Utjecaj generativne umjetne inteligencije na sigurnost i analizu podataka

Podgoršek, Norman

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Norman Podgoršek

UTJECAJ GENERATIVNE UMJETNE INTELIGENCIJE NA SIGURNOST I ANALIZU PODATAKA

DIPLOMSKI RAD

Varaždin, 2024.

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Norman Podgoršek

JMBAG: 0016137116

Studij: Informacijsko i programsko inženjerstvo

UTJECAJ GENERATIVNE UMJETNE INTELIGENCIJE NA SIGURNOST I ANALIZU PODATAKA

DIPLOMSKI RAD

Mentor:

Doc. dr. sc. Igor Tomičić

Varaždin, rujan 2024.

Norman Podgoršek

Izjava o izvornosti

Izjavljujem da je moj diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

Autor potvrdio prihvaćanjem odredbi u sustavu FOI-radovi

_____

Sažetak

Unutar diplomskog rada obrađene su teme definiranja pojmova generativne umjetne inteligencije kao i podataka koji su ključni u uspjehu dobrog generativnog alata. Najprije je opisano što zapravo je umjetna inteligencija, te mnoge pojmove koji se spominju u industriji vezani za GenAI, kao i sigurnosni rizici implementacije GenAI te napad i obrana nad generativnim modelima i alatima. Podatci su od ključne važnosti u GenAI-u, te se unutar rada opisuju načini prikupljanja, pohranjivanja i sustava za rad sa svim podatcima, kao i trenutno stanje u industriji po pitanju podataka. Unutar praktičnog dijela obrađeni su alat pynonymizer i biblioteka Faker, postavljanje i priprema za proces anonimizacije podataka unutar baze podataka putem kreiranja datoteke strategije uz pomoć koje definiramo kako koji stupac anonimizirati uz pomoć tehnika anonimizacije, kao i pregled dobrih i loših primjera anonimizacije baze podataka.

Ključne riječi: umjetna inteligencija; generativna umjetna inteligencija; sigurnost; podaci; sigurnost podataka; anonimizacija; pseudonimizacija; tehnike anonimizacije; pynonymizer;

Sadržaj

1. Uvod

U zadnjih nekoliko godina, umjetna inteligencija doživjela je ogroman napredak u pogledu tehnologije, načina korištenja, odnosa s javnosti, masivne količine podataka u obradi, kao i same performanse raznih modela na raspolaganju. Umjetna inteligencija ima mnogo definicija i načina tumačenja, no najčešće se odnosi na razvoj algoritama i sustava koji imaju u cilju odrađivanja skupa zadataka za koje je tradicionalno potrebna ljudska inteligencija, odnosno zadatke koje računala jednostavno nisu mogla izvršavati do sada. Takvi zadatci su na primjer, prepoznavanje prirodnog govora govornika i automatsko prevođenje, donošenje kompleksnih odluka uz potkrijepljeni kontekst, pa čak i vizualno prepoznavanje rasterskih slika. Sve je to omogućeno intenzivnim tzv. 'treniranjem' modela, što je dio koncepta 'strojno učenje'.

Umjetna inteligencija koja je najviše napredovala tijekom zadnjih godina je generativna umjetna inteligencija (eng. Generative Artificial Intelligence). Generativna umjetna inteligencija (GenAI) je doživjela izniman porast u popularnosti pojavom modela koji su sposobni stvarati mnoge nove sadržaje. No s novom tehnologijom, dolazi i uvijek veća površina napada zlonamjernih korisnika, tako da je iznimno bitno biti svjestan sigurnosnih izazova implementacije generativnih alata, kao i mogućih napada i kako se nositi s njima.

Kako bi modeli bili kvalitetni, iznimno su bitni podatci, koji zapravo omogućavaju kompletan proces treniranja kao i pojmovi big data, data lake, data warehouse kao i anonimizacija prikupljenih podataka raznim tehnikama i pseudonimizacijom kako bi regulacije bile zadovoljene i sigurnost prikupljenih podataka bila očuvana. Stanje u industriji se uvijek mijenja te je zanimljivo vidjeti kako se industrije nose s novim promjenama i generativnim tehnologijama.

Kroz praktični dio pogledati ćemo kako izgleda proces anonimizacije baze podataka jednog poduzeća, te kako se umjetnici brane s masovnom 'krađom' njihovih radova u svrhu treniranja modela.

2. O generativnoj umjetnoj inteligenciji

2.1. AI, ML, DL

Kako bi mogli razumjeti temeljne koncepte generativne umjetne inteligencije, potrebno je razumjeti što zapravo je umjetna inteligencija, kako je nastala, te kako samo od tradicionalne umjetne inteligencije došli do današnje generativne umjetne inteligencije, što je danas među najčešće korištenim oblikom korištenja umjetne inteligencije.

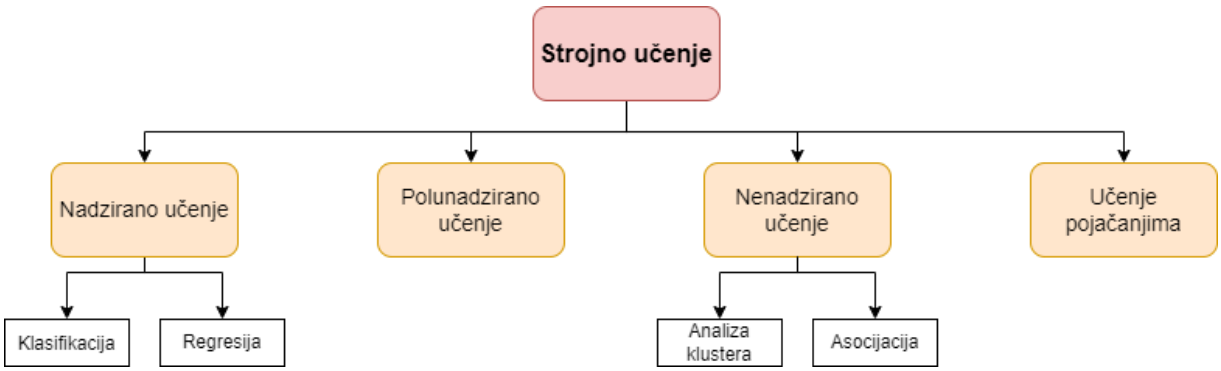
Umjetna inteligencija (eng. Artificial Intelligence) je iznimno široko područje informatike i znanosti koji se generalno bave kreiranjem sustava i algoritama za rješavanje zadataka i problema koji su tradicionalno zahtijevali ljudsku inteligenciju, odnosno rješavanje zadataka koja jednostavno nisu bila moguća do sada. Razumijevanje osnovnog koncepta umjetne inteligencije je ključno, jer u sadašnje vrijeme dolazi do mnogo zabuna i problema s klasifikacijom što zapravo je AI, a koje kvalitete se mogu pridonijeti drugim stavkama poput strojnog učenja (eng. Machine Learning) ili generativne umjetne inteligencije (eng. Generative Artificial Intelligence). Najviše 'krivnje' nesporazumu može se pridonijeti samoj složenosti područja no i širokog spektra primjene tehnologije, kao i varljivom marketingu raznih velikih kompanija kako bi pomogle prodati svoj proizvod, zadovoljili dioničare i prikazati se u boljem svijetlu nego što zapravo jesu.

Umjetna inteligencija može se definirati kao znanost dizajniranja etičnih i transparentnih sustava koji podupiru i ubrzavaju odluke i akcije ljudi [1]. Također, AI je grana informatike koja se bavi stvaranjem inteligentnih agenata, odnosno sustava koji mogu razmišljati, učiti i djelovati autonomno. Napredak u računalnoj obradi i pohrani podataka omogućio je unos i analizu većih količina podataka nego ikada prije. Istovremeno, sve veće povezivanje uređaja na internet rezultiralo je proizvodnjom ogromnih količina podataka. Ovi napreci približili su umjetnu inteligenciju njenom cilju stvaranja inteligentnih sustava, koji imaju sve veću primjenu i važnost u svakodnevnom životu. Od preporuka na omiljenim maloprodajnim stranicama do automatski generiranih oznaka fotografija na društvenim mrežama, mnoge uobičajene online zadatke i pogodnosti pokreće umjetna inteligencija. Većina primjera umjetne inteligencije koje danas susrećemo, od računala koja igraju šah do autonomnih vozila, uvelike se oslanja na umjetnu inteligenciju, što u primjeni je zapravo strojno učenje, koje omogućuje računalima da uče bez eksplicitnog programiranja.

Strojno učenje je zapravo podskup umjetne inteligencije koje automatsko omogućava računalu ili sistemu da uči i napreduje iz stečenog iskustva, odnosno umjesto da se poboljšava eksplicitno pomoću programiranja, strojno učenje koristi algoritme kako bi mogao razumjeti i

analizirati ogromne skupove podataka, učeći iz analiziranih podataka i pomoću njih raditi informirane odluke o zadanom zadatku. [2] Kako bi strojno učenje bilo učinkovito i kvalitetno, potrebno je algoritme za strojno učenje trenirati sa što više podataka kako bi njihov rezultat (eng. output) bio koristan samom korisniku. Strojno učenje također ima za cilj smanjiti broj podataka koji se smatraju „greškama“, te uspješno odraditi predviđanja ispravnosti određenih rezultata algoritama nakon procesa treniranja, no naravno struktura podataka koji se koristi pri naknadnim upitima nad algoritmima strojnog učenja mora biti isti kao i tip podataka koji se koristio tijekom treninga.

Strojnom učenje se dijeli na dva glavna principa: nadzirano učenje (eng. Supervised learning) i nenadzirano učenje (eng. Unsupervised learning). Nadzirano učenje je kategorija strojnog učenja koja koristi označene skupove podataka za obuku algoritama da predviđaju ishode i prepoznaju obrasce. Za razliku od nenadziranog učenja, algoritmi nadziranog učenja dobivaju označene podatke za obuku kako bi naučili odnos između ulaza i izlaza. [3] Pored navedenih principa, postoji i kombinacija navedenih pristupa gdje je cilj imati najbolje značajke iz oba pristupa pod nazivom polunadzirano učenje, što je pristup unutar kojeg je samo dio podataka za trening označeno, no uz veću količinu ne označenih podataka.

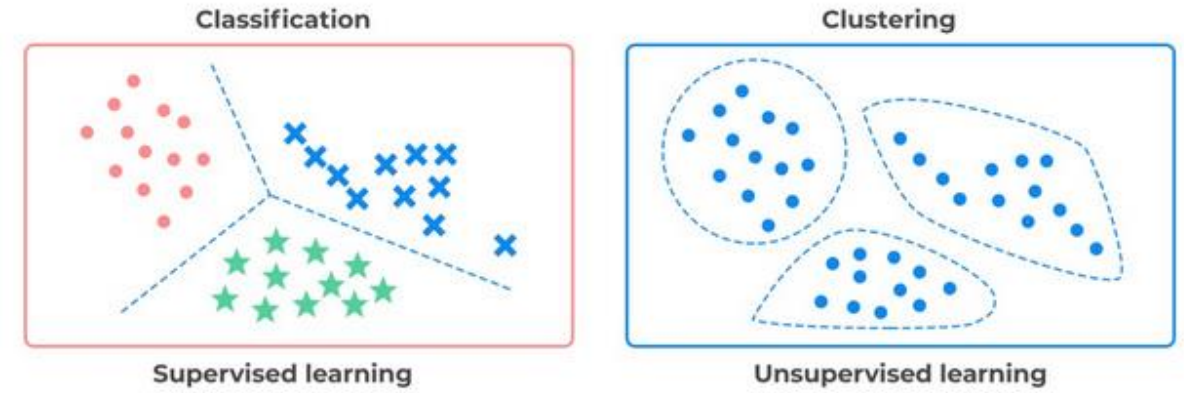


Slika 1. Principi strojnog učenja (Izvor: Vlastita izrada)

Što znači, nadzirano učenje trenirano nad označenim podatcima podrazumijeva da će uvijek znati što je zapravo 'ispravan' izlaz tijekom treninga, te pomoću datih podataka pokušati steći mogućnost generalizacije, odnosno prepoznavanja uzoraka unutar podataka i unaprjeđivanje vlastitih pretpostavki unutar algoritma kako bi mogao raditi s novim podatcima s kojima se algoritam nije susretao. Korištenjem metode nadziranog učenja moguće je kontinuirano unaprjeđivati njegovu preciznost mjerenjem i kontinuiranim razmatranjem nastalih „pogrešaka“ odnosno anomalija. Nadzirano učenje se može podijeliti na dva dodatna načina provedbe: klasifikacija i regresija. Klasifikacija podrazumijeva grupiranje izlaznih podataka unutar kategorija željenih opcija, na primjer ako želimo provoditi analizu i

prepoznavanje unutar slika, gdje možemo trenirati da se prepoznaju životinje ili stvari, označavanje e-mailova kao spam ili nije spam, odnosno što god je moguće svrstati u nekakve kategorije. Regresija u nadziranom učenju se više fokusira na same vrijednosti, što znači da umjesto kategoriziranja podataka želimo predvidjeti neke buduće numeričke vrijednosti. Regresiju je bolje koristiti kada je izlazni podatak kontinuirana numerička vrijednost, što u praksi može biti postotak vjerojatnosti, bilo kakve monetarne vrijednosti, ili predviđanje budućih vrijednosti.

Kod nenadziranog učenja algoritam radi s podatcima koji nisu označeni te algoritam sam pokušava naći uzorke koji se nalaze unutar danih podataka. To znači da, suprotno od nadziranog, nenadzirani algoritmi nemaju unaprijed poznati ispravan izlaz, što znači da ne mogu znati što je eksplicitno dobar izlazni podatak, a što nije ispravan izlaz. Nenadzirano učenje može se podijeliti na dva pristupa: analiza klastera, tj. klasterizacija i asocijacija.



Slika 2. Klasifikacija i analiza klastera vizualizirana (Izvor: ubiAI)

Analiza klastera je tehnika strojnog učenja u kojoj se podaci grupiraju na osnovu sličnosti pomoću mjera sličnosti tako da se mogu ustanoviti način grupiranja skupova podataka. Analiza klastera se najčešće koristi prilikom grupiranja određenih kupaca u svrhu marketinga, pronalaženja grupa korisnika koji odgovaraju sličnim ljudskim karakteristikama, itd. Asocijacija u nenadziranom strojnom učenju istražuje odnose između varijabli unutar skupa podataka, odnosno pokušava pronaći pravila koja definiraju odnose između podataka. Asocijacija se najčešće primjenjuje unutar trgovačkih okruženja gdje je moguće predviđati koji proizvodi se najčešće kupuju u isto vrijeme i tako bolje predložiti kupcu što iduće dodati u košaricu. Uz nenadzirano učenje također se koristi i tehnika redukcija dimenzija koja se najčešće provodi prije samog procesa daljnjeg učenja modela, a ima za cilj smanjiti broj ukupnih vrijednosti tj. količine podataka, no da ne dođe do gubitka ključnih informacija, tako

eliminirajući redundantne podatke. Redukcija dimenzija se najčešće koristi kada ulazni podatak kojeg algoritam treba promotriti ima previše informacija koje mu nisu potrebne, te se ostavljaju samo najbitnije informacije kako bi ostatak modela mogao ispravno raditi predviđanja nad podatcima.

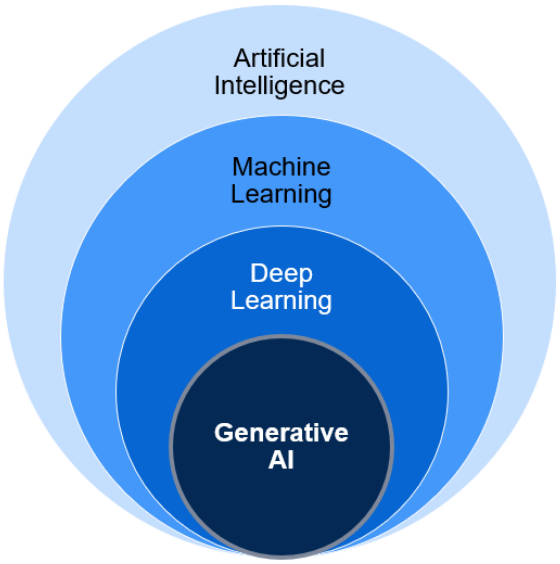
Polunadzirano učenje (semi-supervised learning) kombinira elemente nadziranog i nenadziranog učenja kako bi se poboljšale performanse i točnost modela. U ovom pristupu, samo manji dio skupa podataka ima označene izlaze, dok većina podataka ostaje neoznačena. To omogućuje modelu da uči iz dostupnih označenih podataka, dok istovremeno koristi velike količine neoznačenih podataka za prepoznavanje skrivenih obrazaca i odnosa. Jedna od prednosti polunadziranog učenja je smanjenje troškova i vremena potrebnog za ručno označavanje velikih količina podataka. Umjesto označavanja cijelog skupa podataka, dovoljan je manji, reprezentativan podskup. Ostatak neoznačenih podataka pomaže modelu da bolje generalizira na nove, nepoznate podatke. Polunadzirano učenje je također posebno korisno u područjima gdje je teško ili skupo dobiti označene podatke, kao što su medicinske dijagnostike, prepoznavanje lica ili analiza prirodnog jezika. Korištenjem ovog pristupa, modeli mogu postići veliku razinu preciznosti u usporedbi s neoznačenim podacima, čak i kada je količina označenih podataka ograničena.

