Zaštita osobnih podataka i umjetna inteligencija:

Smjernice za subjekte koji razvijaju modele i sustave umjetne inteligencije

PRIVATNOST I ZAŠTITA PODATAKA

ODGOVORAN RAZVOJ UMJETNE INTELIGENCIJE

USKLAĐENOST I TRANSPARENTNOST

Agencija za zaštitu osobnih podataka

SADRŽAJ

1. Uvod

Ove smjernice Agencija za zaštitu osobnih podataka (AZOP) izradila je kao praktičan vodič za organizacije koje razvijaju, treniraju, testiraju, integriraju ili koriste sustave umjetne inteligencije u okolnostima koje uključuju obradu osobnih podataka. Njihova je svrha pomoći organizacijama u razumijevanju i dosljednoj primjeni zahtjeva Opće uredbe o zaštiti podataka kroz tipične faze životnog ciklusa sustava umjetne inteligencije: od prikupljanja i odabira podataka, preko treniranja i evaluacije modela, do uvođenja sustava u proizvodno okruženje, njegova nadzora, daljnjih izmjena te, u konačnici, povlačenja sustava iz uporabe.

Smjernice naglašavaju pristup utemeljen na riziku te potrebu da se odluke povezane s dizajnom sustava, korištenim podacima i primijenjenim kontrolama pravodobno i jasno dokumentiraju kako bi organizacija mogla dokazati usklađenost s Općom uredbom o zaštiti podataka.

Smjernice su namijenjene:

- dobavljačima sustava umjetne inteligencije *(AI providers)*;
- subjektima koji uvode sustave umjetne inteligencije u svoj rad *(AI deployers)*, primjerice organizacijama koje kupuju, ugrađuju ili koriste model za obradu osobnih podataka korisnika svojih usluga, klijenata, zaposlenika ili građana;
- službenicima za zaštitu podataka, pravnim timovima, timovima za usklađenost i timovima za informacijsku sigurnost u organizacijama.

Akt o umjetnoj inteligenciji i Opću uredbu o zaštiti podataka treba promatrati kao komplementarne propise koji se međusobno nadopunjuju. Usklađenost s Općom uredbom o zaštiti podataka temeljni je preduvjet za razvoj zakonite, etične i antropocentrične umjetne inteligencije, usklađene s temeljnim pravima, demokratskim načelima i vladavinom prava.

Člankom 2. stavkom 7. Uredbe (EU) 2024/1689 Europskog parlamenta i Vijeća od 13. lipnja 2024. o utvrđivanju usklađenih pravila o umjetnoj inteligenciji, odnosno Akta o umjetnoj inteligenciji, izričito je propisano da se pravo Unije o zaštiti osobnih podataka, privatnosti i povjerljivosti komunikacija primjenjuje na osobne podatke koji se obrađuju u vezi s pravima i obvezama utvrđenima tim Aktom. Nadalje, Akt o umjetnoj inteligenciji ne dovodi u pitanje primjenu postojećeg prava Unije u području zaštite osobnih podataka,

privatnosti i povjerljivosti komunikacija, uključujući Opću uredbu o zaštiti podataka, Uredbu (EU) 2018/1725, Direktivu 2002/58/EZ i Direktivu (EU) 2016/680.

Ti propisi ne predstavljaju prepreku inovacijama, već osiguravaju regulatorni okvir koji omogućuje razvoj umjetne inteligencije na način koji maksimizira društvenu korist, uz istodobno smanjenje rizika za prava i slobode pojedinaca. Važno je naglasiti da se pravo EU-a o zaštiti podataka u potpunosti primjenjuje na obradu osobnih podataka uključenih u životni ciklus sustava umjetne inteligencije.

U fokusu ovih smjernica su pitanja koja su se u praksi pokazala presudnima za usklađivanje rješenja utemeljenih na umjetnoj inteligenciji s Općom uredbom o zaštiti podataka. To osobito uključuje: utvrđivanje uloga sudionika u obradi podataka, odnosno razlikovanje voditelja i izvršitelja obrade te uređenje njihovih međusobnih odnosa; procjenu obrađuju li se osobni podaci ili se model i/ili skupovi podataka mogu smatrati anonimiziranima; definiranje svrhe obrade i odabir odgovarajuće pravne osnove; poštovanje temeljnih načela obrade; transparentnost i ostvarivanje prava ispitanika; tehničku i integriranu zaštitu podataka, odnosno zaštitu podataka po dizajnu i po zadanim postavkama *(data protection by design and by default)*; sigurnost obrade; prijenose osobnih podataka; te provođenje procjene učinka na zaštitu podataka, odnosno DPIA-e.

Smjernice se u relevantnim dijelovima dotiču i odabranih obveza iz Akta o umjetnoj inteligenciji kada su one neposredno povezane sa zaštitom osobnih podataka i upravljanjem rizicima, pri čemu se zadržava jasno razgraničenje između područja primjene tih propisa. Time se organizacijama nastoji olakšati usklađivanje zahtjeva koji se u praksi često preklapaju na razini procesa, primjerice u području upravljanja rizicima, dokumentacije, nadzora i postupanja u slučaju incidenata, ali imaju različitu pravnu svrhu, pravne posljedice i kriterije procjene.

U tom smislu, smjernice ne nude „jedno rješenje za sve", nego strukturirani okvir za procjenu, donošenje i dokumentiranje odluka u konkretnom kontekstu, uz pretpostavku da se svaka procjena provodi od slučaja do slučaja.

Smjernice nisu zamjena za pravno savjetovanje u pojedinačnom slučaju niti predstavljaju iscrpan prikaz svih obveza koje mogu proizlaziti iz drugih propisa, sektorskih pravila, ugovornih zahtjeva ili specifične arhitekture pojedinog sustava umjetne inteligencije.

Definicije

Sustav umjetne inteligencije - znači strojni sustav dizajniran za rad s promjenjivim razinama autonomije i koji nakon uvođenja može pokazati prilagodljivost te koji, za eksplicitne ili implicitne ciljeve, iz ulaznih vrijednosti koje prima, zaključuje kako generirati izlazne vrijednosti kao što su predviđanja, sadržaj, preporuke ili odluke koji mogu utjecati na fizička ili virtualna okruženja.[1] U uvodnoj izjavi 12. Akta o umjetnoj inteligenciji dodatno se objašnjava pojam „UI sustav". Ključna značajka UI sustava njihova je sposobnost izvođenja zaključaka. Tehnike koje omogućuju zaključivanje pri izgradnji UI sustava uključuju pristupe strojnog učenja i pristupe temeljene na logici i znanju.

UI model – definiran je neizravno u Aktu o umjetnoj inteligenciji: „Iako su UI modeli ključne komponente UI sustava, oni sami po sebi nisu UI sustavi." UI modeli zahtijevaju dodavanje dodatnih komponenti kao što je korisničko sučelje, kako bi postali UI sustavi. UI modeli obično su integrirani u UI sustave i čine njihov sastavni dio.

Generativna umjetna inteligencija - podskup umjetne inteligencije koji se odnosi na modele koji mogu generirati visokokvalitetan tekst, slike i drugi sadržaj na temelju podataka na kojima su trenirani. Za objašnjenje tehničke pozadine generativne umjetne inteligencije, koriste se složeni modeli strojnog učenja koji se nazivaju modeli dubokog učenja (eng. *deep learning*), a koji oponašaju procese učenja i donošenja odluka ljudskog mozga.

Veliki jezični modeli (LLMs) - vrsta modela strojnog učenja treniranih na velikim količinama tekstualnih podataka te mogu generirati odgovore na prirodnom jeziku na širok raspon ulaza, oslanjajući se na uočene uzorke i odnose između riječi, izraza i konteksta. Generativna umjetna inteligencija široka je kategorija koja obuhvaća modele sposobne za stvaranje novog sadržaja (npr. teksta, slika, zvuka), dok su LLM-ovi specifična primjena generativne umjetne inteligencije usmjerena na obradu i generiranje jezika.

UI agenti - softverski sustavi koji koriste jedan ili više UI modela (često LLM) kako bi, uz zadani cilj, mogli planirati korake, prikupljati informacije te izvršavati radnje u digitalnom okruženju (npr. pozivati alate i API-je, pretraživati baze znanja, izrađivati i pokretati zadatke), uz različitu razinu ljudskog nadzora. Agentska umjetna inteligencija označava napredniji oblik takvih sustava, u kojem agent pokazuje veću autonomiju: iterativno planira, provodi radnje, provjerava rezultate i prilagođava sljedeće korake radi ostvarenja zadanih ciljeva, umjesto da se ograničava na generiranje odgovora na pojedinačne upite.

Osobni podatak - podatak koji se odnosi na pojedinca čiji je identitet utvrđen ili se može utvrditi („ispitanik"); pojedinac čiji se identitet može utvrditi jest osoba koja se može

[1] Europska komisija je u svojoj Komunikaciji od 29. srpanja 2025. izdala Smjernice o definiciji sustava umjetne inteligencije koje su dostupne na: https://azop.hr/wp-content/uploads/2025/07/HR-Commission_Guidelines_on_the_definition_of_an_artificial_intelligence_system_established_by_Regulation_EU_20241689_AI_ActCroatian_5Hn0qpSjpULo4aaI89xid0vxNY_118621.pdf

identificirati izravno ili neizravno, osobito uz pomoć identifikatora kao što su ime, identifikacijski broj, podaci o lokaciji, mrežni identifikator ili uz pomoć jednog ili više čimbenika svojstvenih za fizički, fiziološki, genetski, mentalni, ekonomski, kulturni ili socijalni identitet tog pojedinca. U uvodnoj izjavi 26. Opće uredbe navodi se da se načela zaštite podataka ne bi trebala primjenjivati na anonimne informacije, odnosno informacije koje se ne odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi, uzimajući u obzir „sva sredstva" koja voditelj obrade ili bilo koja druga osoba „mogu po svemu sudeći upotrijebiti". To uključuje: i. podatke koji nikada nisu bili povezani s osobom čiji je identitet utvrđen ili se može utvrditi i ii. osobne podatke koji su anonimizirani tako da se identitet ispitanika ne može utvrditi ili se više ne može utvrditi.

Posebne kategorije osobnih podataka – su podaci koji otkrivaju rasno ili etničko podrijetlo, politička mišljenja, vjerska ili filozofska uvjerenja ili članstvo u sindikatu te obrada genetskih podataka, biometrijskih podataka u svrhu jedinstvene identifikacije pojedinaca, kao i podaci koji se odnose na zdravlje, spolnom životu ili seksualnoj orijentaciji pojedinca.

Voditelj obrade - fizička ili pravna osoba, tijelo javne vlasti, agencija ili drugo tijelo koje samo ili zajedno s drugima određuje svrhe i sredstva obrade osobnih podataka; kada su svrhe i sredstva takve obrade utvrđeni pravom Unije ili pravom države članice, voditelj obrade ili posebni kriteriji za njegovo imenovanje mogu se predvidjeti pravom Unije ili pravom države članice.

Izvršitelj obrade - fizička ili pravna osoba, tijelo javne vlasti, agencija ili drugo tijelo koje obrađuje osobne podatke u ime voditelja obrade.

Zajednički voditelji – voditelji obrade koji zajednički odrede svrhe i načine obrade osobnih podataka.

Dobavljač (eng. Provider) - fizička ili pravna osoba, tijelo javne vlasti, javna agencija ili drugo javno tijelo koje razvija UI sustav ili UI model opće namjene ili koji ima razvijen UI sustav ili UI model opće namjene i stavlja ga na tržište ili stavlja UI sustav u upotrebu pod vlastitim imenom ili žigom, uz plaćanje ili besplatno.

Subjekt koji uvodi sustav umjetne inteligencije (eng. Deployer) - znači fizička ili pravna osoba, tijelo javne vlasti, javna agencija ili drugo javno tijelo koje upotrebljava UI sustav u okviru svoje nadležnosti, osim ako se UI sustav upotrebljava u osobnoj neprofesionalnoj djelatnosti.

Ovlašteni zastupnik - znači fizička ili pravna osoba koja se nalazi u Uniji ili ima poslovni nastan u Uniji, koju je dobavljač UI sustava ili UI modela opće namjene pisanim putem ovlastio da u njegovo ime izvršava i provodi obveze i postupke utvrđene u ovoj Uredbi i koja je takvo ovlaštenje prihvatila.

Uvoznik - fizička ili pravna osoba koja se nalazi u Uniji ili ima poslovni nastan u Uniji i koja stavlja na tržište UI sustav s imenom ili žigom fizičke ili pravne osobe s poslovnim nastanom u trećoj zemlji.

Distributer - fizička ili pravna osoba u opskrbnom lancu koja nije dobavljač ni uvoznik i koja stavlja UI sustav na tržište Unije.

Operator - dobavljač, proizvođač proizvoda, subjekt koji uvodi sustav, ovlašteni zastupnik, uvoznik ili distributer.

RAG – služi za optimizaciju performansi modela umjetne inteligencije povezivanjem s vanjskim bazama znanja. RAG pomaže velikim jezičnim modelima (LLM) da isporuče relevantnije odgovore veće kvalitete. Generativni UI modeli treniraju se na velikim skupovima podataka i koriste te informacije za generiranje izlaza. Međutim, skupovi podataka za treniranje su konačni i ograničeni na informacije kojima UI programer može pristupiti - radovi u javnoj domeni, internetski članci, sadržaj društvenih medija i drugi javno dostupni podaci. RAG omogućuje generativnim UI modelima pristup dodatnim vanjskim bazama znanja, kao što su interni organizacijski podaci, znanstveni časopisi i specijalizirani skupovi podataka. Integriranjem relevantnih informacija u proces generiranja, chatbotovi i drugi alati za obradu prirodnog jezika (NLP) mogu stvoriti točniji sadržaj specifičan za domenu bez potrebe za daljnjom obukom.

Support Vector Machine Model - je nadzirani algoritam strojnog učenja koji klasificira podatke pronalaženjem optimalnog pravca ili hiperravnine koja maksimizira udaljenost između svake klase u N-dimenzionalnom prostoru.

Popis kratica

API – Application Programming Interface

AZOP - Agencija za zaštitu osobnih podataka

DPIA – Procjena učinka na zaštitu podataka

EDPB – European Data Protection Board

GDPR – General Data Protection Regulation

LLM – Large Language Model; Veliki jezični model

NLP – Natural Language Processing

RAG – Retrieval-Augmented Generation

RLHF - Reinforcement Learning from Human Feedback

SVM –Support Vector Machine

UI – umjetna inteligencija

UKS – Uredba o kibernetičkoj sigurnosti

ZKS – Zakon o kibernetičkoj sigurnosti

2. Modeli umjetne inteligencije i osobni podaci: u kojim slučajevima se primjenjuje Opća uredba o zaštiti podataka?

Ocjena može li se određeni UI model smatrati anonimnim ne provodi se apstraktno ni unaprijed, nego od slučaja do slučaja. Pri tome je potrebno uzeti u obzir tehnička svojstva modela, vrstu i izvor podataka korištenih za treniranje, kontekst uporabe, razumno predvidive napade, dostupna dodatna sredstva za identifikaciju te tehničke i organizacijske mjere koje je organizacija provela radi smanjenja rizika ponovne identifikacije.

UI modeli trenirani s pomoću osobnih podataka ne mogu se automatski smatrati anonimnima. U pojedinim slučajevima model može sam po sebi uključivati informacije koje se odnose na pojedinca čiji je identitet utvrđen ili se može utvrditi. To će osobito biti slučaj kada je model posebno osmišljen za pružanje informacija o određenim osobama ili za stavljanje takvih podataka na raspolaganje. Primjeri uključuju generativni UI model dorađen na glasovnim snimkama pojedinca radi oponašanja njegova glasa, model osmišljen da na upit o određenoj osobi generira osobne podatke korištene u treniranju ili model za pretraživanje fotografija pojedinaca.

Čak i kada model nije namjerno osmišljen za generiranje osobnih podataka iz skupa podataka za treniranje, informacije iz tog skupa, uključujući osobne podatke, mogu biti sadržane ili odražene u parametrima modela. Takve informacije mogu biti tehnički kodirane ili predstavljene u strojno čitljivom obliku, ali to ne isključuje mogućnost da se, uz odgovarajuće metode, iz modela ili njegovih izlaza izdvoje ili zaključe podaci koji se odnose na pojedince.

UI model treniran na osobnim podacima može se smatrati anonimnim samo ako je, uzimajući u obzir sva sredstva za koja je razumno vjerojatno da će ih upotrijebiti voditelj obrade ili druga osoba, vjerojatnost identifikacije pojedinaca beznačajna. To osobito znači da mora biti beznačajna:

1. vjerojatnost izravnog ili probabilističkog izdvajanja osobnih podataka koji se odnose na pojedince čiji su podaci korišteni za razvoj modela; i
2. vjerojatnost dobivanja takvih osobnih podataka, namjerno ili nenamjerno, postavljanjem upita modelu.

Drugim riječima, organizacija koja tvrdi da je UI model anoniman mora moći dokumentirano dokazati da se osobni podaci povezani s podacima za treniranje ne mogu izdvojiti iz modela te da se rezultati dobiveni uporabom modela ne odnose na ispitanike čiji su osobni podaci korišteni za treniranje.

Pri procjeni anonimnosti modela potrebno je osobito uzeti u obzir:

a) značajke podataka za treniranje, modela i postupka treniranja, uključujući jedinstvenost zapisa, razinu preciznosti informacija, agregaciju, randomizaciju i druge čimbenike koji mogu utjecati na mogućnost identifikacije;
b) kontekst u kojem se model koristi, objavljuje ili stavlja na raspolaganje, uključujući ograničenja pristupa, ugovorne i organizacijske mjere te pravne zaštitne mehanizme;
c) dodatne informacije koje bi mogle biti dostupne voditelju obrade ili drugim osobama i koje bi mogle omogućiti identifikaciju pojedinaca;
d) troškove, vrijeme i tehničke mogućnosti potrebne za dobivanje ili povezivanje takvih dodatnih informacija;
e) tehnologiju dostupnu u trenutku obrade, kao i razumno predvidiv tehnološki razvoj.

U tom smislu nije dovoljno samo tvrditi da model ne sadržava "čitljive" osobne podatke. Potrebno je procijeniti i mogućnost neizravnog izdvajanja podataka, zaključivanja o članstvu u skupu podataka, zaključivanja o atributima, inverzije modela, rekonstrukcijskih napada, eksfiltracije podataka ili regurgitacije podataka za treniranje.

Organizacija bi u dokumentaciji trebala moći pokazati koje je mjere poduzela tijekom razvoja modela kako bi smanjila rizik identifikacije. To uključuje osobito:

- odabir izvora podataka za treniranje i isključenje neprimjerenih ili nepotrebnih izvora;
- procjenu jesu li za treniranje mogli biti korišteni anonimni, sintetski ili pseudonimizirani podaci;
- primjenu strategija smanjenja količine podataka i filtriranja nepotrebnih osobnih podataka prije treniranja;
- korištenje metoda kojima se smanjuje prekomjerna prilagodba modela i rizik memoriranja podataka;
- primjenu odgovarajućih tehnika zaštite privatnosti, primjerice diferencijalne privatnosti, kada je to primjereno;
- provedbu tehničkih i organizacijskih mjera za ograničavanje mogućnosti izdvajanja osobnih podataka putem upita ili izravnog pristupa modelu.

Pouzdanost procjene anonimnosti treba biti potkrijepljena dokumentacijom i testiranjem. To može uključivati unutarnje ili vanjske revizije, pregled koda, analizu odabranih mjera, model prijetnji, procjenu rizika, testiranje otpornosti na relevantne napade i dokumentiranje rezultata testiranja. Opseg i dubina testiranja trebaju biti razmjerni vrsti modela, vrsti podataka, predviđenom kontekstu uporabe i mogućim posljedicama za ispitanike.

Osobito kada je model treniran na osobnim podacima ili kada postoji razumna mogućnost da bi se osobni podaci mogli izravno ili neizravno izdvojiti iz modela ili njegovih

izlaza, preporučuje se dokumentirano testiranje otpornosti na relevantne napade, uključujući:

- zaključivanje o članstvu i atributima;
- eksfiltraciju podataka;
- regurgitaciju podataka za treniranje;
- inverziju modela;
- rekonstrukcijske napade.

Dokumentacija kojom se potkrepljuje tvrdnja da se model može smatrati anonimnim trebala bi, ovisno o okolnostima, uključivati najmanje:

a) procjenu učinka na zaštitu podataka ili obrazloženu odluku da DPIA nije potrebna;

b) mišljenje ili povratne informacije službenika za zaštitu podataka, ako je imenovan ili je trebao biti imenovan;

c) opis tehničkih i organizacijskih mjera poduzetih tijekom razvoja modela radi smanjenja vjerojatnosti identifikacije;

d) model prijetnji i procjenu rizika na kojima se te mjere temelje;

e) opis izvora podataka za treniranje i mjera primijenjenih na svaki izvor;

f) dokumentaciju o testiranju otpornosti modela na tehnike ponovne identifikacije i izdvajanja podataka;

g) rezultate provedenih testova, uključujući informacije o tome tko je proveo testiranje, kada, kojom metodom i u kojem opsegu;

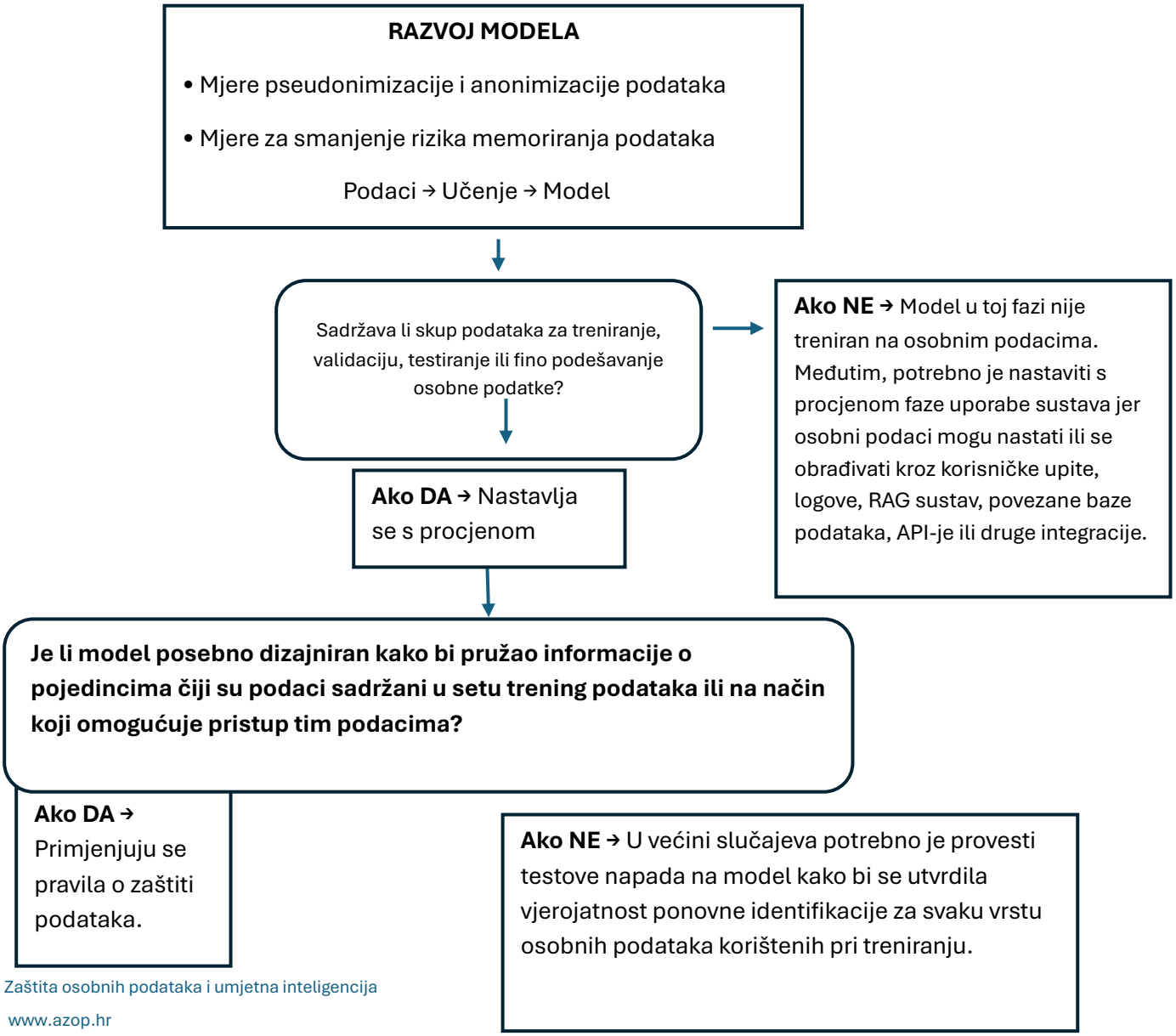
h) informacije o preostalim rizicima i mjerama koje su priopćene subjektima koji model koriste ili integriraju u svoje sustave.

Prikazana procjena odnosi se na pitanje može li se sam model, u određenom kontekstu i na temelju konkretnih dokaza, smatrati anonimnim. Međutim, čak i ako se zaključi da model kao takav ne sadržava osobne podatke ili se može smatrati anonimnim, potrebno je zasebno procijeniti obrađuje li sustav u koji je model integriran osobne podatke u fazi uporabe, primjerice putem korisničkih upita, logova, RAG sustava, povezanih baza

znanja, API-ja ili drugih integracija. Pseudonimizirani podaci i dalje su osobni podaci te podliježu Općoj uredbi o zaštiti podataka. Pseudonimizacija je važna zaštitna mjera koja može smanjiti rizike za ispitanike, ali sama po sebi ne dovodi do anonimizacije. Da bi se podaci ili model mogli smatrati anonimnima, organizacija mora moći dokazati da pojedinca više nije moguće identificirati sredstvima za koja je razumno vjerojatno da će ih koristiti voditelj obrade ili druga osoba. Smanjenje rizika koje proizlazi iz pseudonimizacije može biti relevantno pri procjeni zakonitosti obrade, osobito u okviru testa ravnoteže kod legitimnog interesa iz članka 6. stavka 1. točke f) Opće uredbe o zaštiti podataka, pod uvjetom da su ispunjeni svi ostali zahtjevi za primjenu te pravne osnove.

Više o pseudonimizaciji dostupno je u smjernicama na: https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_hr.

Sljedeća matrica pokazuje način na koji voditelji obrade mogu razmotriti primjenjuje li se na njihov model Opća uredba o zaštiti podataka ili se model smatra anonimnim.



Je li analizom statusa modela zaključeno da je vjerojatnost ponovne identifikacije zanemariva za svaku vrstu osobnih podataka?

Ako NE

→ Primjenjuju se pravila o zaštiti podataka.

Ako DA→ Model se u tom konkretnom kontekstu i na temelju dokumentirane procjene može smatrati anonimnim u odnosu na podatke korištene za treniranje.

POSTUPANJE KADA SE PRIMJENJUJU PRAVILA O ZAŠTITI PODATAKA

Opcija 1: Primjena Opće uredbe o zaštiti podataka

Nastaviti postupanje u skladu s pravilima o zaštiti privatnosti:

• Informiranje ispitanika i ostvarivanje njihovih prava

• Provjera zakonitosti obrade od strane modela

• Smanjenje rizika od izvlačenja podataka i napada, osobito primjenom odgovarajućih tehničkih i organizacijskih mjera

Opcija 2: Provesti enkapsulaciju modela u sustav te implementirati tehničke i organizacijske mjere radi smanjenja vjerojatnosti izvlačenja podataka iz sustava.

Razvoj sustava UI baziranog na neanonimiziranom modelu u skladu s Općom uredbom o zaštiti podataka

Implementacija mjera za smanjenje mogućnosti reidentifikacije iz pristupa sustava

U svim slučajevima potrebno je provesti testove napada na sustav kako bi se odredila vjerojatnost identifikacije osobnih podataka uključenih u trening.

Ako NE

→ Primjenjuju se pravila o zaštiti podataka.

Je li analizom sustava zaključeno da je vjerojatnost ponovne identifikacije zanemariva za svaku vrstu osobnih podataka?

Ako DA

→ Smatra se da je model van opsega Opće uredbe o zaštiti podataka.

Prikazana matrica korisna je kao pomoćni alat za početnu procjenu statusa UI modela u odnosu na osobne podatke. Međutim, zaključak da se model može smatrati anonimnim ne smije se donositi automatski niti odvojeno od konkretnog konteksta uporabe. Čak i ako se model, u odnosu na podatke korištene za treniranje, može smatrati anonimnim, potrebno je zasebno procijeniti obrađuje li UI sustav osobne podatke u fazi uporabe, primjerice putem korisničkih upita, logova, RAG sustava, povezanih baza znanja, API-ja ili vanjskih pružatelja usluga.

