

Frankenstein és Drakula az archívumban

Hozzászólás „A szintetikus dokumentumok és a jövő archívumai” vitához

A vitaindító kérdés arra készlet, hogy az elszabadult frankensteini szörny, az alkotóját messze maga mögött hagyó, mesterséges teremtmény nyomába eredjünk. Ám a szintisztán szintetikus dokumentumok értékelése bajos, mivel ilyenek egyelőre nincsenek. A humán ágencia – létrehozói szándékok vagy felhasználói döntések formájában – detektálható a mégoly emberidegen képződmények létrejöttének, alakváltozásainak folyamatában. Amíg nem tudunk példát hozni arra, amikor kizárólag gépi, szuperhumán intelligencia állít elő és kezel adatokat, csak közelítő esetekre, analógiákra hivatkozva vizionálhatjuk, milyen kihívásokkal szembesít (majd) minket egy ilyen szép új világ (amely, meglehet, afféle párhuzamos univerzumként már létezik, csak emberi beavatkozások és ellenőrzés híján nem tudunk róla).

Az új technológiák nem önmagukban, hanem a világban és annak értelmezésében érzékelhető tendenciák vonatkozásában okoznak fejtörést. A mesterséges intelligencia használatáról egy társadalomtudományi archívumban – a gyűjteménygondozás praktikus szempontjai, valamint a (szerzői, adatvédelmi) jogi aspektusok mellett – a kutatásetikai és -módszertani normák felől is érdemes elgondolkodni. Egyrészt, mert miközben önállósulni látszik a digitális dokumentumok repozitálásra való előkészítésén fáradozó adatgazdászok mint szakma, az is világos, hogy olyan szakemberek képesek megfelelően ellátni ezt a feladatot, akik jártasak a kutatásban, és ismerik az adott diszciplína kereteit, elveit. Másrészt pedig azért, mivel a jelenlegi szakmai-intézményes elvárások adatkezelési tervek készítésére és a kutatási dokumentumok rendszerezett megőrzésére és közzétételére sarkallják (vagy kényszerítik) a kutatókat, akik ezáltal az archiválásban közreműködő adatgazdaként lépnek fel. Az adatfeldolgozás folyamatában – ideális esetben – keverednek egymással a kutatói és archivista szempontok, kompetenciák, szemléletmódok.

Ami egybe fűzi az adatokat, az integritást, az koherenciát biztosít nekik, ezáltal követhetővé téve az eljárásokat, hitelesítve azok eredményeit, az mindkét területen a közös referenciák és egyezményes eljárási szabályok együttese. Kérdés, hogy az olyan normákat, mint hogy mi tekinthető társadalmilag relevánsnak vagy tudományosan megalapozottnak, mennyire forgatja fel a mesterséges intelligencia. Az MI egy technológiai vívmány, amely fordulatot hozott az adatok jellegében, előállításában, kezelésében, értelmezésében. Már voltak ilyen váltások az információs történelemben (sokszorosítás, digitalizáció, internet stb.). Ám az, hogy ez a mostani éppen ennyire fókuszba került, elhomályosítja az előző fordulatok tanulságait és időnként feledtetni a kialakult értelmezési szokásokat. Talán emiatt is tűnik olyan radikálisan újnak.

Holott azok a kutatói szándékok, értelmező eljárások, felhasználói igények, amelyek a mesterséges intelligencia alkalmazásával társíthatók, részben már korábban megjelentek, például a statisztikai szoftverek bevezetése, a digitalizáció vagy

a big data hatására. Egy szóval adatközpontúságként (dataism) jellemezhető az a szellemi irányzat, amelyik a kutatási nyersanyagokat (például interjúkat, képeket, survey-válaszokat) kisebb-nagyobb jelentéstartalmú egységekre, kódokra bontva az eddigieknél határozottabban kiemeli őket az eredeti kontextusukból, hogy azután a lecsupaszított töredékeket különféle kritériumok, hasonlóságok, affinitások alapján új összefüggésekbe helyezve állítson elő belőlük adatot. A transzponálás szerencsés esetben átlátható, szakmailag indokolt és revelatív következményekkel jár, ám az eredeti egységek, összefüggések szétzilálásának problémája akkor is ott kísért. Születhetnek értékes kutatási eredmények, megannyi alternatív világmagyarázat, de elvész a megfigyelések, megnyilatkozások egyedisége, beágyazottsága, történetisége.