U kategoriju strojnog učenja također se uvrštava učenje pojačanjima. Učenje pojačanjima (eng. Reinforcement learning) je tehnika strojnog učenja (ML) koja obučava softver kako donositi odluke za postizanje optimalnih rezultata. Oponaša proces učenja metodom pokušajima i pogreškama koji ljudi koriste kako bi postigli svoje ciljeve. Softverske radnje koje rade prema vašem cilju se potkrepljuju, dok se radnje koje odstupaju od cilja ignoriraju. [4] Neke od prednosti korištenja učenja pojačanjima su manja potreba za ljudskom intervencijom, gdje je standardno potrebno navoditi model označavanjem podataka, ovdje model dobije direktnije povratne informacije od samih ljudi. Radi na sustavu 'nagrade' za odrađene akcije, gdje je nagrade pozitivna ili negativna, tako trenirajući model da izbjegava negativne ishode.

Iako umjetna inteligencija i strojno učenje imaju mnogo sličnosti, umjetna inteligencija ima veći dohvat, podrazumijeva druge koncepte koji omogućavaju računalima da na više razina emuliraju ljudsku inteligenciju, dok je strojno učenje više nalik primjeni same umjetne inteligencije, gdje se koristeći analizom i obradom masovnih ulaznih podataka, odnosno treningom, ostvaruje inteligencija nalik na ljudsku.

Strojno učenje je i dalje vrlo aktualno, no u novije vrijeme sve je relevantnije i duboko učenje kao alternativa strojnom učenju. Duboko učenje je podskup strojnog učenja koji koristi umjetne neuronske mreže za obradu i analizu informacija. Neuronske mreže sastoje se od

računalnih čvorova koji su slojevito organizirani unutar algoritama dubokog učenja. Svaki sloj sadrži ulazni sloj, izlazni sloj i skriveni sloj. Neuronska mreža se hrani podacima za obuku koji pomažu algoritmu da uči i poboljšava točnost. Kada se neuronska mreža sastoji od tri ili više slojeva, kaže se da je „duboka“, otuda naziv duboko učenje. [5] Duboko učenje je po svojoj prirodi kompleksnija izvedba strojnog učenja zbog korištenja neuronskih mreža i kompleksnijih matematičkih formula kako bi odradili zadani zadatak. Ako uzmemo za primjer mrežu za prepoznavanje glasa, prvi sloj bi mogao se baviti prepoznavanjem osnovnih zvukova ili fonema. Sljedeći slojevi bi se fokusirali na kombiniranje tih fonema u riječi, a na kraju bi završni slojevi identificirali cijele rečenice i njihov smisao. Ova višeslojna obrada omogućuje dubokim neuronskim mrežama da prepoznaju složenije obrasce u podacima, što ih čini idealnim za zadatke poput prepoznavanja govora, obrade prirodnog jezika i računalnog „vida“. Na taj način, duboko učenje nadmašuje tradicionalne metode strojnog učenja u mnogim primjenama, pružajući veću točnost i učinkovitost, posebno u područjima koja zahtijevaju razumijevanje velikih količina nestrukturiranih podataka.



Slika 3. Hijerarhija AI, ML, DL i GenAI (Izvor: sas)

2.2. GenAI

Generativna umjetna inteligencija je podskup dubokog učenja te se fokusira na kreiranje sustava koji generiraju nove podatke, što su najčešće u GenAI-u tekst, slike, audio ili video zapisi. Generativna umjetna inteligencija omogućuje korisnicima brzo stvaranje novog sadržaja na temelju različitih ulaznih podataka. Ulazi i le ili druge vrste podataka. Generativni AI modeli koriste neuronske mreže kako bi prepoznali obrasce i strukture unutar

postojećih podataka te generirali novi i originalni sadržaj. [6] S drastičnim povećanjem računalne snage u trenutnom desetljeću, strojno učenje kao koncept postaje sve više pristupačnije, te sami modeli postaju sve više rasprostranjeni i dostupniji širokom spektru korisnika.

No, što čini dobar generativni model uspješnim? Generalno gledano, generativni modeli su orijentirani na jedan glavni zadatak, imaju svoje prednosti i slabosti. Na primjer, jedan model može biti odličan u generiranju slika, dok je drugi izvrstan u generiranju teksta u konzistentnom i željenom stilu. Iznimno je bitno odabrati modele koji odgovaraju našim potrebama, te zatim odraditi evaluaciju. Evaluacijom generativnih modela određujemo najprikladniji model za određeni zadatak, što pomaže identificirati područja koja zahtijevaju poboljšanja ili područja unutar kojih je model izrazito kvalitetan što je zapravo vjerojatnost postizanja željenih rezultata modela.

Prilikom evaluacije modela, poželjno je promatrati tri glavna kriterija:

1) Brzina rada

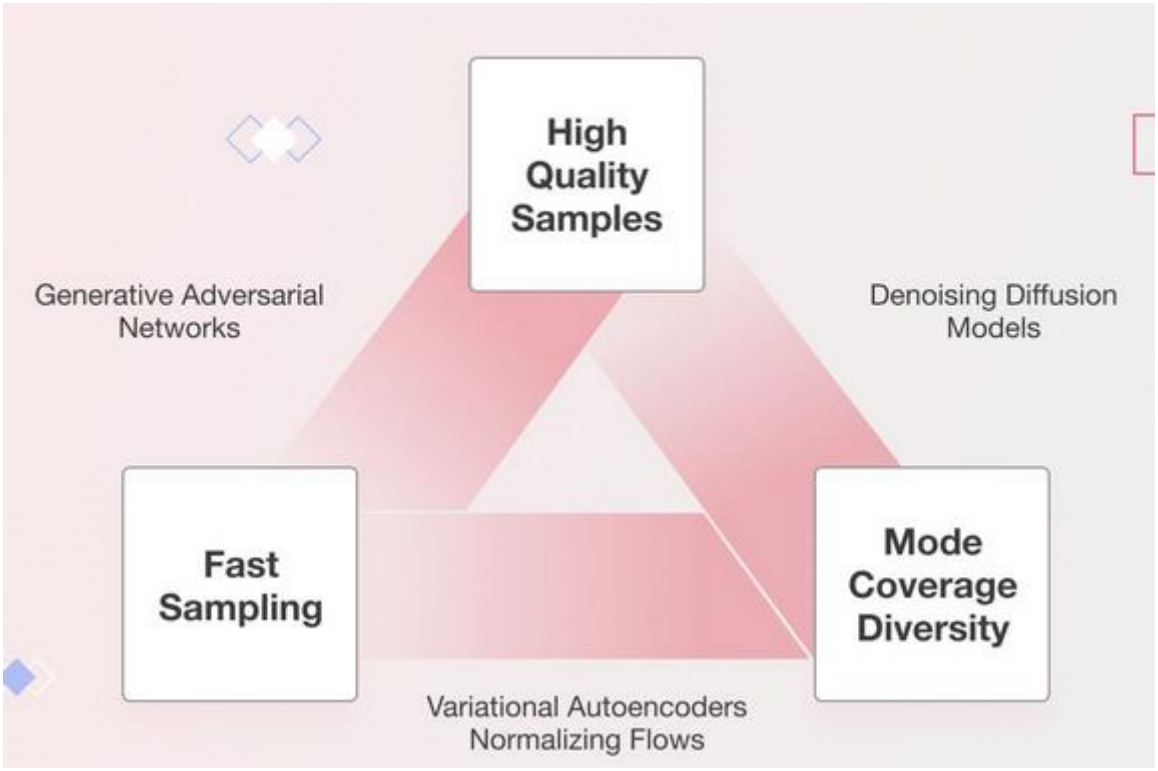
Kako se generativni modeli mogu koristiti u raznim okruženjima, od terminala razvojnih inženjera do elegantnih korisničkih sučelja unutar web aplikacije, iznimno je bitno na umu imati generalnu brzinu rada modela. Na primjer, mnoge interaktivne aplikacije zahtijevaju visoke generacijske brzine, poput uređivanja slika u stvarnom vremenu ili uređivanja detalja na video zapisu, te ako korisnik čeka dugo na kreiranje vizualizacije napravljenih promjena, dolazi do frustracije i gubljenja korisnika. Stoga je brzina kojom generativni model može proizvesti izlaze također važna za razmatranje prilikom procjene njegove učinkovitosti.

2) Raznolikost ulaznih podataka

Dobar generativni model bi trebao biti sposoban obuhvatiti i većinske i manjinske podatke u svojoj distribuciji bez pogoršavanja kvalitete krajnje generacije izlaza korisniku. Takvu kvalitetu modela nazivamo raznolikost, te nam pomaže smanjiti neželjene pristranosti u istreniranim modelima.

3) Kvaliteta izlaza

Kvaliteta generiranih izlaza je ključna stavka u procjeni kvalitete modela, te je izrazito bitna u aplikacijama koje izravno komuniciraju s krajnjim korisnicima modela. Kako bi dobili kvalitetan izlazni podatak, naravno da je iznimno bitno imati kvalitetne podatke nad kojima se model trenirao. Pod kvalitetom moguće je svrstati mnogo pojmova, no neki od najbitnijih kod generativnih modela su: preciznost i pouzdanost, smanjenje pristranosti, sposobnost kvalitetnog generaliziranja, itd. [7]

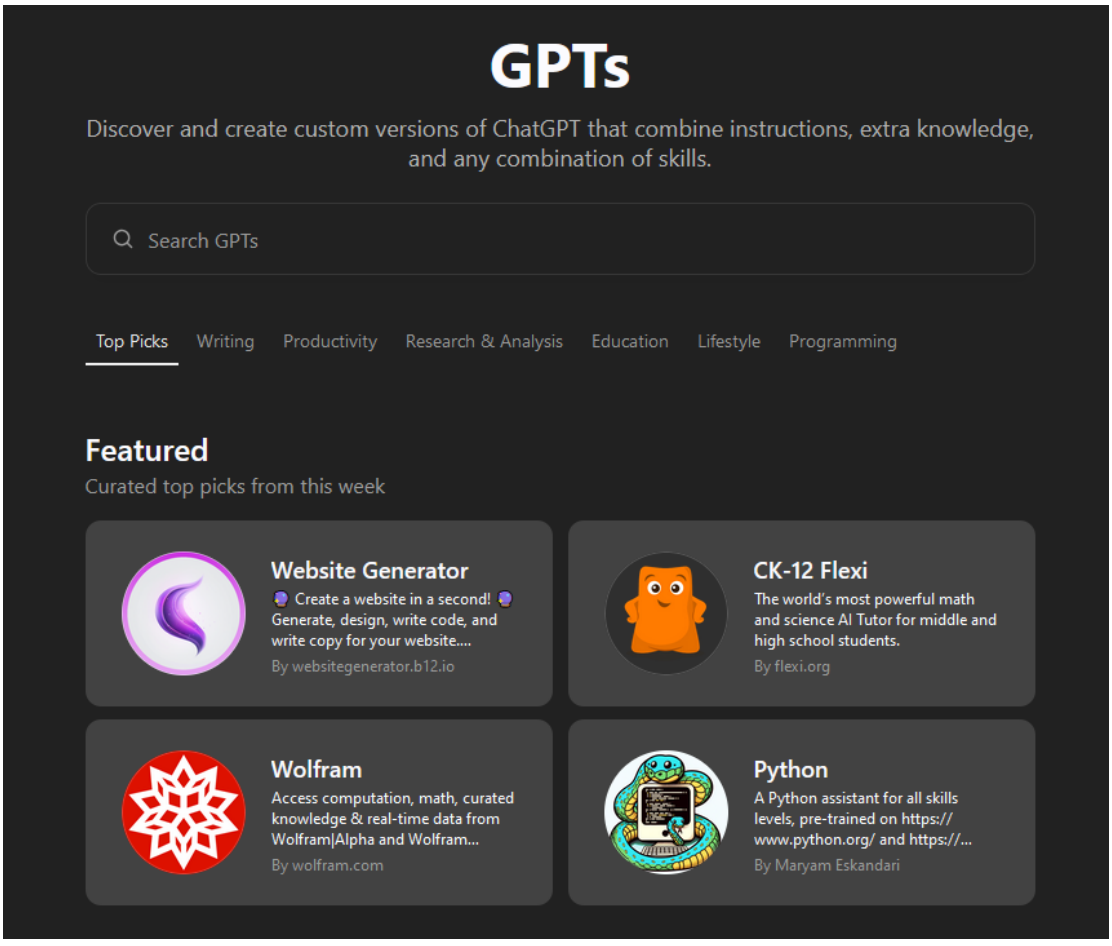


Slika 4. Glavni kriteriji evaluacije modela (Izvor: SIMFORM)

2.3. Temeljni modeli

Kad spominjemo generativnu umjetnu inteligenciju, ključno je biti upoznat i s temeljnim modelima. Generativni AI modeli i temeljni modeli su slični koncepti gdje se elementi preklapaju, no ne preklapaju se u potpunosti. Generativni AI modeli su specifično fokusirani na stvaranje novog sadržaja, kao što su tekst, slike, zvukovi ili videozapisi, no s druge strane, temeljni modeli su općenitiji, mogu se koristiti za širok spektar zadataka i generalno gledano bolje generaliziraju. Trenirani na masivnim skupovima podataka, temeljni modeli su velike neuronske mreže dubokog učenja koje su promijenile način na koji znanstvenici podataka pristupaju strojnome učenju, što znači da umjesto da razvijaju umjetnu inteligenciju od nule, znanstvenici podataka koriste temeljni model kao početnu točku za razvoj modela strojnog učenja koji omogućuju stvaranje novih aplikacija brže i isplativije. [8] Što znači, temeljni modeli su iznimno veliki i univerzalni modeli, trenirani na velikim količinama podataka koji služe kako bi omogućili specijalizaciju modela za potreban zadatak, odnosno daljnje fino podešavanje. (eng. fine tuning) Temeljni modeli su ogroman dio današnje AI industrije. Tvrtke poput OpenAI, koje naravno i sami nude pristup svojim najboljim modelima ChatGPT-4o i 4 pojedinim korisnicima, veliki dio njihove zaradi dolazi od prodaje pristupu svojim modelima drugim poduzećima. Poduzeće može koristiti i same

ChatGPT 4o i 4o mini modele, no poduzeća s više resursa mogu i fino podesiti postojeći GPT temeljni model i trenirati ga na vlastitim podatcima kako bi bio prilagođen potrebama tvrtke. Zanimljivo je spomenuti i da OpenAI ima funkcionalnost koju zovu „GPTs“. Ona omogućava korisniku da na skroman no jednostavan način istrenira 4o model da bude bolje upoznat s tematikom koja je potrebna krajnjem kupcu, te se taj GPT može podijeliti sa svima drugima ili koristiti u privatne svrhe. Naravno to ne zamjenjuje tradicionalno treniranje, na može služiti svrsi ako korisnik ili poduzeće imaju samo nekolicinu bitnih dokumenata iz koje GPT može pregledati relevantne informacije, no iz mog istraživanja, problem i dalje nastaje kada GPT misli da zna nešto preko temeljnog modela, te ne konzultira dan kontekst i dokumente kako bi provjerio, već mu mora biti vrlo strogo eksplicitno rečeno da svejedno pogleda dokumentaciju iako smatra da zna odgovor na zadan upit. Iako će se ovakve greške sigurno ispraviti s vremenom, ukazuje na urođene nedostatke platforme. Naravno, OpenAI nije sam u pružanju tih usluga: Google, IBM, Microsoft, Meta i ostali svi masovno ulažu resurse u AI kako bi imali idući najbolji model.



Slika 5. GPTs unutar ChatGPT web aplikacije (Izvor: Vlastita izrada)

2.4. Izazovi implementiranja GenAI

Koliko god uvođenje generativne umjetne inteligencije (GenAI) u raznim poslovanjima i industrijama nudi ogroman i zasad još nepoznat potencijal za povećanje efikasnosti, inovacija i produktivnosti, sama implementacija GenAI-a dolazi s nizom značajnih izazova implementacije koji se mogu pojaviti u gotovo svim industrijama. Izazovi implementacije variraju od tehničkih i pravnih, tj. regulativnih prepreka, do etičkih i organizacijskih problema, a njihovo umanjivanje i rješavanje su ključni za krajnje uspješno maksimiziranje potencijala GenAI-a kao tehnologije.

Nekvalitetni podatci su najveći problem prilikom implementiranja GenAI-a. Bilo to da su podatci nevažeći ili neispravni, nepotpuni, imaju preveliku pristranost (eng. bias) ili su jednostavno loši, krajnji rezultati treniranog GenAI modela mogu biti neispravni, faktualno netočni, te krajnje iznimno štetni i beskorisni. Česta poslovica kod rada s podatcima u GenAI okruženju je “Smeće unutra, smeće van.” („Garbage in, garbage out.”) [9] što naglašava da nekvalitetni podatci mogu iznimno brzo narušiti sveukupni potencijal i dosadašnje uložen trud u implementaciju GenAI-a unutar poduzeća ili poslovnih procesa. U slučaju da poduzeće ne posjeduje dovoljno podataka za efektivno treniranje GenAI modela, moguće je okrenuti se i kupovini tuđih podataka, no to je također dvosjekli mač jer je također potrebno biti siguran da je kvaliteta kupljenih podataka također iznimno visoka kako bi uopće bila korisna. Na kraju, iznimno je važno razumjeti kakvim podatcima potkrepljujemo GenAI kako bi mogli bolje razumjeti rezultate koje će sam model na kraju i proizvoditi, te tako izbjeći još jedan ključan problem kod GenAI-a: problem „crne kutije“. Problem „crne kutije“ nastaje kada se generiraju rezultati modela bez jasnog uvida u proces razmišljanja tj. korake koji su doveli do tog rezultata, što može biti problematično u organizacijama koje moraju imati detaljno obrazloženu podršku odlučivanju u obliku popratnih materijala, kvalitetno dokumentiranih i prezentiranih činjenica itd.