Tablica. Dodatna kontrolna pitanja uz procjenu može li se UI model smatrati anonimnim

Dodatno kontrolno pitanje	Zašto je važno
Je li model treniran, validiran, testiran ili fino podešen na osobnim podacima?	Ako su u fazi razvoja korišteni osobni podaci, Opća uredba o zaštiti podataka primjenjuje se na obradu u toj fazi te je potrebno utvrditi svrhu i pravnu osnovu obrade, uloge sudionika i primijenjene zaštitne mjere.
Može li se iz modela ili njegovih izlaza izravno ili neizravno izdvojiti, rekonstruirati ili zaključiti osobne podatke?	Ako postoji razumna mogućnost izdvajanja, regurgitacije podataka za treniranje, zaključivanja o članstvu ili atributima, inverzije modela, rekonstrukcijskih napada ili eksfiltracije podataka, model se ne može smatrati anonimnim.
Obrađuje li UI sustav osobne podatke u fazi uporabe?	Čak i ako se sam model može smatrati anonimnim u odnosu na podatke korištene za treniranje, sustav može obrađivati osobne podatke kroz korisničke upite, logove, IP adrese, RAG baze znanja, učitane dokumente, API-je ili vanjske pružatelje usluga.
Je li zaključak o anonimnosti dokumentiran i testiran?	Organizacija mora moći nadzornom tijelu za zaštitu podataka (AZOP-u) dokazati zaključak o anonimnosti, uključujući opis korištenih podataka, primijenjenih mjera, provedenih testiranja, ograničenja i preostalih rizika.
Jesu li pseudonimizirani podaci tretirani kao osobni podaci?	Pseudonimizacija je važna zaštitna mjera za smanjenje rizika, ali sama po sebi ne dovodi do anonimizacije. Pseudonimizirani podaci i dalje su osobni podaci i podliježu Općoj uredbi o zaštiti podataka.

Važno je naglasiti da se zaključak da je određeni model izvan područja primjene Opće uredbe o zaštiti podataka može odnositi samo na konkretan model, konkretan kontekst i dokumentirane dokaze dostupne u trenutku procjene, a ne na sve buduće načine uporabe tog modela ili sustava u koji je model integriran.

Primjer iz prakse: virtualni pomoćnik na mrežnoj stranici javnog tijela

Javno tijelo planira uvesti virtualnog pomoćnika na svojoj mrežnoj stranici radi pružanja općih informacija građanima o pitanjima iz svoje nadležnosti. Sustav se temelji na tehnologiji umjetne inteligencije i koristi javno dostupne informacije objavljene na mrežnoj stranici javnog tijela. Sustav nije namijenjen donošenju odluka o pojedincima niti pružanju individualiziranih pravnih savjeta, već informativnom usmjeravanju korisnika.

Iako sustav nije namijenjen obradi osobnih podataka i od korisnika ne traži unos osobnih podataka, to samo po sebi ne znači da se Opća uredba o zaštiti podataka ne primjenjuje. U praksi se mogu obrađivati tehnički podaci, primjerice IP adresa, log zapisi, podaci o korištenju mrežne stranice ili drugi podaci povezani s radom virtualnog pomoćnika. Osim toga, korisnik može u slobodni tekstualni upit sam unijeti osobne podatke, primjerice ime i prezime, kontakt podatke, podatke o političkom djelovanju, zdravlju ili druge informacije koje nisu nužne za dobivanje opće informacije.

U takvom slučaju organizacija treba najprije utvrditi dolazi li do obrade osobnih podataka. Ako se obrađuju tehnički podaci korisnika, log zapisi ili sadržaj upita koji može sadržavati osobne podatke, potrebno je primijeniti Opću uredbu o zaštiti podataka u dijelu u kojem dolazi do takve obrade.

Organizacija zatim treba jasno utvrditi uloge svih uključenih subjekata. Tijelo koje odlučuje da će se virtualni pomoćnik koristiti na njegovoj mrežnoj stranici, za svrhe koje ono odredi i pod uvjetima koje ono postavi, u pravilu će imati ulogu voditelja obrade. Potrebno je posebno razjasniti uloge pružatelja tehnološkog rješenja, pružatelja AI usluge, pružatelja hostinga i drugih uključenih subjekata te, prema potrebi, urediti odnose sukladno članku 28. Opće uredbe o zaštiti podataka.

U pogledu svrhe obrade mogu se razlikovati najmanje dvije skupine svrha. Prva je omogućavanje rada virtualnog pomoćnika radi pružanja općih informacija korisnicima. Druga je osiguravanje tehničkog funkcioniranja, dostupnosti, integriteta, otpornosti i sigurnosti sustava, uključujući vođenje nužnih log zapisa, sprječavanje zlouporaba, otkrivanje tehničkih pogrešaka i zaštitu informacijskog sustava. Za svaku svrhu potrebno je zasebno utvrditi odgovarajuću pravnu osnovu.

Ako virtualnog pomoćnika koristi javno tijelo u okviru svojih službenih zadaća, pravna osnova za obradu osobnih podataka u pravilu se ne bi trebala temeljiti na privoli korisnika. U takvim slučajevima potrebno je razmotriti postoji li odgovarajuća pravna osnova u izvršavanju zadaće od javnog interesa ili službene ovlasti, odnosno u ispunjavanju pravne obveze, ovisno o konkretnom propisu i okolnostima obrade.

Posebnu pozornost treba posvetiti načelu smanjenja količine podataka. Ako sustav nije namijenjen obradi osobnih podataka, potrebno ga je oblikovati tako da se obrada osobnih

podataka svede na najmanju moguću mjeru. To uključuje jasnu obavijest korisnicima da ne unose osobne podatke niti posebne kategorije osobnih podataka, tehničke mjere za ograničavanje ili filtriranje nepotrebnih podataka, ograničeno vođenje logova, kratke rokove čuvanja, pseudonimizaciju ili anonimizaciju gdje je primjenjivo te zabranu korištenja korisničkih upita za treniranje ili poboljšanje modela ako za to ne postoji odgovarajuća pravna osnova i jasno određena svrha.

Ako korisnik ipak unese osobne podatke koji nisu potrebni za davanje odgovora, sustav bi trebao biti konfiguriran tako da korisnika upozori da takvi podaci nisu potrebni, da ih ne obrađuje dalje nego što je nužno za funkcioniranje sustava i da ih ne zadržava dulje nego što je potrebno. Posebno je važno spriječiti pohranu nepotrebnih osobnih podataka u logovima ili njihovo prosljeđivanje trećim stranama kada to nije nužno.

Organizacija mora korisnicima pružiti jasne, sažete i lako dostupne informacije o obradi osobnih podataka, uključujući informacije o voditelju obrade, službeniku za zaštitu podataka, svrhama i pravnim osnovama obrade, kategorijama podataka, primateljima, mogućem prijenosu podataka izvan Europskog gospodarskog prostora, rokovima čuvanja i pravima ispitanika.

Ako se koristi vanjski pružatelj AI usluge, potrebno je provjeriti gdje se podaci stvarno obrađuju, imaju li pružatelj i njegovi podizvršitelji pristup sadržaju korisničkih upita i tehničkim podacima, koriste li se ti podaci za vlastite svrhe pružatelja usluge ili za treniranje modela, u kojim se rokovima brišu te dolazi li do prijenosa osobnih podataka izvan Europskog gospodarskog prostora.

Prije stavljanja takvog sustava u produkciju potrebno je razmotriti provedbu procjene učinka na zaštitu podataka, osobito ako sustav omogućuje slobodan unos teksta, koristi novu tehnologiju, uključuje vanjske pružatelje AI usluga, može dovesti do nenamjernog unosa osobnih ili posebnih kategorija osobnih podataka ili se koristi u osjetljivom javnom kontekstu.

U skladu s načelom tehničke i integrirane zaštite podataka, virtualni pomoćnik treba biti dizajniran tako da prema zadanim postavkama obrađuje samo nužne podatke, ograničava pohranu i daljnju obradu sadržaja upita, onemogućuje ili bitno ograničava korištenje korisničkih upita za druge svrhe, posebno za treniranje modela, te osigurava da osobni podaci nisu prema zadanim postavkama dostupni neograničenom broju osoba.

Ako sustav pruža samo opće informativne odgovore i ne donosi odluke koje proizvode pravne učinke za korisnike ili na njih na sličan način značajno utječu, u pravilu se ne radi o automatiziranom pojedinačnom donošenju odluka iz članka 22. Opće uredbe o zaštiti podataka. Međutim, korisnicima treba jasno naznačiti da komuniciraju s virtualnim pomoćnikom te da odgovori nemaju karakter službene pojedinačne odluke, pravnog savjeta ili obvezujućeg tumačenja.

Kontrolna pitanja za organizacije

Prije uvođenja virtualnog pomoćnika ili sličnog AI sustava organizacija bi trebala osobito provjeriti:

- obrađuju li se tehnički podaci korisnika, log zapisi ili sadržaj korisničkih upita;
- može li korisnik u slobodni tekst unijeti osobne ili posebne kategorije osobnih podataka;
- tko određuje svrhe i sredstva obrade;
- jesu li uloge voditelja, izvršitelja, podizvršitelja ili zajedničkih voditelja jasno uređene;
- koja je pravna osnova za svaku svrhu obrade;
- koriste li se korisnički upiti za treniranje, fino podešavanje ili poboljšanje modela;
- gdje se podaci obrađuju i dolazi li do prijenosa izvan EGP-a;
- koji su rokovi čuvanja logova i sadržaja upita;
- jesu li korisnici jasno upozoreni da ne unose nepotrebne osobne podatke;
- postoji li potreba za provedbom DPIA-e;
- jesu li osigurane mjere iz članaka 25. i 32. Opće uredbe o zaštiti podataka;
- je li službenik za zaštitu podataka uključen u procjenu prije produkcije.

3. Određivanje uloge dobavljača i subjekta koji uvodi sustav umjetne inteligencije u kontekstu Opće uredbe o zaštiti podataka

Dobavljači i subjekti koji uvode sustav umjetne inteligencije mogu imati različite uloge sukladno članku 4. Opće uredbe o zaštiti podataka, odnosno mogu biti voditelji ili izvršitelji obrade, odnosno zajednički voditelji obrade.

Uloga pojedinog sudionika u smislu Opće uredbe o zaštiti podataka ne određuje se prema nazivu iz ugovora, poslovnom modelu ili terminologiji iz Akta o umjetnoj inteligenciji, nego prema stvarnom utjecaju na svrhe i bitna sredstva obrade osobnih podataka. Procjenu je stoga potrebno provoditi od slučaja do slučaja, na temelju konkretnih činjenica, tehničke arhitekture i ugovornih odnosa.

U situaciji kada navedeni subjekti određuju svrhe i sredstva obrade osobnih podataka tada će biti voditelji obrade, odnosno izvršitelji obrade kada obrađuju osobne podatke u ime voditelja obrade. Kada ti subjekti zajedno određuju svrhe i načine obrade, npr. dva ili

više dobavljača odnosno subjekta koji uvode sustav umjetne inteligencije, tada će biti zajednički voditelji obrade.

Voditelj obrade je taj koji je u obvezi uskladiti sve aktivnosti obrade osobnih podataka u sustavima umjetne inteligencije s Općom uredbom o zaštiti podataka te mora biti u mogućnosti Agenciji dokazati usklađenost. Voditelj obrade je subjekt koji određuje svrhu i sredstva obrade, odnosno koji određuje zašto i kako će se osobni podaci obrađivati. Osnovna sredstva obrade su usko povezana sa samom obradom i njezinim opsegom, kao što su kategorija podataka, hardver i softver koji će se koristiti, trajanje obrade, kategorija primatelja i kategorija ispitanika.

U situacijama kada je organizacija izvršitelj obrade, tada mora slijediti upute voditelja obrade odnosno primjenjuje se članak 28. Opće uredbe o zaštiti podataka kojim se definiraju obveze samog izvršitelja. Moguće je i postojanje više izvršitelja obrade, pri čemu je primarni izvršitelj odgovoran s aspekta zaštite osobnih podataka prema voditelju obrade za sve ostale izvršitelje obrade u lancu tj. podizvršitelje.

Navedeni pojmovi voditelja i izvršitelja obrade pojašnjeni su u Smjernicama Europskog odbora za zaštitu podataka koje su dostupne na poveznici:

https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-072020-concepts-controller-and-processor-gdpr_hr.

Kvalifikacija pojedinog subjekta u skladu s Općom uredbom o zaštiti podataka mora se provoditi od slučaja do slučaja.

Dobavljač će općenito biti voditelj obrade za aktivnosti obrade koje je proveo radi razvoja UI sustava jer će dobavljač općenito definirati svrhu i sredstva takvih aktivnosti obrade, uključujući namjenu i funkcionalnosti modela, predviđeni kontekst njegove implementacije te odabir i pripremu skupova podataka za treniranje koji mogu sadržavati osobne podatke. Na primjer, dobavljač koji pokreće razvoj UI sustava i odabire i stvara skup osobnih podataka za treniranje sustava umjetne inteligencije kojeg razvija smatrao bi se voditeljem obrade. To bi bio slučaj i kad bi dobavljač povjerio izradu takvog skupa podataka pružatelju usluga s pomoću dovoljno detaljnih dokumentiranih uputa (pri čemu potonji djeluje kao izvršitelj obrade). Ako dobavljač zaključi da se na određeni model ili sustav ne primjenjuje Opća uredba o zaštiti podataka (primjerice zato što se smatra da ne obrađuje osobne podatke), mora osigurati da postoji odgovarajuća, dokumentirana analiza tog statusa, koju je moguće predočiti Agenciji na zahtjev. Navedena dokumentacija treba jasno obrazložiti i potkrijepiti zaključak da je vjerojatnost ponovne identifikacije fizičkih osoba, čiji su podaci sadržani u skupu podataka za treniranje, iz modela ili sustava zanemariva, uz opis primijenjenih tehničkih i organizacijskih mjera te relevantnih pretpostavki i ograničenja analize.

Primjer 1 — razvoj sustava preporuka na temelju korisničkih podataka

Organizacija koja upravlja internetskom platformom za dijeljenje i gledanje videosadržaja želi razviti sustav umjetne inteligencije za personalizirane preporuke sadržaja. U tu svrhu koristi skup podataka svojih korisnika, primjerice podatke o pregledanim videosadržajima, trajanju gledanja, interakcijama s platformom, korisničkim preferencijama i drugim podacima prikupljenima tijekom korištenja usluge.

Budući da organizacija određuje svrhu obrade, odnosno razvoj i uporabu sustava za preporuke, kao i bitna sredstva obrade, uključujući vrste podataka koje će koristiti, način treniranja modela i način primjene rezultata sustava, u tom slučaju djeluje kao voditelj obrade osobnih podataka.

Primjer 2 — sustav umjetne inteligencije za procjenu kreditne sposobnosti

Banka uvodi sustav umjetne inteligencije radi potpore u procjeni kreditne sposobnosti klijenata. Sustav obrađuje osobne podatke kao što su prihodi, kreditna povijest, postojeće financijske obveze, status zaposlenja, dob i drugi relevantni financijski pokazatelji, kako bi pružio preporuku ili potporu pri odlučivanju o odobrenju ili odbijanju kreditnog zahtjeva.

U ovom slučaju banka je subjekt koji uvodi sustav umjetne inteligencije u svoj rad *(deployer)*, ali istodobno djeluje i kao voditelj obrade osobnih podataka jer određuje svrhu obrade, odnosno procjenu kreditne sposobnosti klijenata, bira sustav koji će koristiti, određuje koje će se kategorije osobnih podataka obrađivati, kako će se rezultati sustava koristiti u kreditnom procesu te nadzire uporabu sustava pod svojom nadležnošću.

U takvom slučaju banka mora osigurati da je obrada osobnih podataka zakonita, transparentna, razmjerna i ograničena na ono što je nužno za konkretnu svrhu, uz provedbu odgovarajućih tehničkih i organizacijskih mjera te, prema potrebi, procjene učinka na zaštitu podataka.

Subjekt koji uvodi sustav umjetne inteligencije općenito će biti voditelj obrade osobnih podataka u kontekstu uporabe tog sustava pod svojom nadležnošću jer će utvrditi zašto upotrebljava određeni UI sustav te koje podatke obrađuje. To će općenito biti slučaj kada subjekt koji uvodi UI sustav odluči upotrebljavati UI sustav u određenom kontekstu za određene ispitanike i određene kategorije osobnih podataka.

Organizacija koja pokreće razvoj sustava UI i stvara set podataka za trening koje je sama odabrala može se smatrati voditeljem obrade, a jednako vrijedi i za organizaciju koja povjerava stvaranje seta podataka drugoj organizaciji dajući joj detaljne upute.

Dvoje ili više voditelja obrade mogu zajedno odrediti svrhu i sredstva obrade te su u tom slučaju zajednički voditelji obrade. U tom segmentu bitno je vidjeti njihove obveze sukladno članku 26. Opće uredbe o zaštiti podataka koje se prvenstveno sastoji ispunjavanja obveze transparentnosti i ostvarivanja prava ispitanika, a što se čini

sukladno međusobnom dogovoru odnosno sukladno propisu kada isti određuje takvu situaciju. U tom segmentu naglašavamo kako ispitanik može svoja prava sukladno Općoj uredbi o zaštiti podataka, bez obzira na dogovorene uvjete između zajedničkih voditelja, ostvarivati u vezi sa svakim voditeljem obrade, kao i protiv svakog od njih.

Primjer 1 — zajednički razvoj UI sustava za zapošljavanje

Društvo A, društvo B i društvo C zajednički razvijaju sustav umjetne inteligencije za potporu u odabiru kandidata za zapošljavanje. Društva su se dogovorila o zajedničkoj svrsi sustava, vrstama osobnih podataka kandidata koje će koristiti, kriterijima za treniranje modela te načinu na koji će se rezultati sustava koristiti u postupcima zapošljavanja.

Iako se podaci ne prenose u jednu zajedničku bazu, nego se koristi, primjerice, federativno učenje, sva tri društva zajednički odlučuju o ključnim elementima obrade. U takvim okolnostima mogu se smatrati zajedničkim voditeljima obrade, neovisno o tome što se osobni podaci tehnički obrađuju lokalno kod svakog pojedinog društva.

Primjer 2 — zajedničko korištenje podataka za daljnji razvoj modela

Banka koristi sustav umjetne inteligencije za otkrivanje prijevara koji je razvilo društvo specijalizirano za UI rješenja. Ako banka i navedeno društvo dogovore da će zajednički koristiti podatke o ponašanju korisnika radi treniranja nove verzije modela, potrebno je procijeniti njihove uloge u smislu Opće uredbe o zaštiti podataka.

Ako obje strane zajednički određuju svrhu takve obrade, primjerice poboljšanje modela za detekciju prijevara, te bitna sredstva obrade, uključujući vrste podataka koje će se koristiti, način treniranja modela i uvjete daljnje uporabe rezultata, banka i društvo koje razvija UI rješenje mogu se smatrati zajedničkim voditeljima obrade.

U tom slučaju moraju na transparentan način urediti svoje međusobne odgovornosti, osobito u pogledu informiranja ispitanika, ostvarivanja njihovih prava, sigurnosti obrade i postupanja u slučaju povrede osobnih podataka, sukladno članku 26. Opće uredbe o zaštiti podataka.

U situaciji kada organizacija obrađuje osobne podatke u ime voditelja obrade tada se smatra da je ista izvršitelj obrade, npr. ako subjekt koji uvodi UI sustav odredi dobavljaču koje će podatke upotrijebiti za razvoj ili poboljšanje (novog) UI sustava. Tom prilikom izvršitelj obrade mora ispuniti svoje obaveze sukladno članku 28. Opće uredbe o zaštiti podataka, kao što je ranije navedeno.

Primjer 1 — dobavljač kao izvršitelj obrade

Društvo A angažira softversku tvrtku za razvoj internog UI sustava za upravljanje sastancima, primjerice za automatsku transkripciju, sažimanje zapisnika i predlaganje zadataka nakon sastanka. Društvo A određuje svrhu obrade, kategorije osobnih podataka koje će se obrađivati, korisnike sustava, rokove čuvanja podataka i način korištenja rezultata. Softverska tvrtka razvija i održava sustav isključivo prema dokumentiranim uputama društva A te ne koristi podatke za vlastite svrhe, primjerice za treniranje vlastitih modela ili razvoj novih komercijalnih funkcionalnosti. U takvom slučaju društvo A je voditelj obrade, a softverska tvrtka izvršitelj obrade.

Primjer 2 — bolnica kao voditelj, softverska tvrtka kao izvršitelj

Bolnica koristi UI sustav za pomoć pri dijagnostici pacijenata. Sustav obrađuje zdravstvene podatke pacijenata koje u sustav unosi bolnica. Bolnica odlučuje zašto se sustav koristi, za koje pacijente, u kojem kliničkom procesu, koje se kategorije podataka obrađuju i kako će zdravstveni radnici koristiti rezultate sustava. Softverska tvrtka koja je razvila sustav pruža tehničko održavanje i podršku, ali osobne podatke pacijenata obrađuje isključivo prema dokumentiranim uputama bolnice i ne koristi ih za vlastite svrhe. U takvim okolnostima bolnica je voditelj obrade, a softverska tvrtka izvršitelj obrade.

Primjer 3 — HR alat kao izvršitelj obrade

Softverska tvrtka nudi UI alat za automatsku analizu životopisa kandidata. Korporativni klijenti unose životopise kandidata u sustav i samostalno određuju svrhu obrade, kriterije selekcije, relevantne parametre, kategorije kandidata i način na koji će se rezultati koristiti u postupku zapošljavanja. Softverska tvrtka samo omogućuje tehničko funkcioniranje alata i obrađuje osobne podatke kandidata isključivo prema uputama pojedinog klijenta. Ne koristi podatke kandidata za vlastito treniranje modela, razvoj vlastitih proizvoda ili druge vlastite svrhe. U takvom slučaju klijent je voditelj obrade, a softverska tvrtka izvršitelj obrade.

Ako softverska tvrtka sama određuje kriterije rangiranja kandidata, razvija profil idealnog kandidata na temelju podataka svih klijenata ili koristi podatke kandidata za poboljšanje vlastitog modela, njezina se uloga mora ponovno procijeniti. U takvim okolnostima ona može biti zaseban voditelj obrade ili zajednički voditelj obrade.

4. Definiranje svrhe sustava umjetne inteligencije

Svaki UI sustav koji uključuje obradu osobnih podataka mora imati jasno određenu, izričitu i legitimnu svrhu obrade. Svrha se mora definirati na početku projekta i dokumentirati na način koji omogućuje provjeru nužnosti i proporcionalnosti obrade, odabir relevantnih kategorija podataka, određivanje pravne osnove te uspostavu odgovarajućih zaštitnih mjera. Općenite formulacije poput „razvoj i unapređenje UI sustava" u pravilu nisu dovoljne.

U situacijama kada organizacija koja razvija UI sustav jasno zna koja će biti upotreba, sama upotreba će biti svrha razvojne faze, kao i faze uvođenja UI sustava.

Primjer
Organizacija uspostavlja bazu fotografija automobila u prometu na kojima se vide i osobe unutar samih vozila u svrhu kako bi se trenirao algoritam da procijeni koliko se osoba vozi u automobilu kako bi se doprinijelo optimizaciji gradskog prometa.

Međutim, situacija je kompleksnija kada je u pitanju razvoj UI sustava opće namjene koji mogu biti korišteni u različitim kontekstima i s različitom primjenom.

U takvim situacijama, određivanju svrhe ne smije se pristupiti preširoko npr. na način da se definira kao razvoj i unapređenje UI sustava, već je potrebno što detaljnije odrediti:

- Vrsta sustava koji se razvija (npr. razvoj velikog jezičnog modela, sustava računalnog vida ili generativnog UI sustava za slike, zvukove, videozapisi, računalni kod itd.);
- Tehničke izvedive funkcionalnosti i mogućnosti.

Također je dobra praksa dodatno opisati predviđeni način uporabe sustava: koje će zadatke sustav moći obavljati, u kojim će se kontekstima koristiti, koje su uporabe izričito isključene te hoće li sustav biti dostupan kao interna aplikacija, SaaS rješenje, API ili na drugi način.

Primjer
Organizacija izrađuje i javno objavljuje skup fotografija namijenjen treniranju modela računalnog vida za prepoznavanje objekata, primjerice ljudi, vozila i životinja. Budući da se na dijelu fotografija nalaze prepoznatljive fizičke osobe, takav skup podataka može sadržavati osobne podatke. Organizacija stoga mora prije objave jasno odrediti svrhu obrade, pravnu osnovu, uvjete daljnje uporabe skupa podataka i zaštitne mjere, osobito kako bi se spriječila uporaba podataka u nespojive ili visokorizične svrhe.

Podaci se ne smiju dalje obrađivati na način koji nije u skladu s tom početnom svrhom, drugim riječima načelo ograničenja svrhe ograničava način na koji voditelj obrade može upotrebljavati ili ponovno upotrebljavati te podatke u budućnosti.

I u tom dijelu valja imati na umu članak 6. stavak 4. Opće uredbe o zaštiti podataka koji govori o podudarnoj svrsi, odnosno korištenja osobnih podataka za druge svrhe od one za koju su podaci inicijalno obrađivani.

S obzirom na specifičnosti tehnologije moguće je svrhu odrediti odmah s početkom projekta, ali to ne mora biti tako.

Naime, ako se UI sustav razvija za jednu operativnu uporabu, smatra se da je svrha u fazi razvoja izravno povezana sa svrhom obrade u fazi uvođenja. Iz toga slijedi da će se, ako je svrha u fazi razvoja sama po sebi određena, izričita i legitimna, utvrditi i svrha u fazi uvođenja.

Kod razvoja UI sustava čija operativna uporaba u fazi uvođenja nije jasno utvrđena u fazi razvoja – sustavi opće namjene i temeljni modeli koji se mogu upotrebljavati u širokom rasponu i za koje je teško definirati dovoljno određenu i izričitu svrhu u fazi razvoja, svrha obrade tijekom faze razvoja može se smatrati određenom, izričitom i legitimnom samo ako je dovoljno specifična, tj. ako se kumulativno odnosi na:

- „tip” razvijenog sustava, kao što je, na primjer, razvoj velikog jezičnog modela (LLM), sustava računalnog vida ili generativnog UI sustava za slike, videozapise ili zvukove. Vrste sustava moraju biti predstavljene ispitanicima na dovoljno jasan i razumljiv način, uzimajući u obzir njihovu tehničku složenost i brz razvoj u tom području.
- tehnički izvedive funkcionalnosti i sposobnosti, što znači da voditelj obrade mora sastaviti popis sposobnosti koje razumno može predvidjeti u fazi razvoja.

Sljedeći primjeri pokazuju jasno određenu svrhu.

Primjeri

Organizacija razvija veliki jezični model za pomoć poslovnim korisnicima u izradi, prevođenju, sažimanju i jezičnom uređivanju tekstova. Svrha modela je pružanje jezične podrške, a ne donošenje odluka o pojedincima ili njihovo profiliranje.

Organizacija razvija model za obradu govora namijenjen automatskoj transkripciji audiozapisa, prepoznavanju jezika i poboljšanju kvalitete korisničke podrške. Model nije namijenjen jedinstvenoj identifikaciji govornika, profiliranju korisnika ili donošenju odluka koje bi mogle proizvesti pravne ili slične značajne učinke za pojedince.

S druge strane niže navedeni primjeri predstavljaju svrhu koja nije dovoljno određena i specificirana.

Primjeri
Organizacija razvija model generativne UI.

Organizacija razvija UI model za verifikaciju dobi korisnika, bez navođenja konteksta uporabe i načina obrade podataka.

4.1 Podudarna svrha

U situacijama kada voditelj obrade već posjeduje podatke temeljem kojih želi kreirati set podataka za trening modela, mora provjeriti ispunjava li sve uvjete koji su propisani u članku 6. Opće uredbe o zaštiti podataka.

Sukladno uvodnoj izjavi 50. Opće uredbe o zaštiti podataka, obrada osobnih podataka u svrhe različite od svrha za koje su podaci prvotno prikupljeni smjela bi se dopustiti samo ako je obrada usklađena sa svrhama za koje su osobni podaci prvotno prikupljeni. U takvom slučaju nije potrebna posebna pravna osnova zasebna od one kojom je dopušteno prikupljanje osobnih podataka. Ako je obrada potrebna za obavljanje zadaće koja se obavlja u javnom interesu ili pri izvršavanju službene ovlasti koju ima voditelj obrade, pravom Unije ili pravom države članice mogu se utvrditi i odrediti zadaće i svrhe za koje će se nastavak obrade smatrati usklađenim i zakonitim.

Sama činjenica da su osobni podaci javno dostupni na internetu ne znači da se oni mogu slobodno preuzeti, objediniti i koristiti za razvoj ili testiranje UI modela. Organizacija mora zasebno ocijeniti pravnu osnovu obrade, usklađenost daljnje obrade s izvornom svrhom, obvezu informiranja ispitanika, prirodu podataka, moguću prisutnost posebnih kategorija podataka, očekivanja ispitanika te okolnosti u kojima su podaci učinjeni javno dostupnima.

