardisering. (Forebyggelse og beskyttelse) |
|  | Risiko for manglende eller utilstrækkelig uddannelse af centrale medarbejdere. | Hvis medarbejderne ikke er tilstrækkeligt uddannede i at håndtere AI-systemer og dets særegne databeskyttelsesretlige risici, er der risiko for medarbejderfejl og dermed overtrædelse af reglerne. | Artikel 24-25 og artikel 5, stk. 2 | Organisatoriske foranstaltninger | * Uddannelse af relevante medarbejdere, herunder AI-udviklere, projektleder og øvrige nøglepersoner, og dokumentation af uddannelsens indhold. (Præventiv) * Inddragelse af de rette kompetencer fra start og undervejs i AI-systemets livscyklus, herunder AI-udviklere, it-arkitekter, ledelse, proceskonsulenter, DPO’en og personer med domænekendskab (faglige fyrtårne / erfarne sagsbehandlere). (Forebyggelse og beskyttelse, Opdagelse og Respons) |
|  | Risiko for utilstrækkelig kravstillelse, due diligence og tilsyn med eksterne leverandører af AI-løsninger. | Hvis kommunen ikke får stillet tilstrækkelige krav om databeskyttelse til løsningens design, herunder foretager den fornødne forudgående kontrol med systemets overholdelse af reglerne (due diligence), er der risiko for, at løsningen ikke overholder databeskyttelsesreglerne. Tilsvarende gælder, hvis kommunen ikke løbende fører tilsyn med leverandørens overholdelse af kravene under drift. | Artikel 24-25 og artikel 5, stk. 2 | Organisatoriske foranstaltninger | * Dokumenterbare retningslinjer og procedurer for kravstillelse, due diligence af og tilsyn med eksterne leverandører af AI-løsninger. (Forebyggelse og beskyttelse) * Opnåelse af kontraktlige garantier for tredjepartsalgoritmer om, at audit og testning er udført, og at algoritmen er i overensstemmelse med vedtagne standarder. (Forebyggelse og beskyttelse) * Tilsyn af leverandørers overholdelse af kravene. (Forebyggelse og beskyttelse og Opdagelse) |
|  | Risiko for uhensigtsmæssigt lav kvalitet eller utilsigtede hændelser pga. utilstrækkelig testning, revision og overvågning af AI-systemet. | Hvis der ikke foretages tilstrækkelig og dokumenterbare tests, revision og overvågning af AI-systemet både under udvikling og løbende i drift, er der risiko for manglende opdagelse af fejl og utilstrækkelig opfyldelse af kravene til systemet. | Artikel 24-25 og artikel 5, stk. 2 | Organisatoriske foranstaltninger | * Etablering og vedligeholde af dokumenterbare retningslinjer og processer for test af AI-systemet før go live og med et passende interval efter go live samt ved ændringer. Gennemførte tests, resultater heraf samt foranstaltninger truffet i medfør heraf skal kunne dokumenteres. (Forebyggelse og beskyttelse) * Gennemførelse af en algoritmisk audit af AI-systemet, modellen eller kildekoden fra en uafhængig ekstern tredjepart, f.eks. revisor, advokat eller sikkerhedsekspert m.v. Dvs. en audit for at påvise, at systemet rent faktisk fungerer efter hensigten og ikke giver diskriminerende, fejlagtige eller uberettigede resultater. (Opdagelse) |
|  | Risiko for manglende evne til at reagere rettidigt på fejl og hændelser vedrørende AI-systemet. | Der er risiko for overtrædelse af reglerne, hvis kommunen ikke evner at reagere rettidigt på fejl og hændelser og håndtere disse med de fornødne foranstaltninger. Det gælder f.eks. indsigelser eller klager fra de registrerede eller egentlige brud på persondatasikkerheden. | Artikel 24-25, artikel 33-34 og artikel 5, stk. 2 | Organisatoriske foranstaltninger | * Dokumentation af alle hændelser, klager og indsigelser fra de registrerede vedrørende AI-systemet og evaluering heraf. Overvejelse af, om øvrige registrerede kan være påvirkede af fejl. Dokumentation af alle foranstaltninger, der træffes i medfør heraf. (Opdagelse og Respons) * Overvejelse af genoptagelse af sager, hvis der er begået fejl i AI-systemet. (Respons) |

