

Poročilo zaključujem z besedami dr. Tanje Martelanc in s poročilom Komisije za pripravo zaključnih sklepov zborovanja.

Prispevki na 31. zborovanju Arhivskega društva Slovenije so bili izredno kakovostno pripravljene in vsebinsko zanimivi. Prispevali so več novih spoznanj na področju arhivske stroke in znanosti, a hkrati so nas tudi spodbudili, da smo se začeli spraševati, ali je doslej poznana pot pravilna ali nas bo le uporaba novih digitalnih tehnologij privedla do končnega cilja arhivskih ustanov, tj. *»ustreznega ohranjanja, uporabe in promocije arhivskega gradiva«*, tako v fizični kot v digitalni obliki. Morda si nekateri arhivisti še vedno želijo, da bi digitalno gradivo in informacijsko-komunikacijske tehnologije ostale le neka medla senca v prihodnosti, ki se nas ne dotika, a vendar danes vemo, da ni tako. Digitalizacija neizbežno prodira v vse pore našega življenja in arhivske stroke. Arhivisti, ki želimo s pomočjo ohranjanja arhivskega gradiva dokumentirati življenje človeka v nekem prostoru in času, smo primorani ta trenutek »ujeti« v takšni obliki, v kakršni gradivo primarno nastaja. Torej sedaj in v prihodnje zlasti v digitalni obliki. Na podlagi referatov in razprave Komisije za pripravo sklepov zborovanja smo ob koncu zborovanja predlagali naslednje sklepe:

1. sklep: Dilema glede pojma digitalizacija oziroma digitizacija naj se naslovi na Medarhivsko delovno skupino za arhivsko strokovno izrazje. Za mnenje naj skupina, skupaj s predstavljenimi dilemami in poglobljeno analizo problematike, povpraša tudi pristojne ustanove za jezikovno politiko v Republiki Sloveniji, kot je npr. Terminološka svetovalnica Znanstvenoraziskovalnega centra Slovenske akademije znanosti in umetnosti.
2. sklep: Informacijski sistem oziroma aplikacija ScopeArchiv naj bo izključno tista, v kateri se, po arhivskih standardih, popisuje arhivsko gradivo, ne glede na to, kje se hrani. Torej tudi arhivsko gradivo, ki ga poleg javnih arhivov hranijo zasebni arhivi, knjižnice, muzeji in sorodne ustanove.
3. sklep: Skladno s širšimi procesi digitalizacije in odpiranja arhivskih podatkov javnosti predlagamo ponovno preučitev možnosti izboljšanja iskanja po podatkovni zbirki Slovenske javne arhivske službe s pomočjo uveljavljenih spletnih iskalnikov (npr. Google), kar je trenutno zaradi tehničnih in varnostnih razlogov onemogočeno.
4. sklep: Ustvarjalce arhivskega gradiva je treba proaktivno opozarjati, da se pri postopkih digitalizacije upoštevajo standardi in predpisi, ki so skladni z arhivsko zakonodajo. Praksa je namreč pokazala, da digitizirano gradivo, ki ga prevzemajo arhivi, pogosto ne dosega zahtevane kakovosti.
5. sklep: Predlagamo širšo praktično uporabo in razpravo glede obstoječih osnutkov dodatnih strokovno-tehničnih navodil v smeri izboljšanja in standardizacije postopkov priprave in prevzemanja ter popisovanja različnih vrst digitalnega arhivskega gradiva.
6. sklep: Preučiti je treba različne možnosti uporabe umetne inteligence, kot orodja in pomagala pri arhivskem strokovnem delu, vendar se je hkrati treba tudi zavedati njenih prednosti in slabosti.

Zapisal: dr. Žiga Koncilija

POROČILO O OKROGLI MIZI 31. ZBOROVANJA ARHIVSKEGA DRUŠTVA SLOVENIJE

5. in 6. oktobra 2023 je v Eco Hotelu v Bohinjski Bistrici potekalo 31. zborovanje Arhivskega društva Slovenije. Dvodnevni dogodek se je zaključil z okroglo mizo s tematiko **Umetna inteligenca in možnosti vpeljave novih digitalnih procesov v arhive, prenos dobrih praks iz sorodnih institucij,**

strokovna terminologija s področja digitalizacije in druga odprta vprašanja, nakar so bili podani le še sklepi zborovanja.