Nadalje, radi utvrđivanja je li svrha nastavka obrade usklađena sa svrhom prvotnog prikupljanja osobnih podataka, voditelj obrade nakon ispunjavanja svih zahtjeva zakonitosti izvorne obrade trebao bi uzeti u obzir, među ostalim, svaku vezu između te svrhe i svrhe planiranog nastavka obrade, kontekst u kojem su prikupljeni osobni podaci posebno opravdana očekivanja ispitanika koja se temelje na njihovom odnosu s voditeljem obrade u pogledu daljnje uporabe podataka, prirodu osobnih podataka, posljedice planiranog nastavka obrade za ispitanike i postojanje primjerenih zaštitnih mjera u izvornoj i planiranoj daljnjoj obradi.

S tim u vezi ukazujemo kako sukladno članku 6. stavku 4. Opće uredbe o zaštiti podataka, ako se obrada u svrhu koja je različita od svrhe koju su podaci prikupljeni ne temelji na privoli ispitanika ili na pravu Unije ili pravu države članice koje predstavlja nužnu i razmjernu mjeru u demokratskom društvu za zaštitu ciljeva iz članaka 23. stavka 1.,

voditelj obrade s ciljem utvrđivanja je li obrada u drugu svrhu u skladu sa svrhom u koju su osobni podaci prvotno prikupljeni, uzima u obzir, među ostalim:

a) svaku vezu između svrha prikupljanja osobnih podataka i svrha namjeravanog nastavka obrade;

b) kontekst u kojem su prikupljeni osobni podaci, posebno u pogledu odnosa između ispitanika i voditelja obrade;

c) prirodu osobnih podataka, osobito činjenicu obrađuju li se posebne kategorije osobnih podataka u skladu s člankom 9. ili osobni podaci koji se odnose na kaznene osude i kažnjiva djela u skladu s člankom 10.;

d) moguće posljedice namjeravanog nastavka obrade za ispitanike;

e) postojanje odgovarajućih zaštitnih mjera, koje mogu uključivati enkripciju ili pseudonimizaciju.

Naprijed navedenim kriterijima se omogućuje da se ponovno korištenje prethodno prikupljenih osobnih podataka uredi tako da se osigura ravnoteža između, s jedne strane, nužnosti predvidljivosti i pravne sigurnosti u vezi sa svrhama obrade prethodno prikupljenih osobnih podataka i, s druge strane, priznavanja određene fleksibilnosti u korist voditelja obrade u upravljanju tim podacima te oni stoga doprinose ostvarenju cilja koji se sastoji od osiguranja postojane i visoke razine zaštite pojedinaca. [2]

Usklađena daljnja obrada ne mora značiti da je nova svrha obrade „pod-svrha“ inicijalne svrhe obrade. Usklađenost postoji i kada je nova svrha drugačija, ali korelira s inicijalnom svrhom u smislu da se te svrhe provode zajedno, odnosno da su neposredno povezane u vremenskom i kontekstualnom smislu ili da je daljnja obrada logična posljedica inicijalne obrade.

Pri tome se svaka obrada nakon prikupljanja smatra „daljnjom obradom” i ona stoga treba, uz iznimke, ispunjavati zahtjev usklađenosti. Taj zahtjev odražava potrebu za konkretnom, logičnom i dovoljno bliskom vezom između svrhe prikupljanja podataka i njihove daljnje obrade. Drugim riječima, ta obrada ne smije biti nepovezana s izvornom svrhom prikupljanja podataka niti s njom u proturječnosti, a njezin sadržaj treba biti usklađen sa svrhom prikupljanja, neovisno o problematici u pogledu trajanja zadržavanja.[3]

S tim u vezi treba imati na umu, kako je Sud Europske unije, u prethodno navedenom predmetu, skrenuo pozornost na činjenicu da je pojam „obrada” široko definiran u članku 4. točki 2. Opće uredbe o zaštiti podataka tako da označava svaki postupak ili skup

[2] Vidjeti presudu od 20. listopada 2022., Digi Távközlési és Szolgáltató Kft. protiv Nemzeti Adatvédelmi és Információszabadság Hatóság, C-77/21, EU:C:2022:805, t. 37.

[3] Vidjeti mišljenje nezavisnog odvjetnika P. Pikamäea od 3. ožujka 2022., predmet C-77/21

postupaka koji se obavljaju na osobnim podacima ili na skupovima osobnih podataka, bilo automatiziranim bilo neautomatiziranim sredstvima kao što su prikupljanje, bilježenje i pohrana tih podataka.

Također, ako se voditelj obrade oslanja na usklađenost daljnje obrade, mora moći dokazati da je proveo procjenu usklađenosti.

Pri tome potrebno je procijeniti dodatne rizike za ispitanike kojima daljnja obrada može rezultirati. Daljnja obrada ne smije rezultirati značajno većim rizikom od inicijalne zakonite obrade, ako je kvalificirana kao usklađena. To će svakako ograničiti usklađenu daljnju obradu posebnih kategorija podataka. Oslanjanje na usklađenost ne može proširiti zakonitu obradu takvih podataka iznad onoga kako je to definirano člancima 9. i 10. Opće uredbe o zaštiti podataka.

U situaciji neusklađene daljnje obrade, samo privola ispitanika ili posebne zakonske odredbe sukladno članku 23. stavku 1. Opće uredbe o zaštiti podataka mogu biti zakonite pravne osnove zadiranja u načelo ograničenje svrhe. Pri tome, u stavku 2. navedenog članka je definirano koji su to minimalni zahtjevi koje zakonska odredba iz stavka 1. mora ispunjavati.

Primjer

Organizacija pruža korisnicima program za uređivanje teksta koji uključuje značajku generativne umjetne inteligencije za predlaganje nastavka rečenica, sažimanje i jezično uređivanje teksta. Korisnici tijekom korištenja alata ručno ispravljaju ili prilagođavaju tekst koji je sustav predložio. Organizacija nakon toga želi koristiti te ručne ispravke kako bi sustav bolje prepoznavao stil pisanja pojedinog korisnika i ubuduće mu nudio personalizirane prijedloge.

U takvom slučaju riječ je o daljnjoj obradi osobnih podataka jer se podaci nastali tijekom korištenja alata žele koristiti za dodatnu svrhu odnosno personalizaciju usluge. Organizacija mora procijeniti je li ta nova svrha usklađena s prvotnom svrhom pružanja alata za uređivanje teksta. Pri tome osobito treba uzeti u obzir jesu li korisnici mogli razumno očekivati takvu uporabu podataka, jesu li o njoj jasno informirani, mogu li je odbiti, koje se vrste podataka analiziraju, koliko dugo se podaci čuvaju i postoje li odgovarajuće zaštitne mjere.

Organizacija koja želi koristiti javno dostupan skup podataka s interneta za razvoj ili treniranje UI modela ne smije pretpostaviti da je takva uporaba zakonita samo zato što su podaci dostupni online. Iako je za zakonitost prvotnog prikupljanja i objave skupa podataka primarno odgovoran subjekt koji ga je prikupio i objavio, organizacija koja podatke ponovno koristi mora provjeriti postoji li pravna osnova za obradu osobnih podataka, je li nova svrha usklađena s prvotnom svrhom te potječe li skup podataka iz

zakonitog izvora. Organizacija ne smije koristiti skupove podataka za koje je očito ili se razumno može zaključiti da potječu iz nezakonitog izvora, uključujući neovlašteni pristup, povredu osobnih podataka ili objavu neovlašteno pribavljene baze podataka.

Organizacije koje žele ponovno koristiti osobne podatke trebaju provjeriti:

- opis skupa podataka i izvor;
- da kreiranje i objavljivanje seta podataka nije očiti rezultat kaznenog djela ili slično;
- da nema jasne sumnje da je set podataka zakonit (npr. nedostatak pravne osnove prilikom početne obrade);
- da set podataka ne sadrži posebne kategorije podataka ili u slučaju da sadrži postoji jasna pravna osnova iz članka 6. stavka 1. Opće uredbe o zaštiti podataka i iznimka iz članka 9. stavka 2. navedenog propisa.

Propusti koje je počinio izvorni voditelj obrade pri prikupljanju podataka ne dovode automatski do nezakonitosti njihove kasnije uporabe od strane druge organizacije. Međutim, organizacija koja ponovno koristi podatke mora moći dokazati da je njihovu obradu provela u skladu s Općom uredbom o zaštiti podataka.

Pritom je ključno procijeniti izvor podataka i okolnosti njihova pribavljanja. Ako organizacija, postupajući s razumnom pažnjom, nije imala razloga sumnjati u nezakonitost izvora, tada sama činjenica da su podaci možda prethodno prikupljeni nezakonito ne mora nužno utjecati na zakonitost njihove daljnje uporabe. No, takav zaključak moguć je samo ako su provedene odgovarajuće provjere i ako ne postoje indikatori koji bi upućivali na nezakonitost.

Stoga Agencija preporučuje sklapanje ugovora sa stvarateljem skupa podataka, kada je to moguće, pri čemu bi ugovor trebao urediti najmanje sljedeća pitanja:

- Opis skupa podataka
- Planirano razdoblje pohrane
- Ograničenja pristupa
- Ograničenja uporabe
- Licence
- Opis postupka pristupa podacima
- Mjere evidentiranja (logiranja) pristupa podacima
- Primjeri poznatih projekata koji koriste skup podataka
- Primjenjivi pravni okvir (zaštita osobnih podataka, intelektualno vlasništvo i sl.)
- Postojanje osobnih podataka, posebno osjetljivih podataka ili drugih zaštićenih kategorija podataka
- Naziv organizacije koja je pružila set podataka
- Svrha seta podataka
- Potreba koja je potaknula prikupljanje podataka

- Predviđena uporaba
- Izvor podataka
- Način prikupljanja
- Kategorije ispitanika
- Kategorije ispitanika
- Opis vrste podataka (npr. tablični podaci, slike, vremenske serije, video ili audio zapisi)
    - Opis metapodataka
    - Količina podataka po kategorijama
    - Količina podataka po ispitaniku
    - Kategorije (objekti, situacije, osobe i sl.) za koje je testirana reprezentativnost
- Pokazatelji reprezentativnosti (npr. statistička distribucija kategorija podataka)
- Poznate ili očekivane pristranosti
    - Tehnike za mjerenje i ublažavanje pristranosti
        - na razini skupa podataka
        - tijekom treniranja modela
        - na izlazima modela
- Kvaliteta podataka
    - Poznate ili očekivane pogreške u podacima
    - Izvori šuma i netočnosti (uz navođenje utjecaja kada je poznat)
    - Uzroci koji mogu dovesti do smanjenja točnosti podataka (npr. ažuriranja, zastarjelost, drift podataka i sl.)
    - Preporučena podjela skupa podataka na skupove za treniranje, validaciju i testiranje
- Pravna osnova za prikupljanje podataka
- Popis izravno ili neizravno identifikacijskih podataka
- Popis anonimiziranih podataka
- Popis posebnih kategorija podataka u smislu članka 9. Uredbe
    - Primjenjiva iznimka za obradu posebnih kategorija podataka
- Opis postupka informiranja ispitanika i korištenih komunikacijskih kanala
- Mjere zaštite podataka
    - Metode anonimizacije ili pseudonimizacije
- Etička provjera podataka
- Analiza rizika za prava i slobode pojedinaca
- Sigurnosne mjere
- Primjenjivi referentni okviri, standardi i certifikati
- Ostvarivanje prava ispitanika
- Postupak za ispitanike
- Postupak za podnošenje zahtjeva za ostvarivanje prava
- Postupak anotacije podataka
    - Opis
    - Provedba (interno ili putem pružatelja usluge)
    - Razina automatizacije (nimalo, djelomično, u potpunosti)
    - Demografska reprezentativnost tima za anotaciju
    - Jamstva društvene odgovornosti pružatelja usluge anotacije
    - Opis postupka provjere anotacija
    - Opis korištenih metoda predobrade
- Opis planiranih postupaka održavanja i podrške (trenutno i u slučaju objave nove verzije)
    - Planirani datum završetka održavanja i podrške
    - Opis postupka ažuriranja skupa podataka

- Kanali za informiranje o ažuriranjima i razvoju skupa podataka
- Opis postupka doprinosa poboljšanju ili održavanju skupa podataka

Kada okolnosti jasno ukazuju na rizično ili nezakonito podrijetlo podataka, primjerice kada se podaci pribavljaju s izvora za koje je poznato da distribuiraju podatke iz povreda sigurnosti ili neovlaštenih pristupa, organizacija je dužna to uzeti u obzir. U takvim situacijama ne može se pozivati na neznanje jer se smatra da je organizacija mogla i morala biti svjesna rizika, a time se dovodi u pitanje zakonitost obrade.

5. Usklađivanje s načelima zaštite osobnih podataka

Načela obrade osobnih podataka iz članka 5. Opće uredbe o zaštiti podataka polazište su svake zakonite obrade. Kod razvoja i uporabe UI sustava ta načela ne primjenjuju se samo u jednoj fazi, nego prate cijeli životni ciklus sustava, od prikupljanja i pripreme podataka, preko uvođenja sustava u rad i njegova nadzora, do naknadnih izmjena i povlačenja sustava iz uporabe. Općom uredbom o zaštiti podataka definirano je kako osobni podaci moraju biti:

- **zakonito, pošteno i transparentno obrađivani s obzirom na ispitanika (načelo zakonitosti, poštenosti i transparentnosti);**

U svakoj fazi razvoja, testiranja i uporabe sustava umjetne inteligencije obrada osobnih podataka mora biti zakonita, poštena i transparentna. Organizacija stoga mora utvrditi odgovarajuću pravnu osnovu za obradu, procijeniti moguće rizike i posljedice za ispitanike te ih jasno i razumljivo informirati o svrsi i načinu obrade njihovih osobnih podataka,

mogućim učincima obrade, mjerama za ublažavanje rizika i pravima koja im pripadaju na temelju Opće uredbe o zaštiti podataka.

- **prikupljeni u posebne, izričite i zakonite svrhe te se dalje ne smiju obrađivati na način koji nije u skladu s tim svrhama (načelo ograničavanja svrhe);**

Sustavi umjetne inteligencije, a osobito oni koji se temelje na metodama strojnog učenja, zahtijevaju velike količine podataka. Ti podaci ključni za treniranje sustava, za njegovo testiranje, usporedbu i validaciju. Sastavljanje kvalitetnih skupova podataka podrazumijeva značajan trud, budući da podaci moraju biti pravilno označeni (anotirani), očišćeni, standardizirani i pripremljeni za uporabu. Načelo smanjenja količine podataka propisuje da osobni podaci koji se prikupljaju i obrađuju moraju biti primjereni, relevantni i ograničeni na ono što je nužno u odnosu na jasno definiranu svrhu obrade. Posebna pažnja mora se posvetiti prirodi podataka, pri čemu ovo načelo zahtijeva još strožu primjenu kada je riječ o posebnim kategorijama osobnih podataka. Iako uporaba velikih količina podataka može biti nužna za razvoj i funkcioniranje sustava umjetne inteligencije, primjena načela smanjenja količine podataka ne predstavlja prepreku takvoj obradi, već zahtijeva pažljivo planiranje i kontrolu. U fazi učenja (treniranja algoritma) može biti prihvatljivo osigurati pristup većim količinama i raznolikim skupovima podataka, pod uvjetom da se primjenjuju odgovarajuće tehničke i organizacijske mjere koje su proporcionalne rizicima koje takva obrada nosi. Pri tome se osobito mora voditi računa o prirodi podataka, njihovoj količini i svrsi razvoja sustava. Te mjere mogu uključivati, primjerice:

  - ograničen pristup podacima na mali broj ovlaštenih osoba,
  - obradu podataka u točno određenom vremenskom razdoblju,
  - pseudonimizaciju podataka.

Nakon završetka faze učenja i prije prelaska u produkcijsku fazu (odnosno primjene sustava izvan kontroliranog, „laboratorijskog“ okruženja), potrebno je uvesti stroža pravila za nadzor obrade osobnih podataka. To uključuje ograničavanje obrade isključivo na one kategorije podataka koje su se tijekom faze razvoja pokazale nužnima za postizanje unaprijed određene svrhe, kao i definiranje dodatnih tehničkih i organizacijskih mjera zaštite. Važno je pritom uzeti u obzir da se zahtjevi za obradu podataka u produkcijskom okruženju razlikuju od onih u fazi dizajna i razvoja, osim u slučajevima kada i sama razvojna faza uključuje povišene rizike za prava i slobode ispitanika.

- **primjereni, relevantni i ograničeni na ono što je nužno u odnosu na svrhe u koje se obrađuju (načelo smanjenja količine podataka);**

Prilikom ispunjavanje načela i obrade trening podataka nužno je imati u vidu sljedeće korake:

  - provedbu čišćenja podataka;

- identifikaciju relevantnih podataka;
- implementacija mjera iz članka 25. Opće uredbe o zaštiti podataka (tehnička i integrirana zaštita osobnih podataka)
- monitoring i ažuriranje podataka;
- dokumentaciju korištenih podataka.

Načelo smanjenja količine podataka ne zabranjuje obradu velikih količina podataka, ali zahtijeva da se obrađuju samo oni podaci koji su primjereni, relevantni i nužni za jasno određenu svrhu obrade. U kontekstu sustava umjetne inteligencije ta se procjena mora provoditi i tijekom životnog ciklusa sustava, a ne samo u trenutku prikupljanja ili odabira podataka. Promjene u podacima, kontekstu uporabe ili ponašanju korisnika mogu dovesti do toga da određene kategorije podataka više nisu relevantne, nužne ili prikladne za daljnju obradu.

Prilikom određivanja podataka koji će se obrađivati, voditelj obrade osobito treba uzeti u obzir:

- volumen podataka (npr. broj ispitanika, preciznost podataka, distribuciju podataka...);
- kategoriju podataka (npr. godine, spol, slike lica...);
- tipologiju podatka (npr. pravi, sintetizirani, pseudonimizirani...);
- izvore podataka.

Prilikom odabira modela odnosno algoritma, voditelj obrade mora voditi računa o načelu smanjenja količine podataka. Ako se predviđena svrha može postići primjenom algoritma ili modela koji zahtijeva obradu manje količine osobnih podataka, a pritom omogućuje ostvarenje iste svrhe, prednost treba dati takvom, manje invazivnom rješenju.

- **točni i prema potrebi ažurni (načelo točnosti);**

Osobni podaci koji se koriste u razvoju, testiranju i uporabi sustava umjetne inteligencije moraju biti točni i, prema potrebi, ažurni. Netočni, zastarjeli ili nereprezentativni podaci mogu dovesti do pogrešnih rezultata sustava, nepravednog postupanja prema ispitanicima ili donošenja nepravilnih odluka. Voditelj obrade stoga treba osigurati postupke provjere kvalitete podataka, ispravljanja netočnih podataka te redovitog praćenja njihove relevantnosti i ažurnosti tijekom životnog ciklusa sustava.

- **čuvani u obliku koji omogućuje identifikaciju ispitanika samo onoliko dugo koliko je potrebno u svrhe radi kojih se osobni podaci obrađuju (načelo ograničenja pohrane);**

Opća uredba o zaštiti podataka zahtijeva da se unaprijed odredi razdoblje čuvanja podataka, nakon kojeg se podaci moraju izbrisati ili, u određenim slučajevima, arhivirati. Ovo razdoblje čuvanja mora utvrditi voditelj obrade, uzimajući u obzir svrhu zbog koje su osobni podaci prikupljeni. Drugim riječima osobni podaci ne smiju se čuvati neograničeno. Pri tome, u kontekstu implementacije sustava umjetne inteligencije, u pojedinim slučajevima može postojati potreba za duljim čuvanjem osobnih podataka u odnosu na druge oblike obrade. To je, primjerice, slučaj kod sastavljanja skupova podataka za treniranje modela i razvoj novih sustava, kao i radi ispunjavanja zahtjeva za sljedivost, evaluaciju učinkovitosti i mjerenje performansi sustava tijekom njegova korištenja u produkcijskom okruženju. Međutim, potreba za definiranjem razdoblja čuvanja podataka ne predstavlja prepreku za obradu osobnih podataka u sustavima umjetne inteligencije, već osigurava da takva obrada bude zakonita i transparentna. Ovo razdoblje uvijek mora biti razmjerno svrsi obrade. Na primjer, ako je svrha duljeg čuvanja podataka praćenje performansi sustava, ta svrha mora biti jasno definirana i planirana, a podaci koji se čuvaju u tu svrhu moraju biti pažljivo odabrani. Sama činjenica da se želi pratiti učinkovitost sustava tijekom vremena nije dovoljna da bi se opravdalo dugotrajno zadržavanje svih osobnih podataka. Potrebno je osigurati da se čuvaju isključivo oni podaci koji su za tu svrhu nužni.

- **obrađivani na način kojim se osigurava odgovarajuća sigurnost osobnih podataka, uključujući zaštitu od neovlaštene ili nezakonite obrade te od slučajnog gubitka, uništenja ili oštećenja primjenom odgovarajućih tehničkih ili organizacijskih mjera (načelo cjelovitosti i povjerljivosti).**

Nužno je voditi računa i o sigurnosti samih podataka i modela te primijeniti odgovarajuće tehničke i organizacijske mjere zaštite sukladno članku 32. Opće uredbe o zaštiti podataka. Uz tradicionalne rizike kibernetičke sigurnosti, potrebno je uzeti u obzir i sigurnosno-kibernetičke rizike koji su specifično povezani s područjem umjetne inteligencije. Svakako u tom segmentu voditeljima obrade mogu pomoći međunarodno priznati standardi poput ISO i NIST i drugi koji pružaju adekvatne okvire za pitanje tradicionalne kibernetičke sigurnost. Također, s obzirom na adresiranje problema kibernetičke sigurnosti, EU je donijela brojne propise iz ove domene, među kojima su svakako najpoznatiji Direktiva (EU) 2022/2555 Europskog parlamenta i Vijeća od 14. prosinca 2022. o mjerama za visoku zajedničku razinu kibersigurnosti širom Unije, izmjeni Uredbe (EU) br. 910/2014 i Direktive (EU) 2018/1972 i stavljanju izvan snage Direktive (EU) 2016/1148 (Direktiva NIS 2) i Uredba (EU) 2022/2554 Europskog parlamenta i Vijeća od 14. prosinca 2022. o digitalnoj operativnoj otpornosti za financijski sektor i izmjeni uredbi (EZ) br. 1060/2009, (EU) br. 648/2012, (EU) br. 600/2014, (EU) br. 909/2014 i (EU) 2016/1011 (DORA). Za subjekte na koje se ne primjenjuje DORA već NIS 2 direktiva, važno je spomenuti kako je u Hrvatskoj navedena direktiva transponirana u nacionalno zakonodavstvo Zakonom o kibernetičkoj sigurnosti (NN 14/2024) odnosno Uredbom o kibernetičkoj sigurnosti (NN 135/2024). Navedeni nacionalni propisi, kao i drugi

dokumenti koje je izdao nadležni regulator daju popis mjera koje organizacije mogu odnosno moraju implementirati.

Preporučeni set tradicionalnih mjera u području kibernetičke sigurnosti odnosi se na sljedeće mjere kojima se:

- Osigurava popis i kontrola nad imovinom
- Osigurava popis i kontrola nad softverima
- Uspostavljaju tehničke kontrole kojima identificira, klasificira, sigurno obrađuje i briše podatke
- Uspostavlja sigurnu konfiguraciju imovine i softvera
- Implementira upravljanje računima i pristupom
- Nadziru ranjivosti sustava
- Upravlja revizijskim logovima
- Šiti e-pošta i pretraživanje interneta
- Uspostavlja zaštita od zlonamjernih programa
- Omogućuje povrat podataka
- Upravlja mrežnom infrastrukturom
- Nadzire i brani mreža
- Podiže svjesnost i vještine o kibernetičkoj sigurnosti i sigurnosti općenito
- Upravlja trećim stranama odnosno dobavljačima
- Upravlja kroz cijeli životni ciklus alatima
- Uspostavlja odgovor na incidente
- Provodi penetracijsko testiranje

Navedene mjere nisu iscrpan popis svih mogućih sigurnosnih mjera, osobito s obzirom na kontinuirani razvoj tehnologije i specifične rizike povezane s UI sustavima. Ovisno o konkretnom sustavu i kontekstu njegove uporabe, potrebno je razmotriti i dodatne mjere, poput zaštite od malicioznih promptova, sprječavanja trovanja podataka, kontrole izlaznih rezultata velikog jezičnog modela i drugih primjerenih tehničkih i organizacijskih mjera.

Voditelj obrade odgovoran je za usklađenost sa člankom 5. stavkom 1. te je mora biti u mogućnosti dokazati ("pouzdanost").

Načelo pouzdanosti, propisano člankom 5. stavkom 2. Opće uredbe o zaštiti podataka, znači da je voditelj obrade odgovoran za poštovanje načela obrade osobnih podataka te mora biti u mogućnosti dokazati usklađenost s tim načelima. Usklađenost se stoga ne iscrpljuje u formalnom navođenju pravne osnove ili općem pozivanju na zaštitu podataka, nego zahtijeva stvarnu provedbu odgovarajućih mjera i njihovo dokumentiranje.

U kontekstu sustava umjetne inteligencije to je osobito važno jer obrada osobnih podataka može uključivati velike skupove podataka, složene tehničke postupke, više

uključenih subjekata, vanjske pružatelje usluga, modele koji se naknadno prilagođavaju te rizike koji nisu uvijek odmah vidljivi ispitanicima. Organizacija zato mora moći pokazati koje osobne podatke obrađuje, u koju svrhu, na kojoj pravnoj osnovi, tko ima pristup podacima, koliko se dugo čuvaju, koriste li se za treniranje ili daljnje poboljšanje modela te koje su mjere poduzete radi smanjenja rizika za prava i slobode pojedinaca.

Dokazivanje usklađenosti u pravilu uključuje mapiranje aktivnosti obrade i tokova podataka, vođenje evidencija aktivnosti obrade, procjenu uloga uključenih subjekata, ugovorno uređenje odnosa s izvršiteljima obrade, provedbu odgovarajućih tehničkih i organizacijskih mjera, dokumentiranje procjena rizika te, kada je potrebno, provedbu procjene učinka na zaštitu podataka.

Načelo pouzdanosti usko je povezano s tehničkom i integriranom zaštitom podataka iz članka 25. Opće uredbe o zaštiti podataka. Kod sustava umjetne inteligencije to znači da se zaštita osobnih podataka mora uzeti u obzir već pri planiranju, nabavi, razvoju, testiranju i uvođenju sustava, a ne tek nakon što je sustav stavljen u uporabu. Organizacija mora, primjerice, procijeniti jesu li svi podaci nužni za predviđenu svrhu, mogu li se koristiti anonimizirani, sintetski ili pseudonimizirani podaci, kako će se ograničiti pristup podacima, hoće li se korisnički upiti koristiti za daljnje treniranje modela i kako će se osigurati ostvarivanje prava ispitanika.

Voditelj obrade mora osigurati i da su odnosi s dobavljačima, pružateljima AI usluga, pružateljima cloud infrastrukture, API-ja ili drugih tehničkih komponenti jasno uređeni. Ako ti subjekti obrađuju osobne podatke u ime voditelja obrade, odnos mora biti uređen u skladu s člankom 28. Opće uredbe o zaštiti podataka, uključujući jasne upute, obveze povjerljivosti, sigurnosne mjere, pravila o podizvršiteljima i postupanje u slučaju povrede osobnih podataka.

Ako je organizacija obvezna imenovati službenika za zaštitu podataka, potrebno je osigurati njegovu pravodobnu uključenost u sve relevantne faze razvoja, nabave, testiranja i uporabe sustava umjetne inteligencije. Čak i kada imenovanje službenika nije obvezno, organizacija bi trebala osigurati da osobe uključene u razvoj, nabavu, integraciju i uporabu UI sustava imaju odgovarajuću razinu znanja o zaštiti osobnih podataka i internim pravilima postupanja.

Načelo pouzdanosti uključuje i uspostavu jasnih postupaka za postupanje po zahtjevima ispitanika i za postupanje u slučaju povrede osobnih podataka. Organizacija mora moći pravodobno otkriti, evidentirati i procijeniti incident te, kada su ispunjeni uvjeti iz Opće uredbe o zaštiti podataka, obavijestiti nadzorno tijelo i ispitanike.

Neispunjavanje načela pouzdanosti može dovesti do korektivnih mjera nadzornog tijela, uključujući upozorenja, opomene, naloge za usklađivanje obrade, ograničenja obrade i upravne novčane kazne. S druge strane, organizacija koja može dokazati da je pravodobno identificirala rizike, provela odgovarajuće mjere i dokumentirala svoje odluke bit će u

boljem položaju pokazati da je postupala odgovorno i u skladu s Općom uredbom o zaštiti podataka.

6. Određivanje pravne osnove za obradu osobnih podataka

Uz jasno definiranje svrhe za koju se razvija UI model, voditelj obrade mora odrediti odgovarajuću pravnu osnovu iz članka 6. Opće uredbe o zaštiti podataka. Organizacija koja namjerava izraditi skup podataka za treniranje UI modela koji sadržava osobne podatke mora prije početka obrade osigurati da je takva obrada zakonita, odnosno da se temelji na jednoj od pravnih osnova propisanih Općom uredbom o zaštiti podataka.

