Varovanje občutljivih podatkov fizioloških merjenj uporabnikov pametnih storitev

Samo Jean[1], **Marko Kalan**[2], **Andrej Košir**[1]

[1]*Fakulteta za elektrotehniko, Univerza v Ljubljani,* [2]*Svetovalni center za otroke, mladostnike in starše v Ljubljani*
E-pošta: andrej.kosir@fe.uni-lj.si

Protection of sensitive data from physiological measurements of smart service users

In this paper, we introduce and discuss the importance of data protection in our psychological measurements of users. We will address the challenges associated with this type of data collection and explore the legal requirements imposed on data collectors during the execution of such measurements. Furthermore, we will present the legislation that restricts the actions of data collectors and safeguards users against the misuse of their personal and sensitive data. We will examine the current state of technology that enables secure and restricted data collection.

Additionally, we will showcase how these requirements were met through the example of a program designed to collect sensor data from students exposed to challenging mathematical tasks, along with the architecture that ensures safe data collection. We will also demonstrate the necessary data processing steps undertaken to derive meaningful insights while maintaining respect for data holders' privacy. Finally, we will address the challenges encountered throughout this process.

1 Uvod

Problem nadzora nad zbiranjem in hranjenjem podatkov se je znatno povečal z začetkom uporabe spleta s strani uporabnikov leta 1995. Zaradi obsežne razširjenosti in hitrega dostopa do interneta se je povečala količina osebnih podatkov, ki jih podjetja pridobivajo od svojih uporabnikov. Zato je istega leta začela veljati Direktiva o varstvu podatkov DAD [1], ki določa, kako morajo podjetja izvajati zbiranje in hranjenje podatkov ter kdo nosi odgovornost za to.

Skupaj s hitrim razvojem tehnologije se je pojavila tudi potreba po posodobitvi in prilagoditvi teh pravil v skladu s potrebami sodobnih uporabnikov. V ta namen je Evropska unija leta 2018 sprejela Splošno uredbo o varstvu podatkov (GDPR) [2].

Splošna uredba o varstvu podatkov nadgradi Direktivo o varstvu podatkov ter uporabnikom podeli več pravic pri odločanju, kaj se dogaja z njihovimi osebnimi podatki. Prve spremembe v Splošni uredbi se pojavijo tudi pri definiciji osebnih podatkov, saj vključujejo podatke, kot so IP-naslovi, biometrični podatki in drugi. Ta dodatek je sicer otežil delo podjetjem, ki poskušajo pridobiti celovite vpoglede v potrošniške navade svojih uporabnikov, saj uporaba teh podatkov ni dovoljena brez posebne odobritve uporabnika. Med drugim se je poostrilo tudi kaznovanje kršitev, odgovornost pa je postala bolj razpršena, saj se nanaša tudi na podjetja, ki ne le hranijo ali zbirajo podatke, temveč z njimi tudi ravna.

Psihofiziološka merjenja prikazujejo odvisnost med fiziološkimi odzivi ter psihološkimi dejavniki in stanji merjene osebe. Ker spremembe v okolju močno vplivajo na fiziološke dejavnike, je za natančna merjenja potrebno ohranjati okolje čim bolj nespremenjeno. To pomeni, da je najbolj primerno izvajati meritve v nadzorovanem okolju, kjer se izognejo večjim okoljskim spremembam. Merjenje nam omogoča, da se glede na spremembe fizioloških količin kot so elektrodermalna aktivnost, srčni utrip in usmerjenost pogleda, oceni izbrana stanje, kot je nelagodje učenca med obravnavo. Nadalje se to znanje uporabi s strani pedagoga za lažje prilagoditve učnega postopka in s strani učenca za pomoč pri samo-obvladovanju nelagodja.

