AI Sweden – Utvalda juridiska aspekter rörande personuppgifter vid träning av NLP-modeller med patientjournalstexter, 2021-11-20

Juridisk information: Denna artikel utgör allmän information. Artikeln är inte avsedd att vara uttömmande eller tillräcklig som grund för att fatta beslut i konkreta frågor utan är endast ämnad att användas i informationssyfte och utgör inte rådgivning. Vid en specifik rättslig fråga bör en kvalificerad jurist konsulteras.

Tillämpningar inom Artificiell Intelligens (AI) anses ha en stor potential för effektivisering inom och förbättring av hälso- och sjukvården. Ett lovande tillämpningsområde är att använda AI-system som baseras på NLP-modeller som verktyg för kvalitetssäkring eller som beslutsstöd. Syftet med denna artikel är att belysa utvalda juridiska aspekter avseende användning av anonymiserade patientjournaler för träning av NLP-modeller hos en offentlig vårdgivare[1]. Artikeln är begränsad till enskilda juridiska aspekter avseende den allmänna dataskyddsförordningen (GDPR) och patientdatalagen (2008:355, PDL). Offentlighets- och sekretesslagen (2009:400, OSL) nämns uteslutande för att förtydliga regelverkets betydelse i sammanhanget. Artikeln är således inte en uttömmande genomgång av alla juridiska frågor som måste beaktas på området.

Att träna en NLP-modell med patientjournaltexter innebär som utgångspunkt en omfattande behandling av personuppgifter som i sin tur aktualiserar en tillämpning av kraven i GDPR och PDL (som kompletterar GDPR på området hälso-och sjukvård). Eftersom patientjournaler vanligtvis innehåller känsliga personuppgifter omfattas de av särskilt restriktiva krav, som till exempel krav på lagreglerad tystnadsplikt och en strikt reglering av ändamålen för vilka patientjournalerna får användas.[2] Det kan därför finnas ett intresse för offentliga vårdgivare att undvika behandling av personuppgifter vid träning av NLP-modeller.

[1] Med offentlig vårdgivare avses statlig myndighet, region och kommun i fråga om sådan hälso- och sjukvård som myndigheten, regionen eller kommunen har ansvar för, se 1:3 PDL.
[2] Jmf. art. 9.2 h GDPR, 9.3 GDPR och 2:4 PDL. För offentliga vårdgivare regleras tystnadsplikten i OSL.

Enligt GDPR är en personuppgift en uppgift som avser en identifierad eller identifierbar fysisk person. En fysisk person är identifierbar om den kan identifieras direkt eller indirekt.[3] Information som inte hänför sig till en identifierad eller identifierbar fysisk person utgör inte en personuppgift. Personuppgifter som anonymiserats på ett sådant sätt att fysiska personer inte längre är identifierbara, nedan kallat anonymiserade personuppgifter, utgör inte personuppgifter.[4] Behandling av anonymiserade personuppgifter faller utanför GDPR:s och PDL:s tillämpningsområde.[5]

För att anonymisera personuppgifter krävs att personuppgifterna bearbetas så att en enskild individ inte längre kan identifieras med alla hjälpmedel som rimligen kan komma att användas.[6] I praktiken är en av de största utmaningarna med anonymisering av personuppgifter att det inte finns några absoluta kriterier för att fastställa när en uppsättning personuppgifter ska anses vara anonymiserade eller inte. GDPR:s kriterier som avgör huruvida personuppgifter ska anses vara anonymiserade eller inte är allmänt hållna vilket återspeglar lagstiftarens intention att skapa en teknikneutral och framtidssäker lagstiftning.[7] Baserat på dessa kriterier har dataskyddsmyndigheter utvecklat ett riskbaserat angreppsätt för anonymisering som syftar till att sänka risken för en potentiell återidentifering till en acceptabel nivå.[8]

Angreppssättet innebär att riskerna för särskiljbarhet, länkbarhet och inferens sänks till en acceptabel nivå. Med risken för särskiljbarhet avses risken att identifiera en enskild individ genom att kunna särskilja denna från andra individer. Risken för länkbarhet avser risken att använda länkar mellan olika dataset för att identifiera en enskild individ. Risken för inferens avser möjligheten att identifiera en enskild individ med hjälp av kompletterande information eller genom att härleda dennes identitet på något annat sätt.[9]