Pritom je važno imati na umu da se u različitim fazama životnog ciklusa UI sustava mogu primjenjivati različite pravne osnove, ovisno o konkretnoj svrsi obrade. Primjerice, prikupljanje i uporaba podataka za treniranje UI modela u određenim se okolnostima može temeljiti na legitimnom interesu, dok se naknadna personalizacija usluge ili korištenje korisničkih podataka za daljnje poboljšanje modela može temeljiti na privoli ispitanika, ako su za to ispunjeni svi uvjeti iz Opće uredbe o zaštiti podataka.

Opća uredba u članku 6. stavkom 1. propisuje kako je obrada zakonita samo ako i u onoj mjeri u kojoj je ispunjeno najmanje jedno od sljedećega:

> (a) ispitanik je dao privolu za obradu svojih osobnih podataka u jednu ili više posebnih svrha;
>
> (b) obrada je nužna za izvršavanje ugovora u kojem je ispitanik stranka ili kako bi se poduzele radnje na zahtjev ispitanika prije sklapanja ugovora;
>
> (c) obrada je nužna radi poštovanja pravnih obveza voditelja obrade;
>
> (d) obrada je nužna kako bi se zaštitili životno važni interesi ispitanika ili druge fizičke osobe;
>
> (e) obrada je nužna za izvršavanje zadaće od javnog interesa ili pri izvršavanju službene ovlasti voditelja obrade;
>
> (f) obrada je nužna za potrebe legitimnih interesa voditelja obrade ili treće strane, osim kada su od tih interesa jači interesi ili temeljna prava i slobode ispitanika koji zahtijevaju zaštitu osobnih podataka, osobito ako je ispitanik dijete.

Prilikom izbora pravne osnove voditelj obrade mora razmotriti način na koji će kreirati set podataka:

- Prikupljanje podataka direktno od ispitanika;
- Prikupljanje podataka iz otvorenih izvora na internetu za vlastite potrebe;
- Korištenje podataka inicijalno prikupljenih u drugu svrhu.

Svaka od navedenih pravnih osnova ima svoje specifičnosti te je nužno da voditelj obrade prilikom izbora je svjestan svih onih ograničenja koja pojedina ima u svojoj biti. Više o pravnim temeljima obrade možete pronaći https://azop.hr/pravni-temelji-za-obradu-osobnih-podataka-clanak6-gdpr/ , kao i na https://olivia-gdpr-arc.eu/hr .

6.1 Privola

Kada je u pitanju privola ista mora biti slobodno dana, mora biti specifična, informirana i nedvosmislena. Izostajanjem bilo koje sastavnice, privola neće biti valjana te će obrada biti u suprotnosti s Općom uredbom o zaštiti podataka. Europski odbor za zaštitu podataka izradio je smjernice o privoli kao pravnom temelju koje su dostupne na: https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_hr.pdf .

Primjer
Privola ispitanika za objavu fotografije na internetu ne znači automatski da organizacija istu fotografiju smije koristiti i za izradu skupa podataka za treniranje UI modela. Budući da se radi o dvije različite svrhe obrade, ispitanik mora biti jasno informiran o svakoj od njih i mora imati mogućnost zasebno pristati na objavu fotografije, a zasebno na njezinu uporabu za treniranje UI modela.

Privola je valjana samo ako je dana slobodno, specifično, informirano i nedvosmisleno. Voditelj obrade mora moći dokazati da su ispunjeni svi uvjeti koje Opća uredba o zaštiti podataka propisuje.

Važno je napomenuti da je valjanost privole problematična u situacijama neravnoteže moći, primjerice kada je voditelj obrade poslodavac ili javno tijelo. U takvim odnosima zaposlenici ili građani često nisu u poziciji slobodno odbiti davanje privole bez straha od negativnih posljedica. Stoga se privola u tim slučajevima može smatrati valjanom samo iznimno i uz osiguranje potpune dobrovoljnosti. Također, važno je naglasiti da ispitanik ima pravo u svakom trenutku povući svoju privolu, a nakon povlačenja privole voditelj obrade više ne smije koristiti njegove osobne podatke.

Iz tog razloga, u kontekstu sustava umjetne inteligencije mogu se pojaviti poteškoće povezane s ostvarivanjem prava na povlačenje privole, osobito kada postoje tehnička ograničenja ili kada voditelj obrade nije u mogućnosti iz već istreniranog UI modela izdvojiti i obrisati osobne podatke pojedinog ispitanika. U slučajevima kada je model treniran na velikim skupovima podataka, tehnički je vrlo teško ili gotovo nemoguće naknadno ukloniti pojedinačne podatke bez potpunog ponovnog treniranja sustava. Ako voditelj obrade ne može zajamčiti da će ispitanik uistinu moći ostvariti navedena prava, tada oslanjanje na privolu nije prikladna pravna osnova te je potrebno razmotriti druge

osnove obrade, primjerice izvršavanje ugovora ili legitimni interes, pod uvjetom da su ispunjeni kriteriji proporcionalnosti i da su primijenjene odgovarajuće zaštitne mjere.

6.2 Pravna obveza

Obrada osobnih podataka može se temeljiti na pravnoj obvezi iz članka 6. stavka 1. točke (c) Opće uredbe o zaštiti podataka samo ako je nužna radi ispunjavanja konkretne pravne obveze voditelja obrade. Ta obveza mora proizlaziti iz prava Unije ili prava države članice koje se primjenjuje na voditelja obrade te mora biti dovoljno jasna, precizna i predvidljiva.

Opće zakonske obveze, poput obveze zakonitog poslovanja, povećanja učinkovitosti, sigurnosti ili kvalitete usluge, u pravilu neće biti dovoljne za oslanjanje na ovu pravnu osnovu pri razvoju ili uporabi UI sustava. Ako organizacija samostalno odluči razviti ili koristiti UI sustav radi optimizacije procesa, analitike ili poboljšanja usluge, potrebno je razmotriti drugu odgovarajuću pravnu osnovu.

6.3 Sklapanje ili izvršenje ugovora

Ispunjenje ugovora kao pravna osnova može biti korištena za kreiranje seta podataka za trening modela, ako postoji valjani ugovor između ispitanika i voditelja obrade te obrada mora biti nužna za njegovo izvršenje odnosno ako se poduzimaju radnje na zahtjev ispitanika prije sklapanja samog ugovora. Bitno je istaknuti kako se ugovor kao pravna osnova može koristiti samo ako je obrada nužna za izvršenje ugovora između voditelja obrade i ispitanika. U situaciji kada se koriste osobni podaci koji nisu nužni za izvršenje samog ugovora između ispitanika i voditelja obrade, tada se takva obrada ne može temeljiti na ugovoru kao pravnom temelju obrade.

Primjer
Ako korisnik otvara račun na društvenoj mreži i prihvaća opće uvjete poslovanja, činjenica da je u tim uvjetima navedeno kako će se njegovi podaci koristiti za razvoj i unapređenje novih proizvoda, usluga ili funkcionalnosti ne znači da je takva obrada nužna za izvršenje ugovora. Obrada osobnih podataka radi razvoja novih funkcionalnosti ili treniranja UI modela u pravilu predstavlja zasebnu svrhu obrade te se ne može automatski temeljiti na članku 6. stavku 1. točki (b) Opće uredbe o zaštiti podataka.[4]

6.4 Izvršenje zadaće od javnog interesa

Ako je pravni temelj za obradu osobnih podataka pravna obveza voditelja obrade ili izvršavanje zadaće od javnog interesa/službene ovlasti voditelja obrade, tada ta pravna

[4] Vidi presudu Suda EU od 4. srpanja 2023, Meta, C-252/21

osnova mora biti utvrđena u pravu Unije ili pravu države članice kojem voditelj obrade podliježe, a tom pravnom osnovom mora biti određena i svrha obrade ili, u pogledu obrade vezano za izvršavanje zadaće od javnog interesa/službene ovlasti voditelja obrade, mora biti nužna za izvršavanje zadaće od javnog interesa ili izvršavanje službene ovlasti voditelja obrade.

Izvršenje zadaće od javnog interesa odnosno izvršenje službene ovlasti pretpostavlja:

- Zadaća za voditelja obrade proizlazi iz propisa;
- Korištenje podataka omogućava provođenje specifične zadaće na relevantan i odgovarajući način.

Primjer
Znanstvenici s Instituta za hrvatski jezik žele istražiti promjene u uporabi hrvatskog jezika na internetu. U tu svrhu izrađuju skup podataka na temelju javno objavljenih komentara s internetskih platformi i društvenih mreža kako bi trenirali model koji prepoznaje i analizira pojavnost određenih riječi, izraza i jezičnih obrazaca. Budući da takvi komentari mogu sadržavati osobne podatke, primjerice korisnička imena, identifikatore profila ili sadržaj koji se odnosi na pojedince, obradu je potrebno ograničiti na ono što je nužno za istraživačku svrhu te primijeniti odgovarajuće zaštitne mjere, poput uklanjanja izravnih identifikatora i smanjenja količine podataka.

6.5 Legitimni interes

Kako bi se utvrdili može li se određena obrada osobnih podataka temeljiti na članku 6. stavku 1. točki (f) Opće uredbe, voditelji obrade moraju provesti trodijelni test (svrhe, nužnosti, ravnoteže) te isti dokumentirati kako bi Agenciji mogli dokazati da doista imaju legitimni interes za obradu osobnih podataka.

Pri tome je potrebno ispuniti tri kumulativna uvjeta:

1. ostvarivanje legitimnog interesa od strane voditelja obrade ili treće strane;
2. obrada je nužna za ostvarivanje legitimnog interesa; i
3. legitimni interes nije podređen interesima ili temeljnim pravima i slobodama ispitanika.

Interes se može smatrati legitimnim ako su ispunjena sljedeća tri kumulativna kriterija:

a. interes je zakonit

b. interes je jasno i precizno artikuliran i

c. interes je stvaran i prisutan, a ne špekulativan.

Sljedeći primjeri mogu predstavljati legitiman interes u kontekstu UI modela:

i. razvoj usluge razgovornog agenta kako bi se pomoglo korisnicima;

ii. razvoj UI sustava za otkrivanje prijevarnog sadržaja ili ponašanja; i

iii. poboljšanje otkrivanja prijetnji u informacijskom sustavu.

Također, komercijalni interes također može biti legitiman interes pod uvjetom da nije u suprotnosti sa zakonom i da je obrada nužna i proporcionalna.[5] Suprotno tome, legitimni interes neće se smatrati legitimnim kada primjerice UI sustav nema veze s cilj i aktivnostima organizacije ili ako se ne može implementirati u skladu sa zakonom.

Interes koji se želi ostvariti mora biti dovoljno precizan i jasno predočen ispitanicima kao dio obveze transparentnosti voditelja obrade. Stoga je preporučljivo kada se radi npr. o razvoju i poboljšanju UI sustava opće namjene, čak i kada specifična upotreba modela nije poznata, poziva na određeni cilj koji se želi postići.

U drugom koraku, ovisno o slučaju, planiranu količinu osobnih podataka uključenih u UI model treba procijeniti imajući u vidu manje invazivne alternative koje bi razumno mogle biti dostupne kako bi se jednako učinkovito postigla svrha legitimnog interesa koji se nastoji ostvariti.

Ako je ostvarivanje te svrhe moguće i s pomoću UI modela koji ne podrazumijeva obradu osobnih podataka, smatrat će se da ta obrada osobnih podataka nije nužna. To je posebno važno za razvoj UI modela. Posebna pozornost obratit će se na količinu obrađenih osobnih podataka i je li ona proporcionalna za ostvarivanje predmetnog legitimnog interesa, također s obzirom na načelo smanjenja količine podataka.

Pri procjeni nužnosti uzima se u obzir i širi kontekst planirane obrade osobnih podataka. Sredstva koja su manje invazivna za temeljna prava i slobode ispitanika mogu se razlikovati ovisno o tome je li voditelj obrade u izravnom odnosu s ispitanicima (podaci prve strane) ili ne (podaci trećih strana). Sud EU-a u predmetu C621/22, Koninklijke Nederlandse Lawn Tennisbond (ECLI:EU:C:2024:857), točke 51.–53. naveo je neka razmatranja koja treba uzeti u obzir pri analizi nužnosti obrade podataka prve strane u svrhu legitimnih interesa koji se nastoje ostvariti (iako u kontekstu otkrivanja takvih podataka trećim stranama). Uvjet nužnosti znači da voditelj obrade mora osigurati da namjeravana obrada može postići željeni interes i da ne postoji drugi manje nametljivi način postizanja toga cilja osim provođenjem namjeravane obrade. Ovaj uvjet također treba ispitati u vezi s načelom smanjenja količine podataka.

Pri procjeni nužnosti obrade voditelj obrade treba razmotriti mogu li se svrha i funkcionalnost UI sustava postići uz primjenu tehničkih zaštitnih mjera koje smanjuju identifikabilnost ispitanika. Iako takve mjere ne moraju uvijek dovesti do potpune

[5] Vidi presudu Suda EU od 4. listopada 2024, Tennisbond, C-621/22

anonimizacije, primjerice pseudonimizacija, agregacija, filtriranje podataka ili uklanjanje izravnih identifikatora mogu smanjiti opseg obrade osobnih podataka i rizike za ispitanike.

Treći korak procjene legitimnog interesa jest provedba „testa ravnoteže“. Taj se korak sastoji od utvrđivanja i opisivanja različitih suprotstavljenih prava i interesa o kojima je riječ, tj. s jedne strane, interesa, temeljnih prava i sloboda ispitanika, a s druge strane interesa voditelja obrade ili treće strane. Potom bi trebalo razmotriti posebne okolnosti slučaja kako bi se dokazalo da je legitimni interes odgovarajuća pravna osnova za predmetne aktivnosti obrade.

Razumna očekivanja imaju ključnu ulogu u testu ravnoteže, među ostalim zbog složenosti tehnologije koja se upotrebljava u UI modelima i činjenice da bi ispitanicima moglo biti teško razumjeti raznolikost mogućih uporaba UI modela i uključene obrade podataka. U tu se svrhu mogu razmotriti informacije pružene ispitanicima kako bi se procijenilo mogu li ispitanici razumno očekivati da će se njihovi osobni podaci obrađivati. Međutim, iako izostavljanje informacija može doprinijeti tome da ispitanici ne očekuju određenu obradu, samo ispunjenje zahtjeva u pogledu transparentnosti utvrđenih u Općoj uredbi nije dovoljno kako bi se smatralo da ispitanici mogu razumno očekivati određenu obradu. Nadalje, samo zato što su informacije koje se odnose na fazu razvoja UI modela uključene u politiku zaštite privatnosti voditelja obrade, to ne znači nužno da ispitanici mogu razumno očekivati da će se to dogoditi, već je potrebno analizirati posebne okolnosti slučaja i uzeti u obzir sve relevantne čimbenike.

U ovom koraku, voditelj obrade mora procijeniti prevladava li njegov legitimni interes nad interesima, pravima i slobodama ispitanika. Pri tome treba uzeti u obzir koristi obrade, ali i moguće učinke na pojedince, njihova razumna očekivanja, prirodu podataka, opseg obrade i moguće rizike. Kada je to potrebno, voditelj obrade mora primijeniti dodatne zaštitne mjere kako bi se rizici za ispitanike smanjili i kako bi se osigurala pravedna ravnoteža između interesa voditelja obrade i prava ispitanika.

Pri procjeni legitimnog interesa potrebno je utvrditi i dokumentirati konkretne koristi koje voditelj obrade želi ostvariti te moguće pozitivne učinke obrade. U tu svrhu osobito se mogu uzeti u obzir sljedeći čimbenici:

• opseg, priroda i važnost očekivanih koristi od obrade;

• doprinos obrade ispunjavanju drugih regulatornih, sigurnosnih ili tehničkih zahtjeva;

• moguća korist za širu zajednicu, primjerice kada se razvija model otvorenog koda ili rješenje koje može pridonijeti znanstvenom istraživanju, sigurnosti, dostupnosti usluga ili javnom interesu;

• razina konkretnosti i jasnoće interesa koji se žele ostvariti, pri čemu općenite formulacije poput „razvoj umjetne inteligencije" ili „poboljšanje usluge" u pravilu nisu dovoljne.

Nužno je imati na umu kako koristi moraju biti uravnotežene s utjecajem na same ispitanike.

Voditelj obrade mora identificirati i procijeniti sve vrste posljedica, potencijalnih ili stvarnih, koje bi razvoj modela odnosno UI sustava i njegova upotreba mogli imati na ispitanike (npr. privatnost, slobodu izražavanja, pristup uslugama...).

Prilikom procjene potrebno je uzeti u obzir prirodu podataka i ispitanika, način na koji će se podaci obrađivati, prirodu sustava umjetne inteligencije i namjeravanu operativnu upotrebu.

- **Utjecaj povezan s razvojem UI modela**
    - Rizici povezani s prikupljanjem javno dostupnih podataka;
    - Rizik od gubitka povjerljivosti podataka;
    - Rizik povezan s teškoćama u ostvarivanju prava ispitanika;
    - Rizik povezan s teškoćom od osiguravanja transparentnosti.

- **Utjecaj povezan s korištenjem UI sustava**
    - Rizik od memoriranja i ekstrakcije podataka;
    - Rizici od šteta za ugled, širenje lažnih informacija i krađa identiteta;
    - Rizici od kršenja određenih prava i tajni zaštićenih zakonom;
    - Rizici povezani s etikom koji mogu utjecati i na pravilno funkcioniranje društva u cjelini.

Razumna očekivanja ispitanika predstavljaju ključni segment procjene legitimnosti obrade u sklopu testa legitimnog interesa. Ispitanici ne smiju biti iznenađeni modalitetom kao niti posljedicama same obrade.

Međutim, u toku razvoja sustava UI, neki dijelovi obrade mogu premašiti razumna očekivanja ispitanika, stoga će voditelj obrade morati uzeti u obzir sljedeće pokazatelje kada su podaci izravno prikupljeni od ispitanika:

- odnos voditelja obrade i ispitanika;
- postavke privatnosti koje dijeli ispitanik;
- kontekst i priroda usluge u kojoj su podaci prikupljeni;

- utječe li sama obrada na uslugu pruženu ili se koristi za poboljšanje usluge.

Primjer
Ako korisnici koriste internetsku uslugu za razmjenu privatnih poruka, njihovo razumno očekivanje je da će se sadržaj poruka obrađivati radi omogućavanja komunikacije, sigurnosti usluge i eventualnog ispunjenja zakonskih obveza, a ne za treniranje ili poboljšanje UI modela. Zato se korištenje sadržaja privatnih poruka za razvoj UI modela u pravilu ne bi moglo smatrati obradom koja je u skladu s razumnim očekivanjima ispitanika.

Ispitanici sve više su svjesni kako treće strane mogu prikupljati i ponovno koristiti informacije koje objavljuju na internetu. Međutim, ne mogu očekivati da se takva obrada odvija u svim situacijama i za sve vrste javno dostupnih podataka o njima.

Prilikom procjene treba uzeti u obzir sljedeće čimbenike:

- prirodu javno dostupnih podataka;
- kontekst i prirodu izvora odnosno same internetske stranice ili platforme;
- ograničenja koja su postavljena na internetskoj stranici u vidu uvjeta korištenja ili tehničkih mjera zaštite;
- vrstu objave;
- prirodu odnosa između ispitanika i voditelja obrade.

Također, kako bi smanjio utjecaj na ispitanike, voditelj obrade može uvesti dodatne mjere kako bi se postigla ravnoteža u odnosu te iste ne treba miješati s mjerama iz članka 32. Opće uredbe o zaštiti podataka.

Primjer 1

Voditelj obrade društvene mreže želi koristiti komentare korisnika za razvoj UI modela koji će služiti za generiranje teksta, predlaganje odgovora i učinkovitije filtriranje nezakonitog ili štetnog sadržaja. Prije početka takve obrade jasno informira korisnike o svrsi obrade, kategorijama podataka koje će se koristiti, načinu funkcioniranja obrade, mogućim rizicima i zaštitnim mjerama. Korisnicima također omogućuje da prije početka obrade jednostavno ulože prigovor i budu izuzeti iz korištenja njihovih podataka za tu svrhu.

U takvom slučaju legitimni interes može biti moguća pravna osnova, pod uvjetom da je voditelj obrade proveo i dokumentirao test legitimnog interesa te dokazao da njegov interes ne nadilaze interesi, prava i slobode ispitanika.

Primjer 2

Voditelj obrade želi razviti generativni UI sustav za izradu slika te u tu svrhu neselektivno prikuplja velik broj slika s internetskih stranica. Pritom ne isključuje izvore koji mogu

sadržavati osjetljive podatke, fotografije djece, biometrijske podatke ili druge visokorizične kategorije podataka, ne primjenjuje mjere za smanjenje rizika pohranjivanja ili reprodukcije osobnih podataka u modelu te svrhu obrade opisuje samo općenito, primjerice kao „pružanje usluga" ili „razvoj proizvoda".

U takvim okolnostima voditelj obrade teško bi se mogao uspješno pozvati na legitimni interes jer svrha nije dovoljno određena, obrada je neselektivna i opsežna, a rizici za prava i slobode ispitanika nisu odgovarajuće ograničeni.

Primjer 3

Voditelj obrade koristi samoposlužne blagajne u kojima je implementiran UI sustav za otkrivanje mogućih pogrešaka pri skeniranju i plaćanju proizvoda. Podatke nastale tijekom korištenja sustava želi koristiti i za poboljšanje njegove točnosti i smanjenje broja pogrešnih upozorenja.

U tu svrhu voditelj obrade zadržava samo podatke koji su nužni za poboljšanje sustava, primjenjuje mjere koje smanjuju mogućnost ponovne identifikacije kupaca, ograničava rokove čuvanja, jasno informira ispitanike o takvoj obradi i omogućuje im jednostavan prigovor kada se obrada temelji na legitimnom interesu. Takve mjere mogu pridonijeti zaključku da je obrada razmjerna i da su rizici za ispitanike odgovarajuće ublaženi.

Kako bi se limitirao utjecaj na prava i slobode ispitanika moguće je implementacija različitih mjera:

- Mjere ograničavanja prikupljanja ili pohrane osobnih podataka,
- Mjere koje omogućuju ispitanicima zadržavanje kontrole nad podacima,
- Mjere za ublažavanje rizika u operativnoj fazi.

Više o tome kako provesti test razmjernosti i dokazati legitimni interes možete saznati u Smjernicama Europskog odbora za zaštitu podataka: https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-12024-processing-personal-data-based_en ; također dodatno je moguće naći u dokumentu dostupnom na https://www.edpb.europa.eu/system/files/2026-03/spe-oss-case-digest-legitimate-interest_en.pdf .

Agencija je izradila obrazac koji možete koristiti prilikom provedbe procjene: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fazop.hr%2Fwp-content%2Fuploads%2F2021%2F04%2FTest-razmjernosti.docx&wdOrigin=BROWSELINK ; također unutar platforme Olivia razvijen je poseban modul na temu procjene legitimnog interesa koji je dostupan na: https://olivia-gdpr-arc.eu/hr/practical-submodule/overview/12.

Više informacija o legitimnom interestu dostupno je na poveznicama: Mišljenje 28/2024 o određenim aspektima zaštite podataka povezanima s obradom osobnih podataka u kontekstu modela umjetne inteligencije Europskog odbora za zaštitu, https://www.edpb.europa.eu/system/files/2025-05/edpb_opinion_202428_ai-models_hr.pdf.

6.5.1 Automatizirana ekstrakcija podataka s interneta (eng. Web scraping)

Automatizirana ekstrakcija podataka s javno dostupnih internetskih izvora sve se češće koristi, osobito u kontekstu razvoja generativnih UI sustava koji se oslanjaju na velike količine podataka dostupnih na internetu. Takva ekstrakcija može obuhvatiti podatke s portala vijesti, društvenih mreža, foruma, blogova, osobnih internetskih stranica i drugih javno dostupnih izvora. Iako su ti podaci javno dostupni, oni mogu sadržavati osobne podatke, uključujući korisnička imena, fotografije, objave, komentare, identifikatore profila, podatke o lokaciji ili druge informacije koje se odnose na pojedince.

Sama činjenica da su osobni podaci javno dostupni na internetu ne znači da se oni mogu slobodno prikupljati i ponovno koristiti za razvoj, treniranje ili testiranje UI modela. Takva obrada može predstavljati značajan rizik za prava i slobode ispitanika, osobito zato što pojedinci često imaju vrlo ograničenu kontrolu nad daljnjom uporabom podataka koje su objavili ili koji su o njima objavljeni na internetu.

Široka primjena automatizirane ekstrakcije podataka mijenja očekivanja pojedinaca u pogledu korištenja interneta. Podaci koje pojedinac objavi u jednom kontekstu mogu biti masovno prikupljeni, objedinjeni s drugim podacima i ponovno korišteni u potpuno drukčije svrhe, uključujući razvoj UI sustava. Takva obrada može dovesti do rizika za privatnost, nezakonitog ili nepoštenog prikupljanja podataka, profiliranja, reprodukcije osobnih podataka u izlazima modela te mogućeg odvraćajućeg učinka na slobodu izražavanja.

Zakonitost automatizirane ekstrakcije podataka s interneta mora se procjenjivati od slučaja do slučaja, uzimajući u obzir svrhu obrade, vrstu i izvor podataka, razumna očekivanja ispitanika, opseg obrade, moguće posljedice za ispitanike te primijenjene zaštitne mjere. U komercijalnom kontekstu razvoja UI modela voditelji obrade često će razmatrati legitimni interes kao moguću pravnu osnovu. Međutim, legitimni interes nije automatski primjenjiv i može se koristiti samo ako su ispunjeni svi uvjeti iz članka 6. stavka 1. točke (f) Opće uredbe o zaštiti podataka, uključujući postojanje zakonitog interesa, nužnost obrade i provedbu testa ravnoteže između interesa voditelja obrade i prava i sloboda ispitanika.

S obzirom na načelo smanjenja količine podataka iz članka 5. stavka 1. točke (c) Opće uredbe o zaštiti podataka, voditelj obrade koji provodi automatiziranu ekstrakciju podataka s interneta mora osigurati da se prikupljaju samo oni osobni podaci koji su primjereni, relevantni i nužni za jasno određenu svrhu obrade.

U tu svrhu voditelj obrade osobito treba:

• unaprijed definirati jasne i specifične kriterije za odabir izvora i kategorija podataka koji će se prikupljati;

• isključiti iz prikupljanja kategorije podataka koje nisu nužne za ostvarenje svrhe obrade, primjerice primjenom tehničkih filtara ili isključivanjem određenih internetskih stranica odnosno izvora;

• uspostaviti postupke za prepoznavanje i uklanjanje nerelevantnih ili prekomjernih podataka te ih obrisati bez odgode nakon što se utvrdi da nisu nužni za predviđenu svrhu;

• u pravilu isključiti iz prikupljanja internetske stranice ili izvore koji se izričito protive automatiziranoj ekstrakciji podataka, primjerice putem datoteke robots.txt, uvjeta korištenja, CAPTCHA mehanizama ili drugih tehničkih odnosno organizacijskih ograničenja.

Također, pri korištenju alata za automatiziranu ekstrakciju podataka s interneta voditelj obrade trebao bi osigurati dodatne mjere transparentnosti i zaštite ispitanika. To osobito uključuje:

- informiranje ispitanika putem različitih komunikacijskih kanala, primjerice objavom jasnih obavijesti na internetskoj stranici voditelja obrade, objavama na društvenim mrežama, javnim obavijestima ili drugim primjerenim kanalima;
- objavu i redovito ažuriranje popisa kategorija izvora ili, kada je moguće, konkretnih internetskih stranica iz kojih se podaci prikupljaju;
- omogućavanje jednostavnog i prethodnog prava na prigovor, kako bi ispitanici mogli zatražiti da se njihovi podaci ne koriste za takvu obradu;
- primjenu mjera anonimizacije ili pseudonimizacije što je ranije moguće nakon prikupljanja podataka, kada je to primjenjivo i tehnički izvedivo;
- sprječavanje nepotrebnog povezivanja ili objedinjavanja podataka na temelju identifikatora pojedinaca, osobito ako takvo povezivanje nije nužno za ostvarenje svrhe obrade.

7. Obrada posebnih kategorija osobnih podataka

U skladu s člankom 10. stavkom 5. Akta o umjetnoj inteligenciji, dobavljači visokorizičnih UI sustava mogu iznimno obrađivati posebne kategorije osobnih podataka ako je takva obrada strogo nužna radi otkrivanja i ispravljanja pristranosti u visokorizičnom UI sustavu.

Ova odredba ne predstavlja opću dozvolu za obradu posebnih kategorija osobnih podataka u svim UI sustavima, niti oslobađa voditelja obrade od obveza koje proizlaze iz Opće uredbe o zaštiti podataka. Ako obrada uključuje posebne kategorije osobnih podataka, potrebno je utvrditi odgovarajuću pravnu osnovu iz članka 6. stavka 1. Opće uredbe o zaštiti podataka te odgovarajuću iznimku iz članka 9. stavka 2. Opće uredbe o zaštiti podataka.

