

## A történetkutatásnak kell kifejlesztenie az MI korára szabott forráskritikai eszköztárat

*Hozzászólás „A szintetikus dokumentumok és a jövő archívumai” vitához*

### **Az emlékezetőrző intézmények képesek lesznek-e (képesnek kell-e lenniük) a szintetikus dokumentumok felismerésére?**

Levéltárosként a levéltári terület speciális problémájára koncentrálnék. Én – a vitaindító megközelítésének kicsit ellentmondva – levéltárakról beszélnék, és nem archívumokról általában. Igaz, hogy a levéltár szó kimondva a hagyományos írott dokumentumok világát idézi fel, de a hazai intézményrendszerben és jogszabályi környezetben ma már egyértelmű, hogy az „irat” fogalom, hétköznapi jelentésétől függetlenül, megfelel annak, amit a vitaindító a dokumentumfogalom „tág meghatározásaként” értelmez – amennyiben az valamely szerv működése vagy személy tevékenysége során keletkezett. Ez utóbbi szempont az, amely a levéltári anyagot egyértelműen elkülöníti az egyéb emlékezetőrző intézményekétől. A könyvtár, múzeum, filmarchívum, hangarchívum fotóarchívum világában a dokumentum vagy mű önmagában teljes értékű, önállóan kezelhető, önálló nyilvántartást, leírást kívánó entitás, s a gyűjtemények, gyűjteményi egységek ezekből épülnek fel. A levéltári irat viszont, lévén egy iratképző (akár szerv, akár család vagy személy) tevékenységének „mellékterméke”, lenyomata, csak az iratképző anyagának (levéltári fond) részeként, keletkezésének és használatának kontextusában értékelhető. Forrásként való használatának és értékelésének nélkülözhetetlen része ennek a kontextusnak a megértése. Nem csak gazdagítja, és teljesebbé teszi azt – egy eredeti, de kontextusából kiszakított levéltári irat lehet éppolyan megtévesztő, és vezethet félreértelmezésekhez, mint egy hamisítvány. A levéltárnak tehát a dokumentumok megőrzése mellett ezt a kontextust is minél inkább csorbíthatatlanul fenn kell tartania, és láthatóvá kell tennie.

Amennyiben tehát a levéltári iratátvétel és megőrzés fő szempontja az, hogy az iratképző működéséről kontextusában megőrzött képet adjon, és az iratképző anyagának kialakulása, felhalmozódása során a szintetikus dokumentumok annak részévé váltak, illetve hatást gyakoroltak, akkor ezeket nem lehet a megőrzendő dokumentumok sorából kirekeszteni, hiszen éppen ez eredményezne a valóságtól eltérő képet. De miről is beszélünk, amikor a dokumentumoknak a „valósággal való kapcsolatát” kérjük számon? A levéltári anyag hitelessége papíralapú, vagy az MI kora előtti digitális világban sem az iratok külső valóságra vonatkoztatott igazságtartalmán alapult, tehát ezt eddig sem garantálta semmi. A történeti kutatás számára ugyanakkor valamely irat nem kevésbé értékes forrás attól, hogy téves, torzult vagy akár szándékosan hamis információkat tartalmaz. Ha ezeket megfelelő forráskritikával sikerül feltárni, akkor a pontos, tényszerű adatokhoz képest nem kevésbé fontos ismeretekhez juthatunk a korabeli társadalom és az intézmények működéséről, az iratok társadalmi használatáról. A tartalmi igazmondástól meg kell különböztet-

---

ni a levéltári értelemben vett hitelességet, amelynek lényege, hogy az irat akkor és azok által keletkezett, amit „állít magáról”, másrészt az, hogy amit a levéltár átvett, annak az integritását és kontextusát megőrizte, és ezt dokumentálni is képes.

Ami ezen belül a szintetikus dokumentumok helyét illeti, nyilván az lenne az ideális megoldás, hogy a levéltári leírás tartalmazza, hogy a dokumentumok létrejöttében ki (mi) működött közre, így adott esetben a humán közreműködés hiányát, vagy nem megállapítható voltát is. Ahhoz, hogy a levéltári nyilvántartás és leírás ezt eleve rögzítse, ahhoz ennek az információnak az irat levéltárba történő befogadásakor rendelkezésre kellene állnia – utólagos technikai vagy tartalmi elemzéssel ezt aligha lehet hiánytalanul és megbízhatóan megállapítani. (A levéltárak vagy kutatóik persze ehhez is igénybe vehetik az MI-t...) A szintetikus dokumentumnak a vitaindítóban adott meghatározása egyébként eleve nehezen megfogható: kérdés, hogy a humán közreműködésnek vagy hiányának milyen foka szükséges szintetikus és nem szintetikus dokumentumok elválasztásához. Annál a pontnál kezdődik a szintetikus dokumentum, ahol például a kiinduló feladatot humán cselekvő adja, de azt az MI sok folyamatlépésen át és komplexen közvetlen emberi beavatkozás nélkül oldja meg? Vagy akkor számít teljesen szintetikusnak, ha a feladat kitzúzése is emberi közreműködés nélkül történik? Ennek jól felismerhető jegyeit a dokumentumok jelentős része nyilván nem fogja magán hordozni, és az archiváláskor rendelkezésre álló kísérődokumentáció sem fogja tartalmazni, már csak ezért sem lehet arra számítani, hogy a levéltárak, archívumok ennek felismerését megnyugtatóan megoldják, és a felhasználók erre támaszkodva közelíthetnek a dokumentumokhoz.