V okviru projekta ARRS CRP smo razvili aplikacijo za merjenje matematične anksioznosti učencev pri reševanju matematičnih nalog. Aplikacija je namenjena za uporabo s strani psihologa med obravnavo učencev, ki izkazujejo resnejše težave v obliki anksioznosti pri reševanju matematičnih nalog. Aplikacija za delovanje ne potrebuje veliko vnosov s strani psihologa, saj mora ta med merjenjem vnašati samo faze obravnave in posredovati učenčevo interpretacijo svojega počutja. Poleg tega v skladu z aplikacijo delujejo tudi trije psiho-fiziološki senzorji. Dva sta postavljena pred otroka in zbirata podatke s pomočjo optičnih senzorjev (kamera s sledenjem skeleta in sledilnik oči), zadnji pa je nameščen otroku na roko (zapestnica za merjenje fizioloških signalov). Merilne naprave med obravnavo nabirajo podatke o elektrodermalni aktivnosti, srčnemu utripu, temperaturi kože, hitrosti in smeri gibanja roke, pretoku krvi in usmeritvi pogleda, ter razširjenosti zenice in položaju oči. Vse naštete količine spadajo med psiho-fiziološke signale.

V članku bomo podali pomen zaščite podatkov pri fizioloških merjenjih uporabnikov na primeru modeliranja matematične anksioznosti. Obravnavali bomo izzive, povezane s tovrstnim zbiranjem podatkov, ter preučili zakonske zahteve, ki so naložene zbiralcem podatkov med izvajanjem takšnih meritev. Poleg tega bomo predsta-

vili zakonodajo, ki omejuje dejanja zbirateljev podatkov in ščiti uporabnike pred zlorabo njihovih osebnih in občutljivih podatkov.

Predstavili bomo tudi trenutno stanje tehnologije, ki omogočajo varno in omejeno zbiranje podatkov. Poleg tega bomo predstavili, kako so te zahteve izpolnjene na primeru aplikacije, zasnovane za zbiranje fizioloških meritev učencev, izpostavljenih matematičnim nalogam, skupaj z arhitekturo, ki zagotavlja varno zbiranje podatkov. Poleg tega bomo prikazali potrebne korake obdelave podatkov, ki so bili izvedeni za pridobivanje smiselnih vpogledov v matematično anksioznost, hkrati pa so spoštovali zasebnost imetnikov podatkov. Na koncu bomo obravnavali izzive, s katerimi smo se srečali med tem postopkom.

2 Vrste občutljivih podatkov in pravne regulacije

2.1 Vrste občutljivih podatkov

Za boljše razumevanje pravil o nadzoru osebnih podatkov, je najprej potrebno predstaviti definicijo osebnih podatkov ter kako jih razvrščamo. Osebni podatek je vsaka informacija, ki se nanaša na določenega posameznika. Osebni podatek so tudi različni podatki, ki skupaj zbrani lahko vodijo do identifikacije določene osebe. Osebni podatki, ki so bili deidentificirani, šifrirani ali psevdonimizirani, vendar jih je mogoče uporabiti za ponovno identifikacijo osebe, ostanejo osebni podatki in spadajo v področje uporabe GDPR.

Podatki, iz katerih se ne more razbrati identitete posameznika, ne spadajo več med osebne podatke, ampak med anonimne podatke. Da pa dobimo prave anonimne podatke, morajo biti ti nepovratno anonimizirani. Direktiva GDPR varuje osebne podatke ne glede na tehnologijo, uporabljeno za obdelavo teh podatkov – je tehnološko nevtralna in velja tako za avtomatizirano kot tudi ročno obdelavo. Določa tudi, ali so podatki organizirani v skladu z vnaprej določenimi merili (na primer po abecednem vrstnem redu). Prav tako je pomembno, kako so podatki hranjeni – v informacijskem sistemu, prek videonadzora ali na papirju. Ne glede na primer za osebne podatke veljajo zahteve glede varstva, določene v direktivi GDPR

Občutljivi podatki spadajo med posebno vrsto osebnih podatkov, za katere je obdelava prepovedana brez posebnega dovoljenja uporabnikov. Med njih spadajo:

- Podatki ki razkrivajo rasno ali etično poreklo.
- Politično mnenje.
- Versko ali filozofsko prepričanje.
- Članstvo v sindikatu.
- Obdelava genetskih podatkov.
- Obdelava biometričnih podatkov za namene edinstvene identifikacije posameznika.
- Podatkov v zvezi z zdravjem.
- Podatkov s posameznikovo spolno usmerjenostjo ali spolnim življenjem.