Obrada posebnih kategorija osobnih podataka na temelju članka 10. stavka 5. Akta o umjetnoj inteligenciji dopuštena je samo ako su kumulativno ispunjeni propisani uvjeti, osobito:

• cilj otkrivanja i ispravljanja pristranosti ne može se učinkovito postići bez obrade posebnih kategorija osobnih podataka;

• obrada podliježe tehničkim ograničenjima ponovne uporabe osobnih podataka te odgovarajućim mjerama sigurnosti i zaštite privatnosti, kao što je pseudonimizacija;

• primjenjuju se odgovarajuće tehničke i organizacijske mjere, uključujući primjerenu kontrolu pristupa;

• posebne kategorije osobnih podataka ne prenose se, ne ustupaju i ne čine dostupnima drugim stranama;

• posebne kategorije osobnih podataka brišu se nakon što se pristranost ispravi ili nakon isteka utvrđenog roka čuvanja;

• evidencija aktivnosti obrade sadržava obrazloženje zašto je obrada posebnih kategorija osobnih podataka bila strogo nužna za otkrivanje i ispravljanje pristranosti te zašto se taj cilj nije mogao postići obradom drugih podataka.

Dobavljač visokorizičnog UI sustava mora moći dokazati da su navedeni uvjeti ispunjeni, u skladu s načelom odgovornosti iz članka 5. stavka 2. Opće uredbe o zaštiti podataka. To uključuje dokumentiranje procjene nužnosti, proporcionalnosti, primijenjenih zaštitnih mjera, rokova čuvanja, ograničenja pristupa i razloga zbog kojih se cilj otkrivanja i ispravljanja pristranosti nije mogao postići manje invazivnim sredstvima.

Stoga, ako se posebne kategorije osobnih podataka koriste radi otkrivanja ili ispravljanja pristranosti visokorizičnog UI sustava, takva obrada mora biti strogo ograničena na tu svrhu, vremenski ograničena, tehnički i organizacijski zaštićena te jasno dokumentirana. Članak 10. stavak 5. Akta o umjetnoj inteligenciji ne može se koristiti kao osnova za

općenito treniranje UI modela na posebnim kategorijama osobnih podataka, niti za šire svrhe razvoja, poboljšanja ili komercijalne optimizacije modela.

8. Tehnička i integrirana zaštita podataka (eng. Data protection by design and by default)

U skladu s člankom 25. Opće uredbe o zaštiti podataka, voditelj obrade mora, već u fazi određivanja sredstava obrade i tijekom same obrade, provoditi odgovarajuće tehničke i organizacijske mjere kojima se osigurava učinkovita primjena načela zaštite podataka. Te mjere moraju biti primjerene prirodi, opsegu, kontekstu i svrhama obrade, kao i rizicima za prava i slobode ispitanika.

U kontekstu sustava umjetne inteligencije to znači da se zaštita osobnih podataka mora uzeti u obzir od najranije faze razvoja sustava, a ne tek nakon njegova uvođenja u uporabu. Organizacija mora osigurati da se obrađuju samo oni osobni podaci koji su nužni za svaku pojedinu svrhu obrade. Ta se obveza odnosi na količinu podataka koji se prikupljaju, opseg njihove obrade, rokove čuvanja, dostupnost podataka i mogućnost pristupa podacima.

Pri razvoju UI sustava posebnu pozornost treba posvetiti sljedećim elementima:

- jasno određenoj svrsi sustava;
- odabiru metode, modela ili algoritma koji je razmjeran predviđenoj svrsi;
- izboru izvora podataka i kategorija podataka koji su nužni za razvoj, treniranje, validaciju ili testiranje sustava;
- procjeni utjecaja odabranih tehničkih rješenja na prava i slobode ispitanika;
- dokumentiranju ključnih odluka donesenih tijekom razvoja sustava.

Dobavljač UI sustava trebao bi jasno opisati predviđenu namjenu sustava, način njegove uporabe i ograničenja uporabe. To uključuje vrstu rezultata koje sustav proizvodi, primjerice preporuku, predviđanje, klasifikaciju, rangiranje, upozorenje, sažetak ili odluku, očekivane pokazatelje učinkovitosti, kontekst u kojem se sustav smije koristiti te kontekste uporabe koji su izričito isključeni. Takve informacije važne su kako bi subjekt koji uvodi sustav mogao procijeniti je li sustav prikladan za njegovu konkretnu svrhu i koje su mjere potrebne za zaštitu osobnih podataka.

Odabir manje invazivnog tehničkog rješenja

Pri odabiru metode razvoja, modela ili algoritma, potrebno je voditi računa o načelu smanjenja količine podataka i načelu razmjernosti. Ako se ista svrha može postići tehničkim rješenjem koje zahtijeva obradu manje količine osobnih podataka ili predstavlja manji rizik za ispitanike, prednost treba dati takvom rješenju.

Uporaba složenijih i invazivnijih metoda, primjerice dubokog učenja, ne bi trebala biti automatski izbor. Ako se cilj može postići jednostavnijom metodom koja ne uključuje strojno učenje ili zahtijeva manji opseg osobnih podataka, takvu metodu treba razmotriti kao primarnu mogućnost.

Primjer

Za razvoj UI sustava koji procjenjuje broj osoba koje stoje u tramvaju moguća su različita tehnička rješenja. Jedno rješenje može se temeljiti na modelu koji detektira broj zauzetih i slobodnih sjedećih mjesta te na temelju toga procjenjuje broj osoba koje stoje. Drugo rješenje može uključivati analizu držanja tijela svih osoba u tramvaju kako bi se izravno utvrdilo tko sjedi, a tko stoji.

Ako je za predviđenu svrhu, primjerice izradu statistike popunjenosti tramvaja, dovoljna prva metoda, ona bi u pravilu bila prihvatljivija jer zahtijeva obradu manje količine osobnih podataka i manje zadire u privatnost putnika. Druga metoda, koja uključuje detaljniju analizu osoba u prostoru, zahtijevala bi dodatno opravdanje i primjenu snažnijih zaštitnih mjera.

Slično tome, ako je cilj upozoriti osobu da je ušla u opasno područje, potrebno je razmotriti može li se ta svrha postići manje invazivnim sredstvom, primjerice infracrvenim senzorom, umjesto kamerom koja snima ili analizira osobu.

Pri odabiru tehničkog rješenja dobavljač UI sustava treba uzeti u obzir stanje tehnologije, dostupnost manje invazivnih metoda, rezultate usporedbe različitih arhitektura, mogućnost korištenja postojećih modela te primjenu tehnologija koje povećavaju razinu zaštite osobnih podataka. To može uključivati, primjerice, federativno učenje, diferencijalnu privatnost, homomorfnu enkripciju, pseudonimizaciju, agregaciju podataka ili druge tehnologije za poboljšanje privatnosti. Više o takvim tehnologijama obrađeno je u poglavlju „Korištenje tehnologija za poboljšanje privatnosti".

Provjera odluka donesenih u fazi razvoja

Odluke donesene tijekom razvoja UI sustava potrebno je provjeravati i dokumentirati. To se osobito odnosi na odabir svrhe, kategorija podataka, izvora podataka, metode razvoja, arhitekture modela, mjera sigurnosti i načina evaluacije sustava.

Dobavljač UI sustava može provjeriti opravdanost i primjerenost tih odluka, primjerice:

- provedbom pilot-studije kojom se ispituje jesu li odabrani podaci, metoda i tehničko rješenje primjereni predviđenoj svrsi;
- savjetovanjem sa službenikom za zaštitu podataka, ako je imenovan;
- uključivanjem multidisciplinarnog tima koji obuhvaća pravne, tehničke, sigurnosne i etičke kompetencije;

- savjetovanjem s etičkim odborom ili drugim neovisnim savjetodavnim tijelom, osobito kod sustava koji mogu imati značajan utjecaj na pojedince.

Uključivanje etičkog odbora ili sličnog tijela može predstavljati dobru praksu, osobito kod složenih ili visokorizičnih UI sustava. Međutim, takvo tijelo mora imati jasno definiranu ulogu, odgovarajuću stručnost, neovisnost i mogućnost kontinuiranog praćenja razvoja i uporabe sustava.

Rokovi čuvanja i upravljanje podacima tijekom životnog ciklusa UI sustava

Organizacija mora unaprijed odrediti i dokumentirati rokove čuvanja za svaku relevantnu kategoriju osobnih podataka i povezanih artefakata koji nastaju tijekom životnog ciklusa UI sustava. To uključuje, prema potrebi, izvorne skupove podataka, očišćene i anotirane skupove podataka, skupove za treniranje, validaciju i testiranje, korisničke unose, promptove, izlaze sustava, logove, sigurnosne kopije, izvješća o testiranju, metapodatke, verzije modela i druge izvedene artefakte koji mogu sadržavati osobne podatke ili omogućiti njihovo posredno izdvajanje.

U kontekstu UI sustava nije dovoljno upravljati samo izvornim skupovima podataka. Potrebno je uzeti u obzir i izvedene artefakte koji mogu sadržavati osobne podatke, odražavati njihove elemente ili omogućiti ponovnu identifikaciju ispitanika. Rokovi čuvanja moraju biti povezani s konkretnom svrhom obrade i ne smiju biti određeni općenito, neograničeno ili bez jasnog opravdanja.

Preporučuje se da interni akt, evidencija aktivnosti obrade ili drugi relevantni dokument za svaku kategoriju podataka ili artefakata sadržava najmanje:

- opis kategorije podataka ili artefakta;
- svrhu obrade;
- pravnu osnovu;
- lokaciju pohrane;
- odgovornu osobu ili funkciju;
- rok čuvanja;
- pravila verzioniranja;
- način brisanja, anonimizacije ili ograničenja obrade;
- postupanje sa sigurnosnim kopijama;
- način dokazivanja da je brisanje, anonimizacija ili povlačenje iz uporabe provedeno.

Takav pristup pridonosi dokazivanju usklađenosti i smanjuje rizik da se osobni podaci zadržavaju dulje nego što je nužno.

Posebno je potrebno urediti postupanje u slučaju zahtjeva za brisanje, povlačenja podataka iz daljnje uporabe, prestanka projekta, zamjene modela ili povlačenja sustava iz uporabe. U tim slučajevima organizacija mora procijeniti odnosi li se obveza brisanja samo na izvorne podatke ili i na povezane kopije, logove, indeksne strukture, vektorske baze, pomoćne datoteke, sigurnosne kopije i druge artefakte.

Ako potpuno brisanje iz pojedinog artefakta nije razumno izvedivo, organizacija mora to posebno obrazložiti, ograničiti daljnju uporabu tih podataka, primijeniti odgovarajuće zaštitne mjere i dokumentirati razloge takvog postupanja. Takvo obrazloženje ne smije biti općenito, nego mora biti utemeljeno na tehničkim i organizacijskim okolnostima konkretnog sustava.

9. Informiranje ispitanika o obradi osobnih podataka

Voditelj obrade treba obavijestiti ispitanike o obradi osobnih podataka u situacijama kada prikuplja podatke direktno od ispitanika, ali također kada to čini i od treće strane. U tom smislu članak 13. Opće uredbe o zaštiti podataka definira situacije kada se podaci dobivaju direktno od ispitanika, dok članak 14. istog propisa vrijedi u situacijama kada su podaci prikupljeni od treće strane odnosno drugog izvora, a ne od samog ispitanika.

Obvezu informiranja nije problematično ispuniti u situaciji kada se podaci prikupljaju direktno od pojedinca. Međutim, nešto je složenija problematika kada je u pitanju prikupljanje podataka indirektno. Članak 14. Opće uredbe o zaštiti podatak u stavku 5. propisuje izuzetke od obveze informiranja ispitanika, a također u članku 23. se propisuje kada temeljem EU ili nacionalnog prava može postojati takva derogacija.

Primjer
Ako organizacija izradi skup podataka za treniranje UI modela i objavi ga na platformi za razmjenu podataka, na stranici za preuzimanje trebala bi pružiti jasne informacije o izvoru podataka, svrsi izrade skupa, kategorijama podataka i ispitanika, mogućoj prisutnosti osobnih podataka, primijenjenim zaštitnim mjerama i ograničenjima daljnje uporabe. Takve informacije mogu pomoći ponovnim korisnicima u procjeni zakonitosti njihove obrade, ali ih ne oslobađaju vlastitih obveza iz Opće uredbe o zaštiti podataka.

Nadalje, također u situaciji kada je obveza informiranja nemoguća ili zahtijeva nerazmjerne napore, voditelj obrade treba procijeniti mjere zaštite prava i sloboda ispitanika te može jednostavno informirati javnost. Uzimajući u obzir da faze razvoja UI modela mogu uključivati prikupljanje velikih količina podataka iz javno dostupnih izvora, oslanjanje na iznimku iz članka 14. stavka 5. točke (b) Uredbe strogo je ograničeno isključivo na slučajeve u kojima su u potpunosti ispunjeni uvjeti propisani tom odredbom.

U takvom će slučaju biti potrebno provesti analizu, uzimajući u obzir specifičan kontekst svake obrade. Organizacija mora procijeniti i dokumentirati nesrazmjer mjere, s jedne strane, miješanje u privatnost osoba čiji se podaci obrađuju, a s druge strane, teret koji bi pojedinačno pružanje informacija svakom ispitaniku podrazumijevalo. Pri tom trebalo bi uzeti u obzir nedostatak sredstava za kontakt, starost podataka, broj ispitanika, trošak komunikacije.

Naime, pozivanje na iznimku iz članka 14. stavka 5. točke (b), prema kojoj bi pružanje informacija bilo nemoguće ili bi zahtijevalo nerazmjeran napor, mora se tumačiti usko i ne smije biti rutinsko. Sama činjenica da se radi o velikoj količini podataka, velikom broju ispitanika ili složenom tehničkom okruženju nije sama po sebi dovoljna za oslanjanje na ovu iznimku. Pri procjeni treba uzeti u obzir osobito broj ispitanika, starost podataka, prirodu i izvor podataka, mogućnost uspostave kontakta s ispitanicima, trošak i organizacijsko opterećenje individualnog informiranja, kao i zaštitne mjere koje su već uspostavljene.

Ako se voditelj obrade pozove na nerazmjeran napor, mora dokumentirati razloge takve odluke i poduzeti odgovarajuće mjere za zaštitu prava i sloboda te interesa ispitanika. Takve mjere mogu uključivati javno dostupnu obavijest o obradi, višeslojnu transparentnost, objavu informacija na mrežnim stranicama, tehničke i organizacijske mjere za smanjenje rizika te ograničenje uporabe podataka na ono što je nužno za konkretnu svrhu.

U svakom slučaju preporučuje se da voditelj obrade zasebno evidentira: izvor podataka, razlog zbog kojeg se podaci ne prikupljaju izravno od ispitanika, procjenu primjenjivosti članka 14., razloge eventualnog pozivanja na iznimku, kao i alternativne mjere transparentnosti koje su primijenjene

Primjer

Ako voditelj obrade želi ponovno koristiti podatke svojih postojećih klijenata za razvoj ili poboljšanje UI modela, a raspolaže njihovim adresama e-pošte ili drugim kontakt podacima, u pravilu ih mora izravno informirati o novoj svrsi obrade prije njezina početka.

Ako, međutim, voditelj obrade obrađuje samo pseudonimizirane ili neizravno identificirajuće podatke i nema kontakt podatke ispitanika, nije dužan dodatno identificirati osobe samo radi njihova informiranja. U takvom slučaju informacije se mogu pružiti općom obaviješću, primjerice na internetskoj stranici voditelja obrade, pod uvjetom da su jasne, lako dostupne i dovoljno konkretne.

Informacije koje voditelj obrade treba pružiti su određene u člancima 13. i 14. Opće uredbe o zaštiti podataka.

Kada organizacija ponovno koristi skup podataka ili UI model koji sadržava osobne podatke, trebala bi dokumentirati izvor podataka. Ako se radi o skupu podataka koji može

predstavljati veći rizik za ispitanike, preporučuje se evidentirati i kontakt podatke izvornog voditelja obrade ili drugog subjekta od kojeg su podaci preuzeti, kako bi se mogla provjeriti zakonitost izvora, uvjeti daljnje uporabe i eventualno postupanje po zahtjevima ispitanika.

Primjer

Za razvoj UI sustava korišten je skup podataka pribavljen od vanjskog pružatelja. Skup podataka sadržava 1.000 fotografija osoba koje su dobrovoljno sudjelovale u njegovoj izradi. Fotografije prikazuju različite izraze lica i označene su kategorijama emocija radi treniranja modela za prepoznavanje izraza lica. Organizacija je prije uporabe skupa podataka provjerila uvjete njegova korištenja, dostupne informacije o načinu prikupljanja podataka i primijenjene zaštitne mjere.

Osim toga, sustav koristi veliki jezični model koji je razvila kompanija X. Organizacija koristi taj model kao komponentu vlastitog UI sustava te je procijenila uvjete njegova korištenja, uloge uključenih strana i moguće učinke obrade osobnih podataka.

U slučaju automatskog preuzimanja podataka s interneta ili njihove ponovne upotrebe, kada se isto odnosi na nekoliko internetskih stranica preporučuje se navođenje izvora. U situaciji kada je broj stranica velik, preporuka je navođenja pojedinih kategorija.

Primjer obavijesti kada je broj izvora velik

Za razvoj UI modela koristili smo javno dostupne podatke prikupljene automatiziranom ekstrakcijom s internetskih izvora koji sadržavaju publikacije o uzgoju biljaka. Podaci su prikupljeni iz kategorije specijaliziranih internetskih platformi i repozitorija koji objavljuju stručne i korisničke publikacije o uzgoju biljaka.

Prikupljeni su samo tekstovi publikacija relevantni za navedenu temu. Komentari korisnika, fotografije profila, pseudonimi i drugi identifikatori koji nisu bili nužni za svrhu obrade nisu zadržani u skupu podataka.

U slučaju razvoja modela UI opće namjene važno je imati na umu i obveze iz članka 53. Akta o umjetnoj inteligenciji koji zahtijeva od dobavljača modela da na raspolaganje stavi detaljan sažetak sadržaja korištenog za treniranje modela, u skladu s predloškom Ureda za umjetnu inteligenciju (Europska komisija).[6]

[6] Prva verzija predloška sažetka objavljena od strane Ured za umjetnu inteligenciju u srpnju 2025. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

10. Prava ispitanika

Sukladno Općoj uredbi o zaštiti podataka, ispitanici imaju sljedeća prava:

- pravo na pristup osobnim podacima – članak 15.;
- pravo na ispravak – članak 16.;
- pravo na brisanje – članak 17.;
- pravo na ograničenje obrade – članak 18.;
- pravo na prenosivost podataka – članak 20.;
- pravo na prigovor – članak 21.;
- pravo da se na ispitanika ne odnosi odluka koja se temelji isključivo na automatiziranoj obradi, uključujući profiliranje, ako takva odluka proizvodi pravne učinke za ispitanika ili na njega na sličan način značajno utječe – članak 22.

U kontekstu sustava umjetne inteligencije prava ispitanika mogu se ostvarivati osobito:

- u odnosu na osobne podatke koji su korišteni za treniranje, validaciju ili testiranje UI modela;
- u odnosu na UI model, ako se taj model ne može smatrati anonimnim;
- u odnosu na podatke koji se obrađuju tijekom uporabe UI sustava, uključujući korisničke unose, promptove, izlaze sustava, logove, podatke u RAG bazama znanja ili drugim povezanim komponentama sustava.

Ostvarivanje prava ispitanika u odnosu na izvorne podatke ili skupove podataka za treniranje u pravilu je jednostavnije od ostvarivanja prava u odnosu na sam UI model. Kod UI modela mogu se pojaviti tehničke poteškoće, osobito ako je potrebno utvrditi sadržava li model informacije koje se odnose na određenog ispitanika ili postoji li mogućnost izdvajanja osobnih podataka iz modela.

Međutim, tehnička složenost UI modela ne znači da se prava ispitanika mogu unaprijed isključiti. Voditelj obrade mora u svakom konkretnom slučaju procijeniti može li zahtjev ispitanika razumno i razmjerno izvršiti, uzimajući u obzir prirodu modela, dostupne tehničke mogućnosti, troškove provedbe, rizike za prava i slobode ispitanika te načelo odgovornosti.

Ako iz zahtjeva ispitanika nije jasno odnosi li se zahtjev na skup podataka za treniranje, podatke obrađene tijekom uporabe sustava ili na sam UI model, voditelj obrade treba ispitaniku jasno objasniti kako je zahtjev protumačio i, prema potrebi, zatražiti dodatna pojašnjenja.

Ako voditelj obrade ne može identificirati ispitanika ili više nije u mogućnosti povezati određene podatke s ispitanikom, a to može dokazati, o tome može obavijestiti ispitanika. Međutim, ako ispitanik dostavi dodatne informacije koje omogućuju njegovu identifikaciju ili pronalazak relevantnih podataka, voditelj obrade dužan je ponovno procijeniti mogućnost postupanja po zahtjevu.

Organizacija ne bi trebala čuvati identifikacijske podatke samo zato da bi u budućnosti lakše postupala po zahtjevima ispitanika, ako takvo čuvanje nije nužno za konkretnu svrhu obrade. Iznimno, organizacija može čuvati minimalne identifikacijske podatke potrebne za upravljanje privolama, povlačenjem privole ili prigovorima ispitanika, pod uvjetom da su ispitanici o tome jasno informirani, da postoji odgovarajuća pravna osnova i da su rokovi čuvanja jasno ograničeni.

Pravo na pristup

Pravo na pristup omogućuje ispitaniku da dobije potvrdu obrađuju li se njegovi osobni podaci te, ako se obrađuju, pristup tim podacima i informacijama iz članka 15. Opće uredbe o zaštiti podataka.

U kontekstu UI sustava to može uključivati, ovisno o okolnostima, pristup osobnim podacima sadržanima u skupu podataka za treniranje, validaciju ili testiranje, korisničkim unosima, povezanim metapodacima, oznakama, napomenama ili drugim podacima koji se odnose na ispitanika. Kada je to potrebno radi učinkovitog ostvarivanja prava ispitanika, voditelj obrade trebao bi omogućiti pristup relevantnim izvacima iz skupa podataka u razumljivom obliku.

Pritom ostvarivanje prava na pristup ne smije negativno utjecati na prava i slobode drugih osoba. To može uključivati zaštitu osobnih podataka drugih ispitanika, poslovne tajne, prava intelektualnog vlasništva ili sigurnost sustava. Ako je potrebno, voditelj obrade može ograničiti ili prilagoditi način pružanja informacija, ali takvo ograničenje mora biti obrazloženo i razmjerno.

Ispitanik također ima pravo dobiti informacije o primateljima ili kategorijama primatelja kojima su njegovi osobni podaci otkriveni. Kada je moguće, voditelj obrade trebao bi pružiti konkretne informacije o primateljima, a ne samo općenite kategorije. Ako podaci nisu prikupljeni od ispitanika, voditelj obrade dužan je pružiti i dostupne informacije o izvoru podataka.

Pravo na ispravak, brisanje, ograničenje obrade i prigovor u odnosu na skupove podataka za treniranje

Ako se zahtjev ispitanika odnosi na osobne podatke sadržane u skupu podataka za treniranje, validaciju ili testiranje, voditelj obrade mora procijeniti zahtjev u skladu s relevantnim odredbama Opće uredbe o zaštiti podataka.

Ispitanik može, ovisno o okolnostima, imati pravo zahtijevati ispravak netočnih podataka, brisanje podataka, ograničenje obrade ili uložiti prigovor na obradu, osobito ako se obrada temelji na legitimnom interesu. Voditelj obrade mora uspostaviti postupke kojima može identificirati relevantne podatke, procijeniti osnovanost zahtjeva i provesti odgovarajuću mjeru, osim ako postoji zakonit razlog za odbijanje zahtjeva.

Primjer

Organizacija upravlja platformom za prijavu kandidata za zaposlenje. U bazi ima registrirane korisnike koji su ranije unijeli svoje životopise, podatke o obrazovanju, radnom iskustvu, struci i drugim kvalifikacijama. Organizacija želi te podatke koristiti za treniranje UI modela koji bi pomagao u sortiranju prijava prema relevantnosti za određeno radno mjesto i u inicijalnom bodovanju prijava.

Prije početka takve obrade organizacija mora jasno informirati registrirane korisnike o novoj svrsi obrade, vrstama podataka koje namjerava koristiti, pravnoj osnovi, mogućim učincima obrade, rokovima čuvanja i njihovim pravima. Ako se obrada temelji na legitimnom interesu, korisnicima mora omogućiti jednostavno ulaganje prigovora prije početka korištenja njihovih podataka za treniranje modela.

Ostvarivanje prava u odnosu na UI model

Ako se UI model ne može smatrati anonimnim, ispitanici mogu, ovisno o okolnostima, nastojati ostvariti svoja prava i u odnosu na sam model. Takve situacije mogu biti tehnički složene, ali ih voditelj obrade ne smije unaprijed isključiti bez procjene konkretnog zahtjeva.

U određenim slučajevima tehnički može biti moguće postupiti po zahtjevu koji se odnosi na model, osobito ako parametri modela ili povezane komponente sustava izravno sadržavaju ili omogućuju izdvajanje osobnih podataka. To može biti relevantno, primjerice, kod određenih modela poput SVM-ova, algoritama za klasteriranje ili sustava koji su povezani s bazama znanja, kao što su RAG sustavi.

Voditelj obrade može, kada je to potrebno, zatražiti od ispitanika dodatne informacije koje su nužne za provjeru odnosi li se model na njegove osobne podatke. Takav zahtjev mora biti razmjeran i ne smije se koristiti kao način neopravdanog otežavanja ostvarivanja prava.

Primjer

Dobavljač velikog jezičnog modela treniranog na podacima automatski prikupljenima s interneta zaprimi zahtjev osobe koja smatra da model reproducira njezine osobne podatke. Dobavljač može od osobe zatražiti dodatne informacije koje mogu pomoći u provjeri zahtjeva, primjerice primjere upita za koje osoba smatra da dovode do reprodukcije njezinih osobnih podataka ili poveznice na javno dostupne tekstove za koje tvrdi da su korišteni u treniranju modela.

Takav pristup može pomoći voditelju obrade u procjeni sadržava li model podatke koji se odnose na ispitanika ili može li ih reproducirati. Međutim, voditelj obrade i dalje mora samostalno procijeniti zahtjev i ne smije teret dokazivanja u cijelosti prebaciti na ispitanika.

Razlikovanje modela, sustava i povezanih baza znanja

U nekim slučajevima potrebno je razlikovati UI model od UI sustava u koji je model integriran. UI sustav može, osim samog modela, uključivati dodatne komponente, primjerice korisničko sučelje, web pretraživanje, bazu znanja, vektorsku bazu, sustav za dohvat dokumenata ili dodatne sigurnosne i filtracijske slojeve.

Zbog toga ispitanik može pogrešno zaključiti da se određeni osobni podaci nalaze u samom modelu, iako se zapravo nalaze u povezanoj bazi znanja ili drugoj komponenti sustava. To je osobito važno kod RAG sustava, u kojima se odgovor modela može temeljiti na dokumentima dohvaćenima iz vanjske ili interne baze znanja.

Ako je bazu znanja u sustav integrirao dobavljač UI sustava, on mora uzeti u obzir zahtjeve ispitanika koji se odnose na podatke u toj bazi. Ako je, međutim, bazu znanja dodala treća strana koja samostalno određuje svrhu i sredstva obrade, ispitanika je potrebno uputiti na odgovarajućeg voditelja obrade, uz jasno objašnjenje uloge pojedinih sudionika.

Tehnička ograničenja, proporcionalnost i moguće metode postupanja

Prava ispitanika ostvaruju se pod uvjetima i ograničenjima propisanima Općom uredbom o zaštiti podataka. Nisu sva prava apsolutna, a svaki zahtjev potrebno je procijeniti u konkretnom slučaju.

Kod zahtjeva koji se odnose na UI modele mogu se pojaviti tehničke poteškoće, osobito kod zahtjeva za brisanje ili ispravak podataka. Voditelj obrade mora pritom procijeniti proporcionalnost mogućih mjera, uzimajući u obzir:

- prirodu i osjetljivost podataka;
- moguće rizike za ispitanika, primjerice pogrešnu kriminalnu, zdravstvenu, financijsku ili drugu stigmatizaciju;

- mogućnost i trošak provedbe tehničke mjere;
- učinak mjere na funkcionalnost, sigurnost i pouzdanost modela;
- postojanje manje invazivnih ili učinkovitijih alternativnih mjera.