---

[3] Art. 4.4 GDPR.
[4] Skäl 26 GDPR.
[5] Jmf. art. 2 GDPR och 1:1 PDL.
[6] AG-29, Yttrande 05/2014 om avidentifieringsmetoder, WP 216, s. 4-12. Notera att begreppet avidentifiering som används i yttrandet bör läsas som anonymisering. Användning av begreppet avidentifiering beror troligtvis på en felöversättning från det engelska begreppet ”anonymisation”.
[7] Skäl 15 GDPR.
[8] AG-29, Yttrande 05/2014 om avidentifieringsmetoder, WP 216.
[9] AG-29, Yttrande 05/2014 om avidentifieringsmetoder, WP 216, s. 4-12.

Vilken risknivå som är acceptabel avgöras i det enskilda fallet och mot bakgrund av den senaste tekniska utvecklingen. Detta innebär att sannolikheten för att anonymiserade personuppgifter kan komma att återidentifieras ökar över tid på grund av den tekniska utvecklingen. Det krävs således en regelbunden utvärdering av den kvarstående risken för återidentifiering. Följeakteligen kan det vara svårt att fastställa den exakta gränsen för när personuppgifter övergår till anonymiserade personuppgifter. En försvårande omständighet är också att det inte finns särskilt mycket rättspraxis på området.[10] Däremot finns det ett flertal exempel på misslyckade försök på anonymisering av personuppgifter.[11]

Ett tänkbart tillvägagångssätt för en offentlig vårdgivare som vill träna en NLP-modell är att använda anonymiserade journaltexter som träningsdata för modellen. Fördelen med ett sådant tillvägagångssätt, jämfört med att använda icke-anonymiserade journaltexter, är att behandling av sådan träningsdata inte omfattas av reglerna i GDPR och PDL. Nackdelen med tillvägagångssättet är att anonymisering av patientjournaler är komplicerat.

Det finns ett flertal metoder för att anonymisera patientjournaltexter, som exempelvis avidentifiering (de-identification), borttagning av delar av texten (obfuscation) och differentiell integritet (differential privacy).[12] Huruvida dessa metoder möjliggör en effektiv anonymisering av texterna är emellertid omdiskuterat, särskilt när träningsdata omfattar stora mängder text.[13] Att det förs en sådan diskussion innebär dock inte att en sådan anonymisering per automatik är omöjlig.

En viss typ av inferensattack, en så kallad membership inference attack, diskuteras regelbundet som risk för återidentifiering av enskilda vid användning av NLP-modeller. En sådan attack innebär att en kunnig angripare kan använda NLP-modellen för att göra en förutsägelse om huruvida en viss text ingått i modellens träningsdata. Angriparen kan därefter använda denna

[10] AG-29, Yttrande 05/2014 om avidentifieringsmetoder, WP 216, s. 4-12.
[11] AG-29, Yttrande 05/2014 om avidentifieringsmetoder, WP 216, s. 4-12.
[12]Med avidentifiering avses borttagandet av identifierare som till exempel namn, personnummer osv. En avidentifiering innebär inte per automatik att personuppgifter har anonymiserats i den mening som avses i GDPR, men den kan ha samma effekt. En omständighet som försvårar avgränsningen av begreppet är att det engelska begreppet ”anonymisation” regelbundet översätts som avidentifiering, vilket måste betraktas som felaktigt.
[13] Jagannatha, Rawat, Yu, 2021, Membership Inference Attack Susceptibility of Clinical Language Models, arXiv:2104.08305v1.

kunskap för att återidentifiera en enskild individ. En angripare som har tillgång till en enskilds patientjournalstext kan exempelvis använda denna typ av attack för att försöka att bekräfta att texten, eller delar därav, ingått i modellens träningsdata. Olika metoder för att förebygga denna typ av inferensattack finns.[14]