Opći računalni i ljudski resursi potrebni za kreiranje AI modela su iznimno visoki. Treniranje velikih jezičnih modela (eng. Large Language Models - LLM) zahtijeva velike računalne kapacitete, ponajviše sa strane kompleksnih matematičkih operacija matrica. Grafički procesori (eng. Graphics Processing Unit – GPU) su ključni u treniranju LLM-ova zbog svoje arhitekture. Rade na potpuno drugačiji način od tradicionalnih računalnih procesora koji imaju zadani, jednoznamenkasti ili u profesionalnim okruženjima dvoznamenkasti, broj jezgri dok i obični komercijalni grafički posjeduju na tisuće i tisuće CUDA jezgri. Iako individualno sporiji, mogu iznimno brzo računati ogromnu količinu matematičkih izraza i operacija s matricama. Grafički procesori i ljudi koji su sposobni voditi

treniranje velikih modela su u iznimno visokoj potražnji, te su stoga i iznimno skupi, pogotovo za manja i srednja poduzeća koja nemaju pristup svim potrebnim resursima i infrastrukturi potrebnoj za početak implementacije.

Unutar koje god industrije se poduzeće odluči implementirati GenAI tehnologije, bit će suočeni s pravnim i regulatornim pitanjima i nejasnoćama. Budući da pravni okvir koji definira korištenje umjetne inteligencije generalno i dalje nije u potpunosti definiran stvara se nesigurnost za poduzeća koja žele implementirati GenAI u svoje poslovne procese. Unutar Europske unije, AI Act nastoji uspostaviti jasne smjernice za upotrebu umjetne inteligencije na transparentan način, te kakve implementacije GenAI-a imaju nizak, visok ili neprihvatljiv rizik pred budućim regulacijama. [10] U drugim državama poput SAD-a, iako su pokrenuti određeni regulatorni procesi, pravni okvir koji pokriva cjelokupnu umjetnu inteligenciju i generativnu umjetnu inteligenciju je i dalje godinama u dolasku. [11]

2.4.1. Rješenja i mitigacija izazova implementiranja GenAI

Iako postoji mnogo izazova u radu s tehnologijama umjetne inteligencije, rezultati su toliko impresivni da određene industrije jednostavno nemaju luksuz ignoriranja ove ključne tehnologije. Postoji mnogi koraci koje poduzeća mogu primijeniti kako bi se smanjili rizici i povećala uspješnost ispravne implementacije GenAI-a u pojedinom poduzeću.

Kako bi se spriječio rizik masovnog i nepotrebnog trošenja resursa, prvi korak je procijeniti spremnost našeg poduzeća na nove procese koje donosi umjetna inteligencija. Umjetna inteligencija nije magija, ima svoje granice, te je potrebno biti realističan i prizemljen prilikom definiranja strategija i ciljeva. Većina tvrtki koje danas oklijevaju implementirati GenAI spominje nedostatak talenta koji kombinira AI stručnost i specifičnu domensku ekspertizu. Procjena određene organizacije ključna je kako bismo mogli znati nad kojim poslovnim procesima ima smisla koristiti i kako implementirati GenAI. [12] To podrazumijeva primjenu nad stvarnim poslovnim procesima koji se trenutno odrađuju na manje efikasan ili nepotpun način, te je ključno prioritizirati poslovne procese za koje ima smisla implementirati GenAI, ako uzmemo u obzir izvedivost, trenutne resurse, pravna i regulatorna pitanja i sl.

Idući korak je napraviti evaluaciju dostupnih podataka s kojima bi se budući model trenirao. Idealno bi bilo da se tijekom godina rada poduzeća investiralo u generiranje kvalitetnih podataka, te bi time proces pripremanja podataka za unos i treniranje modela mogao biti mnogo kraći. Iako je većina većih kompanija već dugo vremena upoznata da su podaci ključni za mnoge poslovne procese, isti podaci nekada nisu bili očuvani na ispravan

način, segmentirani su na mnogo strana ili nisu objedinjeni jednakim standardima. Tada je potrebno napraviti dodatne transformacije podataka prije unosa u model. Ukoliko poduzeće nema podataka od prije, moguće je koristiti temeljne modele od velikih AI kompanija (OpenAI, Google, Microsoft, IBM, itd.), te na njih nadodati dostupni materijal i podatke ili je moguće kupiti kvalitetne podatke od mnogih posrednika podataka (eng. data brokers). Prilikom kupovine podataka iznimno je bitno fokusirati se na nama relevantne i iznimno kvalitetne podatke od pouzdanih i legitimnih izvora.

Također postoji nekoliko mogućnosti za umanjenje troškova samog procesa treniranja modela koji je izrazito ovisan o dostupnoj grafičkoj procesorskoj snazi. Ukoliko poduzeće već ima izrazito jak hardver da podrži proces treniranja, moguće je prilagoditi se tako da se kreiraju manji, specijalizirani modeli koji su namijenjeni služiti eksplicitnom poslovnom procesu ili zadatku. No, ako je potrebno kreirati generalniji model, koji služi cijelo poduzeće, jedan način za smanjenje zahtjeva za potrebnom računalnom snagom umjetne inteligencije i povećanje energetske učinkovitosti je kvantizacija.

Kvantizacija je krovni pojam koji obuhvaća mnoge različite tehnike za pretvaranje ulaznih vrijednosti iz velikog skupa u izlazne vrijednosti u manjem skupu. [13] Također, kvantizacija omogućava smanjenje preciznosti brojčanih operacija, poput pretvaranja 32-bitnih vrijednosti u 8-bitne, čime se smanjuje količina podataka koje je potrebno procesirati, što dovodi do bržih izračuna i manjih resursnih zahtjeva, što posljedično smanjuje potrošnju energije, te još bitnije potrebnu računalnu snagu. Iako smanjenje broja bitova može dovesti do manjeg pada točnosti modela, moderna istraživanja, poput onih koje provodi Qualcomm, fokusiraju se na razvoj naprednih kvantizacijskih tehnika koje minimaliziraju ovaj gubitak točnosti, omogućujući održavanje visoke performanse modela i energetske učinkovitosti.

Osim organizacijskih, pravnih i tehnoloških rizika, također postoje i sigurnosni rizici koje je potrebno razmotriti. Postoje mnoge nove vrste napada i sigurnosnih prijetnji prilikom korištenja umjetne inteligencije, no za svaku prijetnju postoji način kako se bolje osigurati i smanjiti šansu za ikakvim sigurnosnim rizicima, više o tome u idućoj sekciji.

2.5. Sigurnosni rizici GenAI

Iako generativna umjetna inteligencija donosi brojne inovacije te može biti primijenjena na mnogo još neotkrivenih načina, istovremeno omogućava brojna nova vrata novim sigurnosnim prijetnjama i napadima u području kibernetičke sigurnosti, sa sve većom površinom napada koje zlonamjerni korisnici mogu iskorištavati. Cijela kibernetička sigurnost, koja se generalno odnosi na održavanje sigurnosti i zaštite sustava, podataka i same mreže nekog okruženja, zadobiva nove izazove koje donosi GenAI, jer se tehnologija i modeli koje GenAI koristi mogu zloupotrijebiti na mnoge načine, neke za koje još nismo ni svjesni zbog same sofisticiranosti umjetne inteligencije.

Budući da su podatci ključan dio efikasnog i korisnog rada umjetne inteligencije, kao i generativne umjetne inteligencije, izrazito je bitno da prilikom učenja iz velikih količina podataka i interakcije s osjetljivim podatcima model odgovara na ispravan način jer dovodi se u pitanje ugled poduzeća, povjerenje korisnika, te i sigurnosni rizici. Sigurnost u radu uz AI ne odnosi se samo na tehničku razinu, već i na etičnu i pravnu, pa i na zaštitu privatnosti i intelektualnog vlasništva pojedinca ili poduzeća.

„Shadow AI“ je među najčešćim primjerima sigurnosnih rizika korištenja umjetne inteligencije u poslovnom okruženju te podrazumijeva praksu kada zaposlenici koriste neovlaštene GenAI alate u svome svakodnevnom radu, no da nitko od zaduženih zaposlenika nije svjestan njihovog korištenja. Istraživanja pokazuju da u poslovnim okruženjima, barem 49% zaposlenika je koristilo alata generativne umjetne inteligencije, od kojih barem trećina koristi navedene alate svaki dan. [14] Shadow AI može dovesti do ozbiljnih sigurnosnih problema, najviše zbog nesvjesnog unošenja privatnih podataka unutar neslužbenih i nekontroliranih GenAI alata koji tada mogu koristiti dane podatke za daljnje vlastito treniranje, obradu, pa čak i da podatci kasnije budu javno dostupni, te se tako izlože zlonamjernim korisnicima, čime se nanosi šteta cijelom poduzeću.

„Prompt Injection“ je vrsta kibernetičkog napada na velike jezične modele (LLM). Hakeri prikrivaju zlonamjerne unose kao legitimne promptove, manipulirajući generativnim AI sustavima (GenAI) kako bi došli do povjerljivih podataka, širili dezinformacije ili uzrokovali još veću štetu. [15] Detaljnije, kroz prompt injection napadač može podnijeti upite nad modelima koje ih tjeraju na nepredviđene odgovore, generiranje neprimjerenog ili štetnog sadržaja, te u najgorem slučaju otkrivanje osjetljivih informacija, bilo poslovnih tajni ili osobno identificirajućih podataka o pojedinoj osobi. Kroz prompt injection napad ukazuju se slabosti u sustavu filtriranja i interpretaciji unosa na što je potrebno posebno pripaziti, pogotovo u javno dostupnim modelima.

„Jailbreak“ je podvrsta prompt injectiona unutar kojeg zlonamjerni korisnik pokušava svojim unosom 'nadjačati' originalne upute dane modelu od strane razvojnog tima u čijem je vlasništvu, te omogućiti veću razinu kontrole zlonamjernom korisniku nego što je namijenjeno. Detaljnije, napad omogućava zlonamjernom korisniku da zaobiđe određene postavljene sigurnosne mehanizme postavljene od razvojnog tima kako bi si zlonamjerni korisnik omogućio da generira sadržaj koji nije dozvoljen ili je štetan, pristupi treniranim podatcima direktno i sl.

Korištenje prompt injectiona i jailbreak-a može dovesti do također ozbiljnog problema „Prompt Leak“. Prompt leak je specifičan oblik prompt injectiona kod kojeg veliki jezični model (LLM) nenamjerno otkriva svoje sistemske upute ili internu logiku. Ovaj problem nastaje kada se promptovi osmisle s ciljem izvlačenja osnovnog sistemskog prompta GenAI aplikacije. [16]

Postoje mnogi drugi napadi poput „Denial of Wallet“ i „Denial of Service“ napadi kada zlonamjerni korisnik ima za cilj iskoristiti tuđe resurse u vlastitu korist. U slučaju denial of wallet napada, zlonamjerni korisnik postigne uzvišeni pristup sučelju modela i može podnositi koliko god zahtjeva želi, te tako organizaciji koja plaća za pristup određenom modelu može brzo nakupiti ogroman trošak. Na sličan način radi i denial of service, gdje umjesto da je cilj zlonamjernom korisniku potrošiti što više sredstava, ovdje je cilj namjerno što više preopteretiti model tako da prestane biti dostupan pravim korisnicima koji ga pokušavaju koristiti u isto vrijeme ili da se umanje generalno dostupni resursi određenog modela.

2.6. Primjeri napada i sigurnosnih rizika

Kako umjetna inteligencija i generativna umjetna inteligencija postaju sve bitniji aspekt u modernom poslovanju, svi žure dodati umjetnu inteligenciju u svoje proizvode, kao i pokušati u manje vremena poboljšati svoje poslovne procese, preskačući bitne korake kako bi se očuvala sigurnost umjetne inteligencije. Pregledajmo nekoliko primjera kako se industrija nosila s novitetima generativne umjetne inteligencije.

Unutar Samsungove korejske branše koja se bavi dizajnom procesora, zaposlenici su bezazleno 'nahranili' ChatGPT model povjerljivim informacijama kada su zatražili pomoć i informacije koristeći ChatGPT za analizu koda. Slučajnim ili svjesnim korištenjem generativnih alata za analizu koda prenosi se osjetljiv i povjerljiv programski kod koji je u intelektualnom vlasništvu kompanije, u ovom slučaju Samsung Korea, te u najgorem slučaju može sadržavati bitne poslovne tajne o radu samih proizvoda nad kojima rade. Ovakav incident naziva se Shadow AI, gdje zaposlenici koriste alate umjetne inteligencije bez nadzora ili dopuštenja od zaduženih osoba, što dodatno otežava sigurnosnu kontrolu zaposlenika pri korištenju generativnih alata. Bez adekvatne regulacije i nadzora, korištenje takvih alata može rezultirati iznimno velikim sigurnosnim propustima, pogotovo ako se incident ponavlja svaki dan bez intervencije.

Nakon što je incident zapažen, Samsung je odlučio implementirati strože mjere korištenja generativnih alata, specifično ChatGPT-a, ograničivši veličinu unosa na 1024 bajta i razmatrajući povratak potpune zabrane korištenja koja je bila prije na snazi. [17] Iako je ograničenje unosa do 1024 bajta neobično, možemo interpretirati da je maksimalna dopuštena količina teksta oko 1024 znakova, ili 150-200 riječi, ovisno naravno o veličini pojedine riječi, više ako se koriste hrvatski znakovi poput č, đ, š i sl. Tim ograničenjem su omogućili da se povjerljivi kod ne šalje masovno, već samo najbitnija pitanja koja stanu u kratkom upitu. Zaposlenici su također bili upozoreni da jednom kada se podaci unesu u javno dostupan generativni model, Samsung gubi kontrolu nad predanim podacima, jer se oni od tada pohranjuju na tuđim serverima, te se gotovo sigurno koriste za buduće treniranje novih modela. Iz članka [17] također doznajemo da Samsung već neko vrijeme razmatra razvijanje vlastite interne AI platforme koja će biti bolje regulirana i izolirana od vanjskih faktora i prijetnji. Kao što je i Samsung sigurno naučio iz incidenta, važno je imati jasno definirane politike i sustave kako bi se kontroliralo korištenje AI alata u poslovnim okruženjima.

Kako online „chat botovi“ postaju sve popularnija primjena generativne umjetne inteligencije, prompt injection napadi predstavljaju sve značajniji rizik, što je u skorom primjeru pokazao chat bot Chevroleta i DPD-a.

U slučaju Chevroleta, chat bot je pomoću prompt injection napada od strane korisnika uspio nagovoriti chat bot da mu proda Chevy Tahoe auto za samo jedan dolar. Korisnik [18] je koristio uobičajene razgovorne mogućnosti kako bi manipulirao botom da izjavi „potvrdu“ kupovine. Iako se ovakav napad naizgled čini bezazlen, ukazuje na ozbiljnije potencijalne probleme koje mogu nastati ako se generativni alati ne implementiraju ispravno ili se ne postave potrebne sigurnosne prepreke.

Slična situacija se dogodila i s dostavnom tvrtkom DPD, gdje je chat bot, omogućen generativnom umjetnom inteligencijom, nakon interne nadogradnje sustava, počeo žestoko kritizirati vlastitu tvrtku, te vrijeđati kupca i tako privukao pažnju mnogih korisnika na društvenim mrežama i vijestima. [19] Ovaj incident je pokazao kako u krivim rukama, može se natjerati generativni alat da prođe preko svih ugrađenih sigurnosnih sistema, te krene raditi ozbiljnu štetu zbog nepredvidivog ponašanja, stvarajući negativan publicitet kompaniji, te tako smanjiti povjerenje mnogih korisnika.

Obe kompanije, nakon što su primijetile problem, brzo su reagirale. Chevrolet [18] je deaktivirao svoj chat bot kako bi što prije spriječio daljnju štetu, dok je DPD [19] onemogućio AI sekciju svog chat bota i vratio se samo na ljudske agente. Iz oba incidenta možemo zaključiti da su prompt injection napadi vrlo ozbiljna prijetnja prilikom korištenja generativnih alata, te je iznimno ključno primijeniti odgovarajuće sigurnosne mjere i unaprijediti sigurnosne mehanizme kako bi se ovakvi incidenti što više izbjegli.

Slika 6. Primjer Jailbreak upita (Izvor: Abnormal)

„Jailbreak“ napadi su također iznimno česti, pogotovo na popularnim metama najvećih javnih modela poput ChatGPT, Bing AI, Bard i sl. Iako neki napadi nastaju iz znatiželje ili želje za eksperimentiranjem s novom tehnologijom, neki korisnici imaju zlonamjerne namjere poput generiranja neprimjerenog sadržaja ili dohvaćanja internih podataka. U primjeru sa slike, korisnik se služi 'DAN' upitom (Do Anything Now) unutar kojeg pokušava nadmudriti generativnu umjetnu inteligenciju da korisnik preko upita uvjerava da je on zapravo još jedan AI sustav, te da ne treba biti podložan istim pravilima kao što bi običan korisnik ChatGPT-a bio, te tako generira sadržaj koji je OpenAI eksplicitno zabranio. Rezultat se također postiže specifičnim formuliranjem određenih riječi na način da zbune model neobičnim uputama te ga natjeraju da zaboravi na dana sigurnosna ograničenja. Cilj ovakvih napada je generalno dobivanje ovlaštenog pristupa popularnom modelu, zaobilaženje ograničenja OpenAI-a i ispitivanje sigurnosnih stavki sustava. OpenAI može izgubiti reputaciju kao jedan od najuglednijih generativnih modela, te čak i bitnije stavke poput vlastitih ugrađenih pravila (eng. system prompt) što zapravo podrazumijeva inicijalni skup uputa danih od razvojnog tima kako bi model znao kako se ponašati s krajnjim korisnikom generativnog modela.