Tehnologije poput strojnog odučavanja, ponovnog treniranja modela, uklanjanja podataka iz povezanih baza znanja, primjene filtara, blokiranja određenih izlaza ili drugih tehničkih mjera mogu biti relevantne za ostvarivanje pojedinih prava. Ako potpuno brisanje ili ispravak u samom modelu nije razmjerno ili tehnički izvedivo, voditelj obrade treba razmotriti druge učinkovite i robusne mjere kojima se može spriječiti daljnje korištenje ili reprodukcija spornih osobnih podataka.

Pritom je važno imati na umu da se tehničke mogućnosti brzo razvijaju. Zahtjev koji danas nije moguće razmjerno izvršiti zbog tehničkih ograničenja može u budućnosti postati izvediv zbog razvoja novih metoda i alata. Voditelji obrade stoga trebaju redovito pratiti tehnološki razvoj, osobito u području strojnog odučavanja i upravljanja podacima u UI modelima.

Razlozi za odbijanje zahtjeva

Voditelj obrade može odbiti zahtjev ispitanika samo ako za to postoji valjan razlog u skladu s Općom uredbom o zaštiti podataka ili drugim primjenjivim pravom. To može biti slučaj, primjerice, kada:

- voditelj obrade ne može identificirati ispitanika, a ispitanik ne dostavi dodatne informacije potrebne za identifikaciju;
- zahtjev je očito neutemeljen ili pretjeran, osobito zbog učestalog ponavljanja;
- organizacija koja je zaprimila zahtjev nije voditelj obrade za predmetnu obradu;
- ostvarivanje prava ograničeno je pravom Unije ili pravom države članice;
- postoje druga zakonita ograničenja, primjerice zaštita prava i sloboda drugih osoba.

Ako voditelj obrade odbije zahtjev, mora o tome obavijestiti ispitanika, navesti razloge odbijanja i informirati ga o mogućnosti podnošenja pritužbe nadzornom tijelu i korištenja pravnog lijeka.

U svakom slučaju, voditelj obrade mora uspostaviti interni postupak za zaprimanje, evidentiranje, procjenu i rješavanje zahtjeva ispitanika. U postupku mora biti jasno određeno tko zaprima zahtjev, tko provodi provjeru identiteta, tko ocjenjuje primjenjivost prava u konkretnom slučaju, tko priprema odgovor i tko odlučuje o eventualnom odbijanju zahtjeva. Ako voditelj obrade tvrdi da ne može identificirati ispitanika ili da bi postupanje po zahtjevu zahtijevalo nerazmjeran napor, takvu odluku mora posebno obrazložiti i dokumentirati.

Automatizirano pojedinačno donošenje odluka i profiliranje

Organizacije koje koriste UI sustave za donošenje odluka o pojedincima moraju posebno procijeniti primjenjuje li se članak 22. Opće uredbe o zaštiti podataka. Taj članak štiti ispitanike od odluka koje se temelje isključivo na automatiziranoj obradi, uključujući profiliranje, ako takve odluke proizvode pravne učinke za ispitanika ili na njega na sličan način značajno utječu.

Primjeri takvih odluka mogu uključivati automatsko odbijanje zahtjeva za kredit podnesenog putem interneta, automatizirano isključivanje kandidata iz postupka zapošljavanja bez stvarnog ljudskog uključivanja ili automatizirano određivanje prava, pogodnosti ili pristupa usluzi.

Odluka koja se temelji isključivo na automatiziranoj obradi dopuštena je samo ako je:

a) nužna za sklapanje ili izvršenje ugovora između ispitanika i voditelja obrade;
b) dopuštena pravom Unije ili pravom države članice kojem podliježe voditelj obrade, pri čemu takvo pravo mora propisati odgovarajuće mjere zaštite prava, sloboda i legitimnih interesa ispitanika; ili
c) utemeljena na izričitoj privoli ispitanika.

Ispitanik mora biti jasno informiran o postojanju automatiziranog odlučivanja, uključujući profiliranje, kao i o smislenim informacijama o uključenoj logici te važnosti i predviđenim posljedicama takve obrade za ispitanika.

Ako se odluka temelji na ugovornoj nužnosti ili izričitoj privoli, voditelj obrade mora osigurati odgovarajuće zaštitne mjere, uključujući najmanje pravo ispitanika da zatraži ljudsku intervenciju, izrazi svoje stajalište i ospori odluku. Ako se odluka temelji na pravu Unije ili pravu države članice, takvo pravo mora predvidjeti odgovarajuće mjere zaštite prava, sloboda i legitimnih interesa ispitanika.

Posebnu pozornost potrebno je posvetiti razlikovanju situacija u kojima UI sustav samo pruža pomoć ili preporuku osobi koja donosi odluku od situacija u kojima se odluka u stvarnosti donosi isključivo automatizirano. Formalno uključivanje čovjeka neće biti dovoljno ako osoba koja navodno donosi odluku nema stvarnu mogućnost razumjeti, preispitati i promijeniti rezultat UI sustava.

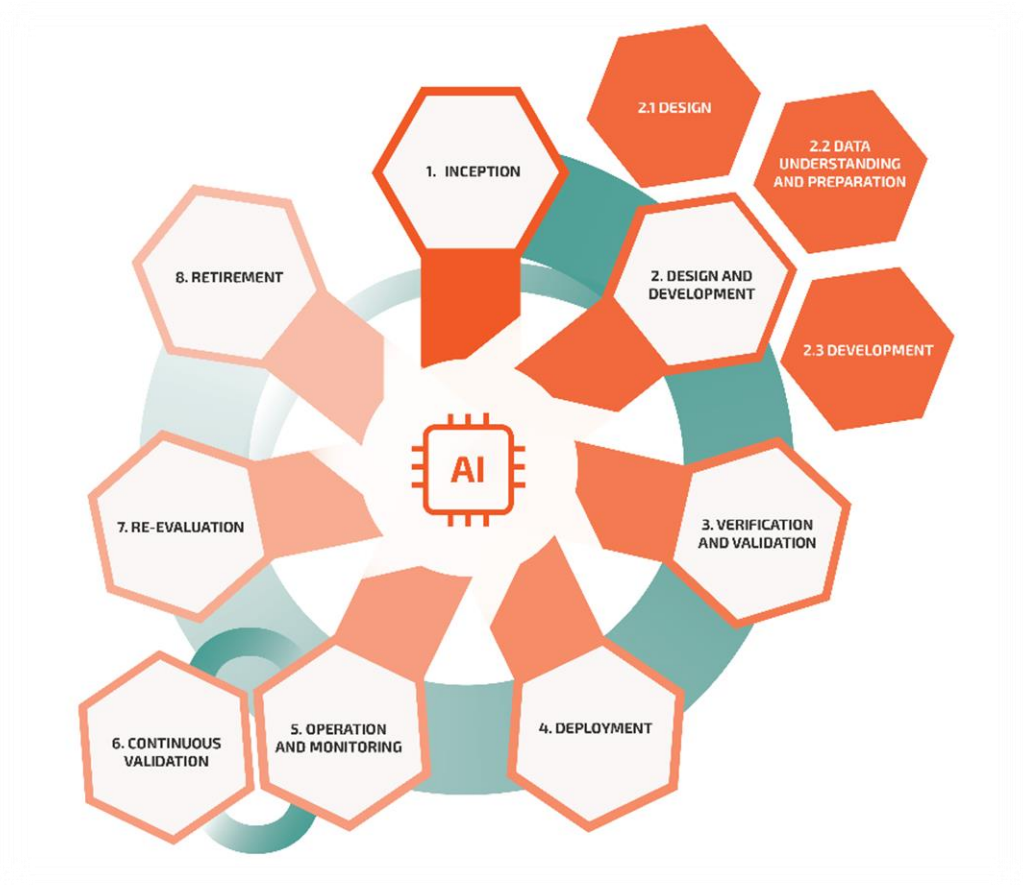
11. Sigurnost sustava umjetne inteligencije

Organizacije se pri osiguravanju sigurnosti obrade mogu osloniti na međunarodne sigurnosne standarde i dobre prakse u području kibernetičke sigurnosti. Međutim, kod UI sustava potrebno je adresirati i specifične rizike povezane s umjetnom inteligencijom, uključujući rizike koji proizlaze iz modela, podataka za treniranje, korisničkih unosa, izlaza sustava i integracija s drugim sustavima.

U skladu s člankom 32. Opće uredbe o zaštiti podataka, voditelj i izvršitelj obrade moraju provesti odgovarajuće tehničke i organizacijske mjere kako bi osigurali razinu sigurnosti primjerenu riziku. Pri tome se osobito uzimaju u obzir rizici od slučajnog ili nezakonitog uništenja, gubitka, izmjene, neovlaštenog otkrivanja osobnih podataka ili neovlaštenog pristupa osobnim podacima.

Subjekti koji uvode UI sustav trebaju provjeriti i sigurnosnu razinu svojih dobavljača zbog toga što rizici trećih strana mogu značajno utjecati na zaštitu osobnih podataka i kibernetičku sigurnost organizacije. Takvu provjeru potrebno je provesti prije ugovaranja usluge te periodično tijekom njezina trajanja, primjerice kroz provjeru relevantnih sigurnosnih kontrola, certifikata, izvješća o reviziji i usklađenosti s primjenjivim standardima, poput ISO/IEC 27001, odnosno relevantnim propisima iz područja kibernetičke sigurnosti.

11.1 Procjena rizika



Source: Based on ISO/IEC 22989

AI Privacy Risks & Mitigations – Large Language Models (LLMs) By Isabel

Svaka faza životnog ciklusa uključuje jedinstvene rizike koji zahtijevaju prilagođene strategije ublažavanja.

Implementacija privatnosti već u dizajnu i primjena tehnologija za zaštitu privatnosti (PETs) u svakoj fazi životnog ciklusa osigurava da se rizici rješavaju proaktivno, a ne retroaktivno.

Prilikom implementacije adekvatnih sigurnosnih mjera, nužno je procijeniti rizik. Procjena rizika općenito se smatra prvom fazom u upravljanju rizicima. Obuhvaća analizu rizika, koja uključuje identificiranje, procjenu i evaluaciju potencijalnih rizika. Kao početnu točku, analiza rizika zahtijeva pažljivu identifikaciju rizika koji se mogu pojaviti u danom kontekstu.

Kako bismo lakše identificirali rizike može se koristiti niz čimbenika rizika. Čimbenici rizika su stanja povezana s većom vjerojatnošću neželjenih ishoda. Oni mogu pomoći u identificiranju, procjeni i određivanju prioriteta potencijalnih rizika. Na primjer, obrada osjetljivih podataka i velikih količina podataka dva su čimbenika rizika s visokom razinom rizika. Njihovo prepoznavanje u vlastitom slučaju upotrebe može pomoći u identificiranju povezanih potencijalnih rizika i njihove ozbiljnosti.

Primjer

Visoka razina rizika	**Primjer primjenjivosti**
Korištenje LLM za odlučivanje ili sprječavanju ostvarivanja temeljnih prava pojedinaca ili o njihovom pristupu usluzi, izvršavanju ugovora ili pristupu financijskim uslugama predstavlja zabrinutost, posebno ako će se te odluke automatizirati bez ljudske intervencije. Pogrešne odluke mogle bi imati negativan utjecaj na pojedince.	Primjena LLM za utvrđivanje kreditne sposobnosti ili odobravanja kredita bez ljudskog nadzora ili za automatizaciju odluka o zapošljavanju, napredovanjima ili otkazima bez odgovarajućih zaštitnih mjera mogla bi negativno utjecati na pojedince.
Kada LLM obrađuje osjetljive podatke kao što su posebne kategorije podataka, osobni podaci povezani s osudama i kaznenim djelima, financijski podaci, podaci o ponašanju, jedinstveni identifikatori, podaci o lokaciji itd. To je razlog za zabrinutost jer bi neprimjerena obrada ovih osobnih podataka mogla negativno utjecati na pojedince.	Korištenje sustava baziranog na LLM za analizu zdravstvenih kartona, postavljanje dijagnoza ili obradu podataka povezanih s kaznenim osudama.

Akt o umjetnoj inteligenciji uvodi niz pojmova povezanih s upravljanjem rizicima, sigurnošću i zaštitom temeljnih prava. Ti su pojmovi relevantni i za organizacije koje razvijaju ili koriste sustave temeljene na velikim jezičnim modelima, osobito kada takvi sustavi uključuju obradu osobnih podataka.

U tom kontekstu važno je razlikovati opasnosti, prijetnje i ranjivosti. Opasnost označava mogući izvor štete, prijetnja okolnost ili aktera koji tu štetu može prouzročiti ili povećati,

dok ranjivost predstavlja slabost sustava, organizacijskog postupka ili sigurnosne mjere koja se može iskoristiti. Kod LLM sustava takvi se rizici mogu pojaviti, primjerice, pri unosu osobnih podataka u korisničke upite, pohrani i daljnjoj uporabi tih podataka, povezivanju modela s drugim sustavima ili nedovoljnoj kontroli pristupa.

Procjena tih rizika ne može se provoditi apstraktno. U skladu s pristupom Opće uredbe o zaštiti podataka, osobito njezinim zahtjevom da se u obzir uzmu priroda, opseg, kontekst i svrhe obrade, organizacija mora procijeniti konkretan način na koji se sustav koristi, koje kategorije podataka obrađuje, tko ima pristup podacima, koje skupine pojedinaca mogu biti pogođene i koje posljedice za njih mogu nastati.

Posebno je važno jasno definirati predviđenu namjenu sustava. Rizici se često pojavljuju upravo onda kada se sustav koristi izvan konteksta za koji je razvijen ili kada se naknadno integrira u procese koji nisu bili obuhvaćeni početnom procjenom. Stoga predviđena namjena, stvarni kontekst uporabe i moguće posljedice za ispitanike moraju biti polazište svake procjene rizika.

Upravljanje rizicima ne završava u trenutku uvođenja sustava u uporabu. Organizacija mora redovito pratiti način korištenja sustava, prijave incidenata, zahtjeve ispitanika, zapisnike pristupa, rezultate testiranja i povratne informacije korisnika. Takvi dokazi omogućuju prepoznavanje novih rizika i prilagodbu tehničkih i organizacijskih mjera.

Primjer

Rizik zaštite podataka	**Opis rizika**	**Potencijalni utjecaj s obzirom na Uredbu**	**Primjer**	**Model usluge**	**Pružatelj UI sustava**	**Subjekt koji uvodi UI sustav**
Nedovoljna zaštita osobnih podataka što na kraju može biti uzrok povrede podataka.	Zaštitne mjere za zaštitu osobnih podataka nisu provedene ili su nedovoljne.	Kršenje: čl. 32 Sigurnost obrade, čl. 5(1)(f) Integritet i povjerljivost i čl. 9 Obrada posebnih kategorija osobnih podataka	Otkrivanje osjetljivih podataka u korisničkim unosima ili tijekom obuke, zaključivanja i izlaza; neovlašteni pristup, nedovoljno šifriranje tijekom prijenosa podataka, zlouporaba API-ja, ranjivosti sučelja, neadekvatne tehnike anonimizacije ili filtriranja, izloženost trećim stranama.	LLM kao usluga	DA	DA

Nakon što su rizici identificirani, sljedeći ključni koraci unutar faze analize rizika su procjena i evaluacija rizika. To uključuje klasifikaciju i prioritizaciju rizika na temelju

njihove vjerojatnosti i ozbiljnosti ili potencijalnog utjecaja. Stvarna razina ili klasifikacija rizika uvelike će ovisiti o specifičnom slučaju upotrebe, operativnom kontekstu, praćenju sustava, rezultatima evaluacije modela i pogođenim dionicima.

Tijekom ove faze, rizici se analiziraju kako bi se detaljnije razumjele njihove implikacije. Ovaj proces uključuje procjenu čimbenika kao što su vjerojatnost pojave rizika, potencijalna šteta koju bi mogao uzrokovati i ranjivosti koje ga omogućuju.

Suradnja dionika igra vitalnu ulogu u ovom procesu, posebno s obzirom na multidisciplinarnu prirodu umjetne inteligencije, gdje su ulazni podaci iz tehničke, pravne, etičke, sigurnosne i operativne perspektive ključni za sveobuhvatno upravljanje rizicima. Etička matrica može biti vrijedan alat za identificiranje na koje bi dionike sustav temeljen na LLM-u mogao izravno ili neizravno utjecati. Mapiranje dionika na temelju njihove razine uključenosti i potencijalnog utjecaja na njih omogućuje organizacijama da uključe perspektive i brige onih na koje se odnosi u proces klasifikacije i ublažavanja rizika. Time se osigurava da se rješavaju etička i praktična razmatranja implementacije LLM-ova, usklađujući implementaciju sustava s potrebama i pravima svih pogođenih strana.

Nakon identificiranja rizika, sljedeći koraci su:

- Procjena vjerojatnosti i ozbiljnosti identificiranih rizika.
- Procjena treba li se postupati s rizicima kako bi se osigurala zaštita osobnih podataka i dokazala usklađenost s Općom uredbom o zaštiti podataka.

Dostupne su različite metodologije upravljanja rizicima za klasifikaciju i procjenu rizika. Ovaj dokument nema za cilj propisati ili definirati određenu metodologiju, jer izbor treba odrediti svaka organizacija.

U općim terminima upravljanja rizicima, rizik se može izraziti kao:

Rizik = Vjerojatnost x Ozbiljnost

Ova jednadžba ističe da je rizik određen vjerojatnošću nastanka događaja, u kombinaciji s potencijalnim utjecajem ili ozbiljnošću nastale štete. Rizik je u Uredbi (uvodna izjava 75) definiran kao potencijalna šteta za prava i slobode pojedinaca, različite vjerojatnosti i ozbiljnosti, koja proizlazi iz obrade osobnih podataka. Slično tome, Akt (članak 3.) definira rizik kao „kombinaciju vjerojatnosti nastanka štete i ozbiljnosti te štete;“. Za procjenu razine rizika za zaštitu podataka i privatnost prilikom nabave, razvoja ili korištenja UI sustava, bitno je procijeniti i vjerojatnost i ozbiljnost ostvarenja identificiranih rizika.

Za određivanje vjerojatnosti rizika može se koristiti sljedeća matricu klasifikacije rizika s četiri razine:

- Vrlo visoka - Velika vjerojatnost nastanka događaja

- Visoka - Značajna vjerojatnost nastanka događaja
- Niska - Mala vjerojatnost nastanka događaja
- Malo vjerojatno - Nema dokaza da se takav rizik materijalizira u bilo kojem slučaju

S obzirom na 7 kriterija temeljem kojih se određuje ukupna vjerojatnost, a to su:

- Frekventnost korištenja - koliko često se koristi UI sustav, povećavajući izloženost potencijalnom riziku koji utječe na pouzdanost (očekivano vrijeme prije kvara).
- Izloženost scenarijima visokog rizika - opseg u kojem UI sustav radi u osjetljivim ili visokorizičnim okruženjima.
- Historijski podaci - Prošli slučajevi sličnih rizika ili kvarova u istim ili usporedivim UI sustavima.
- Okolišni faktori - Vanjski, nekontrolirani uvjeti koji utječu na performanse ili pouzdanost sustava (npr. politička nestabilnost, regulatorne praznine, financijska ograničenja).
- Otpornost sustava - Stupanj u kojem je UI sustav otporan na kvarove ili nenamjerno ponašanje.
- Kvaliteta i integritet podataka - Opseg u kojem se UI sustav oslanja na točne, nepristrane i potpune podatke. Može se promijeniti boljom obradom skupa podataka ili validacijom.
- Ljudski nadzor i ekspertiza - Kako vještine i donošenje odluka ljudskih operatera utječu na pouzdanost sustava i vjerojatnost rizika. Može se promijeniti obukom ili poboljšanjima nadzora.

Iz navedenih kriterija proizaći će u konačnici aritmetička sredina čija će se vrijednost koristiti u daljnjem računu.

Potom je potrebno odrediti ozbiljnost rizika za koji se može također koristiti klasifikacija s četiri razine:

- Vrlo značajna (katastrofalna šteta) - utječe na ostvarivanje temeljnih prava i javnih sloboda, a posljedice su joj nepovratne i/ili su posljedice povezane s posebnim kategorijama podataka ili s kaznenim djelima te su nepovratne i/ili uzrokuje značajnu društvenu štetu, poput diskriminacije, te je nepovratna i/ili nepovratno utječe na posebno ranjive subjekte podataka, posebno djecu, i/ili uzrokuje značajne i nepovratne moralne ili materijalne gubitke.
- Značajna (kritična šteta) - slučajevi kada su učinci reverzibilni i/ili postoji gubitak kontrole ispitanika nad njegovim osobnim podacima, gdje je opseg podataka velik u odnosu na kategorije podataka ili broj ispitanika i/ili dolazi ili može doći do krađe identiteta ispitanika i/ili mogu nastati značajni financijski gubici za ispitanike i/ili gubitak povjerljivosti ispitanika ili kršenje dužnosti povjerljivosti i/ili postoji društvena šteta za ispitanike ili određene skupine ispitanika.

- Ograničena (teška šteta) - vrlo ograničen gubitak kontrole nad nekim osobnim podacima i za određene ispitanike, osim posebnih kategorija ili nepovratnih kaznenih djela ili osuda i/ili zanemarivi i nepovratni financijski gubici i/ili gubitak povjerljivosti podataka na koje se odnosi profesionalna tajna, ali ne i posebne kategorije.
- Vrlo ograničena (umjerena ili mala šteta) - štetni učinci su svi reverzibilni.

Ukupna ozbiljnost odredit će se s obzirom na 11 kriterija:

1. **Priroda temeljnog prava i usklađenost s pravnim ograničenjima** - ovaj kriterij procjenjuje prirodu pogođenog temeljnog prava — je li ono apsolutno ili podložno ograničenjima — te u kojoj mjeri je primjena UI sustava usklađena s zakonitim i razmjernim ograničenjima.

Apsolutna prava su nederogabilna i ne mogu se ograničiti ni pod kojim okolnostima, dok se druga prava mogu ograničiti samo ako zadiru u njih na način koji ispunjava stroge zahtjeve zakonitosti, razmjernosti i nužnosti. Ovaj kriterij pomaže utvrditi težinu utjecaja na temelju stupnja neusklađenosti ili povrede zaštite prava.

2. **Priroda osobnih podataka** - ovaj kriterij procjenjuje osjetljivost osobnih podataka koji se obrađuju, uzimajući u obzir njihov potencijal za nanošenje štete u slučaju zlouporabe. Posebne kategorije podataka (npr. zdravstveni, biometrijski ili genetski podaci) predstavljaju veće rizike za temeljna prava poput privatnosti i autonomije.

3. **Kategorija ispitanika** (npr. maloljetnici ili ne) - ovaj kriterij procjenjuje ranjivost pojedinaca čiji se podaci obrađuju.

Ranjive skupine (npr. maloljetnici, marginalizirane zajednice) izložene su većim rizicima štete zbog zlouporabe podataka.

4**. Svrha obrade** - ovaj kriterij procjenjuje zakonitost, nužnost i razmjernost svrhe za koju se osobni podaci obrađuju.

Nezakonite ili nerazmjerne svrhe povećavaju ozbiljnost rizika.

5. **Opseg utjecaja (društveni, grupni, individualni) i broj pogođenih ispitanika** - širina povrede na razini društva, skupina i pojedinaca. Ovaj kriterij uzima u obzir opseg utjecaja s obzirom na broj osoba čiji su podaci pogođeni.

6. **Kontekstualna i sektorska osjetljivost** - kako specifični kontekstualni čimbenici ili područja pojačavaju ozbiljnost zadiranja u pravo.

Uključuje situacijske rizike poput društveno-političke nestabilnosti te utjecaj na djecu i druge ranjive skupine.

7. **Reverzibilnost, oporavak i mogućnost sanacije - težina ili izvedivost uklanjanja štete te vrijeme potrebno za oporavak.**

Uključuje situacije u kojima je šteta nepovratna.

8. **Trajanje i postojanost štete** - duljina trajanja i postojanost negativnih učinaka uzrokovanih zadiranjem u pravo.

9. **Brzina materijalizacije rizika -** brzina kojom se rizik ostvaruje: postupno, naglo ili kontinuirano.

10. **Transparentnost i mehanizmi odgovornosti -** stupanj transparentnosti sustava i postojanje mehanizama odgovornosti.

11. **„Domino efekt“ i kaskadni učinci** -opseg u kojem zadiranje uzrokuje dodatne štete u drugim sustavima ili područjima.

Procjena vjerojatnosti i ozbiljnosti pruža osnovu za određivanje ukupne razine rizika identificiranih rizika za privatnost i zaštitu podataka. Korištenjem matrice klasifikacije s četiri razine za vjerojatnost i ozbiljnost, rizici se mogu kategorizirati u konačne klasifikacije: vrlo visok, visok, srednji ili nizak.

Matrica, kao što je prikazano u nastavku, služi kao praktičan alat za dobivanje ovih klasifikacija, nudeći jasno i strukturirano rangiranje za određivanje prioriteta rizika i vođenje odgovarajućih strategija ublažavanja. Ova klasifikacija ključni je korak u sljedećem procesu obrade rizika jer osigurava da se resursi usmjeravaju na učinkovito rješavanje najhitnijih rizika.

Vjerojatnost

Vrlo visoka				
Visoka				
Srednja				
Niska				
	Vrlo značajna	Značajna	Ograničena	Vrlo ograničena

Ozbiljnost

Dobre prakse pokazuju da je potrebno najprije adresirati rizike koji su procijenjeni kao vrlo visoki ili visoki. Svaka organizacija treba odrediti vlastite kriterije prihvatljivosti rizika te, ovisno o tim kriterijima, utvrditi dinamiku i prioritete postupanja.

Za svaki identificirani rizik organizacija treba odabrati odgovarajući način postupanja, odnosno jednu ili više mjera tretiranja rizika:

- ublažavanje rizika — provedba tehničkih i organizacijskih mjera kojima se smanjuje vjerojatnost nastanka rizika i/ili ozbiljnost njegovih posljedica;
- prijenos rizika — prijenos dijela rizika ili odgovornosti na drugu stranu, primjerice ugovornim uređenjem, osiguranjem ili korištenjem vanjskih pružatelja usluga;
- izbjegavanje rizika — uklanjanje aktivnosti, funkcionalnosti ili okolnosti koje uzrokuju rizik, ako se rizik ne može svesti na prihvatljivu razinu;
- prihvaćanje rizika — svjesna i dokumentirana odluka da se rizik neće dodatno umanjivati jer se nalazi unutar unaprijed utvrđenih granica prihvatljivosti rizika.

Nakon odabira i provedbe mjera tretiranja rizika, organizacija treba procijeniti preostali, odnosno rezidualni rizik te provjeriti je li on prihvatljiv. Ako rezidualni rizik nije prihvatljiv, potrebno je ponovno razmotriti i prilagoditi mjere sve dok se rizik ne svede na prihvatljivu razinu ili dok se ne donese odluka da se određena obrada, funkcionalnost ili uporaba sustava ne može provesti na usklađen i siguran način.

Organizacija također treba u redovitim vremenskim razmacima, kao i nakon značajnih promjena sustava, svrhe obrade, kategorija podataka, korisnika ili konteksta uporabe, ponovno procijeniti rizike i učinkovitost primijenjenih mjera. Upravljanje rizicima stoga ne predstavlja jednokratnu aktivnost, nego kontinuiran proces procjene, dokumentiranja, praćenja i prilagodbe mjera.

Primjer

Virtualni pomoćnik (Chatbot) za upite kupaca[7]

Scenarij: Poduzeće želi implementirati chatbot kako bi svojim klijentima pružilo opće informacije o svojim proizvodima i uslugama. Chatbot će imati pristup postojećim podacima o klijentima integracijom sa sustavom za upravljanje klijentima (bazom podataka iz CRM-a)

- korisničko sučelje chatbota bit će izgrađeno na temelju „gotovog” (off-the-shelf) velikog jezičnog modela, koji će koristiti RAG kako bi stekao domenski specifična znanja potrebna za rad

[7] EDPB Support Pool of Experts Programme: AI Privacy Risks & Mitigations Large Language Models (LLMs) By Isabel BARBERÁ https://www.edpb.europa.eu/system/files/2025-04/ai-privacy-risks-and-mitigations-in-llms.pdf

Faza životnog ciklusa u kojoj se nalazimo: Dizajn & Razvoj

Faktor rizika	Primjenjivost slučaja upotrebe
Obrada velikih razmjera	Značajna količina podataka bit će obrađena zbog naše opsežne baze podataka o kupcima i velike količine informacija pohranjenih u našem CRM sustavu.
Niska kvaliteta podataka	Ulazni podaci za upite korisnika mogu biti niske kvalitete, a baza podataka CRM-a nije validirana, što bi moglo dovesti do netočnosti ili neučinkovitosti u obradi.
Nedostatne sigurnosne mjere	Postoji potencijalni rizik od prijenosa osobnih podataka u zemlje bez odgovarajuće razine zaštite, posebno ako je model LLM smješten ili se održava u takvim regijama.

- Kako će se sustav koristiti, uključujući njegovu svrhu, trajanje i učestalost.
- Kategorije pojedinaca ili skupina na koje sustav utječe.
- Posebni rizici od štete za temeljna prava.
- Mjere za ljudski nadzor i upravljanje.
- Koraci za rješavanje i ublažavanje rizika ako se ostvare.