Direktiva GDPR opredeljuje biometrične podatke kot posebno kategorijo občutljivih osebnih podatkov, ki izhajajo iz specifične tehnične obdelave, povezane s fizičnimi, fiziološkimi ali vedenjskimi lastnostmi fizične osebe. Biometrični podatki, kot so podatki obraza, šarenice ali prstnih odtisov, potrjujejo nedvoumno identifikacijo osebe. Ta definicija torej privzema, da je za identifikacijo potrebna tehnična obdelava. Na primer, fotografija ne spada v kategorijo biometričnih podatkov, tudi če se uporablja za identifikacijo. Po drugi strani pa se izračun obraznega odtisa na podlagi fotografije šteje za biometrični podatek, saj zahteva tehnično obdelavo. Biometrični podatki so po definiciji posebna kategorija občutljivih podatkov. Razlog za to je, da so biometrični podatki edinstveni in jih ni mogoče zamenjati, v nasprotju z gesli. Zato je potrebna dodatna zaščita podatkov.

2.2 Izbrane relevantne tehologije

Zahteve direktive GDPR so prinesle pomembne spremembe tudi na področju razvoja tehnologije. Ena od ključnih zahtev GDPR-ja je pravica uporabnika, da zahteva vpogled v podatke, ki jih o njem podjetje hrani, ter v primeru neskladnosti ali nezadovoljstva zahteva izbris teh podatkov. To pomeni, da so lahko večja podjetja deležna večjega števila takšnih zahtev v istem časovnem obdobju, zato morajo imeti sistemi in infrastruktura možnost obdelati in odgovoriti na te zahteve.

Poleg nadgradnje sistemov morajo podjetja prepoznati osebne podatke, ki se nanašajo na posameznega uporabnika, ter vzpostaviti sistematičen način komunikacije in reševanja zahtev. Pomembno je tudi imeti učinkovit iskalnik za pregledovanje podatkovnih baz, da se lahko želeni podatki dovolj hitro najdejo.

Poleg tega zahteve uporabnikov glede shranjevanja in še posebej brisanja osebnih podatkov predstavljalo tudi oviro za uporabo tehnologij, kot je strojno učenje. To je posledica dejstva, da nimamo popolnega nadzora nad postopkom učenja strojnih modelov in je optimizacija učenja pogosto bolj intuitivna kot natančno načrtovan postopek. Zahteva po izbrisu podatkov iz začetnega nabora vhodnih podatkov bi lahko v veliki meri uničila delovanje modela strojnega učenja. Enako velja za tehnologije, kot je veriženje blokov (blockchain), kjer bi zahteva po izbrisu podatkov pomenila, da bi moralo podjetje izbrisati pravilne podatke na vsakem členu verige.

Spremembe, ki jih je prinesel GDPR, so torej zahtevale nadgradnje sistemov, večjo pozornost pri ravnanju s podatki in implementacijo boljših iskalnikov. Hkrati pa so postavile izzive glede varovanja osebnih podatkov pri uporabi tehnologij, kot je strojno učenje in veriženje blokov.

2.3 Pravne regulacije

Pred začetkom veljavnosti Splošne uredbe o varstvu podatkov (GDPR, [2]) 25. maja 2018 je kot svetovalna skupina za Direktivo o varstvu podatkov delovala skupina

nacionalnih regulatorjev. Skupaj z novo uredbo se je skupina nadgradila in nastal je Evropski odbor za varstvo podatkov (European Data Protection Board - EDPB) [4].