Pri okrogli mizi, ki je bila hkrati predvajana tudi po spletu, so sodelovali: moderatorka okrogle mize **dr. Zdenka Semlič Rajh** (Zgodovinski arhiv Ljubljana), **dr. Tatjana Hajtnik** (Arhiv Republike Slovenije), **dr. Miroslav Novak** (Pokrajinski arhiv Maribor), **dr. Marko Grobelnik** (Inštitut Jožef Stefan, prisoten po povezavi ZOOM) in **dr. Marko Robnik Šikonja** (Univerza v Ljubljani, Fakulteta za računalništvo in informatiko). Obenem je pri okrogli mizi nepričakovano (ni namreč bila napovedana v napovedniku okrogle mize) sodelovala še ena udeleženka, in sicer sama **umetna inteligenca (UI)** v obliki jezikovnega modela **CHAT GPT 4**. Slednja je svojo priložnost dobila takoj na začetku, ko je po uvodu moderatorke podrobneje predstavila vse udeležence okrogle mize. Na ta način so bili poslušalci že na začetku priča prikazu ene izmed možnosti, ki jih UI ponuja, četudi ta jezikovni model še zdaleč ni popoln. Pri okrogli mizi sta sodelovala tudi informatik **Miroslav Pavleski** iz podjetja AxonV, pogodbeni zunanji izvajalec projekta e-ARH.si: NOO 2022–2025, ki je v ozadju upravljal jezikoven model in s tem omogočil, da smo lahko v živo spremljali uporabo CHAT GPT 4, in **Saša Vrabelj**, ki je s pomočjo aplikacije SLIDO sprejemala vprašanja poslušalcev.

Prvi sklop okrogle mize je zajemal temo **Umetna inteligenca v realnem času**. Dr. Semlič Rajh je diskusijo začela s citiranjem prof. Blaža Zupana, ki je v časopisnem intervjuju¹ izpostavil nevarnost zlorabe UI, a hkrati jasno poudaril, da zaradi tega ne smemo zavračati vseh prednosti, ki jih UI prinaša. Pogovor je nadaljeval dr. Robnik Šikonja s predstavitvijo kontroverzne študije raziskovalcev Univerze Stanford, ki so iskali povezave med spolno usmerjenostjo posameznikov in fotografijami njihovih obrazov, najdenih na družabnih omrežjih. Izpostavil je predvsem etičnost uporabe UI in tehnologije prepoznavanja obrazov. Dr. Grobelnik je prikazal praktičen primer zlorabe jezikovnega modela CHAT GPT 4, s tem, da je poustvaril situacijo, kako več institucij (arhivi, inštituti, univerze) skuša s pomočjo UI negativno vplivati na javno mnenje. Poudaril je potrebo po vzpostavitvi zakonodaje (regulacije), ki bo natančno določala pravila uporabe UI, kar se sicer že pripravlja na ravni Evropske unije. Po dveh ponazoritvah negativne uporabe UI je moderatorka kot protiutež navedla številne primere in področja, pri čemer posameznik in družba ogromno pridobita z uporabo UI. Ob koncu prvega sklopa je bilo zastavljeno tudi prvo vprašanje (spletnih) poslušalcev: Ali bi bilo smiselno trenirati jezikovne modele, ki bi bili namenjeni uporabi v arhivih, in sicer samo na podlagi arhivskega gradiva, ali uporabiti že razvite modele? Dr. Robnik Šikonja je pritrdil, da bi bilo smiselno, saj gre pri arhivskem gradivu za svojevrstno gradivo oziroma vrsto podatkov.

Drugi sklop je nosil naslov **Da se sploh razumemo ...**, pri čemer je moderatorka poprosila profesorja dr. Robnik Šikonja in dr. Grobelnika, da sta na kratko predstavila osnove delovanja UI in terminologijo UI, zaradi katere se tudi arhivisti nemalokrat ne razumemo pri reševanju strokovnih vprašanj. Dr. Grobelnik je poudaril, da četudi se ljudje že več zadnjih desetletij ukvarjajo z razvijanjem UI, je prav lanski november s predstavitvijo jezikovnega modela CHAT GPT 4 prinesel nekakšen preboj oziroma trenutek, ko se je svetovna javnost zavedala, kaj vse tehnologija UI omogoča in v čem vse prekaša človeka. Obenem je predstavil nekatere scenarije, na kakšen način bi lahko UI vstopala na področje arhivskega gradiva. Besedo je nato prevzela moderatorka in poudarila nujnost poenotenja arhivske stroke pri uporabi terminologije s področja procesov digitalizacije v arhivih in dosledni uporabi že opredeljenih pojmov. Z njo se je v nadaljevanju strinjala dr. Hajtnik in kot primer navedla arhivski zakon (ZVDAGA), ki najprej razloži pojme (2. člen), s katerimi potem opredeli

¹ Pogovor z Blažem Zupanom je bil 30. septembra 2023 objavljen v prilogi *V soboto*, ki jo izdaja časnik *Večer*.