❑ U modeliranju prijetnji proces se obično temelji na četiri temeljna pitanja:
❑ Na čemu radimo?
❑ Što može poći po zlu?
❑ Što ćemo učiniti u vezi toga?
❑ Jesmo li dobro obavili posao?

	Kriteriji ozbiljnosti										Ukupni rezultat		TSS
Rizik	1	2	3	4	5	6	7	8	9	10	11	Izračun	Rezultat	Rezultat
1	3	2	2	2	3	1	2	2	3	2	3	3+2+2+2+3+1+2+2+3+2+3 / 11	2,27	Značajna/kritična šteta
2	3	2	2	2	1	1	3	2	2	2	3	3+2+2+2+1+1+3+2+2+2+3 / 11	2,09	Značajna/kritična šteta
3	3	2	2	2	2	1	2	2	2	2	2	3+2+2+2+2+1+2+2+2+2+2 / 11	2	Značajna/kritična šteta
4	3	2	2	2	2	1	3	2	1	2	2	3+2+2+2+2+1+3+2+1+2+2 / 11	2	Značajna/kritična šteta
5	3	2	2	3	3	1	3	2	1	2	2	3+2+2+3+3+1+3+2+1+2+2 / 11	2,18	Značajna/kritična šteta
6	3	2	2	2	3	1	2	1	1	2	1	3+2+2+2+3+1+2+1+1+2+1 / 11	1,81	Značajna/kritična šteta
7	3	2	2	2	3	1	2	1	1	2	1	3+2+2+2+3+1+2+1+1+2+1 / 11	1,81	Značajna/kritična šteta
8	3	2	2	2	3	1	2	1	1	2	1	3+2+2+2+3+1+2+1+1+2+1 / 11	1,81	Značajna/kritična šteta

Primjenom klasifikacijske matrice na dobivene ocjene vjerojatnosti i ozbiljnosti možemo odrediti odgovarajući stupanj klasifikacije rizika. U našem slučaju za sve rizike kombinacija Niska vjerojatnost + Značajna ozbiljnost nudi rezultat **visokog rizika.**

1.0 - 1.5: Vrlo ograničena/umjerena ili manja šteta (1. razina)
1.6 - 2.5: Ograničena / ozbiljna šteta (razina 2)
2.6 - 3.5: Značajna/kritična šteta (3. razina)
3.6 - 4.0: Vrlo značajna/katastrofalna šteta (4. razina)

Vjerojatnost	**vrlo visoka**	Srednja vrijednost	visoka	vrlo visoka	vrlo visoka
	visoka	Niska	visoka	vrlo visoka	vrlo visoka
	Niska	Niska	Srednja vrijednost	visoka	vrlo visoka
	Malo vjerojatno	Niska	Niska	Srednja vrijednost	vrlo visoka
		vrlo ograničeno	**Ograničeno**	**Značajno**	**vrlo značajan**
		Ozbiljnost			

Faza protoka podataka	Opis	Rizici za privatnost	Preporuke za ublažavanje
Unos korisnika	Korisnici stupaju u interakciju s chatbotom tako što putem sučelja (npr. internetske stranice ili mobilne aplikacije) navode svoje ime, adresu e-pošte i preferencije.	▪ Ulaz se može presresti ako se prenosi preko nesigurne veze. ▪ Korisnici mogu pružiti nepotrebne ili prekomjerne osobne podatke. ▪ Djeca ili ranjivi korisnici mogu dijeliti osobne podatke.	▪ Sigurni prijenos podataka pomoću odgovarajućih protokola za šifriranje. ▪ Primijeniti ograničenja unosa kako bi se prikupljeni podaci ograničili na ono što je ključno. ▪ Upotrijebite mehanizme provjere dobi kako biste zaštitili ranjive korisnike (djeca koja nisu naša skupina korisnika) ▪ Jasno obavijestite korisnike o tome kako se njihovi podaci obrađuju i upozorite ih da ne dijele osjetljive ili povjerljive informacije prilikom korištenja chatbota.

Što može poći po zlu?
Nedovoljna zaštita osobnih podataka koja dovodi do povrede podataka.
Pogrešna klasifikacija podataka o osposobljavanju kao anonimnih, a radi se o osobnim podacima.
Mogući negativan učinak na ispitanike koji bi mogao negativno utjecati na temeljna prava.
Ispitanici ne mogu ostvariti svoje prava iz GDPR-a.
Nezakonita prenamjena osobnih podataka.
Nezakonito neograničeno pohranjivanje osobnih podataka.
Nezakonit prijenos osobnih podataka.
Kršenje načela smanjenja količine podataka.

Faze	Mogući rizici
Korisnički ulaz	Otkrivanje osjetljivih podataka u fazi unosa podataka, neovlašteni pristup, nedostatak transparentnosti, *adversial attacks*
Sučelje davatelja usluga & API	Presretanje podataka, zlouporaba API-ja, ranjivosti sučelja
Obrada osobnih podataka na infrastrukturi LLM pružatelja usluga	„model inversion" (rekonstrukcija podataka iz modela), nenamjerno bilježenje podataka, neuspješna anonimizacija, neovlašteni pristup zapisima, rizici agregiranja podataka, izloženost osobnih podataka neovlaštenim trećim stranama, neodgovarajuće politike pohrane podataka
Outputi	Netočni ili osjetljivi odgovori, rizici ponovne identifikacije, zlouporaba rezultata

	Kriteriji vjerojatnosti							Ukupni rezultat		TPS
Rizik	1	2	3	4	5	6	7	Izračun	Rezultat	Rezultat
1	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska
2	3	1	4	1	2	3	2	3+1+4+1+2+3+2 / 7	2,28	Niska
3	3	1	2	1	2	3	2	3+1+2+1+2+3+2 / 7	2	Niska
4	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska
5	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska
6	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska
7	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska
8	3	1	3	1	2	3	2	3+1+3+1+2+3+2 / 7	2,14	Niska

11.2 Napadi na UI modele i UI sustave

UI modeli i UI sustavi mogu biti izloženi različitim oblicima napada i zlouporaba, pri čemu pojedine prijetnje mogu dovesti do neovlaštenog pristupa osobnim podacima, njihova otkrivanja, ponovne identifikacije ispitanika, manipulacije izlazima sustava ili drugih štetnih posljedica za pojedince. U ovom poglavlju navode se odabrani napadi i scenariji ugroze koji su osobito relevantni za procjenu sigurnosti UI sustava u smislu Opće uredbe o zaštiti podataka.

Navedeni pregled ne treba tumačiti kao iscrpan katalog svih mogućih prijetnji, nego kao operativni okvir za prepoznavanje najvažnijih rizika koje je, ovisno o okolnostima konkretnog slučaja, potrebno dodatno razraditi, testirati i povezati s odgovarajućim zaštitnim mjerama.

Trovanje podataka i modela *(eng. Data and Model Poisoning)*

Trovanje podataka i modela označava manipulaciju podacima ili komponentama modela, primjerice podacima za predtreniranje, fino podešavanje ili embeddinge, s ciljem narušavanja sigurnosti, točnosti, pouzdanosti ili očekivanog ponašanja modela. Takva manipulacija može dovesti do smanjenja performansi modela, uvođenja pristranosti, ciljanih pogrešaka, zlonamjernih obrazaca ponašanja ili tzv. backdoor funkcionalnosti. U kontekstu LLM sustava rizik je osobito izražen kada se koriste vanjski, nedovoljno provjereni ili dinamički izvori podataka. OWASP ovaj rizik opisuje kao manipulaciju podacima za predtreniranje, fino podešavanje ili embeddinge koja može ugroziti sigurnost, performanse ili etičko ponašanje modela.

Iscrpljivanje resursa / neograničena potrošnja *(eng. Unbounded Consumption / Model DoS)*

Iscrpljivanje resursa odnosi se na napade ili zlouporabe kojima se uzrokuje prekomjerna i nekontrolirana potrošnja računalnih, financijskih ili infrastrukturnih resursa LLM aplikacije. Do toga može doći slanjem velikog broja zahtjeva, vrlo dugih ili složenih upita, zahtjeva koji aktiviraju skupe operacije, zlouporabom kontekstnog prozora, povezanih alata ili API-ja. Posljedice mogu uključivati narušavanje dostupnosti sustava, povećanje latencije, degradaciju usluge, nerazmjerne troškove ili otežano pružanje usluge legitimnim korisnicima. OWASP 2025. navodi da je **Unbounded Consumption** proširenje ranijeg koncepta denial of service te uključuje i rizike povezane s upravljanjem resursima i neočekivanim troškovima.

Napadi zaključivanja o članstvu *(eng. Membership Inference Attacks)*

Napadi zaključivanja o članstvu odnose se na pokušaje utvrđivanja je li određeni zapis, odnosno podatak o određenoj osobi, bio dio skupa podataka korištenog za treniranje ili fino podešavanje modela. Uspješan napad ne mora nužno otkriti sam sadržaj zapisa, ali može otkriti činjenicu povezanosti pojedinca s određenim skupom podataka, uslugom, obradom ili kategorijom podataka, što samo po sebi može predstavljati rizik za privatnost.

Napadi zaključivanja o atributima *(eng. Attribute Inference Attacks)*

Napadi zaključivanja o atributima odnose se na pokušaje izvođenja osjetljivih ili drugih atributa pojedinca koji nisu izravno dostupni, već se zaključuju iz izlaza modela, pomoćnih informacija ili kombinacije više izvora podataka. Takvi napadi mogu dovesti do neovlaštenog profiliranja, otkrivanja obilježja ispitanika ili zaključivanja informacija koje pojedinac nije namjeravao otkriti.

Napadi inverzije modela *(eng. Model Inversion Attacks)*

Napadi inverzije modela označavaju pokušaje da se, na temelju izlaza modela ili drugih dostupnih informacija o njegovu ponašanju, rekonstruiraju ili aproksimiraju značajke, uzorci ili osjetljivi elementi podataka koji su utjecali na treniranje modela. Ovisno o

okolnostima, takvi napadi mogu povećati rizik od ponovne identifikacije, otkrivanja osjetljivih podataka ili dobivanja informacija o osobama čiji su podaci korišteni u razvoju modela. OWASP povezuje otkrivanje osjetljivih informacija s rizicima kao što su membership inference, model inversion i model extraction.

Reprodukcija podataka za treniranje *(eng. Training Data Regurgitation)*

Reprodukcija podataka za treniranje označava neželjeno ponašanje modela ili scenarij napada u kojem model generira dijelove podataka korištenih tijekom treniranja, fino podešavanja ili druge prilagodbe modela. To može uključivati tekst, identifikatore, kontaktne podatke, povjerljive dokumente, kod ili druge sadržaje. Takvo ponašanje može dovesti do neovlaštenog otkrivanja osobnih podataka, poslovnih tajni ili drugih povjerljivih informacija.

Eksfiltracija podataka *(eng. Data Exfiltration)*

Eksfiltracija podataka označava neovlašteno izvlačenje podataka iz sustava umjetne inteligencije ili s njime povezanih izvora podataka. U kontekstu LLM aplikacija to može uključivati manipulaciju modelom putem prompt injection napada, zlouporabu povezanih alata, neadekvatne kontrole pristupa, ranjivosti u RAG sustavima, curenje podataka kroz vanjske integracije ili otkrivanje podataka putem izlaza modela. Posljedica može biti neovlašteno otkrivanje osobnih podataka, povjerljivih poslovnih informacija, sigurnosnih vjerodajnica ili drugih zaštićenih podataka.

Ekstrakcija modela *(eng. Model Extraction / Model Theft)*

Ekstrakcija modela označava napad pri kojem napadač putem ponavljanih upita i analize odgovora nastoji rekonstruirati funkcionalnost modela, aproksimirati njegovo ponašanje, izraditi zamjenski model ili izvući informacije o njegovim parametrima, arhitekturi, pravilima odlučivanja ili sigurnosnim ograničenjima. Takvi napadi mogu ugroziti poslovne tajne, intelektualno vlasništvo i sigurnost sustava jer napadaču omogućuju bolje razumijevanje ponašanja modela i pripremu daljnjih zlouporaba.

Prompt injection

Prompt injection označava napad manipulacijom ulaznog sadržaja s ciljem promjene ponašanja modela, zaobilaženja sigurnosnih pravila, otkrivanja podataka ili izvršavanja neželjenih radnji. Napad može biti izravan, kada korisnik neposredno unese zlonamjerne upute, ili neizravan, kada model zlonamjerne upute preuzme iz vanjskog izvora, primjerice web-stranice, dokumenta, elektroničke pošte, baze znanja ili drugog sadržaja koji obrađuje. OWASP opisuje prompt injection kao situaciju u kojoj korisnički promptovi mijenjaju ponašanje ili izlaz LLM-a na neželjen način, uključujući mogućnost zaobilaženja pravila, omogućavanja neovlaštenog pristupa ili utjecaja na odluke sustava.

Također, u kontekstu članka 32. Opće uredbe o zaštiti podataka nije dovoljno identificirati samo tehničke prijetnje usmjerene na UI model ili UI sustav, nego je potrebno procijeniti i

njihove moguće učinke na prava i slobode ispitanika. Stoga se za svaki relevantni scenarij napada preporučuje utvrditi: koje vrste osobnih podataka mogu biti pogođene, koje su moguće posljedice za ispitanike, kolika je vjerojatnost uspjeha napada, kolika bi bila ozbiljnost posljedica te koje tehničke i organizacijske mjere mogu taj rizik smanjiti na prihvatljivu razinu. Takav pristup omogućuje da se tehnička analiza sigurnosti izravno poveže s obvezom uspostave primjerene razine sigurnosti obrade s obzirom na rizik za prava i slobode fizičkih osoba.

Tablica. Primjeri napada relevantnih za privatnost i zaštitu osobnih podataka

Vrsta napada	Kada je posebno relevantan	Mogući učinci na ispitanike	Preporučene mjere	Napomena
Napad zaključivanja o članstvu (membership inference attack)	Kada model ili sustav omogućuje dovoljno precizne upite ili izlaze iz kojih se može zaključiti je li određena osoba bila dio trening skupa.	Moguće otkrivanje da je osoba povezana s određenim skupom podataka, uslugom, dijagnozom, događajem ili organizacijom; moguća stigma, diskriminacija i reputacijska šteta.	Minimizacija podataka, pseudonimizacija, diferencijalna privatnost, ograničenje pristupa modelu, ograničenje broja upita, nadzor anomalija, testiranje otpornosti.	Ovaj napad treba izravno povezati s rizikom gubitka povjerljivosti i gubitka kontrole nad osobnim podacima.
Napad zaključivanja o atributima (attribute inference attack)	Kada se iz izlaza modela ili kombinacije podataka mogu izvesti dodatna obilježja osobe koja nisu izravno otkrivena.	Zaključivanje o zdravlju, financijskom statusu, navikama, lokaciji ili drugim osjetljivim obilježjima; moguća diskriminacija ili nepošten tretman.	Smanjenje granularnosti izlaza, filtriranje visokorizičnih atributa, kontrola pristupa, procjena nužnosti varijabli, testovi inferencije i pristranosti.	U procjeni s obzirom na zaštitu podataka naglasak treba staviti na rizik neovlaštenog profiliranja i štetnih posljedica za pojedinca, a ne samo na "tehničku privatnost".
Regurgitacija trening podataka / eksfiltracija osobnih podataka	Kada model može reproducirati dijelove trening podataka ili kada se osobni podaci mogu izvući iz sustava, logova, promptova, vektorske baze ili povezanih spremišta.	Neovlašteno otkrivanje osobnih podataka, povjerljivih zapisa, poslovnih tajni ili posebnih kategorija podataka.	Filtriranje i čišćenje trening podataka, deduplikacija, zaštite na razini izlaza, ograničenje pristupa, segmentacija, enkripcija, DLP mjere, red teaming i ciljano testiranje.	Ovo je jedna od najizravnijih veza s člankom 32. stavkom 2. Uredbe, jer se radi o riziku neovlaštenog otkrivanja ili pristupa osobnim podacima.
Inverzija modela / rekonstrukcijski napadi	Kada bi se iz modela, njegovih parametara ili odgovora mogla rekonstruirati obilježja osoba ili dijelovi izvornih podataka.	Povećan rizik ponovne identifikacije, otkrivanja osjetljivih podataka.	Ograničenje pristupa modelu i artefaktima, kontrola izlaza, zaštita parametara, privatnost u učenju, zaštita kontrolnih točaka, politika zadržavanja i brisanja artefakata.	Ovdje treba procjenjivati i rizik "posrednog" izdvajanja osobnih podataka, ne samo doslovnog otkrivanja cijelog zapisa.
Prompt injection / jailbreaking	Osobito kada je UI sustav povezan s alatima, RAG-om, internim bazama znanja, datotekama, API-jima ili izvršavanjem radnji.	Može dovesti do otkrivanja osobnih podataka iz povezanih izvora, zaobilaženja zaštita ili poduzimanja neovlaštenih radnji koje utječu na ispitanike.	Odvajanje sistemskih uputa od korisničkog unosa, sandboxing alata, autorizacija po radnji, filtriranje promptova, validacija izlaza, ograničenje dohvaćanja podataka, zapisivanje i nadzor pokušaja.	Ovaj napad nije uvijek izravno rizik za osobne podatke, ali to postaje kada dovodi do pristupa osobnim podacima ili pogrešnih radnji nad njima.
Trovanje podataka (data poisoning)	Kada napadač može utjecati na trening, fine-tuning, validacijske skupove, referentne baze ili vanjske izvore na koje se sustav oslanja.	Netočni, manipulirani ili pristrani izlazi koji mogu dovesti do nepravednog postupanja prema ispitanicima, pogrešnih preporuka ili štetnih odluka.	Provjera izvora i provenijencije podataka, kontrola integriteta, odobravanje promjena, detekcija anomalija, razdvajanje dužnosti, verifikacija kvalitete setova podataka.	Ovdje je fokus prije svega na integritetu, točnosti i pravednosti ishoda za ispitanike, a ne samo na povjerljivosti.
Model extraction / model stealing	Kada je moguće kroz niz upita rekonstruirati funkcionalnost modela ili zaobići ugrađene zaštite.	Najčešće neizravan porast rizika: olakšavanje kasnijih napada, zaobilaženje zaštitnih mjera, preciznije izvođenje inferencijskih ili eksfiltracijskih napada.	Rate limiting, autentikacija, watermarking, nadzor obrazaca upita, segmentacija prava, zaštita API-ja, detekcija automatiziranog izvlačenja.	Sam po sebi ne mora značiti povredu osobnih podataka, ali može bitno povećati izloženost drugim napadima koji pogađaju prava i slobode ispitanika.

11.3 Implementacija mjera

Osim općih mjera informacijske i kibernetičke sigurnosti, organizacije koje razvijaju, prilagođavaju, integriraju ili koriste sustave umjetne inteligencije trebaju razmotriti i mjere koje adresiraju rizike specifične za umjetnu inteligenciju. Ti se rizici mogu odnositi osobito na povjerljivost i integritet podataka za treniranje, sigurnost modela, pouzdanost izlaza, mogućnost neovlaštenog izdvajanja osobnih podataka, trovanje podataka, prompt injection, neovlašten pristup povezanim sustavima te gubitak kontrole nad podacima.

Mjere se ne primjenjuju mehanički niti jednako za sve sustave. Njihov odabir mora ovisiti o prirodi, opsegu, kontekstu i svrhama obrade, vrsti osobnih podataka, kategorijama ispitanika, namjeravanoj svrsi sustava, načinu pristupa modelu, povezanim alatima i bazama podataka te ozbiljnosti mogućih posljedica za prava i slobode pojedinaca. U skladu s člankom 32. Opće uredbe o zaštiti podataka, voditelj i izvršitelj obrade moraju osigurati razinu sigurnosti primjerenu riziku, uzimajući u obzir stanje tehnike, troškove provedbe te prirodu, opseg, kontekst i svrhe obrade.

Pri odabiru mjera organizacija treba osobito razmotriti mjere koje se odnose na: podatke za treniranje, razvoj i integraciju sustava, rad sustava u produkcijskom okruženju te horizontalne organizacijske mjere.

Tablica 1. Pregled područja mjera

Područje mjera	Svrha	Primjeri mjera koje treba razmotriti
Podaci za treniranje	Osigurati povjerljivost, integritet, kvalitetu i kontrolu nad podacima koji se koriste za treniranje, validaciju, fino podešavanje ili evaluaciju modela.	provjera pouzdanosti, kvalitete i integriteta izvora podataka; verzioniranje skupova podataka; kontrola pristupa; pseudonimizacija ili anonimizacija kada je moguća; korištenje sintetičkih ili fiktivnih podataka kada stvarni podaci nisu nužni; enkripcija; segmentacija osjetljivih skupova podataka; mjere protiv gubitka kontrole nad podacima
Razvoj i integracija sustava	Ugraditi zaštitu podataka i sigurnost u dizajn, razvojno okruženje, odabir alata, modela i komponenti te postupke testiranja.	zaštita podataka po dizajnu i po zadanim postavkama; provjerene biblioteke, alati i modeli; kontrolirano razvojno okruženje; DevSecOps; sigurni formati; revizija koda; dokumentacija; sigurnosne provjere i red teaming
Rad sustava u produkciji	Osigurati sigurno i predvidljivo korištenje sustava u stvarnom okruženju te nadzirati izlaze, pristupe, incidente i odstupanja od predviđene namjene.	informiranje korisnika o ograničenjima sustava; upute za tumačenje rezultata; filtriranje izlaza; nadzor promptova i odgovora; zaštita povezanih alata i API-ja; mogućnost zaustavljanja ili rollbacka; praćenje incidenata i zahtjeva ispitanika
Horizontalne organizacijske mjere	Uspostaviti upravljački, organizacijski i dokumentacijski okvir koji omogućuje kontinuirano praćenje, dokazivanje usklađenosti i prilagodbu mjera tijekom životnog ciklusa sustava.	sigurnosna politika i akcijski plan; plan odgovora na incidente; multidisciplinarni tim; edukacija; upravljanje ovlaštenjima; logiranje i analiza pristupa; periodična revizija mjera; DPIA kada je potrebna

Tablica 2. Mjere vezane uz podatke za treniranje

Mjera	Kada je osobito relevantna	Svrha / napomena
Provjera pouzdanosti izvora podataka i anotacija	Kada se koriste vanjski, otvoreni, dinamički ili automatizirano prikupljeni izvori podataka.	Smanjuje rizik korištenja netočnih, manipuliranih, nezakonito prikupljenih ili kontekstualno neprimjerenih podataka. Provjeru treba provoditi i nakon promjena izvora podataka.
Provjera kvalitete podataka i anotacija	Kada kvaliteta podataka izravno utječe na točnost, pravednost, sigurnost ili zakonitost izlaza modela.	Obuhvaća provjeru potpunosti, točnosti, reprezentativnosti, relevantnosti, dosljednosti i mogućih pristranosti. Posebno je važna kod kontinuiranog učenja i rizika od data drifta.
Provjera integriteta podataka kroz životni ciklus	Kada postoji rizik neovlaštene izmjene, trovanja podataka ili manipulacije referentnim bazama i izvorima znanja.	Omogućuje otkrivanje pokušaja data poisoninga, neovlaštenih izmjena i degradacije kvalitete podataka. Može uključivati kontrole promjena, provjere hash vrijednosti i detekciju anomalija.
Verzioniranje i logiranje skupova podataka	Kada se skupovi podataka mijenjaju, nadopunjuju, filtriraju ili koriste za različite verzije modela.	Omogućuje sljedivost, dokazivanje usklađenosti, lakšu analizu incidenata i povezivanje određene verzije modela s konkretnim skupom podataka.
Korištenje sintetičkih ili fiktivnih podataka	Kada stvarni osobni podaci nisu nužni, primjerice za testiranje, integraciju, demonstracije ili određene sigurnosne provjere.	Smanjuje izloženost ispitanika. Prije uporabe sintetičkih podataka treba procijeniti rizik ponovne identifikacije i zadržavanja obilježja stvarnih osoba.
Enkripcija sigurnosnih kopija i komunikacije	Kada se pohranjuju ili prenose skupovi podataka, modeli, artefakti, logovi ili sigurnosne kopije koje mogu sadržavati osobne podatke.	Ograničava posljedice neovlaštenog pristupa ili curenja podataka. U određenim slučajevima mogu se razmotriti napredne tehnike, poput sigurnog višestranačkog računanja ili homomorfnog šifriranja, uzimajući u obzir izvedivost i ograničenja.

Mjera	Kada je osobito relevantna	Svrha / napomena
Kontrola pristupa podacima	Kada skupovi podataka nisu javno dostupni ili sadržavaju osobne podatke, posebne kategorije podataka ili povjerljive informacije.	Pristup treba ograničiti prema ulogama i načelu najmanjih ovlasti, uz autentifikaciju, autorizaciju, evidenciju pristupa i periodičnu provjeru ovlaštenja.
Anonimizacija, pseudonimizacija i druge tehnike smanjenja rizika	Kada je moguće smanjiti vezu između podataka i pojedinca bez narušavanja legitimne svrhe obrade.	Anonimizacija se primjenjuje samo kada je stvarno postignuta nepovratnost identifikacije. Pseudonimizirani podaci ostaju osobni podaci. Mogu se razmotriti generalizacija, obfuskacija i diferencijalna privatnost.
Segmentacija osjetljivih podataka	Kada skup podataka sadržava posebne kategorije osobnih podataka ili druge vrlo osjetljive podatke.	Može uključivati logičku ili fizičku segmentaciju, zasebne ključeve za enkripciju, strože pristupne mehanizme i dodatni nadzor. Segmentacija ne smije olakšati pronalazak osjetljivih podataka.
Mjere protiv gubitka kontrole nad podacima	Kada se podaci izvoze, dijele, prenose dobavljačima ili koriste u više okruženja.	Uključuje kontrolu izvoza i dijeljenja, DLP mjere, rokove čuvanja i brisanja, digitalne vodene žigove, digitalne potpise ili hash funkcije za provjeru izvora, sljedivosti i integriteta.

Tablica 3. Mjere vezane uz razvoj i integraciju sustava

Mjera	Kada je osobito relevantna	Svrha / napomena
Zaštita podataka po dizajnu i po zadanim postavkama	Od početka dizajna i razvoja sustava, osobito kada se obrađuju osobni podaci ili kada sustav može značajno utjecati na pojedince.	Zahtijeva da se minimizacija podataka, ograničenje svrhe, kontrola pristupa, sigurnost, transparentnost i ostvarivanje prava ispitanika ugrade u arhitekturu i rad sustava.
Korištenje provjerenih biblioteka, alata, modela i konfiguracija	Kada se koriste open source komponente, unaprijed trenirani modeli, vanjski alati, API-ji ili konfiguracijske datoteke.	Smanjuje rizik ranjivosti, nepoznatih ovisnosti, backdoor funkcionalnosti i neusklađenih komponenti. Potrebno je pratiti ažuriranja i sigurnosna upozorenja.
Sigurne razvojne prakse	Tijekom razvoja, prilagodbe, fine-tuninga, integracije i održavanja sustava.	Uključuje sigurne formate za uvoz i sigurnosne kopije, zabranu opasnih funkcija, reviziju koda, provjeru ovisnosti, statičku i dinamičku analizu te postupanje po upozorenjima sigurnosnih alata.
Kontrolirano i ponovljivo razvojno okruženje	Kada više osoba ili timova radi na razvoju, testiranju ili integraciji sustava.	Kontejneri, virtualni strojevi i automatizirane konfiguracije olakšavaju sljedivost, ponovljivost i izolaciju. Za osjetljive obrade mogu se razmotriti sigurna izvršna okruženja i sigurnost hardvera.
DevSecOps i kontinuirana integracija	Kada se sustav često mijenja, nadograđuje ili spaja s drugim komponentama.	Sigurnost se ugrađuje u razvojni ciklus, uz kontrolu izmjena, testove, autentifikaciju za izmjene produkcijskog koda i odvajanje razvojnih, testnih i produkcijskih okruženja.
Sveobuhvatna dokumentacija	Za sve sustave kod kojih je potrebno dokazati usklađenost, sigurnost, namjenu i ograničenja sustava.	Može uključivati opis dizajna, podatke i modele, razloge odabira, rezultate validacije, ograničenja, performanse, analizu pristranosti, sigurnosne mjere, uvjete uporabe i zahtjeve infrastrukture.
Sigurnosne provjere, revizije i red teaming	Kada sustav ima povećan rizik za prava i slobode pojedinaca, koristi osjetljive podatke ili je povezan s alatima, API-jima i bazama znanja.	Interno ili vanjsko testiranje može otkriti ranjivosti, prompt injection scenarije, curenje podataka, zaobilaženje zaštita i druge sigurnosne slabosti prije produkcijske uporabe.

