

INFORMAČNÉ INTERAKCIE 2023 – INFORMAČNÁ ETIKA, UMELÁ INTELIGENCIA A INFORMAČNÁ TVORIVOSŤ: SÚVISLOSTI A INTERAKCIE

PhDr. Jakub Fázik, PhD.; jakub.fazik@cvtisr.sk; (Centrum vedecko-technických informácií SR)

V piatok 24. novembra 2023 sa na pôde Katedry knižničnej a informačnej vedy Filozofickej fakulty Univerzity Komenského v Bratislave uskutočnil už 9. ročník konferencie so zahraničnou účasťou spojenej s doktorandským metodologickým seminárom *Informačné interakcie*. Nosnou témou bola *informačná etika v kontexte umelej inteligencie*, na ktorú boli sústredené všetky hlavné príspevky troch pozvaných hostí vrátane panelovej diskusie, a súviseli s ňou aj príspevky zamestnancov a doktorandov usporiadateľskej katedry.

<http://doi.org/10.52036/1335793X.2023.3-4.65-67>

Konferenciu otvorili prof. Jela Steinerová, predsedníčka programového výboru a zodpovedná riešiteľka VEGA 1/0360/21, a prodekanka Filozofickej fakulty UK doc. Lucia Lichnerová.

Mgr. Juraj Podroužek, PhD., výskumník Kempelenovho inštitútu inteligentných technológií (KInIT), sa v nosnom príspevku (keynote) „Etika v umelej inteligencii“ zaoberal otázkami *dôveryhodnosti umelej inteligencie* (AI), ktorej systémy by mali byť dostatočne integrálne, technicky pevné a v súlade s etickými normami (model pieskovej duny Luciana Floridiho). Etika ako sociálny (mäkký – soft) regulačný mechanizmus AI však nie je jediným, ďalšími sú právo (tvrdý spôsob), (auto)regulácia z pohľadu trhu či priamo regulácia architektúry/dizajnu systémov AI. V akýchkoľvek úvahách o etike AI by mal byť preto podľa Podroužeka kladený dôraz primárne na ľudí, ktorí AI-systémy vyvíjajú a nasadzujú, etika autonómnych prístrojov a etika ostatných entít je až sekundárnou témou. Autor predstavil dve možnosti riešenia etických problémov či dilem v súvislosti s prevádzkovaním systémov AI – ex post (riešenie existujúcich problémov) a ex ante (nazeranie na problémy ešte predtým, než nastanú; identifikácia rizík). Hoci sa druhý z uvedených prístupov javí byť elegantnejším, podlieha tzv. Collingridgeovej dileme, v ktorej úvodné snahy o ovplyvnenie či kontrolu vývoja technológie sa po čase prejavujú na jednej strane nárastom prediktability, no zároveň poklesom miery kontroly na strane druhej. Spôsob, ako čeliť tejto metodologickej dileme, je

v mentálnom nastavení ľudí (mindset); pri voľbe prístupu ex ante dôkladné analýzy ľudí a tokov dát umožňujú včasnú definíciu etických rizík a návrhy protioopatrení.

Podroužek ďalej varoval pred hrozbou *antropomorfizácie systémov AI* (dopad AI dopad na ľudskú prirodzenosť) a domnieva sa, že na korekciu správania určitého LLM (Large Language Models – veľké jazykové modely) systému AI je treba iný AI-systém. V diskusii o otázke rozdielu medzi ChatGPT 3.5 a verziou 4.0 je podľa neho najväčším problémom neznalosť súboru dát, na ktorých je táto novšia verzia z dielne OpenAI natrénovaná, čo oprávnene so sebou prináša riziko ochrany súkromia a osobných dát. Možnosťami riešenia týchto problémov je tvrdá regulácia (právo, napr. AI Act), mäkké (soft) možnosti (priamy kontakt s vývojármi, dizajnérmi, dodávateľmi) či sociálna regulácia (tlak používateľov). Bezbrannosť ľudí spôsobená absenciou kontroly a regulácie sa explicitne prejavila vo fenoméne sociálnych sietí, ktoré nazval v neskoršej panelovej diskusii „najcynickejším experimentom na ľudstve“. V panelovej diskusii tiež prirovnal systémy AI k „predĺženiam ľudských schopností a aktivít“, no nemožno ich antropomorfizovať, aby sme hovorili o ich zodpovednosti a spoľahlivosti; treba rozlišovať mieru morálnosti medzi ľudskými morálnymi agentmi a inými entitami, ktoré nemožno považovať vôbec za morálnych agentov. Pripomenul, že generatívna AI produkuje nový obsah len na úrovni štatistiky, bez porozumenia, čo prirovnal k správaniu papagájov.