### **Az archívumoknak, ha felismerik, meg kell-e őrizniük a szintetikus dokumentumokat? Ha igen, hogyan?**

Ha azt az esetet tekintjük, hogy a mesterséges intelligencia akár emberi feladatkitűzéssel, akár anélkül összetett feladatsort önállóan hajt végre, és ezt olyan kontextusban végzi, amely illeszkedik egy irat- és dokumentumkezelési rendszerbe, akkor miért ne tudna a saját termékének archiválásáról is dönteni? Tehát – legalábbis az esetek számottevő részében – aligha a levéltárosok lesznek döntési helyzetben. Amikor igen, és fel is ismerik a dokumentumok szintetikus voltát, az első kérdésre adott válaszban jelzett szempontok szerint dönthetnek igennel. Nem kérdés, hogy ilyen dokumentumok be fognak kerülni a levéltárak anyagába.

Ha ezek tömegesen termelődnek, és eljutnak a levéltárakba, akkor bizonyára arra is kifejlesztenek MI-eszközöket, hogy tartalmi és technikai elemzéssel segítsenek a szintetikus eredetet megállapítani. Ezzel azonban vélhetően hasonló lesz a helyzet, mint általában a kibervédelemmel, azaz e területen is állandó lesz a napi küzdelem és a versenyfutás a rosszindulatú MI-vel szemben, sőt a rossz- és jóindulatú folyamatok, entitások felismerése is nagy kihívás lesz.

A tapasztalat az, és ez aligha fog megváltozni, hogy a digitális iratok, dokumentumok, információk hiteles és jól értelmezhető megőrzésére a társadalom (állam, üzleti szféra stb., tetszés szerint behelyettesíthető) addig szánja rá a szükséges, igen komoly erőforrásokat, ameddig ezek gyakorlati célja belátható. Csak úgy általában a

jövőnek szolgáló megőrzésre nem. Ha ez így marad, akkor arra sem lehet számítani, hogy a levéltárak, archívumok lesznek a leginkább robusztusan ellátva a rosszindulató MI-felismeréséhez szükséges eszközökkel...

### **A szintetikus dokumentumok elszaporodása megváltoztatja-e majd az archívumok „hiteleshely” mivoltát a közfelfogásban?**

Mivel a mesterséges intelligencia arra is alkalmas, hogy elmúlt korok forrásait tömegesen hamisítsák vele, első ránézésre a szakemberek számára is nehezen megkülönböztethető módon, az eredetieket őrző levéltárak hiteleshelyi szerepe akár fel is értékelődhet.

Másrészt, ha általánosságban nem is fogjuk tudni biztosítani annak megállapítását, hogy mi készült humán közreműködéssel, és mi szintetikus (ami, újfent hangsúlyozom, nem feleltethető meg a valós/hamis tengelynek), lehetnek olyan projektek, állományrészek, amelyek esetében ez sikerülni fog, és akkor ezek transzparens folyamat eredményeként, jó minőségű leírással és dokumentációval maradnak az utókorra. Az is lehet, hogy ez a hiteles, illetve megbízhatóan hitelesíthető állományrész épp olyan töredéke lesz az összes keletkező iratnak/információnak, mint amennyire töredékes a mi megbízható információnk a történelem sötét korszakairól, mondjuk, a népvándorlás koráról. Ettől még lehetséges a népvándorlás korára vonatkozóan is tudományos érvényű megállapításokat tenni...

Ha a levéltárak nem is tudják majd eleve megmondani minden anyagukról, mi az, ami keletkezését tekintve szintetikus dokumentum és milyen mértékben, ha tudnak olyan szakértelmet koncentrálni, amely arra azért képes lesz, hogy bizonyos anyagokra nézve megalapozottan kijelentse, hogy ismerjük az eredetét és előállítási módját, bizonyíthatóan megőriztük az integritását, az már érték lesz.