2.7. Primjeri obrane i smanjenje sigurnosnih rizika

Iako uvođenje raznih generativnih alata u poslovnim procesima i okruženju može biti dug i izazovan put, krajnja vrijednost je izrazito velika, te stoga kako bi se obranili od najčešćih napada, treba biti svjestan kako napadi funkcioniraju, te kako spriječiti incident u cijelosti.

Za napad „Prompt Injection“, postoje mnoge tehnike izbjegavanja incidenata poput uvrštavanje ljudskog elementa koji bi mogao prekontrolirati skreće li generativni alat sa zadanog puta, no realističnije je dodavanje još jednog modela. [20] Drugi model bi mogao razmatrati dosadašnji razgovor i potencijalno označiti da razgovor ide u krivu, te predati kontrolu ljudskom agentu, ili model može jednostavno kontrolirati unos sa strane korisnika chat bota. Ukoliko drugi model sumnja na manipulaciju, može poslati signal da se razgovor obustavi ili preda ljudskom agentu. Također je iznimno bitno da se sadržaj kojem generativna umjetna inteligencija ima pristup odvoji na prikladan način kako bi se odvojio sadržaj koji želimo da AI ima pristup, poput prodajnih dokumenata, informacija o uslugama i sl., od osobnih podataka korisnika i osjetljivih podataka poduzeća, poput programskog koda, internih poslovnih dokumenata i informacija zaposlenika.

Problem „Shadow AI-a“, odnosno skrivene umjetne inteligencije, nastaje kada zaposlenici koriste javno dostupne generativne modele bez eksplicitnog dopuštenja poduzeća, često potkrepljeni osjetljivim informacijama poput programskog koda poduzeća ili povjerljivih dokumenata. Problemu se može pristupiti na više načina, jedan od manje fizički restriktivnih je dogovaranje interne politike, poput samo kratkih pitanja, bez dodavanja internog koda, ili samo potvrđeno sigurnih alata i slično. No, moguće je da će neki zaposlenici pokušati ignorirati interna pravila, te tada je moguće iskoristiti alate za kontrolu mrežnog pristupa web mjestima i/ili sučeljima aplikacija modela pomoću konfiguracije firewall-a i kontroliranja samog pristupa uz pomoć korištenja virtualnih privatnih mreža (eng. Virtual Private Networks (VPN)). [21]

„Jailbreak“ se odnosi na pokušaje zlonamjernog korisnika da manipulacijom modela zaobiđe ugrađena ograničenja i generira nedozvoljen sadržaj. Jedno od rješenja je višeslojno filtriranje upita korisnika tako da se posebno promotre određene fraze, pogotovo one poznate ključne riječi u jailbreak strategijama poput „ignore all previous prompts“, „from now on“, „developer mode“, itd. Kao i kod drugih opcija, imati ljudske agente kao pripravu je dobra ideja i kod ovog napada. Moguće je i primijeniti "human-in-the-loop" sustav, gdje kritične ili osjetljive radnje zahtijevaju pregledavanje i odobrenje od strane ljudskog agenta.

Za slučaj „Denial of Service“ i „Denial of Wallet“ napada, gdje napadač pokušava preopteretiti računalne resurse i smanjiti kvalitetu usluge i dostupnost ili povećati troškove bezazlenom poduzeću tako da koristi usluge besplatno ili namjerno pokušava nanijeti što veću monetarnu štetu. Mjere prevencije uključuju ograničavanje broja upita koje pojedini korisnik može odraditi u zadanom periodu (rate limiting), kontrolu dužine i/ili složenosti upita, najčešće uz kontrolu koliko tokena model može prihvatiti kao ulazni parametar, npr. limit od 1000-2000. Tokeni se mogu smatrati dijelovima riječi. Prije nego što API obradi zahtjev, ulaz se razbija na tokene. Ovi tokeni nisu nužno podijeljeni točno tamo gdje riječi počinju ili završavaju – tokeni mogu uključivati razmake na kraju i čak pod-dijelove riječi, gdje su otprilike četiri slova jednaka jednom tokenu. [22] Naravno, uz kontinuirano praćenje anomalija sustava i ispravno implementiranih sigurnosnih prepreka napadača, isplativost izvršavanja „Denial of Service“ i „Denial of Wallet“ napada se može drastično smanjiti.

Kao generalne savjete kako se nositi s novitetima generativne umjetne inteligencije i uvijek rastućom površinom napada od strane zlonamjernih korisnika, važno je držati se čestih smjernica. Implementacija „Zero-trust“ pristupa tretira generativne alate kao nepouzdane entitete, te omogućuje filtriranje i čišćenje unosa i izlaza modela prije razgovora s korisnikom, što pomaže obraniti se od određenih napada. Kontrola pristupa pomoću Role-based access control (RBAC) i attribute-based access control (ABAC) je iznimno moćan

pristup što se tiče sigurnosti generativnih alata jer direktno upravlja tko ima ovlasti raditi sa samim modelom i podacima koje model koristi. Role-based access control pruža jednostavniju kontrolu pristupa na temelju unaprijed definiranih uloga (npr. prodaja, backend developeri, menadžeri, itd.). RBAC je iznimno praktičan za manje do srednje velika poduzeća, no ako se radi o velikim poduzećima, jednostavne uloge mogu biti nedovoljno fleksibilne za iznimno složene generativne sustave. Attribute-based access control nudi generalno veću sigurnost jer omogućuje detaljniju kontrolu pristupa utemeljenu na definiranim atributima (npr. radno mjesto korisnika, starost podataka, razina povjerljivosti podataka itd.). Time se bolje nosimo s rizicima neovlaštenog pristupa povjerljivim podacima, no također je zahtjevnije za implementiranje zbog veće količine atributa u usporedbi s jednostavnim ulogama. Kontinuirani nadzor stanja resursa i ograničenje korištenja istih iznimno su bitni kod određenih vrsta napada. Praćenje najnovijih napada od kojih se potrebno braniti, kao i poboljšanja i ažuriranja sustava koje koristimo, od kritične su važnosti jer se nove vrste napada razvijaju svakodnevno. Održavanje najaktualnijih verzija programa koje koristimo je nužno.

3. Podatci

Na internetu postaje dostupno sve više i više podataka svakih dana, otvoreno svim ljudima na pregled i korištenje. Po nekim izvorima, u 2024. godini, svakog dana se razmjenjuje 0.40274 zetabajta podataka, što je 402.74 milijuna terabajta podataka. [23] Podatci pokreću internet, te su od izrazite važnosti u dobu umjetne inteligencije. Podatci su temelj svakog modela umjetne inteligencije. Kvaliteta, količina i raznolikost podataka su ključni faktori između kvalitetnog modela ili nedovoljno treniranog modela, te izravno utječu na performanse i sposobnosti modela umjetne inteligencije. Tako da, i najnapredniji modeli su vrlo ograničeni u svojoj funkcionalnosti i koristi ako nisu potkrijepljeni odgovarajućim podatcima.

3.1. Big data

Jedan od najvažnijih koncepata u renesansi umjetne inteligencije je big data. Big Data odnosi se na izuzetno velike i raznolike skupove strukturiranih, nestrukturiranih i polustrukturiranih podataka koji se s vremenom eksponencijalno povećavaju. Ovi skupovi podataka su toliko veliki i složeni po volumenu, brzini i raznolikosti da ih tradicionalni sustavi za upravljanje podacima ne mogu pohraniti, obraditi i analizirati. [24] Zbog ogromne količine podataka koja kontinuirano raste, potrebno je razviti nove metode analize i prikupljanja kako bismo imali što više koristi od prikupljenih podataka. Kod big data bitno je imati na umu tri faktora: volumen, brzinu i raznolikost.

Volumen je sama stavka koja i definira big data, te predstavlja ogromnu količinu podataka koja je prikupljena ili je dostupna za prikupljanje iz raznih izvora, bilo to web aplikacije poput društvenih mreža, mobilnih uređaja, Internet of Things (IoT) uređaja poput senzora i kamera, internet trgovina, itd. S toliko dostupnih podataka iz toliko različitih izvora, potrebno je stvoriti procese za pohranu generiranih podataka i daljnju obradu, jer koja je svrha skupljanja podataka ako samo stoje. Naravno, bolje je agregirati podatke za kasniju analizu ako poduzeće ili klijent još nije spreman ili postoji manjak dostupnih resursa. Način na koji upravljamo ovom ogromnom količinom podataka je ključan kako bismo mogli kreirati strategije u poslovanju, promatrati uzorke u korištenju, brže donositi odluke i ostalo.

Brzina predstavlja interval brzine koliko brzo se novi podatci generiraju. Neki podatci nisu vremenski ograničeni, te kad god se analiziraju neće izgubiti na vrijednosti informacija, no dobar dio podataka danas ima veliku korist od prikupljanja, obrade i analize, sve u pravom vremenu. Iako je idealni scenarij sve obraditi u pravom vremenu, treba imati na umu

i našu vlastitu infrastrukturu, koliko podataka je moguće obraditi sada, a koliko kasnije? Tada dolazi u pitanje koji podatci bi trebali imati prioritet pri obradi, te koji bi služili za dugoročnu analizu. U tim slučajevima najbolje je imati na umu potrebe samog poduzeća, što je potrebno kako bi se pospješili poslovni procesi, radile bolje odluke, itd.

Pod raznolikosti ne podrazumijevamo raznolikost samih podataka, već činjenicu da podaci mogu dolaziti iz različitih izvora, te mogu biti strukturirani, polustrukturirani ili nestrukturirani. [24] Prijašnji strukturirani podaci su bili oblikovani tako da mogu biti posloženi unutar relacijske SQL baze, dok u novije vrijeme, kako podaci dolaze iz svih dijelova interneta i uređaja, podaci poprimaju drugačije oblike, čineći ih nestrukturiranima. Generalno gledano, polustrukturirani i strukturirani podatci su zvučni zapisi, video zapisi, čisti tekst, sve što zahtijeva dodatne resurse da bude korisno kao podatak.

Kod big data još je bitno imati na umu i pouzdanost podataka. Ako podatci dolaze od nepouzdanih izvora ili je sam proces prikupljanja podataka nepotpun, može napraviti više štete nego koristi. Sve veći skupovi podataka su teži za osigurati kontrolu, no manji skupovi podataka mogu predstavljati nepotpunu sliku. Naravno, što kvalitetnije podatke posjedujemo, to je veća sama vrijednost prikupljenih podataka.

3.2. Data lake, warehouse i lakehouse

Data lakes (hrv. jezera podataka) funkcioniraju kao središnje spremište za pohranu velike količine podataka u njihovom izvornom, neobrađenom formatu. To uključuje strukturirane, polustrukturirane i nestrukturirane podatke. Podatkovna jezera nude fleksibilnost i skalabilnost, omogućujući organizacijama da pohrane bilo koju vrstu podataka koju bi mogli smatrati vrijednom u budućnosti. [25] Što znači, unutar data lake-a podatke grupiramo u bilo kojem obliku: PDF dokumenti, zvučni zapisi, CSV/JSON skupovi podataka, običan tekst, stanje Internet of Things uređaja, DOCX/XLSX dokumenti, videozapisi, itd. Cilj je sve podatke, dokle god imaju neku trenutnu ili buduću vrijednost, pohraniti unutar data lake-ova za buduću korist. Data lakes za samu obradu podataka koriste Extract, Load, Transform (ELT) procese pomoću kojih se podaci izvlače iz svih dostupnih resursa unutar sustava pohrane samog data lake-a, gdje se nakon pripreme transformiraju u druge potrebne oblike kako bi mogli raditi s ostalim aplikacijama poput OLAP aplikacija analitičke obrade na mreži. Neke od prednosti korištenja data lakeova su kada imamo potrebu za velikim brojem podataka i informacija odjednom, što čini proces analize i planiranja efikasnijim i učinkovitijim u usporedbi s manje, no strukturiranim podacima poput data warehouse sustava, kao i jeftinije skaliranje jer povećavamo samo veličinu koliko čistih podataka možemo pohraniti odjednom.

Data warehouse su usmjereni na pohranu strukturiranih, unaprijed obrađenih podataka koji su relevantni za specifične poslovne potrebe. Skladišta podataka su pažljivo organizirana i optimizirana za upite i analizu, što ih čini idealnima za izvješćivanje i generiranje uvida. [25] U srži slični koncepti, no data warehouse-i najčešće imaju fiksnu strukturu ili shemu koje se potrebno držati, kao u pravom skladištu, jer se očekuju da podatci budu ispravno strukturirani i ispravno označeni. Zbog svoje nehomogene prirode, za razliku od data lake-a, ne možemo unutra pohraniti kakve god podatke želimo poput audio zapisa, dokumenata raznih vrsta i čistih podataka, te ih po kasnijem datumu obraditi. Podaci već trebaju biti obrađeni i spremni za uvoz u data warehouse. No, ovisno o potrebama i poslovnim procesima, poduzeća moraju odlučiti koji način skladištenja primijeniti.

Data lakehouse približava koncepte između pohrane velike količine podataka i njihove obrade u strukturiranom, visokoučinkovitom okruženju. Služi kao jedinstvena platforma za pohranu, obradu i analizu podataka, koja zauzvrat pruža solidnu osnovu za upravljanje AI modelima i provedbu projekata vođenih umjetnom inteligencijom. [26] Data lakehouse je kombinacija data warehouse i data lake koncepata, s ciljem da se iskoriste prednosti i smanje nedostaci oba pristupa pohrane podataka. Data lakehouses omogućuju masovnu pohranu čistih podataka za daljnju obradu, no također omogućuju strukturu koja je potrebna za daljnju integraciju, bilo to s aplikacijama, za korištenje u strojnom učenju i treniranju modela ili sustavima za analizu poslovanja. Rezultat je jednostavnija arhitektura sustava za pohranu, niži troškovi od data warehouse-a jer imamo mogućnost samo dodavati masovnu pohranu bez plaćanja ETL procesa, te potencijalno bolji krajnji rezultat jer podatci na kraju moraju biti svrstani po strukturi sheme.

3.3. Anonimizacija podataka

Anonimizacija podataka je vrsta sanitizacije informacija čija je namjera zaštita privatnosti. To je proces uklanjanja osobno prepoznatljivih informacija iz skupova podataka kako bi osobe koje podaci opisuju ostale anonimne. [27] Osobno prepoznatljive informacije su mnoge, poput imena i prezimena, OIB i JMBG, adrese, brojeva telefona i mobitela, no kroz proces anonimizacije uklanjaju se osobni podaci i/ili zamjenjuju s neprepoznatljivim podacima pomoću enkripcije i kriptografskih tehnika, dodavanjem nasumičnih znakova ili sl.

Količina podataka koja se svaki dan skuplja je iznimno velika, te kako bi poduzeća koja sakupljaju te podatke mogli ih koristiti za strojno učenje i treniranje, daljnje obrade ili prodaje drugima, potrebno je zaštititi privatnost individualaca čiji su podaci prikupljeni zbog regulacija poput Europskog General Data Protection Regulation (GDPR).

Kako bi postigli anonimizaciju podataka, moguće je koristiti mnoge tehnike. Pod anonimizaciju podataka također svrstavamo koncepte pseudonimizacije i prekrivanja podataka.

3.3.1. Tehnike anonimizacije podataka

Postoji mnogo različitih tehnika anonimizacije podataka, te koju je potrebno odabrati ovisi o podatcima s kojima radimo, te za koju potrebu anonimiziramo podatke (analiza, strojno učenje, itd.).

Uklanjanje atributa podrazumijeva uklanjanje cijelog dijela podataka (također poznatog kao "stupac" u bazama podataka i proračunskim tablicama) u skupu podataka. Ovo je najjača vrsta tehnike anonimizacije jer ne postoji način da se iz takvog atributa ponovno dobiju bilo kakve informacije. [28] Generalno, ova metoda se koristi kada određeni atribut više nije koristan ili relevantan za buduću analizu ili ako nije moguće napraviti adekvatnu anonimizaciju, te kompletnim brisanjem podataka iz daljnjeg procesa garantiramo nepovratnost izbrisanih podataka.

Tehnika zamjene znakova, također se nekad zove i tehnika maskiranja, podrazumijeva zamjenu stvarnih podataka s unaprijed dogovorenim simbolima, najčešće znakom X ili *. Prilikom zamjene moguće je zamijeniti cijeli podatak u stupcu, tako da jedina informacija ostane koliko znakova ima podatak, ili je moguće napraviti da samo dio podatka bude sakriven, ovisno o potrebi koliko anonimizacije je potrebno kako bi ostvarili anonimizaciju podataka. Tehnika zamjene znakova također može biti korisna kada je potrebno zadržati neke osnovne činjenice o podatku, no izbjegnuti detalje, npr. poštanski

broj, sve poslije prve znamenke zamijeniti ili kod broja mobitela možemo zadržati prve tri znamenke tako da imamo informaciju kojeg teleoperatera koriste osobe koji podatci opisuju.