PhDr. Jan Černý, Ph.D. z Fakulty informatiky a štatistiky Katedry informačných technológií Vysokej školy ekonomickej v Prahe prezentoval súbor výhod a príležitostí v *zavádzaní AI do edukačného procesu*, neopomenul však ani súvisiace riziká. Príspevok „AI, etika a trendy ve vzdeláváníí“ začal prehľadom aktualít a trendov z oblasti vývoja AI, do ktorého zaradil customizáciou modelov GPT, systém Grok, personálne manévry spoločnosti OpenAI, nové možnosti Google Pixel 8 či vývoj AGI (Artificial General Intelligence), ale i žaloby tvorcov obsahu, na ktorom sa pravdepodobne trénujú systémy AI spoločností OpenAI a Microsoft. Konštatoval (spolu s Mary Cummings), že modely generatívnej AI poskytujú len obsah priemernej kvality. V rozvoji systému formálneho vzdelávania, ktorého prostriedkom i cieľom by malo byť kritické myslenie jednotlivca, však môžu nástroje AI prispieť personalizáciou učiva a výučby (napr. asistované, interaktívne a adaptívne učenie), sprístupnením výučby väčšiemu počtu žiakov a študentov. K ďalším príležitostiam autor zaraďuje rozvoj vzdelávacích kurikúl, preklad širokého spektra učebných materiálov z rôznych cudzích jazykov s vysokou mierou presnosti, automatizované vyhodnocovanie žiackych/študentských prác a zefektívnenie administratívnych procesov v edukácii. Určite dôjde k prehodnoteniu rolí pedagóga a k ich zmene na facilitátora (podporovateľa) vzdelávania, ale i rolí žiaka/študenta – v prípade študentov už nebude napríklad kľúčovou produkcia školských/akademických písomných útvarov, ale skôr schopnosť diskutovať o téme či probléme. Viaceré pracovné pozície budú nahradené AI, Černý predpokladá napríklad vysokú mieru generovania textov AI aj v novinárskej profesii, čím zároveň vznikne ale napríklad väčší priestor pre realizáciu reportáží priamo v teréne. Dôležitým problémom generatívnej AI je úmyselná „injekcia závadného obsahu“ do cvičných (využívaných) dát, za kľúčové preto považuje rozvoj kompetencií informačnej a mediálnej gramotnosti jednotlivca v edukačnom procese. Na autorovom pracovisku sa výskumne venujú AI jednak v oblasti vzdelávania (bibliografický výskum v databáze ERIC), ale i eliminácii halucinácií AI a otázke kyberbezpečnosti (testovanie Kali LINUX nástrojov – trolling, kyberškana, kyberzločiny, dark web...).

Na prezentovanú tému Jana Černého voľne nadviazal RNDr. Michal Černý, Ph.D. z Katedry informačných štúdií a knihovníctva Filozofickej fakulty Masarykovej univerzity v Brne príspevkom „Nová konceptualizace fenoménu informační gramotnosti v kontextu rozvoje generativní umělé inteligence“. *Informačná gramotnosť* ako kompetencia žiť v každodennom informačnom prostredí podľa neho predstavuje základný epistemický nástroj s významnými ontologickými a etickými implikáciami,