1. Oversigt over mulige databeskyttelsesretlige skadevirkninger for de registrerede

| Type | Kategori | Beskrivelse | Eksempler |
| --- | --- | --- | --- |
| Individuel skade | Finansiel skade | Uagtsomt, bevidst eller med vilje at bane vejen for, at økonomisk skade kan ske. | * Brud, der fører til svindel * Påvirkning af kreditvurdering * Afpresning ved brug af personlige data * Målrettet markedsføring over for personer med spilafhængighed med spilreklamer * Tab af indkomst/medarbejdere på grund af skade på omdømmeskade |
| Fysisk skade | Uagtsomt, bevidst eller med vilje at bane vejen for, at fysisk skade kan ske. | * Selvmord eller anden selvskade * Personlige oplysninger, der anvendes til at spore nogens lokation, og som fører til overfald * Medicinsk fejlbehandling forårsaget af uagtsomhed eller fejl |
| Omkostninger ved at undgå eller afhjælpe en skade | Omkostninger i tid eller penge, som er påkrævet for at undgå eller afhjælpe skade eller sårbarheder relateret til databeskyttelse. | * Tid brugt på at undgå skade/risiko for skade * Sikkerhedsomkostninger forbundet med beskyttelse af personlige oplysninger |
| Diskrimination | Skader, der opstår på grund af diskrimination eller fordomme/bias (enten bevidste eller ubevidste). | * Indgroede fordomme/bias i automatiserede afgørelser * Prisdiskrimination |
| Uønsket indtrængen/kontakt | Uønsket kontakt eller indtrængen, der forstyrrer eller afbryder aktiviteter, eller øger risikoen for, at andre skader vil ske. | * Uønsket målrettet reklame * Generende opkald eller spam * Uberettiget overvågning |
| Tab af kontrol over personlige oplysninger | Skader forårsaget af misbrug, anvendelse af data til nye formål, uønsket opbevaring eller fortsat behandling og deling af personlige oplysninger, herunder mangel på datakvalitet og transparens. | * Skade på sindsro og evnen til at håndtere risici * Begrænsninger i evnen til at tilgå eller kontrollere behandlingen af personlige oplysninger * Ikke lovlig genbehandling af personoplysninger, der fører til en psykisk belastning/stress |
| Tab af autonomi; manipulation og indflydelse | Begrænsning, tvang eller manipulation af menneskers valgmuligheder eller deres evne til at træffe et informeret valg. | * ”Nudging”, der fører til dårlige beslutninger * Begrænsning af valgmuligheder på grund af magt- og informationsasymmetri |
| Psykologiske skader | Uagtsomt, bevidst eller med vilje at bane vejen for forekomsten af følelsesmæssig stress eller forstyrrelse (forlegenhed, angst, frygt). | * Skade på mental sundhed * Tab af identitetsfølelse eller kontrol af identitet * Belastede relationer * Tab af selvtillid * Tab af omdømme/tab af anseelse * Chikane eller mobning |
| Frygt for en potentiel skade | Tilbageholdenhed ved brug af services eller aktiviteter på grund af en faktisk eller anticiperet risiko for potentielle skader. Og dermed en reduktion i de fordele, der kunne være opnået. | * Tilbageholdenhed ved brug af produkter og services, der indebærer deling af data * Nedgang i aktiviteter, der vil kræve en god kreditvurdering for at komme videre |
| Negative konsekvenser for grundlæggende rettigheder og frihedsrettigheder | Negative konsekvenser for grundlæggende rettigheder og frihedsrettigheder. | * Begrænsninger i forhold til databeskyttelsesrettigheder * Begrænsninger i forsamlingsfriheden * Menneskers tilbageholdenhed med at bruge deres ytringsfrihed |
| Samfundsmæssig skade | Skade på lov og retfærdighed | Begrænsning af eller underminering af retsstaten, civil- og straffeprocessen samt retsplejen. | * Handlinger, der skaber grundlag for omfattende underminering af lovgivningen * Ofre eller vidners tilbageholdenhed og frygt på grund af manglende retssikkerhed. |
| Skade på medier, demokrati, informationsfriheden og den offentlige debat | Negative konsekvenser for medier, demokrati, informationsfriheden og den offentlige debat på et samfundsmæssigt niveau. | * Mistillid til håndteringen af valgroller, der påvirker valg eller vælgerdeltagelse * Udbredt mistillid, der fører til frygt for og tilbageholdenhed med at bruge ytringsfriheden |
| Skade på folkesundheden | Skader, der resulterer i negative sundhedsmæssige konsekvenser for samfundet. | * Mistillid til håndtering af sundhedsdata, der smitter af på borgernes lyst til at gøre brug af landet sundhedsydelser |
| Skade på økonomien | Negative konsekvenser for økonomien, herunder særligt på enten lokalt, regionalt eller nationalt niveau, eller for en specifik sektor. | * Borgernes mistillid på grund af udbredt misbrug af privatlivets fred, der påvirker borgernes lyst og tillid til at anvende visse tjenester/services, der behandler personoplysninger * Misbrug af personlige data, der fører til urimelige konkurrencefordele |
| Skade på miljøet | Negative konsekvenser for miljøet, enten direkte eller indirekte som følge af misbrug af data eller i forbindelse med håndtering af en tilknyttet risiko. | * Højt energiforbrug forbundet med dataindsamling, -opbevaring og -deling * Tab af biologisk mangfoldighed (biodiversitet) og/eller grønne områder på grund af, at land/marker mv. anvendes til serverfarme |

*Kilde: Information Commissioner’s Office: Overview of Data Protection Harms and the ICO Taxonomy, april 2022.*

1. *Feature engineering* omfatter processen med at skabe nye inputvariabler fra rå data for at forbedre maskinlæringsmodeller. [↑](#footnote-ref-1)
2. Modelkort er en form for dokumentation, som beskriver AI-modellens egenskaber, ydeevne, anvendelsesområder og begrænsninger. [↑](#footnote-ref-2)
3. *Least privilege*-princippet indebærer, at en bruger med adgang til data skal have så lidt adgang som muligt for at de stadig kan udføre deres arbejde. [↑](#footnote-ref-3)
4. Zero-trust betyder, at der etableres og verificeres den nødvendige tillid, før der gives adgang til data. [↑](#footnote-ref-4)