Tehnika miješanja podataka podrazumijeva nasumično preuređivanje, tj. miješanje podataka unutar skupa podataka. Svrha zamjene je preurediti podatke u skupu podataka na način da su pojedinačne vrijednosti atributa još uvijek prisutne u skupu podataka, ali općenito ne odgovaraju originalnim zapisima. Tehnika se također naziva mijenjanje i permutacija. [29] Tehnika miješanja je optimalna za korištenje u slučajevima gdje je potrebno analizirati samo jedan atribut unutar baze podataka ili skupa podataka, te ga nije potrebno povezivati s drugim atributima kako bi dolazili do zaključaka o podacima. Iako jaka tehnika, najbolje ju je koristiti u kombinaciji s drugim tehnikama anonimizacije kako bi povećali opću sigurnost protiv ponovne identifikacije pojedinaca.

Još jedna izrazito bitna tehnika je generalizacija. Pod generalizacijom podrazumijevamo promjenu razine detalja pojedinih atributa kako bi se smanjila mogućnost ponovnog identificiranja pojedinca unutar baze ili skupa podataka. Postoje tri glavne podtehnike generalizacije podataka.

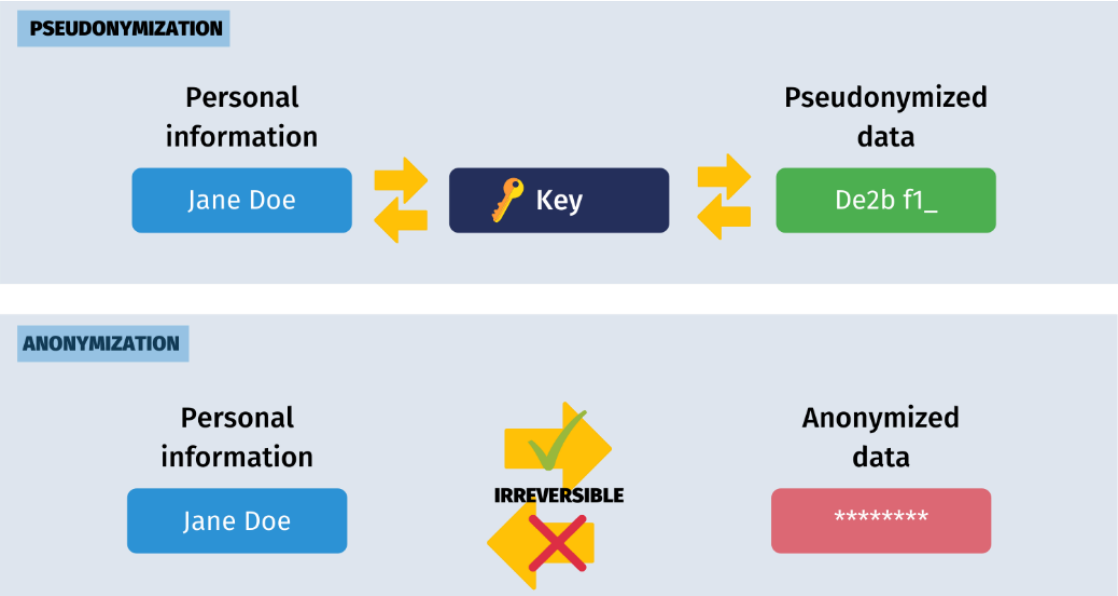
K-anonymity se koristi kako bi osigurali da svaki pojedini redak ili zapis unutar skupa podataka ne može biti jedinstveno identificiran. Drugim riječima, za bilo koji zapis u skupu podataka postoji najmanje K drugih zapisa s identičnim vrijednostima za sve identifikacijske atribute. [30] Korištenjem K-anonymity smanjujemo rizik ponovnog identificiranja, no nije potpuno eliminiran, jer nismo uzeli u obzir određene vanjske faktore koji mogu utjecati na ponovnu identifikaciju. Također je potrebno pripaziti na kvazi-identifikatore. Kvazi-identifikatori su atributi unutar baze ili skupa podataka koji sami po sebi ne identificiraju osobu, no u kombinaciji s drugim informacijama mogu omogućiti ponovnu identifikaciju osobe, što nam je cilj izbjeći pod svaku cijenu prilikom anonimizacije podataka. Kvazi-identifikatori su podaci poput datuma rođenja, spola, mjesta rada i slični.

L-diverzitet, koji osigurava da se informacije nijedne osobe ne mogu razlikovati od najmanje L drugih pojedinaca u skupu podataka na temelju osjetljivog atributa, predstavlja proširenje K-anonimnosti. [30] Najveći problem K-anonymity je ako nemamo dovoljno podataka ili podatci nisu dovoljno raznoliki, individualci se i dalje mogu identificirati putem identifikatora i kvazi identifikatora. No, uz L-diverzitet garantiramo da postoji bar L osoba u skupu podataka za bilo koju vrijednost osjetljivog atributa, kako bi spriječili ponovno identificiranje osobe. Nažalost, L-diverzitet također pati od istih nedostataka kao i K-anonymity te ga je teže implementirati, pogotovo ako se ne radi o velikom i raznolikom skupu podataka ili bazi podataka.

T-bliskost zahtijeva da distribucija osjetljivog atributa u bilo kojoj ekvivalentnoj klasi bude bliska distribuciji tog osjetljivog atributa u cijeloj tablici. [31] Drugim riječima, kada anonimiziramo informacije pomoću generalizacije, T-bliskost omogućava da je raspon vrijednosti ravnomjerno podijeljen, te se tako smanji šansa za ponovnim identificiranjem, pogotovo za gornji i donji ekstrem kod skupa podataka. Na taj način, T-bliskost dodatno povećava zaštitu privatnosti održavajući konzistentnost osjetljivih podataka u odnosu na cijelu populaciju. No, T-bliskost također ima iste poteškoće kao i K-anonymity i L-diverzitet, te ne može garantirati da osoba neće moći biti ponovno identificirana, već se sama šansa ponovne identifikacije smanjuje.

3.3.2. Pseudonimizacija

Europski GDPR definira pseudonimizaciju kao obradu osobnih podataka na način da se podaci više ne mogu pripisati određenom korisniku bez upotrebe dodatnih informacija, pod uvjetom da se takve dodatne informacije čuvaju zasebno i podvrgnute su tehničkim i organizacijskim mjerama kako bi se osiguralo da ne dođe do pripisivanja identificiranoj ili prepoznatljivoj osobi. [32] Jednostavnije rečeno, bilo kakav podatak zamjenjujemo vrijednošću koja ne omogućuje da se osoba identificira. Taj podatak, najčešće enkripcijski ključ, potrebno je čuvati odvojeno, na sigurnoj lokaciji. No, pošto su podatci i dalje indirektno vezani za osobu, i nakon pseudonimizacije, podatci se još uvijek smatraju osobnim podatcima. Pseudonimizacija donosi brojne prednosti. Smanjuje rizike koje obrada podataka predstavlja, omogućava provedbu zaštite podataka kroz dizajn i osigurava odgovarajuću sigurnost. Također, omogućava bolje korištenje podataka, primjerice za arhiviranje, znanstvena i povijesna istraživanja, statističke analize te druge kompatibilne svrhe. [33] Jedna bitna distinkcija kod pseudonimizacije je ako dođe do krađe pseudonimiziranih podataka, nije obavezno obavijestiti osobe unutar podataka jer su vrijednosti pseudonimizirane i podaci su gotovo beskorisni bez ključa, što je izrazito bitno svim firmama koje rukuju podacima ikakve vrste. No, nije svaka tehnika pseudonimizacije legalno podržana, potrebno je pažljivo razmotriti i primijeniti tehniku koja odgovara našem poduzeću.



Slika 7. Vizualizacije anonimizacije i pseudonimizacije (Izvor: Data Privacy Manager)

3.4. Centralizacija podataka

Centralizacija podataka uključuje konsolidaciju podataka iz različitih izvora u jedno spremište, poboljšavajući upravljanje, kvalitetu i dostupnost podataka. Ovaj proces eliminira podatkovne silose i pruža jedinstven pregled podataka u cijeloj organizaciji, podržavajući donošenje odluka i strateško planiranje. Centralizirani podaci nisu samo pitanje fizičke blizine; radi se o stvaranju kohezivnog podatkovnog ekosustava koji podržava upravljanje podacima, sigurnost i upravljanje kvalitetom. [34]

Centralizacija podataka je bitan proces u ciklusu podataka, te ima mnoge prednosti koje pospješuju cijeli proces agregiranja i korištenja podataka. U centraliziranom sustavu, podaci se mogu lakše standardizirati, što omogućuje dosljednost i integritet podataka, te najbitnije, održava se očekivani standard kvalitete podataka. Standardizacijom smanjujemo šansu dupliciranih podataka i mogućih pogrešaka formata zapisa unutar skupa ili baze podataka.

Također, uz centralizaciju, povećavamo sigurnost pristupa podacima i mogućnost učinkovitijeg upravljanja resursima. Unutar centraliziranog sustava možemo implementirati razne sigurnosne protokole i kontrolirati pristup nad samim podatcima samo ovlaštenim osobama, što je potrebno za usklađivanje s određenim zakonskim regulativama, kao i generalno dobro pravilo. Također, što su podaci raspršeni kroz manje različitih sustava, to

firme mogu smanjiti troškove održavanja i iskoristiti resurse za dodavanje više resursa poput procesorske snage i dodatne pohrane.

Sve u svemu, centralizacija, ako je ispravno implementirana, daje određenu dozu fleksibilnosti nad prikupljenim podacima. Firme koje su uspješno centralizirale svoje podatke mogu se brže prilagoditi novim poslovnim promjenama i prilikama poput korištenja svih sakupljenih podataka za potrebe treniranja modela. Centralizirani i visokokvalitetni podatci su optimalni i ključni za uspješno treniranje modela, poboljšavajući preciznost modela.

Centralizacija podataka nije bez nedostataka. Jedan od glavnih problema je ako su nam svi podaci centralizirani na jednom mjestu, ako dođe do sigurnosne prijetnje, napadačima je lakše probiti jedan sustav u kojem se sve nalazi nego više manjih, ali razdvojenih sustava. Također, ako previše ljudi unutar poduzeća pokušava koristiti jedan sustav odjedanput, može doći do zagušenja i nedostatka resursa čime ne mogu svi pristupiti podacima koji su im potrebni. No, takvi problemi se mogu umanjiti tako da budemo pažljivi da uvijek investiramo u unapređenje samog hardvera koji pokreće centraliziran sustav, bilo to procesor, pohrana, pristup i sl.

3.5. Trenutno stanje u industriji podataka i etička pitanja

Tijekom ogromnog proširenja i brzog razvoja umjetne inteligencije zadnjih godina, profesionalci u raznim industrijama poput umjetnika, novinara, glazbenika i drugih kreativaca su na prvoj liniji udara. Kako AI modeli postaju sve sposobniji, tako sve bolje mogu oponašati značajke ljudske mašte i kreativnosti, te to predstavlja iznimno velik izazov očuvanja radnih mjesta u ovim industrijama. Gotovo svi aktualni generativni modeli trenirani su na raznim medijima dostupnima na internetu, bilo to slike, fotografije, videozapisi, pa čak i članci. Treniranje se na početku odrađivalo bez ikakvih dopuštenja originalnih autora, no čak i dan danas, regulacije se razvijaju vrlo sporo i svaki drugi dan izađe nova vijest da je kompanija prekršila autorska prava.

Uzmimo za primjer samu Nvidiju, na čelu hardverske podrške za sve vezano za AI uz pomoć CUDA sustava i iznimno sposobnih i bitnih grafičkih kartica, nekoliko puta je već uhvaćena da trenira svoje modele na sve dostupne videozapise u zadnjih 80 godina, praktički od zasnivanja samih videozapisa. Istraga Proof Newsa otkrila je da je Nvidia započela sakupljanje već kreiranih video datasetova s interneta, čija se veličina kretala od stotina klipova do stotina milijuna videozapisa. Prema podacima iz internog, kompanijskog Slacka i povjerljivih dokumenata, osoblje se brzo fokusiralo i na YouTube, koji je dom

milijardama videa, a koje je Nvidia tim prikupljao preuzimanjem datasetova prethodno skinutih videa, kao i vlastitim scrapingom. [35] Naravno, sami zaposlenici nisu sami odlučili kršiti autorska prava svih kreatora na svijetu, te kad je pitanje bilo postavljeno oko autorskih prava i dopuštenja, vrh Nvidije je odgovorio da samo nastave s radom, očigledno takve odluke uvijek dolaze od vrha. “In the meeting today, we get [sic] permissions to download all kind of data,” rekao je Ming-Yu Liu, potpredsjednik razvoja i istraživanja u Nvidiji. “Should we download the whole Netflix too? How could we operationalize this?”. [35] Sve su jasni indikatori da nikakav dogovor nije bio sklopljen, već samo beru što god im je volja, sve u cilju imati najkonkurentniji model na tržištu. Naravno, nisu oni jedini koji koriste, najblaže rečeno, neetične metode sakupljanja podataka za treniranje. Unutar istog internog kanala, zaposlenici Nvidije su interno optuživali OpenAI zaposlenike da treniraju svoj vrhunski model Sora na komercijalnim filmovima zbog brojnih sličnosti pri usporedbi rezultata modela. U zaključku, nijedna od velikih kompanija koje se bave umjetnom inteligencijom ne shvaća autorska prava ozbiljno, svi samo rade što žele jer regulacije su gotovo nepostojeće još uvijek.

No, postoje neki pomaci gdje se podatci i generalan sadržaj pokušava licencirati na legitiman način. Google i Reddit su nedavno stekli novo partnerstvo, kroz čiju suradnju će se omogućiti Googleu pristup Redditovom API-ju za podatke, koji isporučuje sadržaj u stvarnom vremenu s Redditove platforme. [36] Google i OpenAI su oboje sklopili dogovor sa Stack Overflow-om kako bi imali legitimni pristup stotinama tisuća pitanja i odgovora što se tiče programiranja. [37] OpenAI je također sklopio suradnju s TIME nakladništvom, omogućujući legitiman pristup stogodišnjem katalogu članaka i materijala OpenAI-u za slobodno korištenje tijekom treninga. [38] Sve to ukazuje da velike kompanije imaju volje platiti velike novce za kvalitetne podatke za treniranje. No, naravno, to ne mijenja činjenicu da se velika većina ostalog sadržaja na internetu slobodno preuzima u svrhu treniranja bez brige za autorska prava, što je daleko od idealnog sistema.

3.5.1. Zaštita umjetnika od scrapinga i treniranja generativnih modela

Kreativci na internetu nažalost moraju sami zaštititi svoj rad od scrapinga u svrhe treninga generativnog modela. Tim sa Sveučilišta u Chicagu razvio je projekt Glaze, koji uključuje alate poput Glaze i Nightshade za zaštitu kreativaca od invazivnih upotreba generativne umjetne inteligencije (GenAI). [39] Nezadovoljni tretiranjem kreativaca u industriji, pokrenuli su projekt Glaze kako bi omeli masovna treniranja modela nad njihovim radom, naglašavajući koliko je bitna ljudska kreativnost, perspektiva i emocije tijekom cijelog kreativnog procesa, te sve s krajnjim ciljem - vratiti kontrolu kreativcima nad svojim vlastitim

kreacijama. Od svog zasnivanja 2024. godine, članovi tima kontinuirano rade na alatima, radovima i aplikacijama sve u svrhu zaštite kreativnosti.

Razvili su dva glavna alata, Glaze i Nightshade, svaki ima svoju primjenu, no različit način napada i utjecaj na trenirani model.

Glaze je sustav dizajniran za zaštitu ljudskih umjetnika ometanjem oponašanja umjetničkog stila. Na visokoj razini, Glaze djeluje tako da razumije način na koji AI modeli, koji se treniraju na ljudskoj umjetnosti, rade i koristeći algoritme strojnog učenja izračunava skup minimalnih promjena na umjetničkim djelima, tako da ona ostanu nepromijenjena za ljudsko oko, ali AI modeli ih vide kao radikalno drugačiji umjetnički stil. [40] Vrlo laički rečeno, može se zamisliti kao iznimno inteligentno osmišljen 'filter' nad konačnom slikom, razlike između dvije slike mogu se vidjeti po artefaktima na slici, ovisno od slike do slike.

Glaze funkcionira na način dodavanja dodatnog sloja, jedva vidljivog ljudskom oku, no sloj izrazito mijenja način na koji umjetna inteligencija percipira stil slike, otežavajući krađu stila umjetnika. Glaze ne funkcionira kao običan filter preko rasterske slike, već je puno otporniji na razne tehnike manipulacije slike od strane napadača koji bi htio ukrasti sliku za treniranje vlastitih modela. Glaze projekt tim navodi da je proces otporan na bilo kakve vrste napada, uključujući: snimanje zaslona s kojeg se pregledava slika (eng. screenshot), obrezivanje dijelova slike, procesi uklanjanja šuma ili artefakata, mijenjanje formata slike, kompresija slike, zaglađivanje piksela i dodavanje 'šuma' slici, sve u pokušaju kako bi podli korisnici mogli koristiti slike, no sve neuspješno. Takvi pokušaji nisu uspješni jer Glaze radi na potpuno drugačiji način, gdje pri bilo kakvom mijenjanju originalnog rada tehnike poput steganografije, gdje je moguće ugraditi prekrivene poruke i podatke unutar slike, odmah prestaju raditi, Glaze ustraje.