ktoré v prípade etiky majú emocionálnu povahu; ako komponent sociálnej etiky informácií sa prejavuje napr. problémami informačnej chudoby, izolácie a identitarizmu (rešenia identity). Nedostatočná úroveň tohto typu gramotnosti má priamy vplyv napr. na zhoršenie zdravia, študijného prospechu či na prežívanie šťastia – je totiž „kľúčom k premýšľaniu“ jednotlivca nad svojim konaním. Analogicky k takému chápaniu informačnej gramotnosti môžeme odvodiť pojem „AI literacy“ ako kompetencie využívať nástroje umelej inteligencie v každodennom živote; v tomto prípade sú jej (negatívne) konzekvencie odlišné a jednotlivec ňou nenadobudne pravú intelektuálnu nezávislosť, ktorú obsahuje informačná gramotnosť. V rámci úvah o vzťahoch oboch typov gramotností umožňuje podľa Michala Černého informačná gramotnosť posudzovať nástroje AI a overovať výsledky AI. Na informačnú gramotnosť autor vo svojej najnovšej publikácii nazerá metaforou série poviedok Agaty Christie o Herculovi Poirotovi, kde každá z troch hlavných postáv príbehov – Poirot, Japp a Hastings – predstavuje určitý spôsob uvažovania: a) formalistický spôsob podľa pravidiel – metafora deterministických algoritmov a informačných systémov (inšpektor Japp), b) analytický typ poznania, všíma si detaily (Hercules Poirot), c) komplexné, vtelené myslenie (kapitán Hastings).

Po hlavnom bloku príspevkov nasledovala živá panelová diskusia k téme etike umelej inteligencie, ktorú viedla Jela Steinerová a pozvala do nej prednášateľov J. Podrožeka, M. Černého a J. Šušolu.

V úvode bloku príspevkov zamestnancov KKIV FiF UK pripravila Mgr. Marta Špániová, PhD. prehľad vybraných etických problémov a dilem z katalogizačnej praxe v rámci vyhlásení etického kódexu *katalogizácie* prijatého medzinárodnou katalogizačnou komunitou v roku 2021 (ALA). V príspevku „Etika katalogizácie: etický kódex katalogizácie a etické dilemy katalogizátora v praxi“ upozornila, že katalogizátori majú etickú zodpovednosť i „moc“ pri rozhodovaní a určovaní vyhľadateľnosti informácií, cieľom by mala byť používateľsky orientovaná katalogizácia. Hoci prvé katalogizačné pravidlá vznikli už v stredoveku, človek (používateľ) sa v ich stredobode objavil až v 1841 (91 pravidiel – Anthony Panizzi – dilema medzi rýchlosťou katalogizácie a jej presnosťou). Pohodlie používateľa by malo byť prioritné pred pohodlím katalogizátora. Viac sa téme venovala autorka v samostatnom príspevku v špeciálnom vydaní časopisu *ITlib* v roku 2022.

Prof. PhDr. Jaroslav Šušol, PhD. sa zaoberal etickými otázkami *informačnej bezpečnosti a kyberbezpečnosti* (v prí-

pade sieťových IKT), ktoré – ako tvrdí v príspevku „Etické dimenzie informačnej bezpečnosti – aktuálne problémy a výzvy“ – presahujú rámec jednoduchého dodržiavania zákonov a predpisov a zahŕňajú širšie spektrum morálnych povinností a spoločenskej zodpovednosti. Ku kľúčovým etickým rozmerom problematiky zaradil v zmysle „triády CIA“ z dielne National Institute of Standards and Technology: a) dôvernosť – ochranu súkromia, b) zabezpečenie integrity údajov, c) dostupnosť systémov, ale tiež aj udržiavanie spoľahlivých mechanizmov kontroly a zodpovednosti za tieto systémy (vlastníctvo, autenticita a užitočnosť). Podľa kritikov však uvedená triáda (či hexáda) je cieľom bezpečného systému a nevymedzuje rozdiel medzi bezpečným a nezabezpečeným informačným systémom. Preto sa v súčasnosti vo výskumoch venuje pozornosť aj aplikácii etických prístupov v doméne informačnej bezpečnosti, problémom v konkrétnych oblastiach informačnej bezpečnosti/kyberbezpečnosti a kultúre informačnej bezpečnosti. V tomto zmysle predstavujú otázky ochrany osobných údajov, strojového učenia (AI), udržateľnosti prostredia, zdravotných dôsledkov využívania IKT alebo infodémie (dátové zbrane – data weaponization) také výzvy či dilemy, ku ktorým musíme pristupovať citlivo a najmä eticky.

