Izzivi varstva zasebnosti na področju umetne inteligence

mag. Andrej Tomšič
Informacijski pooblaščenec RS

13. POSVET DOLENJSKIH IN BELOKRANJSKIH INFORMATIKOV

21.april 2022

Tri slike namesto uvoda…

Predicted: wolf
True: wolf

Predicted: husky
True: husky

Predicted: wolf
True: wolf

Predicted: husky
True: husky

Predicted: wolf
True: wolf

Predicted: wolf
True: husky


LEFT REARWARD VEHICLE CAMERA
MEDIUM RANGE VEHICLE CAMERA
RIGHT REARWARD VEHICLE CAMERA
MOTION FLOW
LANE LINES
LANE LINES
ROAD FLOW
IN-PATH OBJECTS
ROAD LIGHTS
OBJECTS
ROAD SIGNS


TayTweets
@TayandYou
@mayank_jee can i just say that im stoked to meet u? humans are super cool
23/03/2016, 20:32
TayTweets
@TayandYou
@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody
24/03/2016, 08:59
TayTweets
@TayandYou
@NYCitizen07 I [redacted] hate feminists and they should all die and burn in hell
24/03/2016, 11:41
TayTweets
@TayandYou
@brightonus33 Hitler was right I hate the jews.
24/03/2016, 11:45
gerry
@geraldmellor
Follow
"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI
6:56 AM - 24 Mar 2016
12,933
10,614

AI in tveganja za zasebnost (IWGDPT)

- Pristranskost algoritmov
- Transparentnost algoritmov (na razumljiv način)
- Erozija privolitve („click fatigue“, informirane odločitve)
- Maksimizacija obdelave proti temeljnemu načelu minimizacije obdelave
- Erozija načela namenskosti (npr. posnetki naprav, ki se upravljajo z zvokom, uporaba za druge namene)
- Analiza podatkov lahko razkrije občutljive podatke (npr. prej neznani vzorci, ki nakazujejo zdravstveno stanje…)
- Tveganje ponovne identifikacije
- Varnostno-informacijska tveganja

Primer: Nizozemska (UI „Systeem Risico Indicatie“ in goljufije pri socialnih transferjih)

Priporočila varuhov zasebnosti

Deklaracija Mednarodne konference varuhov zasebnosti o etiki in varstvu osebnih podatkov pri umetni inteligenci (International Conference of Data Protection and Privacy Commissioners, 40. konferenca, Bruselj, 2018):

Oblikovanje, razvoj in uporaba umetne inteligence (AI) mora potekati ob upoštevanju temeljnih načel:

- poštenosti in spoštovanja temeljnih človekovih pravic,
- odgovornosti in previdnosti,
- transparentnosti in razumljivosti,
- vgrajene in privzete zasebnosti,
- opolnomočenja posameznikov in
- nediskriminatornosti ter izogibanja pristranskim odločitvam.

Podobno **IWGDPT**: Working Paper on Privacy and Artificial Intelligence (64th Meeting, 29-30 November 2018, Queenstown, New Zealand)

Splošna uredba (člen 22) - Avtomatizirano sprejemanje posameznih odločitev, vključno z oblikovanjem profila

→ **Smernice EDPB**

Posameznik ima pravico, da zanj ne velja **odločitev, ki temelji zgolj na avtomatizirani obdelavi, vključno z oblikovanjem profilov**, ki ima pravne učinke v zvezi z njim ali na podoben način nanj znatno vpliva.
Avtomatizirano odločanje in profiliranje **dopustno**, če je odločitev:

a) **nujna za sklenitev ali izvajanje pogodbe** med posameznikom in upravljavcem;
b) **dovoljena v pravu Unije ali pravu države članice + zaščitni ukrepi**, ali
c) **utemeljena z izrecno privolitvijo posameznika**.

Pravica do:

- **osebnega posredovanja** (**human intervention*) **upravljavca**, (nevarnost **privzetega človeškega potrjevanja strojnih odločitev** + ni pravice do informiranja!)
- do **izražanja lastnega stališča in**
- **izpodbijanja odločitve**.

Avtomatizirane odločitve ne temeljijo na *posebnih vrstah OP* (možne izjeme – npr. javni interes!).

VB: predlaga umik 22. člena (sept. 2021)-> "i**n the name of promoting a data-driven economy and society**", ker naj bi bil tako kot Člen 15 Direktive 1995/45: „a second-class data protection right - ***rarely enforced, poorly understood and easily circumvented***".