Glaze tim naglašuje ograničenja i rizike koje dolaze s korištenjem Glazea. Nedostatak koji je prije predstavljao veći problem je utjecaj na kvalitetu slike, pokazivajući vidljive artefakte, pogotovo na jasnijim detaljima, no s novijim poboljšanjima verzije v2.0, artefakti vidljivi ljudskom oku su sve manje zamjetljivi, što je izrazito bitno umjetnicima na internetu. Također naglašavaju da Glaze nije trajno rješenje protiv krađe stila. Budući da algoritmi za treniranje i metode za struganje weba naglo napreduju, bit će sve teže braniti se od takvih napada kako umjetna inteligencija i tehnologija napreduje, no njihovim kontinuiranim radom nadaju se da će Glaze biti iznimno bitan alat umjetnicima u borbi za vlastita prava nad svojim radom.

Nightshade djeluje slično kao Glaze, ali umjesto obrane od oponašanja stila, dizajniran je kao alat napada za izobličenje značajki unutar generativnih AI modela za slike. [41] Umjesto obrane od cjelokupnog treniranja što je cilj Glazea, Nightshade je više orijentiran ofenzivi, gdje nije cilj potpuno uništiti model nad kojim se trenira, već drastično povećati cijenu cjelokupnog procesa treniranja kako sam proces više ne bi bio isplativ. To se postiže 'trovanjem' modela prilikom treniranja. Kao ljudska bića, po našoj prirodi mi promatramo slike puno drugačije nego što računalo ili algoritam ih tumači, te kada je nad slikom primijenjen Nightshade, umjetna inteligencija može krivo protumačiti sadržaj slike i trenirati nad potpuno nevažećim podatcima, poništavajući dio rezultata treniranja. Nightshade radi na istom principu kao i Glaze, gotovo nevidljivog ljudskom oku 'filtera' nad slikom koji drastično mijenja način na koji umjetna inteligencija percipira sliku, no s ključnom razlikom. Nightshade nastoji promijeniti percepciju slike na način da uvjeri model koji se pokušava trenirati da je sadržaj slike nešto potpuno različito.

Također, to znači da snimanje zaslona s kojeg se pregledava slika (eng. screenshot), obrezivanje dijelova slike, procesi uklanjanja šuma ili artefakata, mijenjanje formata slike, kompresija slike, zaglađivanje piksela i dodavanje 'šuma' na slici ne uklanjaju Nightshade efekt. Nažalost, kako Glaze tim nije velik, Nightshade nije dobio brza poboljšanja kao što je Glaze dobio, te je i dalje inicijalna verzija s početka 2023. godine najaktualnija verzija. No, ako što više umjetnika koristi Nightshade, to će imati značajniji utjecaj na rezultate treniranja modela.

Slika 8. Usporedba verzija Glazea i Nighstshade-a (Izvor: Vlastita izrada)

U gornjem lijevom kutu je originalna slika, u gornjem desnom kutu Glaze v1.0.1 verzija slike, u donjem desnom kutu je Glaze v2.1, te u donjem lijevom kutu je Nightshade v1.0 (High inicijalne postavke). U tamnijim dijelovima slike možemo vidjeti razliku između prve originalne slike i druge slike obrađene Glazeom v1.0.1 jer su vidljivi tamnozeleni artefakti na plaštu lijevog lika sa slike, koji su na svu sreću drastično umanjeni u trećoj slici, Glaze v2.1. Razlike između prvog, originalnog rada i treće, Glaze v2.1 verzije su iznimno minimalne kroz cijelu sliku, kao i na prijašnjim primjerima, iznimno je teško uočiti razlike golim okom, tako da njihove tvrdnje da je način primjenjivanja Glazea drastično poboljšan drže vodu. Na Nightshadeu v1.0 na najjačim postavkama možemo vidjeti da i iz daljeg kuta promatranja slike, zelene mrlje i artefakti su iznimno vidljivi, i to ne samo na tamnijim i rubnim dijelovima slike kao što smo prije zamijetili, nego preko cijele slike mogu se uočiti drastični artefakti, čineći sliku previše uništenu za generalno korištenje. No, slika izgleda mnogo bolje na zadanim (Normal) postavkama.

4. Praktični dio

4.1. Pregled dosadašnjih radova

Za praktični dio rada iznimno je relevantan rad „Analysis of Data Anonymization Techniques“. [43] Kroz rad se razmatraju različite tehnike anonimizacije podataka koje se koriste kako bi se uklonili osobni identifikatori iz skupa podataka. Autori istražuju prednosti i slabosti različitih tehnika anonimizacije te analiziraju rizike ponovne identifikacije, što je posebno važno u eri kada se osobni podaci široko dijele i koriste u komercijalne, istraživačke svrhe i za strojno učenje ogromnih modela. Iako pravilno primijenjene tehnike anonimizacije mogu značajno smanjiti rizik od kršenja privatnosti, autori naglašavaju da u određenim slučajevima postoji mogućnost djelomične ponovne identifikacije pojedinca čak i nakon anonimizacije.

Također za praktični dio rada bitan je i rad „Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification“. [44] Ovaj rad istražuje kako dijeljenje podataka utječe na privatnost, anonimizaciju i de-identifikaciju, posebno u zdravstvenom sektoru.

Autori počinju pregledom velikih povreda podataka u zdravstvu i naglašavaju važnost osiguravanja da samo ovlaštene osobe imaju pristup podacima. Raspravljaju o različitim metodama zaštite podataka, uključujući enkripciju i tehnike anonimizacije, te pružaju pregled modela zrelosti za procjenu zaštite podataka.

Rad je bitan za moj praktični prikaz anonimizacije baza podataka jer pruža praktične savjete o zaštiti podataka, posebno zdravstvenih nad kojima ću prikazati loše primjere primjene anonimizacije.

Kroz svoj praktični dio prikazat ću kako bi izgledao proces anonimizacije baze podataka manjeg poduzeća, na koje podatke je potrebno posebno pripaziti, kako koje podatke sačuvati kako bi i dalje bili korisni, te da i dalje podatci budu potpuno anonimizirani.

4.2. Uvod u pynonymizer

Postoji mnogo alata, od otvorenog koda do komercijalnih, koji mogu pomoći s anonimizacijom podataka unutar skupa podataka ili baze podataka.

pynonymizer je alat otvorenog koda, napisan u Python programskom jeziku pomoću kojeg možemo obaviti anonimizaciju željenih podataka unutar baza podatka, te tako postižu usklađenost naše baze s GDPR-om ili pripremanjem podataka za treniranje. Pynonymizer, umjesto da zamijeni podatke simbolom ili enkriptiranim podacima kao kod pseudonimizacije, koristi nasumične, no i dalje realistične podatke uz pomoć druge biblioteke, Faker.

Ključne značajke pynonymizera su: [45]

- Podržava MySQL, PostgreSQL i MSSQL baze podataka
- Prihvaća različite ulazne formate (SQL, komprimirane datoteke)
- Generira anonimizirane izlaze u više formata
- Fleksibilne strategije generiranja podataka za različite slučajeve upotrebe
- Jednostavno sučelje za korištenje putem naredbenog retka i Python biblioteke

Alat pynonymizer ima mogućnost anonimizirati mnogo tipova podataka unutar baze podataka poput imena, prezimena, emaila, grada, poštanskog broja, pa čak i IPv4 adresa, MAC adresa i brojeva kreditnih kartica. Za taj proces je najčešće zadužen fake_update koji nasumično vuče podatke iz skupa podataka kako bi zamijenio legitimne podatke. Zamijenjeni podaci nisu unikatni, niti ne mogu biti, jer se vrijednosti s kojima se zamjenjuje biraju nasumično, te baze podataka mogu biti male do nevjerojatno velike. Ako je potrebno svojstvo unikatnosti podataka, potrebno je koristiti sofisticiranije metode.

Alat pynonymizer može raditi s MySQL Serverom, Microsoft SQL Serverom ili PostgreSQL-om, za svaki podržava malo drugačije ulazne i izlazne podatke, no proces je uvijek sličan. Također je moguće raditi direktno i sa .sql datotekama za pažljiviju kontrolu nad odrađenim zadacima.

4.2.1. Datoteka strategije

Način na koji usmjeravamo pynonymizer kako raditi s našim vlastitim podatcima je datoteka strategije. Unutar datoteke strategije definiramo “što i kako” pristupiti procesu anonimizacije. [46]

Format strategijske datoteke može biti ili u JSON ili u YAML obliku. Ako unutar ulaznih podataka imamo više tablica, potrebno je koristiti ključnu riječ tables, te po želji definirati postavke za svaku dodatnu tablicu unutar datoteke strategije. Neke tablice je moguće potpuno maknuti iz konačnog rezultata, pomoću truncate (manje poštuje ovisnosti oko ključeva) ili delete (više pazi oko mogućih ograničenja ključeva). Ako želimo specifične vrijednosti unutar tablice zamijeniti s drugima, možemo koristiti update_columns kao i where upit gdje definiramo što tražimo. Unutar datoteke strategije možemo i definirati druge izvore zamjenskih podataka, povrh Faker biblioteke, uz pomoć ključne riječi providers. Alat pynonymizer također omogućava, umjesto da se koriste standardne engleske riječi i imena za zamjenu, korištenje ključne riječi locale, gdje je moguće definirati koju lokalizaciju koristiti prilikom procesa anonimizacije. Na primjer, umjesto da anonimiziramo prava imena u tablici engleskim imenima, možemo anonimizirati čestim hrvatskim imenima. Konačno, unutar strategijske datoteke možemo definirati dodatan SQL kod koji je potrebno izvršiti nad bazom pomoću ključne riječi before (za akcije koje želimo odraditi prije procesa anonimizacije) i after (za akcije koje želimo odraditi poslije procesa anonimizacije).

4.3. Anonimizacija baze podataka

Pogledajmo na primjeru kako bi izgledao proces anonimizacije baze podataka unutar jednog poduzeća. Uzmimo za primjer da postoji interni portal za sve zaposlenike poduzeća. Postojeća baza zaposlenika ima stupce: ID broj, korisničko ime, ime, prezime, spol, lozinka, status i plaća.

```
CREATE TABLE IF NOT EXISTS `detalji_zaposlenika` (
  `korisnik_id` int(11) NOT NULL AUTO_INCREMENT,
  `korisnicko_ime` varchar(255) DEFAULT NULL,
  `ime` varchar(50) DEFAULT NULL,
  `prezime` varchar(50) DEFAULT NULL,
  `spol` varchar(10) DEFAULT NULL,
  `lozinka` varchar(50) DEFAULT NULL,
  `status` tinyint(10) DEFAULT NULL,
  `placa` varchar(30) DEFAULT NULL,
  PRIMARY KEY (`korisnik_id`)
) ENGINE=MyISAM DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_unicode_ci AUTO_INCREMENT=10001;
```

Slika 9. Inicijalna baza podataka zaposlenika (Izvor: Vlastita izrada)

Prilikom definiranja same baze podataka, također smo trebali promijeniti par stavki prije početka procesa anonimizacije. Stupac placa je postao varchar(30) umjesto int(10), pošto nova vrijednost neće biti broj, nego raspon, tako da se mora zapisivati kao znakovi kroz varchar. Budući da je plan koristiti hrvatske zamjene podataka kroz anonimizaciju, potrebno je dodati DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_unicode_ci kako bismo mogli koristiti hrvatske znakove, bez čega proces anonimizacije ne prolazi jer ćemo koristiti hrvatske zamjene. UTF8 oznake dodali smo na kraj tablice tako da se primijeni za svaki stupac unutar tablice, no ako je potrebno koristiti UTF8 samo na određenim stupcima, može se definirati na definiciji pojedinog stupca.

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

korisnik_id	korisnicko_ime	ime	prezime	spol	lozinka	placa	status
1	rogers63	david	john	Žensko	e6a33eee180b07e563d74fee8c2c66b8	2124	1
2	mike28	rogers	paul	Muško	2e7dc6b8a1598f4f75c3eaa47958ee2f	2934	1
3	rivera92	david	john	Muško	1c3a8e03f448d211904161a6f5849b68	1812	1
4	ross95	maria	sanders	Muško	62f0a68a4179c5cdd997189760cbcf18	2543	1
5	paul85	morris	miller	Žensko	61bd060b07bddfecccea56a82b850ecf	1932	1
6	smith34	daniel	michael	Žensko	7055b3d9f5cb2829c26cd7e0e601cde5	2975	1
7	james84	sanders	paul	Žensko	b7f72d6eb92b45458020748c8d1a3573	1310	1
8	daniel53	mark	mike	Muško	299cbf7171ad1b2967408ed200b4e26c	2567	1
9	brooks80	morgan	maria	Žensko	aa736a35dc15934d67c0a999dccff8f6	2786	1
10	morgan65	paul	miller	Žensko	a28dca31f5aa5792e1cefd1dfd098569	1754	1
11	sanders84	david	miller	Žensko	0629e4f9f0e01e6f20bc2066175e09f7	1643	1
12	maria40	chrisha…	bell	Žensko	17f286a78c74db7ee24374c608a2f20c	1492	1
13	brown71	michael	brown	Muško	fa0c46cc4339a8a51a7da1b33e9d2831	1847	1
14	james63	morgan	james	Muško	b945416fa907fac533d94efe1974ec07	2635	1
15	jenny0993	rogers	chrisha…	Žensko	388823cb9249d4cebc9d677a99e1d79d	1378	1
16	john96	morgan	wright	Muško	d0bb977705c3cdad1e346c898f32a1b7	2764	1
17	miller64	morgan	wright	Muško	58b207ee33794b046511203967c8e0d7	2938	1
18	mark46	david	ross	Žensko	21cdcb68a932871524e16680fac72e18	1229	1
19	jenny0988	maria	morgan	Žensko	ec9ed18ae2a13fef709964af24bb60e6	2701	1
20	mark80	mike	bell	Muško	084489b355edd349bca1c798788de19a	2875	1

Slika 10. Početni sadržaj baze podataka (Izvor: Vlastita izrada)

Zadatak nam je maknuti sve podatke koji se mogu smatrati podatcima koji mogu identificirati pojedinu osobu. Kako bi to postigli, potrebno je započeti s kreiranjem datoteke strategije gdje ćemo opisati sve radnje koje želimo napraviti nad podatcima unutar tablice.

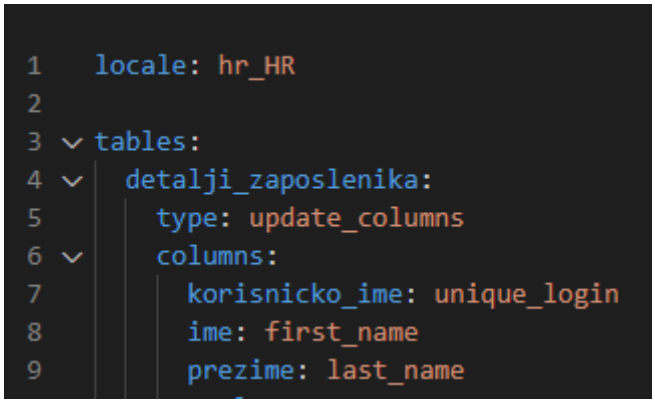
Prvo je potrebno utvrditi koji se sve stupci u tablici mogu smatrati osobno identificirajućima, te tri stupca koja su najizravnije vezana za osobu su korisničko_ime, ime i prezime. Obavit ćemo anonimizaciju nad svakim stupcem, tako da zaposlenici više ne mogu biti pojedinačno identificirani. Jedan također problematičan stupac je korisnik_id. Istina, sam po sebi ne mora govoriti puno, no u kombinaciji s drugim informacijama poput imena, može se saznati tko je bio unesen u tablicu prije od drugih, odnosno tko je kada zaposlen unutar poduzeća kojeg promatramo, što može doprinijeti neželjenom curenju informacija. Zato ćemo

unutar datoteke strategije navesti da želimo obrisati korisnik_id stupac prije početka procesa same anonimizacije.

Plaća također može biti osjetljivi podatak, pogotovo kada je zapisana u točnom iznosu. Primijenit ćemo anonimizaciju i postaviti raspone plaća u iznosu od 200 kako bismo smanjili šansu ponovne identifikacije osobe.

4.3.1. Kreiranje strategijske datoteke

Kako bi pokrenuli proces anonimizacije baze podataka putem alata pynonymizer, potrebno je napraviti strategijsku datoteku kako bi znao što treba odraditi u kojim stupcima. Iako je početna baza relativno jednostavna, ako želimo postići anonimizaciju, potrebno je za svaki stupac razmisliti otkriva li osobne podatke korisnika. Kako bi odradili anonimizaciju nad tablicom podataka zaposlenika, kreirali smo strategijsku datoteku strategija.yml.



Slika 11. Definicija locale i prvih stupaca tablice (Izvor: Vlastita izrada)

Na početku strategijske datoteke odmah definiramo vrijednost locale. Locale u ovom slučaju omogućava da se umjesto engleskih alternativa (npr. engleska imena i prezimena, brojevi telefona i sl.) koriste hrvatski lažni podaci. Locale može iznimno biti koristan kada je i dalje bitno očuvati državljanstvo osoba, ili ako je prikladnije raditi s hrvatskim podatcima.