GDPR uveljavljajo in nadzorujejo nadzorni organi (Supervisory Authorities - SA) individualno znotraj vsake države. V primeru, da naletijo na mejni primer ali pride do spora pri interpretaciji zakona, se problem naslovi na krovni nadzorni organ (Lead Supervisory Authority - LSA). Če se tudi po sklepu nadzornega organa obe strani ne strinjata s sklepom, je problem napoten na EDPB.

Naloge EDPB ujejo:

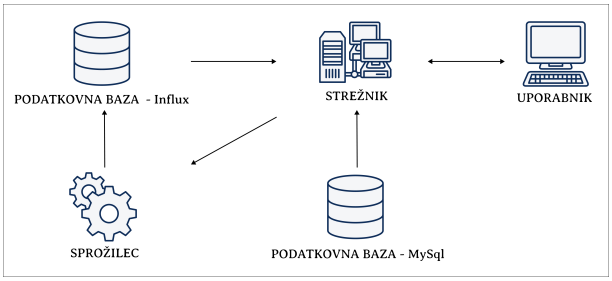
- Nadzor in uveljavljanje GDPR zahtev na področju Evropske Unije.
- Pomoč pri reševanju mejnih primerov, kjer nadzorni organi sami ne morejo rešiti spora.
- Določanje splošnih smernic za pojasnitev in širjenje skupnega razumevanja zakonodaje.
- Odgovarjanje na vprašanja in sprejemanje mnenj, ki so naslovljeni na Evropsko komisijo ali nacionalni nadzorni organ.
- Zagotavljanje doslednosti dejavnosti nacionalnih organov in izdajanje skupnih mnenj z Evropskim nadzornim organom za varstvo podatkov (European Data Protection Supervisor - EDPS). Z odobritvijo organov lahko sledi sprejem zavezujoče odločitve.
- Spodbujanje in podpiranje sodelovanja med nacionalnimi nadzornimi organi, no z izvrševanjem.
- Zagotavljanje sekretariata za Koordinirani nadzorni odbor (Consistency and Cooperation - CSC).
- Prizadevanje v koordiniran nadzor večjih IT sistemov in EU teles.

3 Primer rešitve varovanja občutljivih podatkov

V tem poglavju podajamo rešitev varovanja podatkov za projekt merjenja matematične anksioznosti, pri kateri smo upoštevali zahteve regulacij za varovanje podatkov ter njihove posledice [3]. Projekt in študijo je odobrila Komisija Univerze v Ljubljani za Etiko v Raziskavah, ki uje Delo z Ljudmi (KERL UL).

3.1 Projekt merjenja matematične anksioznosti

Matematična anksioznost je definirana kot občutek napetosti in strahu, ki vpliva na zmožnost učenca pri reševanju matematičnih nalog. To lahko vpliva na učenčevo željo po izpostavljenosti matematičnim nalogam in posledično pusti dolgoročne posledice pri njegovem izobraževanju in poklicu pozneje v življenju. V okviru izdelana aplikacija je bila načrtovana za uporabo terapevta med obravnavo z učencem, ki izkazuje resnejše težave pri reševanju matematičnih nalog. Aplikacija za delovanje potrebuje minimalno interakcije s strani terapevta, saj mora ta med merjenjem vnašati samo faze obravnave in preko enega klika posredovati učenčevo interpretacijo svojega počutja. Aplikacija zbira podatke fizioloških senzorjev. Dva sta postavljena pred otroka in zbirata podatke s pomočjo optičnih senzorjev, zadnji pa je nameščen otroku na roko. Merilne naprave med obravnavo merijo podatke o elektro-dermalni aktivnosti, srčnemu utripu, temperaturi kože, hitrosti in smeri gibanja roke, pretoku krvi in usmeritvi pogleda, ter razširjenosti zenice in položaju oči. Vse naštete količine spadajo med psihofiziološke podatke.



Slika 1: Arhitektura aplikacije

3.2 Arhitektura zajema podatkov

Na sliki 1 je prikazana arhitektura aplikacije za merjenje matematične anksioznosti.