A helyzet emlékeztet a valóság ismeretlen eredetű, szintetikus dokumentumok általi elidegenedett, széttöredezett reprezentációjának víziójára – annyi kiegészítéssel, hogy itt a fragmentáció, mondhatni, a dokumentumon belül megy végbe. Az újraértelmezett adatokat a kontextusvesztésen kívül az új kötések gyengesége teheti sérülékennyé és egyúttal idegenszerűvé, kvázi mesterségesé. A feltörekvő technológiák bűvöletében az ember hajlamos eltekinteni alapvető módszertani elvárásoktól, ismeretelméleti elvektől. Az eredeti összefüggéseiből kiragadott, kutatói kérdésfeltevéstől, elméletektől függetlennek vélt és ekként tárgyiasított (objektívnek tekintett) adatból mintha nem is kellene következtetéseket levonni, hiszen magában hordozza, emanálja a jelentését. Legfeljebb gondozni, tisztogatni, strukturálni szükséges, de az értelmezésére nincs feltétlenül igény.

A problémát tehát abban látom, hogy a gépi megoldások előtérbe kerülése nem teljesen új keletű, egészében véve mégis rendhagyó módszertanokat inspirál, olyanokat, amelyek alapján a kutató a hagyományos eljárásokat, bevett igazolási mechanizmusokat, hermeneutikus értelmezési köröket megtörve ízekre szétszedi a vizsgálat tárgyát, majd az elemeket adatmenedzsment eszközökkel valamilyen rendbe rakja, analógiákat keres, tükröz, sablonokkal operál. Leltárba veszi a dolgokat. Megrázva a kaleidoszkópot az új rend különféle alakzatokat vesz fel. A kontextus, amely egyben tartja, immár maga az eszköz: a henger és a prizma. Nem kell ahhoz mesterséges intelligencia, hogy az adatot és így a világot kiforgassa a sarkából: az ember is képes rá.

Úgy gondolom, hogy a mostani bizalmi (és ismeretelméleti) válságra az adatok keletkezését övező és az adatkezelési eljárások háttéréül szolgáló kontextusok helyreállítása, adott esetben a személyesség visszanyerése jelenthet gyógyírt. Némileg hasonlóan ahhoz a politikai elnyomás által előidézett állapothoz a kritikai társadalomkutatás kezdetén, amikor a (személytelen) adatok megbízhatatlansága, torzításai, hiányossága váltotta ki azt a kutatói igényt, hogy terepre járva, személyes beszélgetések útján (Losonczi Ágnes szavaival) az „adatok mögé nézzenek”, feltárják az azokat létrehozó összefüggéseket. A korszak jelentős interjúk vizsgálatainak anyagát őrző 20. Század Hangja Archívum és Kutatóműhely¹¹ munkatársai a gyakran publikációig el sem jutó projektek nyersanyagaiból, töredékeiből rekonstruálják a kutatásokat, életműveket, iskolákat, ehhez kutatásokat végeznek, interjúznak, leírá-

¹ <https://20szadhangja.tk.hu/>

sokat készítenek, metaadatolnak. És persze bőszen digitalizálnak, konvertálnak és adattisztítanak. Azt hiszem, egy webes tartalmakra, mondjuk, mémekre, álhírekre specializálódott archívum is lényegében hasonlóan kell, hogy eljárjon, ha értelmet akar adni a válogatásának.

A hangsúly nem önmagában a megőrzésre (annak spekulatív szükségességére) kerül, hanem a felhasználás elősegítésére. Ehhez össze kell szedni az értelmezést segítő kontextuális tényeket, mintegy életre kelteni a dokumentumokat olyan, a létrejöttükre, előtörténetükre vonatkozó információk megadásával, melyeket, mint valami Drakula, kiszívott belőlük az idő. A történeti rekonstrukció mellett az összegyűjtött és közzétett adatok másik funkciója, hogy egyáltalán megtalálhatóvá tegyék az anyagokat, tudni lehessen a létezésükről, gépíleg össze lehessen őket kapcsolni másutt tárolt dokumentumokkal.

Voltaképpen ezek a tartalmat, formát, terjedelmet leíró metaadatok tükrözik az adott archívum profilját, szelekciós elveit. Az eredetről azt gondolom, hogy nem feltétlenül perdöntő, és nyilván gyakran nem is kideríthető. Sok tekintetben (de nem mindegyikben!) legalábbis mindegy, ki készítette a fotót Brezsnyev és Honecker érzelmes csókjáról. A szerzőség szellemi alkotásoknál érdekes, de semmire sem garancia, ami a lelet valóságosságát, igazságtartalmát illeti. A relevancia mint olyan inkább a következményeken múlik, a társadalmi, szellemi hatásra vonatkozik, és bár előre nem megjósolható (és ezért nyilván vannak és lesznek is veszteségek), úgysis a jelenből nézve számít, mit kívánunk átmenteni a múltból a jövőnek: ilyen a kulturális örökség természete.

Kérdés persze, hogy továbbra is társadalomtudományból és információtudományból egyaránt kiképzett, felkent archivisták töltik-e majd be a kapuőr szerepét. Amennyiben nem, a válogatás kritériumainak metaadatokban rögzített szempontjai akkor is tovább élnek, ha az archívumok gépi szüreteléssel gyűjtenek adatokat online repozitóriumokból vagy a weben ügködő „mikroarchivistáktól”. De vajon elszabadulhat-e a metaadat, hogy láncait letépvé, szintetikussá átlényegülve random szippantson fel tartalmakat az archívumba, szétfeszítve annak fizikai és szellemi határait?