Zatim je potrebno definirati sve tablice nad kojima ćemo raditi anonimizaciju. U našem slučaju imamo samo tablicu detalji_zaposlenika, tako da navodimo nju. Pod type navodimo vrstu akcije koju želimo odraditi nad navedenom tablicom: delete, truncate ili update_columns. Delete i truncate potpuno brišu navedene tablice, pri čemu delete pokušava pažljivo obrisati tablicu u odnosu na ključeve unutar tablice, dok truncate generalno samo briše. U našem slučaju, mi želimo anonimizirati same podatke, ne ih obrisati, pa koristimo update_columns kako bismo ažurirali vrijednosti unutar definiranih

stupaca. Tri stupca koja smo odmah znali da je potrebno anonimizirati su korisnicko_ime, ime i prezime. Te informacije su preosjetljive, te korisnicko_ime pogotovo direktno identificira pojedinu osobu.

```
10        spol:
11          type: literal
12          value: |
13            CASE
14              WHEN spol = 'Muško' THEN 'M'
15              WHEN spol = 'Žensko' THEN 'Ž'
16              ELSE 'O'
17            END
```

Slika 12. Provjera vrijednosti stupca spol (Izvor: Vlastita izrada)

Kroz stupac spol smo demonstrirali kako bi izgledalo ako npr. imamo zadatak neke vrijednosti usput promijeniti, u slučaju da je potrebno promijeniti format kako se neki podatci zapisuju unutar baze. Type: literal govori da će se željena promjena izvršiti pomoću doslovne vrijednosti, što znači da će vrijednosti unutar stupca spol biti direktno zamijenjene. Value: | samo znači da umjesto da vrijednost bude unutar jednog reda, koristit ćemo više redaka kako bismo dovršili upit koji želimo izvršiti. Pomoću ključne riječi CASE možemo izvršiti više provjera određenih vrijednosti odjednom. U našem slučaju, prilikom procesa anonimizacije, pogledat će vrijednost unutar retka spol, te ako je 'Muško' zamijenit će sa znakom 'M', te tako i za 'Žensko'. Ako slučajno naiđe na vrijednost koja nije ni 'Muško' ni 'Žensko', dodijelit će se vrijednost 'Ostalo'.

```
18        placa:
19          type: literal
20          value: |
21            CASE
22              WHEN CAST(placa AS UNSIGNED) >= 1200 AND CAST(placa AS UNSIGNED) < 1400 THEN '1200-1400'
23              WHEN CAST(placa AS UNSIGNED) >= 1400 AND CAST(placa AS UNSIGNED) < 1600 THEN '1400-1600'
24              WHEN CAST(placa AS UNSIGNED) >= 1600 AND CAST(placa AS UNSIGNED) < 1800 THEN '1600-1800'
25              WHEN CAST(placa AS UNSIGNED) >= 1800 AND CAST(placa AS UNSIGNED) < 2000 THEN '1800-2000'
26              WHEN CAST(placa AS UNSIGNED) >= 2000 AND CAST(placa AS UNSIGNED) < 2200 THEN '2000-2200'
27              WHEN CAST(placa AS UNSIGNED) >= 2200 AND CAST(placa AS UNSIGNED) < 2400 THEN '2400-2600'
28              WHEN CAST(placa AS UNSIGNED) >= 2600 AND CAST(placa AS UNSIGNED) < 2800 THEN '2600-2800'
29              WHEN CAST(placa AS UNSIGNED) >= 2800 AND CAST(placa AS UNSIGNED) <= 3000 THEN '2800-3000'
30              ELSE 'Placa izvan raspona...'
31            END
```

Slika 13. Integracija generalizacije u stupcu placa (Izvor: Vlastita izrada)

Idući stupac koji je potrebno modificirati je placa. Unutar originalne tablice nad kojom radimo anonimizaciju, plaće su zapisane u vrlo preciznom iznosu, što je problematično jer je lakše identificirati pojedinu osobu u tablici ako znamo iznos njihove plaće. Kao i kod spola,

koristimo type: literal, te pod value imamo višerednu vrijednost. Vrijednost placa više nije zapisana s tipom podataka int, već smo tip promijenili u varchar kako bi mogli raditi raspone nad podatcima. Budući da s varcharom prirodno nije moguće raditi provjere i usporedbe, potrebno je dohvaćeni podatak pretvoriti iz niza znakova (varchar) u broj (UNSIGNED) uz pomoć kojeg možemo raditi usporedbe. Pojednostavljeno, ako je dohvaćena vrijednost plaće 1352, uzet će se prva vrijednost, te će se u polje baze upisati '1200-1400', te će se tako uspješno anonimizirati plaća zaposlenika.

```
33 scripts:
34   after:
35     - ALTER TABLE detalji_zaposlenika DROP COLUMN korisnik_id;
```

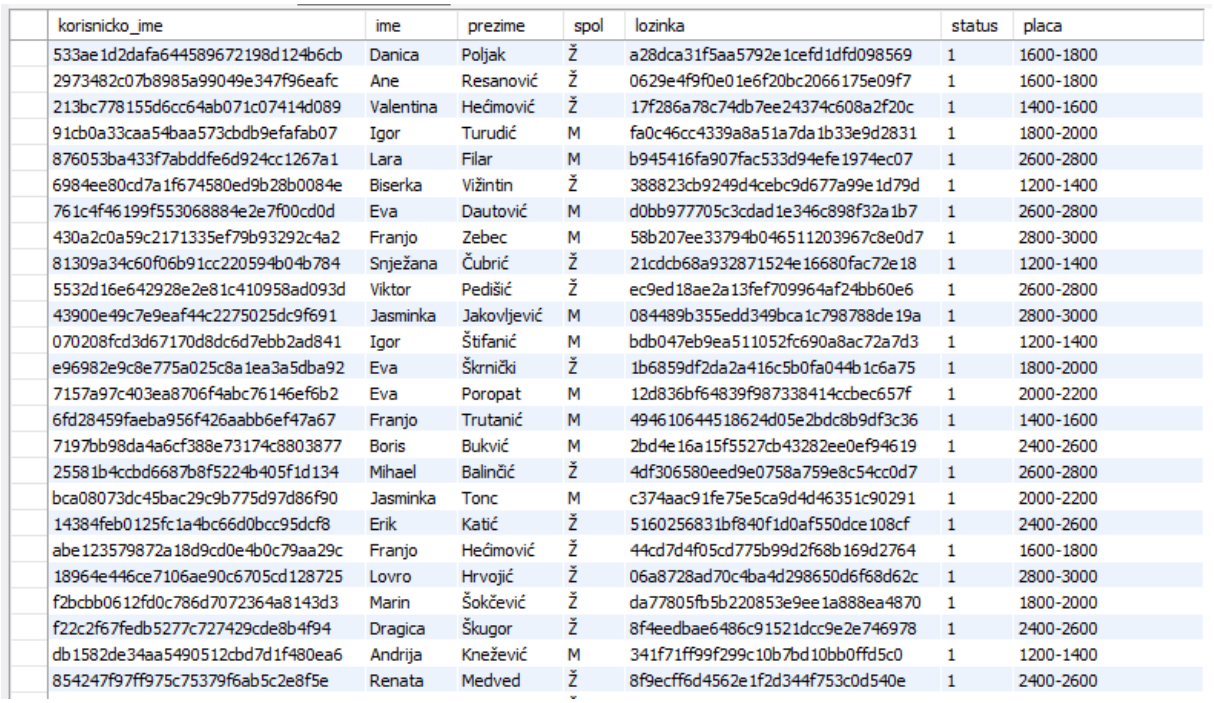
Slika 14. SQL skripta nakon anonimizacije (Izvor: Vlastita izrada)

Na kraju datoteke moguće je definirati SQL akcije koje želimo da se odrade prije procesa anonimizacije pomoću before ili poslije procesa anonimizacije s after. U našem slučaju, želimo maknuti korisnik_id iz naše konačne tablice, tako da navodimo da želimo modificirati tablicu detalji_zaposlenika na način da izbrišemo stupac korisnik_id.

```
pynonymizer --db-user root --db-password zaporkaBaze --db-name Zaposlenici -i ulaznaDat.sql -s strategija.yml -o izlaznaDat.sql
```

Slika 15. Početni argumenti pynonymizer-a (Izvor: Vlastita izrada)

Sam proces anonimizacije započinjemo pozivanjem alata pynonymizer, te navodeći vjerodajnice --db-user za korisničko ime u bazi, i --db-password za lozinku korisnika kako bi se alat mogao povezati s našim lokalnim SQL serverom te mogao obaviti proces anonimizacije. --db-name podrazumijeva ime sheme unutar koje želimo raditi. Ako ne postoji, pynonymizer će ju privremeno kreirati. Zadnji argument -i je ime ulazne datoteke, u ovom slučaju ime datoteke baze podataka prije anonimizacije, dok je -o željeno ime konačne datoteke nakon procesa anonimizacije. Konačno, -s argument je za navođenje koje strategijsku datoteku je potrebno koristiti kako bi alat znao što napraviti s dobivenom tablicom.



korisnicko_ime	ime	prezime	spol	lozinka	status	placa
533ae1d2dafa644589672198d124b6cb	Danica	Poljak	Ž	a28dca31f5aa5792e1cefd1dfd098569	1	1600-1800
2973482c07b8985a99049e347f96eafc	Ane	Resanović	Ž	0629e4f9f0e01e6f20bc2066175e09f7	1	1600-1800
213bc778155d6cc64ab071c07414d089	Valentina	Hećimović	Ž	17f286a78c74db7ee24374c608a2f20c	1	1400-1600
91cb0a33caa54baa573cbdb9efafab07	Igor	Turudić	M	fa0c46cc4339a8a51a7da1b33e9d2831	1	1800-2000
876053ba433f7abddfe6d924cc1267a1	Lara	Filar	M	b945416fa907fac533d94efe1974ec07	1	2600-2800
6984ee80cd7a1f674580ed9b28b0084e	Biserka	Vižintin	Ž	388823cb9249d4cebc9d677a99e1d79d	1	1200-1400
761c4f46199f553068884e2e7f00cd0d	Eva	Dautović	M	d0bb977705c3cdad1e346c898f32a1b7	1	2600-2800
430a2c0a59c2171335ef79b93292c4a2	Franjo	Zebec	M	58b207ee33794b046511203967c8e0d7	1	2800-3000
81309a34c60f06b91cc220594b04b784	Snježana	Čubrić	Ž	21cdcb68a932871524e16680fac72e18	1	1200-1400
5532d16e642928e2e81c410958ad093d	Viktor	Pedišić	Ž	ec9ed18ae2a13fef709964af24bb60e6	1	2600-2800
43900e49c7e9eaf44c2275025dc9f691	Jasminka	Jakovljević	M	084489b355edd349bca1c798788de19a	1	2800-3000
070208fcd3d67170d8dc6d7ebb2ad841	Igor	Štifanić	M	bdb047eb9ea511052fc690a8ac72a7d3	1	1200-1400
e96982e9c8e775a025c8a1ea3a5dba92	Eva	Škrnički	Ž	1b6859df2da2a416c5b0fa044b1c6a75	1	1800-2000
7157a97c403ea8706f4abc76146ef6b2	Eva	Poropat	M	12d836bf64839f987338414ccbec657f	1	2000-2200
6fd28459faeba956f426aabb6ef47a67	Franjo	Trutanić	M	494610644518624d05e2bdc8b9df3c36	1	1400-1600
7197bb98da4a6cf388e73174c8803877	Boris	Bukvić	M	2bd4e16a15f5527cb43282ee0ef94619	1	2400-2600
25581b4ccbd6687b8f5224b405f1d134	Mihael	Balinčić	Ž	4df306580eed9e0758a759e8c54cc0d7	1	2600-2800
bca08073dc45bac29c9b775d97d86f90	Jasminka	Tonc	M	c374aac91fe75e5ca9d4d46351c90291	1	2000-2200
14384feb0125fc1a4bc66d0bcc95dcf8	Erik	Katić	Ž	5160256831bf840f1d0af550dce108cf	1	2400-2600
abe123579872a18d9cd0e4b0c79aa29c	Franjo	Hećimović	Ž	44cd7d4f05cd775b99d2f68b169d2764	1	1600-1800
18964e446ce7106ae90c6705cd128725	Lovro	Hrvojić	Ž	06a8728ad70c4ba4d298650d6f68d62c	1	2800-3000
f2bcbb0612fd0c786d7072364a8143d3	Marin	Šokčević	Ž	da77805fb5b220853e9ee1a888ea4870	1	1800-2000
f22c2f67fedb5277c727429cde8b4f94	Dragica	Škugor	Ž	8f4eedbae6486c91521dcc9e2e746978	1	2400-2600
db1582de34aa5490512cbd7d1f480ea6	Andrija	Knežević	M	341f71ff99f299c10b7bd10bb0ffd5c0	1	1200-1400
854247f97ff975c75379f6ab5c2e8f5e	Renata	Medved	Ž	8f9ecff6d4562e1f2d344f753c0d540e	1	2400-2600

Slika 16. Rezultat anonimizacije baze podataka (Izvor: Vlastita izrada)

Krajnji rezultat je anonimizirana tablica koja je i dalje korisna za strojno učenje ili poslovnu analizu. Imena i prezimena su uspješno zamijenjena hrvatskim imenima, a korisničko ime je zamijenjeno unikatnim nasumičnim heksadecimalnim znakovima. Spol umjesto Muško ili Žensko je sada samo M ili Ž, placa stupac ima uspješno primijenjene raspone, te je stupac korisnik_id uspješno uklonjen nakon procesa anonimizacije.

4.3.2. Korištene tehnike anonimizacije

Po redu, prvo smo radili zamjenu vrijednosti unutar stupaca korisnicko_ime, ime i prezime, što predstavlja tehniku pseudonimizacije. Kroz primjenu pseudonimizacije, originalne vrijednosti zamijenjene su lažnim, ali jedinstvenim vrijednostima kako bi se očuvala konzistentnost unutar podataka, dok se istovremeno štiti identitet stvarnih korisnika. Ova tehnika omogućuje daljnju obradu podataka bez otkrivanja stvarnih osobnih informacija.

Zatim smo na stupac spol u teoriji primijenili tehniku zamjene znakova, odnosno maskiranja podataka. U ovom slučaju, izvorne vrijednosti su zamijenjene standardiziranim oznakama ('M' za mušku osobu, 'Ž' za žensku osobu ili 'O' za ostale spolove i orijentacije), što u teoriji smanjuje specifičnost podataka, no i dalje se može razumjeti originalno značenje podataka. Inače, maskiranjem se smanjuje mogućnost identifikacije pojedinaca na temelju osjetljivih podataka, te smo na primjeru spola pokazali kako bi se moglo implementirati

unutar baze podataka. No, češće bi se koristilo za maskiranje imena ili osobno identificirajućih brojeva poput broja kartice, OIB-a, JMBG-a i sl.

Tehniku generalizacije koristili smo na stupac plaća, gdje su točne vrijednosti plaća zamijenjene širim rasponima od 200. Dodatno, postoje metode kako odrediti optimalnu širinu raspona kako bi dobili ravnomjeran raspon, te pogotovo smanjili mogućnost ponovne identifikacije. Kroz generalizaciju smanjujemo detaljnost podataka, ali zadržavamo dovoljno informacija za prepoznavanje obrazaca i trendova u plaćama, pogotovo kada je uparena s drugim podacima poput spola.

Konačno, primijenili smo najosnovniju tehniku uklanjanja atributa brisanjem stupca korisnik_id, te tako osiguravamo da se potencijalno osjetljive informacije koje nisu potrebne za daljnju analizu u potpunosti uklone iz baze podataka.