Druhý blok uzavrela *prof. PhDr. Jela Steinerová, PhD.* komplexným príspevkom „Informačná etika a informačná tvorivosť: možné smery výskumov“, v ktorom prezentovala výsledky Delfskej štúdie a bibliometrických analýz aktuálne riešeného projektu na KKIV FiF UK (VEGA 1/0360/21). Cieľom prezentácie však bola aj konceptualizácia fenoménu informačnej skúsenosti, ktorou sa *informačná etika s informačnou tvorivosťou* premostujú, a fenoménov informačného správania a informačnej gramotnosti, ktoré sú popri etike vyhľadávania informácií a etike organizácie poznania základnými piliermi informačnej etiky. Dôležitými výstupmi výskumu sú preto aj nové konceptuálne modely analyzovaných fenoménov v kontexte informačnej infraštruktúry a prostredia. Kľúčovými výskumnými trendmi podporujúcimi rozvoj informačnej etiky v súčasnosti sú etika umelej inteligencie, dátová etika a problémy dezinformácií v digitálnom prostredí; k ďalším perspektívnym výskumným témam autorka zaradila vzťahy človeka a umelej inteligencie (napätie a spolupráca), krízu etiky informácií a etiky technológií, vzdelávanie v oblasti informačnej etiky a etiky AI (vývojári, manažéri) či spoluprácu v rámci profesii a disciplín (informatika, informačná veda, filozofické a sociálne vedy, inštitúcie). Získané poznatky možno aplikovať napríklad do hodnotovo senzitívneho (re) dizajnu informačných systémov, služieb a priestorov (nielen) akademických knižníc, podporujúcich vedu, vzdelávanie,

verifikáciu informácií, kolaboráciu, kreatívnu a etickú exploráciu a stanovujúcich pravidlá pre využívanie nástrojov AI. Konečným výsledkom všetkých vzdelávacích a výskumných aktivít by mala byť etická senzitivita spoločnosti pri práci s informáciami v digitálnom prostredí.

V rámci metodologického seminára (tretí blok) boli prezentované témy súvisiace s riešením dizertačných projektov a dizertačných prác doktorandov KKIV FiF UK v Bratislave. Východiskom práce *Mgr. Kristýny Kalmárovej* je *výskumná a publikačná činnosť* ako dynamický a iteratívny súbor procesov. V príspevku „Mapování životního cyklu výzkumu: Přístupy a modely“ predstavila 15 modelov mapujúcich fázy výskumného procesu, v rámci ktorých identifikovala ich 4 typy: clustrové, lineárne, cyklické a hierarchické modely. Prezentácia *Mgr. Tomáša Mirgu* „Dezinformácia v digitálnom veku: trolie farmy a ich vplyv na verejnú mienku a politiku“ je detailne spracovaná a publikovaná v aktuálnom vydaní časopisu *ITlib*. Pojmy *zdravotnej gramotnosti* a *zdravotnej informačnej gramotnosti* porovnával *Mgr. Jakub Zábajník* v prezentácii „Zdravotná informačná gramotnosť v kontextoch tvorivosti a umelej inteligencie“. Keďže významná časť ľudskej populácie disponuje obmedzenými kompetenciami zdravotnej (informačnej) gramotnosti, nastupujúcim trendom sú inovatívne a tvorivé prístupy k vzdelávaniu v oblasti zdravia (napr. gamifikácia), a patrí k nim aj využívanie umelej inteligencie; podľa uskutočnených štúdií však nástroje AI vyhodnocujú informácie z oblasti zdravia a posudzujú ich kvalitu nedostačujúco. Doktorandské fórum zakončila *Mgr. Ekaterina Tverdokhlebova* „Konceptuálnou analýzou benchmarkingu“ – príspevkom, v ktorom priblížila históriu *benchmarkingu* (pôvodne mal súvis so strelivom a výberom zbrane), jeho dnešný význam, typológie, procesy, limity, možnosti aplikácie do prostredia knižníc (príklady, príbeh), a napokon predstavila tému jej pripravovaného výskumu, ktorého výsledkom by mal byť návrh benchmarkingu pre akademické knižnice na Slovensku.

*Zborník abstraktov*¹ a ďalšie jednotlivé prezentácie z aktuálneho ročníka konferencie i jej predošlých ročníkov sú dostupné na webe informationinteractions.weebly.com/. Podujatie bolo realizované v rámci riešenia projektu VEGA 1/0360/21 Sociálne reprezentácie etických výziev digitálnej informačnej revolúcie a prispelo k spolupráci odborníkov z rôznych oblastí pri riešení aktuálnych tém etiky informácií a etiky umelej inteligencie.

POZNÁMKA:

¹ http://informationinteractions.weebly.com/uploads/8/1/8/5/8185625/inf_int_23_zbornik_abstraktov.pdf