Aplikacija je zasnovana na tehnologiji Node.js in programskim jezikom JavaScript ter sledi Model - View - Controller (MVC) vzorcu. MVC vzorec pomeni, da je aplikacija razdeljena na tri dele: *angl. Model*, ki predstavlja podatke, *angl. View*, ki predstavlja prikaz, ter *angl. Controller*, ki vzpostavlja povezavo med njima. Grafični del aplikacije je narejen s pomočjo *render engine Handlebars*, kar pomeni da, se stran sestavi v zalednem sistemu aplikacije. Programi za sprožitev in delovanje zunanjih senzorjev so napisani v programskem jeziku Python. Glavna Node.js aplikacija požene tri Python skripte, ki poganjajo programe za sprejemanje podatkov v neskončni zanki. Aplikacija je povezana z dvema ločenima podatkovnima bazama, kjer ena skrbi za shrambo uporabniških podatkov, druga pa skrbi za shranjevanje merilnih podatkov senzorjev. Za hrambo uporabniških podatkov je uporabljena relacijska podatkovna baza MySql. Za hrambo merjenjih količin iz senzorjev je uporabljena časovna podatkovna baza InfluDB, ki je prilagojena hitremu prenosu podatkov senzorjev.

3.3 Senzorji in shranjevanje v realnem času

Pri projektu merjenja matematične anksioznosti smo uporabili tri senzorje za zbiranje podatkov v realnem času: Tobii Eye Tracker Spark, Empatica Embrace Plus in kamero Lumix Oak-D. Empatica Embrace Plus je zapestnica, opremljena s senzorji za merjenje elektrodermalne aktivnosti, pospeškomerom, žiroskopom, termometrom za kožo, fotopletizmogramom, senzorjem za merjenje srčnega utripa in intervalov med utripi srca ter senzorji za zaznavanje korakov, ur počitka ter preverjanja, ali je naprava pravilno nameščena na koži.

S pomočjo teh senzorjev ima zapestnica sposobnost

zaznavanja indikatorjev več različnih bolezni, med katerimi so napadi epilepsije, kapi, sladkorne bolezni, depresije in drugih. Med zbranimi količinami iz senzorja Empatice ima za določanje matematične anksioznosti poseben pomen signal merjenja elektrodermalne aktivnosti. Tobii Eye Tracker Spark je senzor, ki ga namestimo pod površino, na katero se osredotočamo s pogledom. S tem senzorjem zbiramo podatke o pogledu osebe ter določamo lokacijo posameznih delov obraza, kot so oči, nos in ušesa. Poleg tega senzor zaznava tudi velikost zenice, kar je pomemben pokazatelj različnih stanj merjenje osebe (kognitivna obremenitev, osredotočenost na trenutno aktivnost). Tretji senzor je kamera Luxonis OAK-D. uje dva optična senzorja, ki zbirata podatke v sivinskih odtenkih za boljšo predstavo globine, ter eno barvno kamero za zbiranje barvnih informacij. S tem senzorjem smo zajemali podatke o lokaciji oči, ušes, ramen, nosu ter skupaj z informacijami iz Tobii senzorja pridobili boljše predstave o skeletu osebe. Pri zbiranju podatkov za izračun matematične anksioznosti je pomembna tako oblika kot tudi postavitev senzorjev, saj bi preveč invazivni senzorji lahko vplivali na končne rezultate meritev.

3.4 Tipi podatkov in arhitektura hrambe podatkov

Po izvajanju prve faze meritev na posameznikih smo pridobili nabor surovih podatkov iz vseh treh senzorjev. Poleg tega smo beležili tudi vnose terapevta, ki je vnašal stanje obravnave ter počutje učenca med merjenjem. Nabor merjenih podatkov fizioloških senzorjev je razviden v poglavju 3.3.

Da bi ohranili anonimnost udeležencev, smo namesto osebnih podatkov učencev uporabili unikatne številke, ki so poznane le ustanovi, kjer so bile meritve izvedene. To je zagotovilo, da identifikacija učencev na podlagi pridobljenih podatkov ni mogoča.