*Okrogla miza o umetni inteligenci
(Foto: Dejan Hvala)*

področja in dejavnosti v ostalih členih. Za razjasnitev pojmov »digitalizacija« in »digitizacija« se je obrnila na CHAT GPT 4 in prišla do enakega zaključka, da je pomembno, da kakor koli se že stroka dogovori glede nekega termina, da se potem to dosledno uporablja in ne vnaša zmede v strokovne debate. Sledila so zanimiva vprašanja poslušalcev: Ali je za arhivista/arhivistko v zunanji službi relevantno vprašanje, ali je informacijo, ki jo prevzema, generirala UI ali človek? Dr. Novak je podal mnenje, da je težko odgovoriti na to vprašanje, četudi bi po njegovem mnenju to morala biti pomembna informacija, tj. izvor določenega dokumenta. Dr. Semlič Rajh je poudarila, da je naša naloga predvsem, da gradivo od ustvarjalca prevzamemo, četudi za gradivo ne moremo zatrdno vedeti, ali ni znotraj tudi gradiva tudi določeno število nepristnih dokumentov. Pri tem je besedo povzel dr. Gregor Jenuš, ki je prav tako zastavil vprašanje in zagovarjal mnenje, da bi moralo biti arhivsko gradivo, ki bi nastalo s pomočjo UI, tudi označeno (npr. v metapodatkih) kot gradivo, ki ga je ustvarila UI. Nadaljeval je z vprašanjem, ali je UI zagotovo uporaben pripomoček v arhivski stroki, zanimalo ga je predvsem, kakšno je pravno tveganje napačne rabe in kdo je odgovoren. Dr. Novak je na tem mestu omenil srednjeveške listine, za katere je danes znano, da jih je precej ponarejenih, a da seveda ne moremo teh listin popravljati, saj gre za arhivsko gradivo, ki je zaščiteno, medtem ko je vprašanje pustil odprto, saj naj bi bilo nanj odgovorjeno med nadaljnjo diskusijo. Naslednje vprašanje se je glasilo: Ali so jezikovni modeli najzanimivejša tehnologija UI za arhivsko gradivo, ker je večino arhivskega gradiva v obliki besedil? Dr. Grobelnik je odgovoril, da zaradi hitrega razvoja UI to niti ni tako bistveno, saj UI že zdaj lahko razbira informacija iz besedil in fotografij/slik, razvija se tudi že tehnologija prepoznavanja informacij iz videov, tako da bo sčasoma lahko UI zbiral podatke iz virov ne glede na to, na kakšnem nosilcu je ustvarjen neki zapis. Sledilo je vprašanje dr. Jožeta Škofljanca: Kakšna bi bila minimalna začetna količina podatkov za treniranje jezikovnega modela? Dr. Robnik Šikonja je odgovoril, da so že kar nekaj slovenskih jezikovnih modelov ustvarili in da bi bilo bolj smiselno, če govorimo o arhivskih jezikovnih modelih, da bi že obstoječe modele dopolnili z arhivskim gradivom, saj se je treba zavedati, da gre pri vseh podatkih, ki jih imajo npr. slo-

venski arhivi, za zelo majhno količino podatkov, če jih primerjamo s količinami podatkov, na katerih so bili naučeni največji jezikovni modeli. Dr. Škofljanec je zastavil še eno vprašanje: Ali je smiselna in možna zaporedna in/ali vzporedna uporaba več različnih jezikovnih modelov? Dr. Grobelnik in dr. Robnik sta se strinjala, da so najnovejši modeli že tako močni, da so sposobni rešiti večino vprašanj, o katerih se danes pogovarjamo, in da so ti jezikovni modeli že sestavljeni iz več prejšnjih manjših modelov. Moderatorka je podala tudi dve mnenji, ki ju je s pomočjo aplikacije SLIDO posredoval okrogli mizi dr. Matevž Košir, ki se je navezal na vprašanje označevanja načina izvora arhivskega gradiva, pri čemer je predvideval, »da bo evropska direktiva določila, da mora biti navedeno, da je besedilo generirano z umetno inteligenco, kar je že praksa pri nekaterih objavah« in »mislim, da so možnosti UI pri generiranju arhivskih evidenc in podobno precejšnje«. Od prisotnih poslušalcev se je k besedi priglasila Branka Molnar in podala svoj pogled na razlikovanje terminov »digitalizacija« in »digitizacija«, ki ga je utemeljila z jezikovnega stališča.