Blotta förekomsten av risken för en sådan inferensattack tolkas av vissa som att en anonymisering av patientjournaltexter enligt GDPR är omöjlig vid träning och senare användning av NLP-modeller.[15] Det finns dock ett mer nyanserat synsätt att se på frågan. Att det föreligger en risk för en inferensattack av detta slag innebär inte per automatik att en anonymisering enligt GDPR är omöjlig. Tvärtom är kravet som ställs för att uppnå en anonymisering att risken för en potentiell återidentifering sänks till en acceptabel nivå. Att en angripare kan göra en förutsägelse om huruvida en text ingått i NLP-modellens träningsdata kan inte per automatik likställas med en framgångsrik återidentifiering av en enskild individ, särskilt när texten som ingått i förutsägelsen inte på något sätt kan kopplas till den enskilde individen. Givetvis finns det inget hinder för att föra ett resonemang om att en sådan förutsägelse trots allt skulle kunna utgöra en personuppgift. Att utveckla ett sådant resonemang skulle dock gå utöver ramarna i denna artikel.

Nästa fråga är själva anonymiseringsprocessen. Indata till anonymiseringen är icke-anonymiserade patientjournalstexter som innehåller personuppgifter. Att anonymisera texterna innebär således en behandling av personuppgifter som omfattas av reglerna i GDPR och PDL.

En av de viktigaste aspekterna som måste beaktas är att varje behandling av personuppgifter kräver en rättslig grund och att principen om ändamålsbegränsning efterlevs.[16] Rättsliga grunder för behandling av personuppgifter regleras i GDPR. Det finns en uppsättning av rättsliga

[14] Jagannatha, Rawat, Yu, 2021, Membership Inference Attack Susceptibility of Clinical Language Models, arXiv:2104.08305v1.
[15] Jagannatha, Rawat, Yu, 2021, Membership Inference Attack Susceptibility of Clinical Language Models, arXiv:2104.08305v1.
[16] Denna princip kräver att personuppgifter ska samlas in för särskilda, uttryckligt angivna och berättigade ändamål och inte senare behandlas på ett sätt som är oförenligt med dessa ändamål (se art. 5.1 b GDPR).

grunder som kan användas som stöd för behandling av personuppgifter inom hälso- och sjukvården.[17]

PDL begränsar för vilka ändamål personuppgifter får behandlas inom hälso- och sjukvården.[18] Framtagning av NLP-modeller med anonymiserade patientjournaltexter som träningsdata bör kunna rymmas inom ändamål som rör verksamhetsutveckling och kvalitetssäkring.[19] Eftersom sådana NLP-modeller kan klassas som statistik är det även tänkbart att modellens framtagning kan rymmas inom ändamålet att framställa statistik om hälso- och sjukvården.[20] Om och i vilken omfattning framtagning av sådana NLP-modeller kan rymmas inom de övriga ändamålen som anges i PDL är en intressant fråga som bör utredas vidare. Notera att denna diskussion är begränsad till framtagning av NLP-modeller som tränats med anonymiserade patientjournaltexter.

Ett annat sätt att se på anonymiseringsprocessen är att betrakta den som en ytterligare behandling av personuppgifter som är förenlig med det ursprungliga ändamålet för vilka personuppgifterna behandlas.[21] Utifrån den synviklen kan det finnas ett utrymme för att stödja anonymiseringsprocessen på samma rättsliga grund som den ursprungliga personuppgiftsbehandlingen. Detta gäller särskilt om anonymiseringsprocessen kan likställas med en radering av personuppgifter.[22] Huruvida en ytterligare behandling av personuppgifter är förenlig med det ursprungliga ändamålet avgörs i en förenlighetsprövning.[23] Huruvida anonymiseringsprocessen kan anses vara en ytterligare behandling i GDPR:s mening är en intressant fråga som bör utredas vidare. I samband med detta är det även intressant att undersöka hur PDL och annan relevant nationell lagstiftning ser på konceptet ytterligare behandling.