Za nadaljnjo obdelavo smo shranili tako strukturo zlepkov kot tudi interpolirane vrednosti podatkov s frekvenco 10 Hz. To nam je omogočilo nadaljnjo analizo in uporabo podatkov za strojno učenje modela matematične anksioznosti.

3.5 Načrt upravljanja podatkov

Pred pričetkom zbiranja podatkov je bilo potrebno sestaviti načrt upravljanja podatkov (angl. Data management plan, DPM), ki bolj podrobno določi vse aspekte upravljanja s podatki. Načrt mora v celoti vključevati naše namene za uporabo podatkov ter določiti kako bomo z podatki ravnali. Bolj podrobno lahko načrt za upravljanje s podatki določimo z naslednjimi vidiki:

- Tip podatkov: Vir, format, lastnosti in količina podatkov.
- Kontekstualne podrobnosti: Kako se bo podatke dokumentiralo in opisalo.
- Hranjenje in varovanje: Kako se bo podatke hranilo in kako bodo zaščiteni.
- Določbe glede pokrivanja težav: Katere težave glede shranjevanja podatkov je potrebno nasloviti in kako se lotimo reševanja teh problemov.
- Ponovna uporaba istih podatkov: Ali se podatke sme uporabiti tudi pri drugih raziskovalnih nalogah in kdo točno lahko te podatke uporablja.
- Dostop in deljenje: Kako je omogočen dostop do podatkov in kako bodo bodoči uporabniki izvedeli za obstoj teh podatkov.
- Arhiviranje in deljenje dostopa: Kako se dobi dostop do podatkov in kako bodo potencialni uporabniki izvedeli za obstoj teh podatkov.

Poleg shranjenih podatkov smo pripravili tudi dokumentacijo, ki je opisala delovanje senzorja in individualno popisala vse signale, ki jih je posamezen senzor zbiral. Opisani so bili podatki individualno glede na fazo obdelave podatkov. Namen te dokumentacije je predstavitev podatkov tudi raziskovalcem, ki bodo podatke obdelovali v prihodnosti in na projektu niso sodelovali. Zaradi tega je bilo pomembno tudi uskladiti poimenovanje map v dokumentaciji in dejanski hrambi podatkov.

Oblika hranjenja podatkov je bila opisana v poglavju 3.3. Pomembno je dodati, da bodo podatki uporabljeni tudi v prihodnosti in da morajo biti v tem času shranjeni varno z onemogočenim dostopom nepooblaščenih oseb. Podatki aplikacije so se shranili lokalno na zaščiten lokalni strežnik, do katerega imajo dostop samo uporabniki, ki jim bil dovoljen vpogled do podatkov tudi med izvajanjem projekta.

4 Zaključki in razprava

Problem varovanja občutljivih podatkov bo z razvojem novih tehnologij in čedalje hitrejšo in bolj dostopno strojno opremo postal ključen pri zagotavljanju varnosti uporabnikov. V članku smo predstavili izbrane tehnološke in regulatorne (pravne) vidike varovanja občutljivih osebnih podatkov ter predstavili rešitev na primeru merjenja matematične anksioznosti učencev.

Literatura

[1] Direktiva o varstvu osebnih podatkov. https://eur-lex.europa.eu/legal-content/SL/TXT/?uri=CELEX Accessed: 20.7.2023.

[2] Splošna uredba o varstvu podatkov (gdpr) evropske unije. https://eur-lex.europa.eu/legal-content/SL/TXT/HTML/?uri=CELEX:02016R0679-20160504from=EN. Accessed: 20.7.2023.

[3] Wu He He Li, Lu Yu. The impact of gdpr on global technology development. https://www.tandfonline.com/-doi/full/10.1080/1097198X.2019.1569186. Accessed: 20.7.2023.

[4] Laima Jančiūtė. European data protection board: a nascent eu agency or an 'intergovernmental club'? https://academic.oup.com/idpl/article-abstract/10/1/57/5681448?redirectedFrom=PDF. Accessed: 20.7.2023.