Predlagana regulacija na EU ravni

- Evropska strategija - Bela knjiga o umetni inteligenci
- Predlog Uredbe EU za umetno inteligenco
  – 21. april 2021
- Usklajen načrt EU za umetno inteligenco - revizija 2021

Predlog Uredbe EU za umetno inteligenco

- Predlog uredbe v ospredje postavlja državljane:
  - varstvo splošnih interesov, zdravja, varnosti, človekovih pravic in temeljnih svoboščin posebej poudarjeno
  - Podobno kot Splošna uredba – pristop, ki temelji na tveganju: **večja tveganja, večje regulatorne zahteve za AI sisteme**
  - Tri ključne skupine
    - Prepovedani sistemi AI
    - Visoko tvegani sistemi AI
    - Ostali sistemi AI

Prepovedani sistemi AI

Določeni AI sistemi naj bi bili zaradi posebej visokih tveganj **prepovedani**:

- AI sistemi, ki uporabljajo **subliminalne tehnike**, nezaznavne s strani posameznika, in pomembno vplivajo na obnašanje posameznika na način, ki lahko povzroči fizično ali psihološko škodo tej ali drugim osebam;

- AI sistemi, ki **izkoriščajo posameznikove slabosti** na osnovi starosti ali posebnih potreb;

- AI sistemi, ki se lahko uporabijo za **rangiranje posamezikov** na osnovi njihovega družbenega obnašanja oziroma položaja („social scoring“);

- Sistemi za **biometrijsko prepoznavo v realnem času na javnih prostorih** (razen ob strogo določenih izjemah, npr. pri iskanju žrtev kaznivih dejanj ali pogrešanih oseb ter za pregon storilcev (težjih?) kaznivih dejanj).

Visoko tvegani sistemi AI

- **Osrednje področje** regulacije (dobra polovica pravil naj bi veljala za te)
- Dovoljeni samo **pod določenimi pogoji** (primeri v Prilogi 3 uredbe)

Sem naj bi sodili med drugim:

- **Sistemi biometrijske prepoznave obrazov v posebnih kritičnih okoljih** (npr. upravljanje prometa, upravljanje z vodnimi viri, energetika, dostopni sistemi);
- Dostopni sistemi pri univerzitetnem preverjanju znanja, izbira kadidatov za zaposlitev, prediktivni nadzor kriminalitete, verifikacija izjav prič in druge uporabe na področju policije in sodstva;
- **Sistemi za dostop do določenih storitev v javnem in zasebnem sektorju** (najverjetneje npr. centralni kreditni registri, upravičenost do socialnih transferjev).

Visoko tvegani sistemi AI

Visoko tvegani sistemi AI naj bi zahtevali pomembne obveznosti za ponudnike in uporabnike, med drugim:

- Zahteve glede kakovosti podatkov
- Preglednost do uporabnikov
- Obstoj človeškega nadzora nad delovanjem
- Testiranje in certifikacija skladnosti
- Vzpostavitev ustreznih mehanizmov upravljanja;
- Uporaba relevantnih, reprezentativnih, točnih in ažurnih ter popolnih podatkovnih nizov;
- Obvladovana tehnična dokumentacija s splošnim opisom delovanja, vključno s podatki o testiranju in validaciji ter informacije o delovanju, vključno z metriko točnosti;
- …

Pogoji za visoko tvegane sisteme AI

- Ustrezna **preglednost**, da lahko posamezniki razumejo, kako AI sistem proizvede rezultate, z **opisom splošne logike delovanja in opisom podatkov**, na katerih se sistem uči;
- Sistemi morajo omogočiti **nadzor uporabnikov za minimizacijo tveganj** - pretirano zanašanje na rezultate z vidika pristranskosti ter **možnost človeškega nadzora nad odločanjem**;
- Sistemi morajo imeti **ustrezen nivo točnosti** glede na namen uporabe ter **odpornost na napake, pomote in nekonsistentnosti**, vključno z obrambo pred zlonamernimi napadalci ali izrabo ranljivosti;
- **Ključna obveznost: ponudniki morajo izvesti teste skladnosti z zahtevami uredbe.**
- Po prodaji/začetku uporabe morajo ponudniki imeti **vzpostavljene nadzorne sisteme za zagotovitev trajne skladnosti v delovanju**. Stalno učeči se sistemi morajo biti podvrženi novim testom skladnosti v primeru večjih sprememb (odgovornost ponudnika ali uporabnika).

Ostali sistemi AI

- AI sistemi z nizkim tveganjem so sistemi, ki so namenjeni interakciji z ljudmi, kot npr. avtomatski klepetalniki, sistemi za prepoznavo čustev/odzivov.
- Informacije o njihovem delovanju morajo biti pregledne za uporabnike.
- Ostali AI sistemi so preostali sistemi z minimalnimi tveganji, ki ne sodijo v ostale skupine. Zanje se uporabljajo splošna pravila o varstvu osebnih podatkov brez dodatnih/novih obveznosti.