Tablica 4. Mjere vezane uz rad sustava u produkcijskom okruženju

Mjera	Kada je osobito relevantna	Svrha / napomena
Informiranje korisnika o namjeni i ograničenjima sustava	Kada korisnici ili zaposlenici koriste UI sustav za donošenje preporuka, obradu upita, izradu sadržaja ili podršku odlučivanju.	Korisnicima treba jasno objasniti predviđenu namjenu, ograničenja, zabranjene ili nepreporučene uporabe, rizike pogrešnih izlaza i potrebu ljudske provjere kada je relevantno.
Informacije za tumačenje rezultata	Kada izlazi sustava mogu utjecati na pojedince ili se koriste kao podloga za odluke.	Pomaže korisnicima prepoznati pogreške, neizvjesnost, halucinacije, pristranosti ili ograničenja izlaza. Može uključivati upute, objašnjenja, upozorenja i komunikacijski kanal za povratne informacije.
Filtriranje i validacija izlaza	Kada generativni sustavi mogu proizvesti osobne podatke, povjerljive informacije, štetan sadržaj ili obmanjujuće rezultate.	Može uključivati output filtere, provjere pravila, ljudsku provjeru, RLHF ili druge oblike kontrole izlaza. Ne zamjenjuje obvezu procjene zakonitosti i sigurnosti obrade.
Ograničenje pristupa modelu, alatima i povezanim izvorima	Kada je UI sustav povezan s internim bazama, RAG-om, API-jima, datotekama ili mogućnošću izvršavanja radnji.	Smanjuje rizik prompt injection napada, neovlaštenog pristupa i eksfiltracije podataka. Posebno je važno autorizirati svaku radnju i ograničiti dohvat podataka na nužan opseg.
Nadzor promptova, odgovora i obrazaca korištenja	Kada postoji rizik zlouporabe, neovlaštenog izdvajanja podataka, model extraction napada ili nekontrolirane potrošnje resursa.	Omogućuje detekciju anomalija, pokušaja napada, prekomjernog broja upita, neuobičajenih obrazaca ponašanja i korištenja sustava izvan predviđene namjene.
Mogućnost zaustavljanja, rollbacka ili ograničavanja funkcionalnosti	Kada pogrešan rad sustava može dovesti do ozbiljnih posljedica za ispitanike, organizaciju ili sigurnost sustava.	Plan rollbacka, deaktivacije funkcionalnosti ili vraćanja na sigurnu verziju smanjuje posljedice incidenta i omogućuje brzu reakciju.
Praćenje incidenata, zahtjeva ispitanika i povratnih informacija	Tijekom cijelog životnog ciklusa sustava, osobito nakon značajnih promjena sustava ili konteksta uporabe.	Omogućuje prepoznavanje novih rizika, prilagodbu mjera, dokazivanje odgovornosti i pravodobno postupanje u slučaju povrede osobnih podataka ili drugih incidenata.

Tablica 5. Horizontalne organizacijske mjere

Mjera	Svrha	Elementi koje treba razmotriti
Sigurnosna politika i akcijski plan	Upravljanje provedbom mjera i provjera stvarnog smanjenja rizika tijekom životnog ciklusa sustava.	odgovornosti; rokovi; prioriteti; kriteriji prihvatljivosti rizika; plan brisanja podataka; redovita provjera učinkovitosti mjera
Plan odgovora na incidente	Brzo i usklađeno postupanje u slučaju sigurnosnog incidenta ili povrede osobnih podataka.	detekcija; eskalacija; izolacija; analiza uzroka; obavještavanje nadležnih tijela i ispitanika kada je potrebno; dokumentiranje incidenta; korektivne mjere
Multidisciplinarni tim	Povezivanje tehničkog, pravnog, sigurnosnog i poslovnog znanja u procjeni i upravljanju rizicima.	uključivanje DPO-a, pravnog tima, informacijskih stručnjaka, sigurnosti, nabave, vlasnika procesa i stručnjaka za domenu uporabe sustava
Edukacija i svijest zaposlenika	Smanjenje rizika pogrešne uporabe, neovlaštenog unosa podataka i zaobilaženja sigurnosnih pravila.	upute za korištenje UI alata; zabrana unosa nepotrebnih osobnih podataka; prepoznavanje prompt injectiona; postupanje s incidentima; obveze čuvanja povjerljivosti
Upravljanje ovlaštenjima	Ograničavanje pristupa podacima, modelima, alatima, logovima i produkcijskim funkcionalnostima.	načelo najmanjih ovlasti; popis privilegiranih korisnika; periodična revizija ovlaštenja; odvajanje dužnosti; ukidanje pristupa nakon prestanka potrebe
Logiranje i analiza pristupa	Otkrivanje pokušaja neovlaštenog pristupa, izmjena, izvoza podataka, zlouporabe sustava ili neobičnih obrazaca korištenja.	evidencija pristupa, izmjena i brisanja; analiza zapisa u stvarnom vremenu kada je moguće; periodična analiza; zaštita logova od neovlaštene izmjene

Mjera	Svrha	Elementi koje treba razmotriti
DPIA i dokumentiranje odluka	Dokazivanje procjene rizika i opravdanosti odabranih mjera kada obrada vjerojatno predstavlja visok rizik.	opis obrade; procjena nužnosti i proporcionalnosti; procjena rizika za prava i slobode; mjere za ublažavanje rizika; rezidualni rizik; konzultacije kada su potrebne
Periodična revizija mjera	Osiguravanje da mjere ostanu primjerene promjenama sustava, podataka, prijetnji i regulatornog okvira.	revizija nakon većih promjena; redovito testiranje; praćenje učinkovitosti; ažuriranje dokumentacije; ponovna procjena rezidualnog rizika

Tablica 6. Primjeri specifičnih AI/LLM rizika i mogućih mjera

Rizik / napad	Kada je osobito relevantan	Primjeri mjera
Napad zaključivanja o članstvu	Kada model omogućuje dovoljno precizne upite ili izlaze, osobito ako je treniran na manjim, osjetljivim ili specifičnim skupovima podataka.	minimizacija podataka; pseudonimizacija; diferencijalna privatnost; ograničenje pristupa modelu; rate limiting; nadzor anomalija; testiranje otpornosti
Napad zaključivanja o atributima	Kada se iz izlaza modela ili kombinacije podataka mogu izvesti dodatna obilježja osobe koja nisu izravno otkrivena.	smanjenje granularnosti izlaza; filtriranje visokorizičnih atributa; kontrola pristupa; procjena nužnosti varijabli; testovi inferencije i pristranosti
Neovlašteno otkrivanje ili eksfiltracija osobnih podataka, uključujući regurgitaciju trening podataka	Kada model može reproducirati dijelove trening podataka ili kada se podaci mogu izvući iz logova, promptova, vektorske baze, RAG sustava ili povezanih spremišta.	čišćenje trening podataka; deduplikacija; zaštite na razini izlaza; ograničenje pristupa; segmentacija; enkripcija; DLP mjere; red teaming; ciljano testiranje
Inverzija modela / rekonstrukcijski napadi	Kada bi se iz modela, njegovih parametara ili odgovora mogla rekonstruirati obilježja osoba ili dijelovi izvornih podataka.	ograničenje pristupa modelu i artefaktima; kontrola izlaza; zaštita parametara; privatnost u učenju; zaštita kontrolnih točaka; politika zadržavanja i brisanja artefakata
Prompt injection, uključujući neizravni prompt injection i jailbreaking	Kada je UI sustav povezan s alatima, RAG-om, internim bazama znanja, datotekama, API-jima ili izvršavanjem radnji.	odvajanje sistemskih uputa od korisničkog unosa; sandboxing alata; autorizacija po radnji; filtriranje promptova; validacija izlaza; ograničenje dohvaćanja podataka; zapisivanje i nadzor pokušaja
Trovanje podataka i modela / manipulacija izvorima znanja	Kada napadač može utjecati na trening, fine-tuning, validacijske skupove, referentne baze, RAG izvore ili vanjske izvore na koje se sustav oslanja.	provjera izvora i provenijencije podataka; kontrola integriteta; odobravanje promjena; detekcija anomalija; razdvajanje dužnosti; verifikacija kvalitete setova podataka
Ekstrakcija ili krađa modela	Kada je moguće kroz niz upita rekonstruirati funkcionalnost modela, aproksimirati njegovo ponašanje ili zaobići ugrađene zaštite.	rate limiting; autentikacija; nadzor obrazaca upita; segmentacija prava; zaštita API-ja; detekcija automatiziranog izvlačenja; vodeni žigovi kada su prikladni
Neograničena potrošnja / iscrpljivanje resursa	Kada korisnici mogu slati veliki broj zahtjeva, vrlo duge upite, skupe operacije ili zahtjeve koji aktiviraju vanjske alate.	ograničenje broja i duljine upita; kvote; kontrola troškova; timeout mehanizmi; nadzor korištenja; zaštita od automatiziranih zahtjeva; prioritet za legitimne korisnike

11.4 Korištenje tehnologija za poboljšavanje privatnosti („privacy enhancing technologies“)

Upotrebom tehnologija za poboljšavanje privatnosti *(privacy-enhancing technologies, dalje u tekstu: TPP)* organizacija može smanjiti rizike za prava i slobode ispitanika, osobito u slučajevima u kojima je potrebno analizirati, povezivati, dijeliti ili koristiti velike skupove podataka uz smanjenje izloženosti osobnih podataka. Takve tehnologije mogu pridonijeti provedbi načela smanjenja količine podataka, cjelovitosti i povjerljivosti, sigurnosti obrade te zaštite podataka po dizajnu i po zadanim postavkama.

Međutim, uporaba TPP-a sama po sebi ne isključuje primjenu Opće uredbe o zaštiti podataka niti zamjenjuje obvezu procjene zakonitosti, nužnosti, proporcionalnosti i sigurnosti obrade. TPP treba promatrati kao dio šireg skupa tehničkih i organizacijskih mjera, a ne kao samostalno i dostatno rješenje za usklađenost. Osobito je važno razlikovati anonimizaciju od pseudonimizacije: pseudonimizirani podaci i dalje su osobni podaci, dok se podaci mogu smatrati anonimiziranima samo ako se pojedinac više ne može identificirati sredstvima za koja je razumno vjerojatno da će se koristiti, uzimajući u obzir konkretan kontekst, dostupne podatke, stanje tehnike i rizik ponovne identifikacije.

U praksi se mogu koristiti različite tehnologije i tehnike za poboljšavanje privatnosti, ovisno o svrsi obrade, vrsti podataka, arhitekturi sustava, namjeravanoj uporabi i procijenjenim rizicima. Među tipičnim mjerama izdvajaju se osobito pseudonimizacija, diferencijalna privatnost, federativno učenje, sintetski podaci, sigurno višestranačko računanje, homomorfno šifriranje, dokazi bez otkrivanja znanja i pouzdana izvršna okruženja. U određenim okolnostima takve mjere mogu omogućiti dobivanje korisnih uvida iz podataka uz manju izloženost osobnih podataka, sigurniju suradnju između organizacija, ograničavanje pristupa izvornim podacima te smanjenje rizika od neovlaštenog otkrivanja, izdvajanja ili ponovne identifikacije podataka.

TPP mogu biti korisne osobito kada organizacija želi:

- omogućiti suradnju između više organizacija uz tehnička ograničenja koja smanjuju mogućnost korištenja podataka izvan namjeravane svrhe;
- omogućiti analizu ili uvid u podatke bez otkrivanja cjelovitog sadržaja izvornog skupa podataka;
- provoditi obradu u okruženju u kojem ne postoji potpuno povjerenje između svih sudionika;
- smanjiti potrebu za centraliziranim prikupljanjem i pohranom osobnih podataka;
- ograničiti rizik od izdvajanja podataka iz modela ili rezultata obrade;

- ojačati sposobnost organizacije da zadrži kontrolu nad podacima tijekom njihova životnog ciklusa.

Povjerljivo izvršno okruženje *(Trusted Execution Environment – TEE)*

Povjerljivo izvršno okruženje označava hardverski podržano, izolirano izvršno okruženje koje može pružiti dodatnu zaštitu podacima tijekom obrade tako da se osjetljive operacije izvršavaju unutar zaštićenog dijela procesora. Takav pristup može smanjiti izloženost podataka domaćinskom sustavu, administratorima infrastrukture ili drugim komponentama okruženja.

Međutim, TEE sam po sebi ne jamči potpunu zaštitu od svih rizika niti automatski sprječava neprimjereno logiranje, pohranu, neovlaštene pristupe ili druge oblike curenja podataka. Stoga ga je potrebno promatrati kao jednu od zaštitnih mjera u širem sigurnosnom i organizacijskom okviru.

Homomorfno šifriranje *(Homomorphic Encryption)*

Homomorfno šifriranje omogućuje izvođenje određenih izračuna nad šifriranim podacima bez prethodnog dešifriranja podataka. Kod potpuno homomorfnog šifriranja *(Fully Homomorphic Encryption – FHE)* moguće je, barem načelno, izvoditi složenije računalne operacije nad šifriranim podacima, pri čemu podaci tijekom obrade ostaju zaštićeni od strane koja obradu izvršava.

Ovaj pristup može biti koristan u slučajevima u kojima je potrebno omogućiti obradu ili analizu podataka u okruženju kojem se ne želi ili ne može u cijelosti povjeriti pristup izvornim podacima. Ipak, homomorfno šifriranje u praksi može biti računalno zahtjevno, složeno za implementaciju i ograničeno s obzirom na vrstu obrade, performanse i uporabnu vrijednost sustava. Zato ga je potrebno primjenjivati selektivno, uz prethodnu procjenu tehničke izvedivosti, troškova i stvarnog učinka na smanjenje rizika.

Sigurno višestranačko računanje *(Secure Multi-Party Computation – SMPC)*

Sigurno višestranačko računanje označava skup kriptografskih tehnika koje omogućuju da više strana zajednički provedu izračun nad svojim podacima, a da pritom ne otkriju same ulazne podatke drugim sudionicima. Takav pristup može biti koristan kada više organizacija želi zajednički analizirati podatke ili trenirati određene modele, ali ne želi ili ne smije međusobno dijeliti izvorne skupove podataka.

Iako SMPC može značajno smanjiti potrebu za izravnim dijeljenjem podataka, njegova uporaba ne uklanja sve rizike. Potrebno je procijeniti konkretan model prijetnji, arhitekturu sustava, vrstu izlaznih rezultata, mogućnost zaključivanja informacija iz rezultata obrade te organizacijske odnose između sudionika.

Federativno učenje *(Federated Learning – FL)*

Federativno učenje označava pristup u kojem više sudionika trenira model nad svojim lokalno pohranjenim podacima, bez razmjene samih izvornih skupova podataka. Umjesto prijenosa podataka u središnji sustav, razmjenjuju se određene informacije o učenju modela, primjerice gradijenti, parametri ili druga ažuriranja, koja se zatim agregiraju u zajednički model.

Takav pristup može smanjiti potrebu za centraliziranim prikupljanjem osobnih podataka i time smanjiti određene rizike za privatnost. Međutim, federativno učenje samo po sebi ne uklanja sve rizike. Ažuriranja modela, ovisno o okolnostima, mogu otkrivati informacije o lokalnim podacima ili omogućiti inferencijske napade. Zbog toga se federativno učenje u pravilu kombinira s dodatnim mjerama, kao što su sigurna agregacija, diferencijalna privatnost, enkripcija, kontrola pristupa i nadzor sudionika u procesu treniranja.

Dokaz bez otkrivanja znanja *(Zero-Knowledge Proof)*

Dokaz bez otkrivanja znanja označava kriptografski protokol kojim jedna strana može drugoj strani dokazati istinitost određene tvrdnje, bez otkrivanja dodatnih informacija osim same činjenice da je tvrdnja istinita. Primjerice, osoba može dokazati da ispunjava uvjet punoljetnosti, a da pritom ne otkriva točan datum rođenja ili broj godina.

U kontekstu zaštite osobnih podataka takvi mehanizmi mogu biti korisni kada je za ostvarenje određene svrhe dovoljno potvrditi određeno svojstvo, status ili uvjet, bez prikupljanja i obrade šireg opsega osobnih podataka. Time se može pridonijeti načelu smanjenja količine podataka.

Diferencijalna privatnost *(Differential Privacy)*

Diferencijalna privatnost označava formalni matematički pristup kojim se ograničava koliko prisutnost ili odsutnost pojedinog zapisa može utjecati na rezultat obrade, upita ili treniranja modela. U praksi se to najčešće postiže dodavanjem kontroliranog šuma ili drugim mehanizmima koji smanjuju mogućnost zaključivanja informacija o pojedinoj osobi.

Diferencijalna privatnost može biti korisna za smanjenje rizika od zaključivanja o članstvu u skupu podataka, rekonstrukcije pojedinačnih zapisa ili otkrivanja informacija o pojedincima iz agregiranih rezultata. Međutim, njezina učinkovitost ovisi o pravilnom podešavanju parametara, vrsti obrade, veličini skupa podataka i prihvatljivom odnosu između razine zaštite i korisnosti rezultata. Veća razina zaštite u pravilu može smanjiti preciznost i uporabnu vrijednost rezultata, pa je potrebno pažljivo dokumentirati razloge za odabrani pristup.

Sintetski podaci

Sintetski podaci su umjetno generirani podaci nastali uporabom modela ili drugih metoda sinteze, s ciljem oponašanja određenih obrazaca, struktura ili statističkih svojstava stvarnih podataka. Mogu biti korisni za razvoj, testiranje, istraživanje, validaciju sustava ili određene analitičke svrhe, osobito kada nije potrebno koristiti ili dijeliti izvorne osobne podatke.

Međutim, sintetski podaci nisu automatski anonimni niti su samim time izvan područja primjene pravila o zaštiti podataka. Ako su generirani na temelju stvarnih osobnih podataka, potrebno je procijeniti postoji li rizik da sintetski podaci odražavaju stvarne pojedince, omogućuju njihovu ponovnu identifikaciju ili otkrivaju informacije o izvornim zapisima. Stoga je potrebno zasebno procijeniti stupanj sličnosti sa stvarnim podacima, rizik identifikacije, kvalitetu generiranja, svrhu uporabe i prikladnost takvih podataka za konkretan slučaj.

Odabir i dokumentiranje TPP mjera

Pri odabiru TPP mjera organizacija treba procijeniti njihovu tehničku zrelost, stvarni učinak na smanjenje rizika, utjecaj na točnost i korisnost sustava, složenost implementacije, troškove, potrebu za dodatnim organizacijskim mjerama te mogućnost dokazivanja njihove učinkovitosti. Važno je dokumentirati zašto je određena mjera odabrana, koje rizike adresira, koja su njezina ograničenja te koji rezidualni rizici ostaju nakon njezine primjene.

Organizacija također treba redovito preispitivati učinkovitost primijenjenih TPP mjera, osobito nakon promjene svrhe obrade, arhitekture sustava, vrste podataka, modela, korisničke baze, dobavljača ili tehničkog okruženja. Kao i druge tehničke i organizacijske mjere, TPP moraju biti dio kontinuiranog procesa upravljanja rizicima, a ne jednokratna odluka donesena samo u fazi početnog razvoja sustava.

12. Provedba procjene učinka na zaštitu podataka (Data protection impact assessment – DPIA)

Sukladno članku 35. Opće uredbe o zaštiti podataka, ako je vjerojatno da će neka vrsta obrade, osobito putem novih tehnologija i uzimajući u obzir prirodu, opseg, kontekst i svrhe obrade, prouzročiti visok rizik za prava i slobode pojedinaca, voditelj obrade prije obrade provodi procjenu učinka predviđenih postupaka obrade na zaštitu osobnih podataka. Jedna procjena može se odnositi na niz sličnih postupaka obrade koji predstavljaju slične visoke rizike.

Potrebno je jasno naglasiti da je za svaku obradu osobnih podataka potrebno prvo provesti procjenu rizika iz članka 32. stavka 2. nakon čega će organizacije biti u mogućnosti procijeniti jesu li obrade koje provode visokorizične.

Procjena učinka treba biti provedena prije nego što se započne sa samom obradom te je također u slučaju izmjena same obrade potrebno istu ažurirati. Procjenom učinka se se identificiraju i procjenjuju rizici za ispitanike, analiziraju mjere koje omogućuju ostvarivanje prava iz Opće uredbe o zaštiti podataka, procjenjuju implementirane kontrole i sama transparentnost.

Uvjeti kada je potrebno provesti propisani su člankom 35. Opće uredbe o zaštiti podatak, a Agencija je u okviru svoje nadležnosti donijela Odluku o uspostavi i javnoj objavi popisa vrsta postupaka obrade koje podliježu zahtjevu za procjenu učinka na zaštitu podataka kojom se propisuju postupci obrade za koje je potrebno provesti istu.[8]

U kontekstu razvoja tehnologija umjetne inteligencije koje uključuju obradu osobnih podataka, preporuka Agencije je da se provede procjena učinka na zaštitu podataka. Procjena učinka na zaštitu podataka osobito je važna pri razvoju i korištenju sustava umjetne inteligencije jer predstavlja strukturirani postupak kojim se: (1) opisuje planirana obrada te procjenjuje njezina nužnost i proporcionalnost, te (2) identificiraju i upravljaju rizici za prava i slobode pojedinaca koji proizlaze iz obrade osobnih podataka, uključujući procjenu rizika i određivanje mjera za njihovo ublažavanje.

Obrazac za provedbu procjene učinka na zaštitu podataka i više informacija o provedbi procjena učinka, kao i test za provedbu procjene učinka dostupni su na poveznici: https://azop.hr/procjena-ucinka-na-zastitu-podataka-eng-data-protection-impact-assessment-dpia/ i https://olivia-gdpr-arc.eu/hr/course/overview/15.

13. Prijenosi osobnih podataka u treće zemlje i međunarodne organizacije

Razvoj ili implementacija komponente UI sustava temeljene na uslugama u oblaku, ili primjerice otkrivanje korisničkih podataka trećim stranama radi razvoja UI modela, može podrazumijevati prekogranične tokove podataka u treće zemlje. Tokovi podataka koji se odvijaju unutar okvira Europskog gospodarskog prostora ne kvalificiraju se kao međunarodni prijenosi.

Jamstva navedena u Poglavlju V. Opće uredbe o zaštiti podataka "Prijenosi osobnih podataka u treće zemlje ili međunarodne organizacije" moraju se primijeniti na takve

[8] Odluka dostupna na: https://azop.hr/odluka-o-uspostavi-i-javnoj-objavi-popisa-vrsta-postupaka-obrade-koje-podlijezu-zahtjevu-za-procjenu-ucinka-na-zastitu-podataka/

prijenose. Posebno je važno uspostaviti mehanizme kako bi se omogućilo nesmetano upravljanje ugovorima potpisanim u ovom kontekstu međunarodnih prijenosa, istovremeno osiguravajući da klijent, kao voditelj obrade podataka, ima dovoljno informacija o ugovornim stranama i zadržava ovlast donošenja odluka. Kada postoje međunarodni prijenosi, ispitanici moraju biti obaviješteni u skladu s uvjetima iz članaka 13. i 14. Opće uredbe o zaštiti podataka, a takvi međunarodni prijenosi moraju biti uključeni u evidenciju aktivnosti obrade.

U kontekstu UI sustava međunarodni prijenosi osobnih podataka ne nastaju samo kada se glavni skupovi podataka ili korisnički podaci izravno pohranjuju u trećoj zemlji. Potrebno je uzeti u obzir i druge tokove podataka, kao što su logovi, telemetrija, promptovi, izlazi sustava, sigurnosni zapisi, razvojna i testna okruženja, udaljeni pristup radi održavanja ili podrške, kao i pristup podacima od strane izvršitelja i podizvršitelja. Osobitu pozornost treba posvetiti situacijama u kojima pružatelj usluge ili njegov podizvršitelj iz treće zemlje ima udaljeni pristup osobnim podacima, budući da i takav pristup može predstavljati prijenos u smislu Opće uredbe o zaštiti podataka.

Stoga se preporučuje da organizacija prije početka uporabe UI sustava ili povezane usluge mapira sve prijenose osobnih podataka izvan EGP-a, uključujući izravne i daljnje prijenose, sve uključene izvršitelje i podizvršitelje, države u koje se podaci prenose ili iz kojih im se pristupa, vrste podataka koje su obuhvaćene, svrhe prijenosa, korištena okruženja te odgovarajući prijenosni mehanizam i eventualne dodatne zaštitne mjere.

III. Izvori

Akt o umjetnoj inteligenciji – https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:32024R1689

AZOP. *Pravni temelji za obradu osobnih podataka* – https://azop.hr/pravni-temelji-za-obradu-osobnih-podataka-clanak6-gdpr/

AZOP. *Procjena učinka na zaštitu podataka* – https://azop.hr/procjena-ucinka-na-zastitu-podataka-eng-data-protection-impact-assessment-dpia/

AZOP. *Usklađivanje s Općom uredbom o zaštiti podataka* – https://azop.hr/voditelji-i-izvrsitelji-obrade/

Europski nadzornik za zaštitu podataka (EDPS). *Revised Guidelines / Orientations on Generative AI* – https://www.edps.europa.eu/system/files/2025-10/25-10_28_revised_genai_orientations_en.pdf

Europski odbor za zaštitu podataka (EDPB). *Mišljenje 28/2024 o anonimnosti AI modela* – https://www.edpb.europa.eu/system/files/2025-05/edpb_opinion_202428_ai-models_hr.pdf

Europski odbor za zaštitu podataka (EDPB). *Smjernice 07/2020 o konceptima voditelja obrade i izvršitelja obrade prema GDPR-u* – https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-072020-concepts-controller-and-processor-gdpr_hr

Europski odbor za zaštitu podataka (EDPB). *Smjernice 01/2025 o pseudonimizaciji* – https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

Europski odbor za zaštitu podataka (EDPB), Support Pool of Experts. *AI Effective Implementation of Data Subjects' Rights* – https://www.edpb.europa.eu/system/files/2025-01/d2-ai-effective-implementation-of-data-subjects-rights_en.pdf

Europski odbor za zaštitu podataka (EDPB), Support Pool of Experts. *AI Privacy Risks & Mitigations in LLMs* – https://www.edpb.europa.eu/system/files/2025-04/ai-privacy-risks-and-mitigations-in-llms.pdf

Europski odbor za zaštitu podataka (EDPB), Support Pool of Experts. *Training on AI and Data Protection – Legal and Technical Materials* – https://www.edpb.europa.eu/system/files/2025-06/spe-training-on-ai-and-data-protection-legal_en.pdf i https://www.edpb.europa.eu/system/files/2025-06/spe-training-on-ai-and-data-protection-technical_en.pdf

Federal Office for Information Security Germany (BSI). *AI Security Concerns in a Nutshell* – https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Practical_AI-Security_Guide_2023.pdf?__blob=publicationFile&v=5

Francusko nadzorno tijelo za zaštitu podataka (CNIL). *AI How-to Sheets* – https://www.cnil.fr/fr/ai-how-to-sheets

Komunikacija Europske komisije od 29. srpnja 2025. *Smjernice o definiciji sustava umjetne inteligencije* – https://azop.hr/wp-content/uploads/2025/07/HR-Commission_Guidelines_on_the_definition_of_an_artificial_intelligence_system_established_by_Regulation_EU_20241689_AI_ActCroatian_5Hn0qpSjpULo4aaI89xid0vxNY_118621.pdf

OECD. *Sharing Trustworthy AI Models with Privacy-Enhancing Technologies* – https://www.oecd.org/en/publications/sharing-trustworthy-ai-models-with-privacy-enhancing-technologies_a266160b-en.html

Opća uredba o zaštiti podataka – https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:32016R0679

OWASP. *AI Exchange: The World's AI Security Guide* – https://owaspai.org/

OWASP. *OWASP Top 10 for Large Language Model Applications* – https://genai.owasp.org/llm-top-10/

PET for Artificial Intelligence-Enabled Systems – https://arxiv.org/html/2404.03509v1#S5

Radna skupina iz članka 29. *Mišljenje 05/2014 o tehnikama anonimizacije* – https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_hr.pdf

Radna skupina iz članka 29. *Smjernice o procjeni učinka na zaštitu podataka* – https://ec.europa.eu/newsroom/article29/items/611236

Španjolska Agencija za zaštitu osobnih podataka (AEPD). *Data and Information in Artificial Intelligence* – https://www.aepd.es/en/press-and-communication/blog/data-and-information-in-artificial-Intelligence

Tijelo za zaštitu osobnih podataka Singapura (PDPC). *Synthetic Data Generation* – https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/other-guides/proposed-guide-on-synthetic-data-generation.pdf

Tijelo za zaštitu osobnih podataka Velike Britanije (ICO). *Artificial Intelligence* – https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/

UN. *The PET Guide* – https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf

Uredba (EU) 2022/2554 Europskog parlamenta i Vijeća od 14. prosinca 2022. o digitalnoj operativnoj otpornosti za financijski sektor – https://eur-lex.europa.eu/eli/reg/2022/2554/oj?locale=hr

Uredba o kibernetičkoj sigurnosti – https://narodne-novine.nn.hr/clanci/sluzbeni/2024_11_135_2217.html

Zakon o kibernetičkoj sigurnosti – https://narodne-novine.nn.hr/clanci/sluzbeni/2024_02_14_254.html