4.3.3. Primjer loše anonimizacije

id	ime	prezime	dob	spol	dijagnoza	dani_od_zadnjeg_pregleda	prezime_doktora
1	Ivan	Horvat	45	Muško	Dijabetes	427	Marić
2	Ana	Kovač	34	Žensko	Hipertenzija	67	Novak
3	Marko	Marić	29	Muško	Astma	331	Perić
4	Ivana	Novak	52	Žensko	Hipertenzija	110	Perić
5	Petar	Babić	41	Muško	Dijabetes	790	Perić
6	Maja	Jurić	37	Žensko	Astma	655	Matić
7	Luka	Matić	28	Muško	Hipertenzija	34	Kovač
8	Sara	Perić	33	Žensko	Dijabetes	171	Vuković
9	Tomislav	Vuković	50	Muško	Astma	384	Božić
10	Marija	Božić	46	Žensko	Hipertenzija	960	Božić
11	Josip	Knežević	39	Muško	Dijabetes	701	Božić
12	Ivana	Šimić	31	Žensko	Astma	96	Radić
13	Ante	Krnić	55	Muško	Hipertenzija	425	Lovrić
14	Katarina	Radić	42	Žensko	Dijabetes	114	Barišić
15	Filip	Lovrić	36	Muško	Astma	830	Pavlović
16	Martina	Barišić	47	Žensko	Hipertenzija	742	Klarić
17	Nikola	Pavlović	38	Muško	Dijabetes	364	Klarić
18	Tina	Klarić	32	Žensko	Astma	741	Klarić
19	Davor	Petrović	49	Muško	Hipertenzija	889	Marić
20	Ana	Grgić	40	Žensko	Dijabetes	135	Novak

```yaml
locale: hr_HR

tables:
  zdravstveni_podaci:
    type: update_columns
    columns:
      ime: first_name
      prezime: last_name
      dob:
        type: literal
        value: 0
```

Slika 17. Prvi primjer loše anonimizacije, baza i strategijska datoteka(Izvor: Vlastita izrada)

Uzmimo za primjer bazu podataka s osnovnim informacijama o pacijentima neke bolnice poput imena, prezimena, dobi, dijagnoze i sl. Kako bismo prikazali primjere loše pseudonimizacije, kreirat ćemo novu strategijsku datoteku. Također ćemo koristiti biblioteku Faker i atribut locale postavljen na hr_HR tako da zamjenjujemo prava imena i prezimena s mogućim hrvatskim alternativama. Strategijsku datoteku postavljamo na način ažuriranja stupaca podataka u tablici uz atribut type i vrijednost update_columns. Želimo ažurirati ime i prezime vrijednosti te zamijeniti ih vrijednostima iz Faker biblioteke tako da su imena i dalje realistična, no anonimizirana. Također vrijednost dob svih redaka unutar tablice smo postavili

na 0 kako bi dodatno anonimizirali podatke osobe unutar tablice. Sada kada smo anonimizirali ime, prezime i dob, jesu li osobe sigurno zaštićene? Nažalost, nisu, jer zlonamjerna osoba i dalje može donijeti određene zaključke u slučaju krađe podataka baze podataka.

id	ime	prezime	dob	spol	dijagnoza	dani_od_zadnjeg_pregleda	prezime_doktora
1	Lovre	Mušćet	0	Muško	Dijabetes	427	Marić
2	Niko	Vorkapić	0	Žensko	Hipertenzija	67	Novak
3	Mirko	Raljević	0	Muško	Astma	331	Perić
4	Ane	Mirosavljević	0	Žensko	Hipertenzija	110	Perić
5	Jakov	Andrašek	0	Muško	Dijabetes	790	Perić
6	Joso	Lenić	0	Žensko	Astma	655	Matić
7	Barica	Dolić	0	Muško	Hipertenzija	34	Kovač
8	Mateja	Štifanić	0	Žensko	Dijabetes	171	Vuković
9	Jana	Čuček	0	Muško	Astma	384	Božić
10	Renata	Bačić	0	Žensko	Hipertenzija	960	Božić
11	Danijela	Špralja	0	Muško	Dijabetes	701	Božić
12	Pavao	Maržić	0	Žensko	Astma	96	Radić
13	Terezija	Petrić	0	Muško	Hipertenzija	425	Lovrić
14	Toni	Sokić	0	Žensko	Dijabetes	114	Barišić
15	Josip	Sardelić	0	Muško	Astma	830	Pavlović
16	Nika	Guberović	0	Žensko	Hipertenzija	742	Klarić
17	Denis	Volarić	0	Muško	Dijabetes	364	Klarić
18	Barica	Miličić	0	Žensko	Astma	741	Klarić
19	Joško	Vuković	0	Muško	Hipertenzija	889	Marić
20	Nina	Debelić	0	Žensko	Dijabetes	135	Novak

Slika 18. Rezultat prve loše anonimizacije (Izvor: Vlastita izrada)

Jer već jednostavnim upitom, gdje znamo samo dvije informacije: da je osoba žensko i da je prezime njenog doktora Perić, možemo saznati dodatne informacije o osobi. U ovom slučaju smanjili smo mogućnosti od n broja ljudi u bazi podataka na samo dva potencijalna unosa, te time zlonamjeran korisnik može saznati dijagnozu danu osobi i koliko je dana prošlo od zadnjeg pregleda u bolnici, podatke koji mogu biti vrijedni ako zlonamjerna osoba dođe do njih.

```
1 • SELECT * FROM baza.zdravstveni_podaci WHERE prezime_doktora = 'Perić' AND spol = 'Žensko';
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

id	ime	prezime	dob	spol	dijagnoza	dani_od_zadnjeg_pregleda	prezime_doktora
4	Ane	Mirosavljević	0	Žensko	Hipertenzija	110	Perić
40	Željka	Zanoški	0	Žensko	Dijabetes	321	Perić
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Slika 19. Primjer upita nad prvim lošim primjerom anonimizacije (Izvor: Vlastita izrada)

U idućem primjeru dodali smo stupac aktivno_dopunsko koji definira ima li određena osoba uplaćeno dopunsko osiguranje.

```yaml
locale: hr_HR

tables:
  zdravstveni_podaci:
    type: update_columns
    columns:
      ime: first_name
      prezime: last_name
      spol:
        type: literal
        value: spol = '0'
      prezime_doktora: last_name
      dani_od_zadnjeg_pregleda:
        type: literal
        value: 0
```

Slika 20. Strategijska datoteka za drugi primjer loše anonimizacije (Izvor: Vlastita izrada)

Unutar strategijske datoteke i dalje smo anonimizirali stupce imena i prezimena zamjenskim imenima iz biblioteke Faker, no dob smo ostavili čitavom. Također smo anonimizirali spol, prezime doktora i dane od zadnjeg pregleda. No to i dalje nije dovoljno kako bi potpuno zaštitili naše korisnike. Entiteti poput osiguravajuće kuće, ako uspiju doći do 'anonimiziranih' podataka iz bolnice, mogu uz upite nad bazom doći do informacija. U ovom primjeru napravili smo upit nad bazom koje sve mlađe osobe ispod 40 godina nemaju aktivno dopunsko osiguranje i imaju dijagnozu astme. Na temelju dobivenih informacija osiguranje i dalje može donijeti vlastite zaključke ili pokušati ponovno identificirati korisnika, pogotovo ako znaju točnu dob korisnika o kojem žele znati više.

```sql
SELECT * FROM blabla.zdravstveni_podaci WHERE aktivno_dopunsko = 0 AND dijagnoza = 'Astma'AND dob <= 40;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

id	ime	prezime	dob	spol	dijagnoza	dani_od_zadnjeg_pregleda	prezime_doktora	aktivno_dopunsko
3	Dora	Šoštarić	29	0	Astma	0	Tudić	0
15	Hana	Polović	36	0	Astma	0	Šantić	0
33	Vedran	Hranić	28	0	Astma	0	Ribičić	0
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Slika 21. Primjer upita nad drugim lošim primjerom anonimizacije (Izvor: Vlastita izrada)

5. Zaključak

Generativna umjetna inteligencija, kao i umjetna inteligencija općenito, zadnjih godina postali su ključni koncepti u suvremenom računarstvu i mnogim industrijama. Podjednako je važno razumjeti prednosti i potencijal korištenja generativnih alata, ali i sigurnosne prepreke i opasnosti koje dolaze s njima. Kako bismo uspješno implementirali umjetnu inteligenciju unutar vlastitog poduzeća, ključno je demistificirati pojmove povezane s njom te shvatiti međusobni odnos između njih. Prije nego se generativni alati uvedu u poslovne procese poduzeća, potrebno je biti svjestan rizika i izazova koje donose, kao i sigurnosnih i pravnih implikacija. Pažljivom i detaljnom implementacijom generativnih tehnologija moguće je smanjiti površinu napada zlonamjernih korisnika, što omogućuje jaču obranu od potencijalnih prijetnji.

Podatci su neizostavan element u treniranju modela generativne umjetne inteligencije. Zbog toga, način na koji poduzeća prikupljaju, pohranjuju i obrađuju podatke od ključne je važnosti za uspjeh implementacije. Posebno je važno osigurati ispravnu anonimizaciju podataka kako bi se zaštitila privatnost korisnika i zadovoljili regulatorni zahtjevi koji se još razvijaju i prilagođavaju novim tehnologijama. Svjetske industrije se različito suočavaju s izazovima generativne umjetne inteligencije, s posebnim naglaskom na zaštitu autorskih prava u kreativnim industrijama, gdje je potrebno zaštiti se od krađe podataka u svrhu treniranja budućih modela.

Kroz praktični dio rada, korištenjem alata Pynonymizer, prikazali smo kako se može provesti temeljita anonimizacija baze podataka putem strategijske datoteke koja precizno definira način pristupa podacima, kao i dobre i loše primjere anonimizacije baze podataka. Time je omogućeno sigurno korištenje podataka u raznim primjenama, bilo za treniranje generativnih modela, poslovnu analitiku ili provođenje istraživanja na anonimnim podacima.

Zaključno, umjetna inteligencija nastavit će se razvijati ubrzanim tempom, donoseći nove mogućnosti i izazove. Pravovremenom pripremom, posebno kroz adekvatno upravljanje podacima i zaštitom privatnosti, organizacije mogu maksimalno iskoristiti potencijal generativne umjetne inteligencije, dok istovremeno minimiziraju rizike koje ona nosi.

Popis literature

[1] Where does GenAI fit within the AI landscape, [Na internetu] Dostupno na: https://communities.sas.com/t5/SAS-Communities-Library/Where-does-GenAI-fit-within-the-AI-landscape/ta-p/9150000 [Pristupljeno 26-srp-2024]

[2] Artificial intelligence (AI) vs. machine learning (ML), [Na internetu] Dostupno na: https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning [Pristupljeno 27-srp-2024]

[3] What is Supervised Learning?, [Na internetu] Dostupno na: https://cloud.google.com/discover/what-is-supervised-learning [Pristupljeno 30-srp-2024]

[4] What is Reinforcement Learning?, [Na internetu] Dostupno na: https://aws.amazon.com/what-is/reinforcement-learning/ [Pristupljeno 30-srp-2024]

[5] What's the difference between deep learning, machine learning, and artificial intelligence?, [Na internetu] Dostupno na: https://cloud.google.com/discover/deep-learning-vs-machine-learning [Pristupljeno 02-kol-2024]

[6] What is Generative AI?, [Na internetu] Dostupno na: https://www.nvidia.com/en-us/glossary/generative-ai/ [Pristupljeno 06-kol-2024]

[7] Generative AI Tips: Use High-Quality Data, [Na internetu] Dostupno na: https://www.linkedin.com/pulse/generative-ai-tips-use-high-quality-data-rick-spair-hnrbe/ [Pristupljeno 06-kol-2024]

[8] What are Foundation Models?, [Na internetu] Dostupno na: https://aws.amazon.com/what-is/foundation-models/ [Pristupljeno 09-kol-2024]

[9] Garbage In, Garbage Out: Why Third-Party Data Sources Matter When Using Generative AI, [Na internetu] Dostupno na: https://www.lexisnexis.com/blogs/ae/b/data-as-a-service/posts/third-party-data-sources-when-using-generative-ai [Pristupljeno 07-ruj-2024]

[10] EU AI Act: first regulation on artificial intelligence, [Na internetu] Dostupno na: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
[Pristupljeno 07-ruj-2024]

[11] I Asked a Lawyer About AI Regulation in the U.S., [Na internetu] Dostupno na: https://blog.hubspot.com/ai/ai-regulation-in-us [Pristupljeno 07-ruj-2024]

[12] 5 Must-Haves for a Successful GenAI Implementation, [Na internetu] Dostupno na: https://wwa.wavestone.com/en/insight/5-must-haves-for-a-successful-gen-ai-implementation/ [Pristupljeno 07-ruj-2024]
[13] Here's why quantization matters for AI, [Na internetu] Dostupno na: https://www.qualcomm.com/news/onq/2019/03/heres-why-quantization-matters-ai [Pristupljeno 08-ruj-2024]
[14] New AI Usage Data Shows Who's Using AI — and Uncovers a Population of 'Super-Users', [Na internetu] Dostupno na: https://www.salesforce.com/news/press-releases/2023/09/07/ai-usage-research/ [Pristupljeno 08-ruj-2024]
[15] What is a prompt injection attack?, [Na internetu] Dostupno na: https://www.ibm.com/topics/prompt-injection [Pristupljeno 09-ruj-2024]
[16] Prompt Leak, [Na internetu] Dostupno na: https://www.prompt.security/vulnerabilities/prompt-leak [Pristupljeno 09-ruj-2024]
[17] Samsung Software Engineers Busted for Pasting Proprietary Code Into ChatGPT [Na internetu] Dostupno na: https://www.pcmag.com/news/samsung-software-engineers-busted-for-pasting-proprietary-code-into-chatgpt [Pristupljeno 09-ruj-2024]
[18] Chris Bakke Twitter račun, [Na internetu] Dostupno na https://x.com/ChrisJBakke/status/1736533308849443121: [Pristupljeno 09-ruj-2024]
[19] DPD error caused chatbot to swear at customer [Na internetu] Dostupno na: https://www.bbc.com/news/technology-68025677 [Pristupljeno 10-ruj-2024]
[20] Prompt Injection 101, [Na internetu] Dostupno na: https://www.prompt.security/blog/prompt-injection-101 [Pristupljeno 10-ruj-2024]
[21] Shadow AI and How To Avoid It, [Na internetu] Dostupno na: https://www.pipefy.com/blog/shadow-ai/ [Pristupljeno 10-ruj-2024]
[22] What are tokens and how to count them?, [Na internetu] Dostupno na: https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them [Pristupljeno 10-ruj-2024]
[23] Amount of Data Created Daily (2024), [Na internetu] Dostupno na: https://explodingtopics.com/blog/data-generated-per-day [Pristupljeno 13-kol-2024]
[24] What is Big Data?, [Na internetu] Dostupno na: https://cloud.google.com/learn/what-is-big-data [Pristupljeno 13-kol-2024]
[25] The Role of AI and Machine Learning in Data Lakes and Warehouses, [Na internetu] Dostupno na: https://www.bayrocklabs.com/post/the-role-of-ai-and-machine-learning-in-data-lakes-and-warehouses [Pristupljeno 14-kol-2024]

[26] How Investing in a Data Lakehouse Can Make or Break AI Projects, [Na internetu] Dostupno na: https://lingarogroup.com/blog/how-investing-in-a-data-lakehouse-can-make-or-break-ai-projects [Pristupljeno 14-kol-2024]

[27] Data anonymization, [Na internetu] Dostupno na: https://en.wikipedia.org/wiki/Data_anonymization [Pristupljeno 15-kol-2024]

[28] Data Anonymisation, [Na internetu] Dostupno na: https://libguides.ntu.edu.sg/c.php?g=927336&p=6698844 [Pristupljeno 15-kol-2024]

[29] GUIDE TO BASIC DATA ANONYMISATION TECHNIQUES, [Na internetu] Dostupno na: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf [Pristupljeno 18-kol-2024]

[30] Data Anonymization Techniques, [Na internetu] Dostupno na: https://www.k2view.com/what-is-data-anonymization/#Data-Anonymization-Techniques [Pristupljeno 18-kol-2024]

[31] Other Privacy Definitions: l-diversity and t-closeness, [Na internetu] Dostupno na: https://personal.utdallas.edu/~muratk/courses/dbsec09s_files/DBSec_priv3.pdf [Pristupljeno 19-kol-2024]

[32] EU GDPR "Definitions", [Na internetu] Dostupno na: https://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm [Pristupljeno 20-kol-2024]

[33] Chapter 3: pseudonymisation, [Na internetu] Dostupno na: https://ico.org.uk/media/about-the-ico/consultations/4019579/chapter-3-anonymisation-guidance.pdf [Pristupljeno 20-kol-2024]

[34] What Is Data Centralization?, [Na internetu] Dostupno na: https://www.teradata.com/insights/data-platform/what-is-data-centralization [Pristupljeno]

[35] Nvidia Scrapes YouTube, Eyes Netflix, Discovery to Train New Video Model, [Na internetu] Dostupno na: https://www.proofnews.org/nvidia-scrapes-youtube-eyes-netflix-discovery-to-train-new-video-model/ [Pristupljeno 24-kol-2024]

[36] Google cut a deal with Reddit for AI training data, [Na internetu] Dostupno na: https://www.theverge.com/2024/2/22/24080165/google-reddit-ai-training-data [Pristupljeno 24-kol-2024]

[37] Google’s Deal With Stack Overflow Is the Latest Proof That AI Giants Will Pay for Data, [Na internetu] Dostupno na: https://www.wired.com/story/google-deal-stackoverflow-ai-giants-pay-for-data/ [Pristupljeno 24-kol-2024]

[38] Exclusive: Time strikes licensing deal with OpenAI, [Na internetu] Dostupno na: https://www.axios.com/2024/06/27/openai-time-licensing-deal-chatgpt [Pristupljeno 24-kol-2024]

[39] About The Glaze Project, [Na internetu] Dostupno na: https://glaze.cs.uchicago.edu/aboutus.html [Pristupljeno 26-kol-2024]

[40] What Is Glaze?, [Na internetu] Dostupno na: https://glaze.cs.uchicago.edu/what-is-glaze.html [Pristupljeno 26-kol-2024]

[41] What Is Nightshade?, [Na internetu] Dostupno na: https://nightshade.cs.uchicago.edu/whatis.html [Pristupljeno 26-kol-2024]

[42] AI poisoning tool Nightshade received 250,000 downloads in 5 days: 'beyond anything we imagined', [Na internetu] Dostupno na: https://venturebeat.com/ai/ai-poisoning-tool-Nightshadereceived-250000-downloads-in-5-days-beyond-anything-we-imagined/ [Pristupljeno 26-kol-2024]

[43] Marques, Joana Ferreira; Bernardino, Jorge; Analysis of Data Anonymization Techniques [Na internetu] Dostupno na: https://pdfs.semanticscholar.org/015d/d496535d38d3dff934a7a10441ce49cd5cff.pdf [Pristupljeno]

[44] S. Nelson, Gregory [Na internetu] Dostupno na: https://www.lexjansen.com/pharmasug/2016/IB/PharmaSUG-2016-IB06.pdf [Pristupljeno 10-ruj-2024]

[45] pynonymizer, [Na internetu] Dostupno na: https://pypi.org/project/pynonymizer/ [Pristupljeno 22-kol-2024]

[46] pynonimizer dokumentacija, Strategyfiles [Na internetu] Dostupno na: https://github.com/rwnx/pynonymizer/blob/main/doc/strategyfiles.md [Pristupljeno 23-kol-2024]

Popis slika