Obseg in področje uporabe

Po široki definiciji vključuje progamsko opremo, ki temelji na:

a) strojnemu učenju, vključno z nadzorovanim, nenadzorovanim in spodbujevanim učenjem, pri katerih se uporabljajo najrazličnejše metode, vključno z globokim učenjem;

b) pristopih, ki temeljijo na logiki in znanju, vključno s predstavitvijo znanja, induktivnim (logičnim) programiranjem, bazami znanja, inferenčnimi in deduktivnimi sistemi, (simbolnim) sklepanjem in strokovnimi sistemi;

c) statističnih pristopih, Bayesovo ocenjevanje, metode iskanja in optimizacije.

- Podobno kot pri Splošni uredbi naj bi veljala **za vse, ki uporabljajo AI sisteme napram EU državljanom**, tako v **javnem kot v zasebnem sektorju** in **tudi za ponudnike izven EU**, če se sistem uporablja v EU ali vpliva na državljane EU.
- Zavezanci so tako razvijalci kot uporabniki (podjetja, ustanove/javni organi), ki uporabljajo zadevne sisteme.
- Povsem **zasebna raba naj bi bila izvzeta**.

Nadzor in sankcije

- Vsaka DČ naj bi imenovala (vsaj en) **nadzorni organ**, ne nujno ločen oz. nov organ
- Ustanovitev **European Artificial Intelligence Board** (predseduje EK, po en član vsakega nadzornega organa, predstavnik EDPB):
  - izdaja mnenj in smernic,
  - deljenje dobrih praks,
  - razvoj usklajenih tehničnih standardov,

  * za razliko od EDPB večja vloga oz. primat EK.

- Predlog uredbe predvideva še **višje kazni kot Splošna uredba**:

  - **Upravne globe do 30 mio EUR ali do 6%** skupnega svetovnega letnega prometa, odvisno od tega, kateri znesek je višji („penalties of eye-catching severity“;)

Očitane pomanjkljivosti

- **Izkoriščevalna oz. manipulativna narava sistemov** (=prepovedana kategorija) naj bi bila predmet presoje regulatorjev;
- **Tehnološki giganti („GAFAM“) praktično niso naslovljeni**: algoritmi družbenih omrežij, spletnih trgovin, tržnic aplikacij in operacijskih sistemov (kljub možni izkoriščevalni ali manipulativni naravi);
- **Pomanjkljive obveznosti glede preglednosti** – brez generalne obveznosti po informiranju oseb, ki jih zadevajo odločitve AI sistemov. Ni obveze po informiranju glede uporabe AI pri ocenjevanju primernosti za kredite, socialne transferje, izobraževnje, zaposlitev.
- **Premalo pozornosti „poštenosti algoritmov“** v samih določbah glede na obsežne uvodne izjave;
- Novo-zahtevani **testi skladnosti naj bi bili zgolj interni procesi**, ki niso deležni javnega pregleda ali nadzora regulatorjev; niti ne dokument, zgolj „**izjava o skladnosti**“.
- Omejitve se kot kaže **ne nanašajo na biometrijske ukepe na že zbranih podatkih** (ne v realnem času) – Clearview, naknadna prepoznava protestnikov ipd.?
- **Ni izrecne zahteve po ocenah učinkov** glede uporabe posebnih vrst podatkov;
- **Podatki o obvladovanju pristranskosti** ne nujno na voljo uporabnikom, le nadzornim organom na njihovo zahtevo.

Namesto zaključka

https://www.youtube.com/watch?v=Y2wQQ-xSE4s

Viri

- IWGDPT: Working Paper on Privacy and Artificial Intelligence; https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/publikationen/working-paper/2018/2018-IWGDPT-Working_Paper_Artificial_Intelligence.pdf
- Machines learn that Brussels writes the rules: The EU's new AI regulation, https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/
- Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS; COM/2021/206 final; https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206
- The EU-AI Regulation - Part 1: Overview and structure, https://www.fieldfisher.com/en/insights/the-eu-ai-regulation-part-1
- Digitalizacija Družbe, https://www.gov.si/teme/digitalizacija-druzbe/
- BEUC: Regulating AI To Protect The Consumer, Position Paper on the AI Act; https://www.beuc.eu/publications/beuc-x-2021-088_regulating_ai_to_protect_the_consumer.pdf
- Paul De Hert and Guillermo Lazcoz: Radical rewriting of Article 22 GDPR on machine decisions in the AI era; https://europeanlawblog.eu/2021/10/13/radical-rewriting-of-article-22-gdpr-on-machine-decisions-in-the-ai-era/
- Monitor: https://www.monitor.si/novica/nizozemski-algoritmi-ki-so-unicevali-zivljenja/

- Viri slik: https://becominghuman.ai/its-magic-i-owe-you-no-explanation-explainableai-43e798273a08; https://www.quora.com/What-is-Microsoft-AI-Tay-bot

Hvala za pozornost!

andrej.tomsic@ip-rs.si