Biztosan lesznek jó szelekciós elvek és módszerek is, amelyek ma is vannak, viszont a létező intézményi és társadalmi környezetben ezeket erősen részlegesen sikerül alkalmazni, és attól tartok, ez a jövőben, az MI meghatározta korszakban sem lesz másképp. Minden értékesnek tartott anyag megőrzéséért érdemes küzdeni, mert lehet, hogy éppen ez lesz sikeres, és akkor ez nagyon fontos lehet az utókorok számára – éppen azért, mert szerintem a siker erősen részleges lesz. Tehát – a jelenhez képest régi korokhoz hasonlóan – a sors fogja eldönteni, mi marad meg olyan formában, hogy a hitelességét és az integritását hosszú távon is igazolni lehet.

Hozzátennem, hogy a megőrzési probléma súlyossága nem az MI-vel kezdődik, hanem magával az elektronikus irattermeléssel, a digitális információ általánosság válásával. Ezt a kérdést sokszor arra redukálják, hogy lesz-e elég tárhely, és hogy fogjuk a fájlokat megnyitni, elolvasni. Holott azzal, hogy elolvastuk, még szinte semmit sem csináltunk. Egy analóg iratnak számos olyan külső és belső jege van, amely alapján a hitelességéről évszázadok múlva is mondani lehet valamit, ezt kellene elérni elektronikus környezetben is.

Erre eszközök és megoldások is rendelkezésre állnak, szabvány, jó gyakorlat ma már bőségesen van, tehát a módokat illetően nincs akadály. A digitális információ hosszú távú hitelesítésére nem találtak ki jobb megoldást, és valószínűleg nem is

---

fognak, mint azt, hogy mindent dokumentáljunk arról, amit a levéltár átvett, és a levéltári kezelésének minden mozzanatát addig a pillanatig, amíg a kutató – aki nem humán aktor is lehet – találkozik vele. A mindezekre vonatkozó metaadatok a megőrzendő állománnyal együtt alkotják az úgynevezett archív információs csomagot, tehát nem csak az elsődleges irat, de ez is a ez a megőrzés tárgyát képezi. Ezzel az a legfőbb probléma, hogy ezek erőforrásigényes dolgok, és azok is maradnak. És mint azt az előző kérdésnél említettem, az információ megőrzésére nemigen szán erőforrást a társadalom. Lehetnek tehát elveink arról, hogyan szelektáljunk, de valójában sokféle véletlen tényező összejátszása fogja szelektálni azt, ami végül tényleg hiteles és kontextusával együtt jól értelmezhető formában az utókorra marad. Az én nem túl optimista jövőképem szerint, ami hosszú távon, mondjuk, több száz év múlva is ilyen formában hozzáférhető lesz arról a korról, amikor már szinte minden irat és információ digitálisan keletkezik – és analóg módon teljes komplexitásában nem is reprodukálható –, az annyira töredékes lesz, mint amennyit mi birtokolunk a történelem azon korszakairól, amelyekre nem a forrásbőség jellemző.

### **Milyen társadalmi következményekkel járhat a szintetikus dokumentumok tömegessé válása és mesterséges intelligencia általi felhasználása?**

Ehhez először azt a kérdést kellene tisztázni, hogy ahol jellemzően szintetikus dokumentumok tömeges felhasználásával betanított mesterséges intelligencia használja fel tömegesen a levéltári dokumentumokat, ott mennyiben beszélhetünk még emberi társadalomról és kultúráról? Ennek megválaszolására most nem vállalkoznék.

Az előző kérdésekre adott válaszaimból kitűnik, nem számítok arra, hogy a levéltáraknak megnyugtató megoldása lesz arra, hogy a kutatók számára minden esetben átláthatóvá, felismerhetővé tegyék a dokumentumok szintetikus voltát. A történetkutatásnak kell kifejlesztenie az MI korára szabott forráskritikai eszköztárat. Ha ez nem sikerül, a történettudomány nem fog tudni fennmaradni abban a formában, ahogy ma ismerjük. Bár a történettudományos üzem működése sem jelent feltétlen garanciát arra, hogy a társadalom reális alapokon nyugvó és a társadalmi integráció kívánatos szintjét hathatósan előmozdító múltképpel rendelkezzen, azért ennek mégiscsak esélyét és alapját jelenti, tehát felbomlásának következményeit e helyütt talán nem szükséges tovább fejtegetni.

---

**Sipos András** az ELTE BTK történelem-levéltár szakán végzett, 1988 óta dolgozik Budapest Főváros Levéltárában. Fő kutatási területe a 19–20. század várostörténete. Számos levéltári digitalizálási, online közzétételi és elektronikus levéltári projektben vett részt.

**Így hivatkozzon erre a cikkre:**

<https://dx.doi.org/10.22503/inftars.XXVI.2026.1.3>