Tema **Umetna inteligenca v sodobnem času** je predstavljala naslednji (3.) sklop okrogle mize. Moderatorka dr. Semlič Rajh je oba strokovnjaka s področja računalništva in informatike izzvala z vprašanjem, kako bi ocenila realno stanje tehnologij, ki jih s skupnim imenom označujemo kot UI in strojno učenje, oziroma predvsem njihovo praktično uporabnost na področju varstva kulturne dediščine ter kje vidita priložnost in hkrati omejitve teh sistemov na splošno in na področju varstva kulturne dediščine. Dr. Grobelnik je odgovoril v smislu, da gre pri uporabi UI na področju varstva kulturne dediščine za široke možnosti uporabe, pri čemer je imel v mislih kulturno dediščino v najširšem smislu in v vseh oblikah (npr. na Inštitutu Jožef Stefan trenutno poteka projekt shranjevanja vonjev). Poudaril je, da je UI obvladovanje kompleksnosti in da je sposobnost UI v tem, da s pomočjo senzorjev pridobiva podatke iz realnega ali virtualnega okolja ter te iste podatke obdeluje iz ogromnih količin informacij, ki jih človek za časa svojega življenja sam nikoli ne bi uspel niti prebrati niti sprocesirati. Na drugi strani je predstavil, kaj UI pomeni v svetovnem merilu,



*Udeleženci okrogle mize od leve proti desni Saša Vrabelj, dr. Tatjana Hajtnik, doc. dr. Zdenka Semlič Rajh, dr. Miroslav Novak in dr. Marko Robnik Šikonja
(Foto: Dejan Hvala)*

pri čemer poteka huda tekma predvsem med Kitajsko in Združenimi državami Amerike oziroma med podjetji teh držav, ki razvijajo tehnologijo UI. Njegovo misel je nadaljeval dr. Robnik Šikonja s prepričanjem, da predstavlja UI tisto orodje, ki bo lahko pomagalo obvladovati ogromne količine podatkov, ki danes nastajajo, in katerih del bodo morali arhivisti ohraniti za zanamce. Moderatorica je nato zastavila vprašanje dr. Novaku: Kakšno je trenutno stanje na tem področju v okviru arhivske teorije in prakse? Vprašani je odgovoril, da lahko praktičnim primerom uporabe UI v arhivistiki sledimo že precej let, četudi je javnost šele zadnjih nekaj let postala ozaveščena o prisotnosti UI. Kot primer je navedel Public Record Office v Victorii v Avstraliji, kjer so pred približno 10 leti poskušali ovrednotiti elektronsko pošto, ki je bila zapisana na 67.000 trakovih (okoli 28 petabajtov podatkov), pri čemer so ugotovili, da je kar 40 % omenjene elektronske pošte podvojene. Izpostavil je prizadevanja za izdelavo programov, ki omogočajo transkribiranje dokumentov in iskanje posameznih besed znotraj dokumentov. Nadalje je omenil primer ameriške vladne agencije NARA, kjer so razvijali program za »samoopisovanje« dokumentov in program za prepoznavanje obrazov na fotografijah. Mednarodni arhivski svet je dejaven pri promoviranju novih tehnologij za uporabo in zaščito arhivskega gradiva, v Sloveniji enake cilje zasleduje Arhivsko društvo Slovenije. Na evropski ravni poteka projekt »Time machine«, v katerega je vključenih več kot 400 institucij, ki poskušajo digitalizirati ogromno količino podatkov ter jih s pomočjo UI obdelati in interpretirati. Z enakimi izzivi se spoprijemajo tudi slovenski bibliotekarji (uporaba jezikovnih modelov v knjižničarstvu) in zaposleni v Narodni galeriji (spreminjanje oseb na fotografijah v avdio-video oblike, koloriranje fotografij, izboljšanje digitalizatorov).