---

[17] Vid behandling av personuppgifter för hälso- och sjukvårdsändamål är det främst art. 6.1 c GDPR (rättslig förpliktelse) eller 6.1 e GDPR (allmänt intresse eller myndighetsutövning) som kan vara tillämpliga. För behandling av känsliga personuppgifter inom hälso- och sjukvård är art. 9.2.h GDPR tillämplig under förutsättning att det finns en lagreglerad tystnadsplikt enligt artikel 9.3 GDPR.
[18] 2:4 PDL.
[19] 2:4 1 st. 4 p. PDL
[20] 2:4 1 st. 6 p. PDL
[21] Jmf. art. 5.1 b och art. 6.4 GDPR. Se även WP 216 s. 7-8.
[22] Principen om lagringsminimering (art. 5.1 e GDPR) kräver inte uttryckligen att personuppgifter ”raderas” utan att personuppgifter ”inte förvaras i en form som möjliggör identifiering av den registrerade under en längre tid än vad som är nödvändigt”.
[23] Jmf. art. 6.4 GDPR. Se även AG-29, Yttrande 03/2013 om ändamålsbegränsning, WP 203, s. 20-36.

En näraliggande fråga är att använda extern expertis för framtagning av NLP-modeller. Utifrån GDPR kan en leverantör som involveras i ett sådant arbete betraktas som personuppgiftsbiträde, givet att den behandlar personuppgifter på vårdgivarens uppdrag. Detta kan exempelvis bli aktuellt när leverantören ska vara behjälplig vid anonymisering av patientjournaltexter.[24] GDPR ställer särskilda krav på ett sådant biträdesförhållande som bland annat ett avtal som reglerar personuppgiftsbiträdets hantering av personuppgifterna.[25] Samtidigt ställer GDPR krav på lagreglerad tystnadsplikt vid behandling av personuppgifter inom hälso- och sjukvården.[26] Bestämmelser om tystnadsplikt och sekretess inom den offentliga förvaltningen finns i huvudsak i OSL.

Enligt Justitieombudet (JO) kan en pseudonymisering av personuppgifter medföra att det saknas risk för att en enskild person ska lida skada eller men, och att uppgifterna därmed kan lämnas ut i enlighet med OSL.[27] Det är emellertid viktigt att en myndighet även vid utlämnande av uppgifter i pseudonymiserad form gör en noggrann sekretessprövning utifrån förhållandena i varje enskilt fall.[28] JO:s beslut avsåg uppgifter ur patientjournaler som en offentlig vårdgivare lämnat ut till ett personuppgiftsbiträde i pseudonymiserad form. Utvecklingen som sker på området är dock dynamisk. En offentlig vårdgivare som vill anlita ett personuppgiftsbiträde i samband med framtagning av en sådan NLP-modell som beskrivs ovan bör därför göra en noggrann prövning i varje enskilt fall utifrån rättsläget som gäller vid den aktuella tidpunkten.

Sammanfattningsvis konstateras att juridiken avseende användning av anonymiserade patientjournaler för träning av NLP-modeller hos en offentlig vårdgivare är komplicerad. Det finns inga enkla svar utan det krävs nyanserade juridiska analyser och bedömningar i varje enskilt fall. Vidare konstateras att mer arbete för att analysera gränsdragningen mellan personuppgifter och icke-personuppgifter i sammanhanget patientjournalstexter behöver göras. För att uppnå

[24] Art. 4.8 GDPR.
[25] Jmf. art. 28 GDPR.
[26] Jmf. art. 9.2.h GDPR och 9.3 GDPR.
[27] Med behandling av personuppgifter på ett sätt som innebär att personuppgifterna inte längre kan tillskrivas en specifik registrerad utan att kompletterande uppgifter används, under förutsättning att dessa kompletterande uppgifter förvaras separat och är föremål för tekniska och organisatoriska åtgärder som säkerställer att personuppgifterna inte tillskrivs en identifierad eller identifierbar fysisk person, se art. 4.5 GDPR. Till skillnad från anonymiserade personuppgifter omfattas pseudonymiserade personuppgifter av GDPR och PDL.
[28] JO dnr 6794-2017 och 6864-2017.

regelefterlevnad bör juridisk kompetens involveras löpande i NLP-projekt. Slutligen visar ovanstående sammanställning att anonymisering av personuppgifter endast är en av många aspekter som behöver beaktas vid träning av NLP-modeller.