Mennyiségi okokból és a technológiai fejlesztések következtében egyre erősebb a készítés az archívumi munka legalább részleges automatizálására. Egy nemrég az ONTOLISST projekt keretében (ONTOLISST) folytatott vizsgálatunkból azonban az derült ki, hogy nagy európai társadalomtudományos archívumok még éppen csak a kísérletezés fázisánál tartanak, ami a mesterséges intelligencia bevonását illeti a metaadatolásba, különös tekintettel az állomány tartalmi feltérképezését lehetővé tévő, mélyreható és sűrű tárgyszavazásra. A szakértelem hiányán túl a késlekedés annak is betudható, hogy nem állnak rendelkezésre olyan összehangolt sztenderdek és széles körben elfogadott, ellenőrzött szókészletek, amelyek révén általánosan kimutathatóvá lehetne tenni a tematikus relációkat. Márpedig enélkül a célkitűzés nélkül bele sem érdemes vágni a dologba. Mit ér a saját állományban rendet vágni, ha az azután zárványként elérhetetlen vagy legalábbis tökéletesen felderítetlen marad a gépi keresőmotorok számára? A jövő (és a jelen) archívumától elvárható, hogy kielégítse az adatok láthatóvá tétele és összekapcsolhatósága (interoperabilitása) iránti felhasználói igényeket, ezzel is előmozdítva az átjárhatóságot, átláthatóságot,

a tudomány nyitottságát. Ez a törekvés azonban csak részlegesen valósul meg, és ennek oka talán a mesterséges intelligenciától való idegenkedés. Lehetne szimplán bizalmatlanságnak, maradiságnak tekinteni, hogy az archívumok őrei nem tudnak vagy nem akarnak lemondani a humán kontrollról. Érzésem szerint viszont ennél többről van szó: azt akarják, hogy értelmezési (és ezáltal szelekciós) ügyekben az övék maradjon az utolsó szó.

Nincs kétségem afelől, hogy a tematikus metaadatolásban mutatkozó restanciát vészes gyorsasággal leküzdének az archívumok, mihelyt kizárólag gépi módszerekre hagyatkoznának. Elvégre léteznek annotálásra kiválóan megfelelő szöveg- és képelemző szoftverek, és bizonyára egy-kettőre kerülne olyan jolly joker fogalomrendszer is, amelyik alkalmas a világ feltérképezésére. Csakhogy amíg emberek és intézmények is részt vesznek a munkálatokban a maguk előéletével és előítéleteivel, szokásaival, hagyományaival, elképzeléseivel és törekvéseivel, addig egyezkedések, küzdelmek terepe marad az értelmezés, és sokrétű a jelentés. Márpedig egyelőre úgy tűnik, nem szállnak ki a történetből. A standardizációval, mechanikus megoldásokkal kapcsolatos várakozásokat, igényeket ellentételezi az emberi sokszínűség, a reprezentációs harc. A valóság és annak képe tagolt és egyszerismind kohézív marad: egyfelől jelenségek partikuláris emberi érdekek által vezérelt, részrehajló megjelenítéséből, másfelől statisztikailag megalapozott és fényesre csiszolt, de ugyancsak tökéletlen, gép által létrehozott információmorzsákból tevődik össze. Ugyanez a kettség jellemzi a teljes adatsiklust is, amelybe a kutatás, az archiválás és a további felhasználások egyaránt beletartoznak. A művi megoldásokkal, hamis látszatokkal teli világ, legalábbis az a része, amelyikről tudomást szerzünk, már csak ezáltal, a kogníciónk révén is mélyen emberi.

Hivatkozások

ONTOLISST. „Thematic ontologies in social science research data.” Utolsó hozzáférés: 2026. március 26.

<https://oscars-project.eu/projects/ontolisst-thematic-ontologies-social-science-research-data>

Vajda Róza az ELTE-n angol szakos diplomát, majd a New York-i New School for Social Research szociológia tanszékén MA-fokozatot szerzett. Jelenleg az ELTE TK Kutatási Dokumentációs Központban működő, a magyar társadalomtudományos kutatások digitális repozitóriumát kezelő és fejlesztő 20. Század Hangja Archívum és Kutatóműhely munkatársa. Számos hazai és nemzetközi projektben működött közre kutatóként. Fő területei a gender, etnicitás, a kisebbségek kutatása, az egyenlő bánásmód politika, társadalmi mozgalmak, valamint a hazai szociológia története.

Így hivatkozzon erre a cikkre:

<https://dx.doi.org/10.22503/inftars.XXVI.2026.1.6>