Moderatorica je zatem zastavila isto vprašanje še dr. Tatjani Hajtnik, le da jo je pri tem zanimal informacijski vidik. Slednja je omenila, da je znotraj Arhiva Republike Slovenije precej živahno na tem področju (projekt e-ARH.si), kjer se odpirajo novi projekti iz sredstev Evropske unije, ki zahtevajo implementacijo UI in strojnega učenja v sistem e-ARH.si. Posebej je izpostavila eno izmed projektnih nalog, pri čemer je cilj, da s pomočjo UI digitalizirajo postopek vrednotenja in odbiranja arhivskega gradiva.

Četrty sklop okrogle mize je nosil naslov **Izzivi in pasti umetne inteligence pri našem delu**. Moderatorica je dr. Grobelniku in dr. Robnik Šikonji zastavila vprašanje, kako ocenjujeta pojav t. i. »globokega ponarejanja vsebin« in razpoznavanje teh pojavov, kaj menita o trenutni stopnji pristranosti rezultatov, ki jih generira UI, in ali so ti procesi skladni z etičnimi in moralnimi normami. Dr. Robnik Šikonja ocenjuje, da bo zaradi pojava »globokega ponarejanja vsebin« težko razpoznavati izvor gradiva oziroma vsebin in da bo na tem mestu vedno več nejasnosti iz zmede. Opozoril je na problematiko pristranosti, pri čemer UI lahko črpa iz podatkov, ki morda niso relevantni. Glede etičnosti uporabe UI je bil mnenja, da so te vrednote sedaj ogrožene in da se bo moral vzgojno-izobraževalni sistem prilagoditi novi situaciji z uvajanjem znaj in vedenj o digitalni pismenosti. Dr. Grobelnik se je s kolegom popolnoma strinjal in ga dopolnil, da procesa pojava UI nima smisla zavirati, ker nam omogoča ogromno novih orodij in tehnologij, in da moramo zasledovati predvsem pozitivne vidike pojava UI. Dr. Hajtnik je podala primer iz junija 2023, ko sta dva odvetnika v Združenih državah Amerike tožila jezikovni model CHAT GPT 4, ker sta na sodišču pri nekem procesu v sodno vlogo vključila izmišljeno pravno raziskavo, ki je nastala z UI. Sama je ob prebrani novici bila začudena, saj je že od pojava interneta pred tremi desetletji splošno znana resnica, da vsem spletnim informacijam ne moremo zaupati in da je zato potrebno dodatno preverjanje informacij iz drugih virov. Tudi glede CHAT GPT 4 je – potem ko ga je dr. Hajtnik vprašala – odgovoril, da informacijam, ki jih poda, ne smemo popolnoma verjeti, ampak je vedno zaželeno kritičnost pridobljenih podatkov. Dr. Grobelnik je ob tem dopolnil, da na Inštitutu Jožef Stefan zbirajo podatke o tem, koliko incidentov se zgodi zaradi

UI, in da so slednji eksponentno narasli v zadnjem letu, a da zaradi omenjenih incidentov ne umre nič več ljudi, vendar jih je veliko užaljenih. Ob koncu sklopa je bilo še prebrano vprašanje poslušalcev, in sicer je dr. Škofljanca zanimalo, ali obstajajo (glede na odprtost arhivskih podatkov) metode, ki bi jih arhivi lahko uporabljali za preprečitev zlorabe teh podatkov s pomočjo orodij UI, pri čemer bi lahko bila velika težava deanonimizacija podatkov. Tako dr. Robnik Šikonja kot dr. Grobelnik sta se strinjala, da je velika verjetnost, da bi to lahko bila težava, saj anonimizacija podatkov zahteva več ukrepov, ki jih lahko UI pozneje zaobide, če črpa podatke iz konteksta.

Moderatorica je zaradi pomanjkanja časa napovedala, da bo peti sklop okrogle mize, na katerega je bilo sicer podanih že veliko odgovorov med samo okroglo mizo, izpuščen, in da bodo podani samo še zaključki okrogle mize. Ti so bili kakor predstavitev sodelujočih na začetku ponovno generirani s pomočjo UI na ta način, da so bili vsi štirje slišani sklopi posneti, nato pretvorjeni iz zvočka v besedila in obdelani z jezikovnim modelom CHAT GPT 4. Dr. Semlič Rajh je prebrala generirane zaključke, ki so bili kljub posameznim napakam precej natančni. Na koncu se je v vlogi moderatorke zahvalila vsem sodelujočim pri okrogli mizi.

Zapisala: Dejan Hvala in Branislava Majerič Bolka