

## A szintetikus dokumentumok és a jövő archívumai

### *Vitaindító*

A szintetikus dokumentumok előállítása és használata egyre elterjedtebb, és várhatóan növekedni fog a jövőben. Ezek a dokumentumok elérik, és egyre inkább el fogják érni az emlékezetőrző intézményeket, amelyek mandátuma a jelen lenyomatának megőrzése a közelebbi vagy távolabbi jövő számára. Ez a jelenség új, vagy legalábbis ismert, de megoldatlan problémákat vet fel az archívumok számára – vitánk, eszmecsserénk ezekre a jelenségekre és problémákra fókuszál.

Szintetikus dokumentumokon a rövidség és idézhetőség kedvéért a mesterséges intelligencia (MI) által, autonóm módon, közvetlen humán kezdeményezés vagy közreműködés nélkül született adatokat, iratokat, képeket, videókat, tudományos vagy irodalmi műveket, híreket stb. (összefoglalóan: dokumentumok) értem, amelyeknek a valósághoz, humán entitásokhoz való viszonya külső szemlélő számára nem ellenőrizhető, nem visszafejthető. E tágan értelmezett szintetikus dokumentumfogalom nem tévesztendő össze a szintetikusadat-fogalommal, amelynek ismeretes formája az adatkezelők által az adatvédelmi előírások alóli feltételezett mentesség elérése céljából valós személyes adatok mintájára generált látszólagos személyes adatok (amelyek azonban egyetlen személlyel sem hozhatók kapcsolatba), de ilyenek tekinthetők egy korábbi technológiai környezetben a relációs adatbázisok tesztelésére, kezelőinek betanítására használt, létező adatállományok elemeinek keverésével, avagy utcajegyzékek, utónévtárak adatainak felhasználásával készült állományok. Hasonló célúak az MI-rendszerek tanítására, öntanulására szolgáló, létező dokumentumok mintájára generált szintetikus adatok, dokumentumok is.

A fentiek értelmében „dokumentumon” is egy igen tág fogalmat értek, ami azonban a vitánk tárgya és tárgyalásának szintje szempontjából megfelelő; amennyiben ennél pontosabb meghatározásra lesz szükség, elvégezzük a vita során. Ez a tág dokumentumfogalom úgy határozható meg, hogy amennyiben az *információt* megfeleltetjük három egyszerű követelménynek, nevezetesen rögzítjük, strukturáljuk és visszakereshetővé tesszük, akkor *adat* jön létre belőle (ami önmagában jelentés nélküli, csak az eredeti vagy új kontextusba helyezéssel nyer jelentést); az olyan adatok együttese pedig, amelyek formai vagy tartalmi szempontból együtt kezelendők, *dokumentumot* hoz létre. Ebből a szempontból másodlagos a dokumentum formája, műfaja, hordozója: lehet szövegalapú, képi, mozgóképes, hangfelvétel, számítógépes adatállomány, multimédia stb. (megjegyezve ugyanakkor, hogy egyes dokumentumok meghatározása, kiterjedése, pillanatnyi állapota, önálló részekre bonthatósága nem mindig problémamentes).

Nem „irat” megnevezést használok itt, egyrészt azért, mert a magyar nyelv képszerűsége és a levéltári hagyományok miatt az irat papíralapú és írással kapcsolatos

---

képzeteket kelt, másrészt felfogásom szerint a „dokumentum” tágabb fogalom, mint az irat, még a papíralapú világban is. Az „irat” az angol (főleg amerikai) *record*-fogalomnak feleltethető meg, amelynek meghatározott követelményei vannak, például ismerni kell az irat keletkeztetőjét és a keletkeztetés idejét. Ezeket a követelményeket a „recordness” fogalma tartalmazza. Ez a különbségtétel a törvényi kötelezettség alapján közintézmények iratait átvevő és megőrző levéltárak számára nem releváns, azonban olyan magánarchívumok számára, amelyek magánszervezetektől vagy magánszemélyektől vesznek át történeti értékű adományokat, a dokumentumoknak legalább akkora jelentőségük van, mint a szoros értelemben vett iratoknak. Egy harctéri körülmények között talált kézírásos levél nyilvánvalóan nem „irat”, de fontos történeti dokumentum lehet.

Hasonló okok miatt archívumokról írok és nem levéltárakról: a levéltár fogalma erősen kötődik a papíralapú iratokhoz, dokumentumokhoz (Jakab Elek már az 1874-ben tartott akadémiai székfoglalójában is megjegyezte a nemzeti nyelvű fogalom nem teljes körű lefedését), és ma is használjuk a filmarchívum, webarchívum, fotóarchívum stb. kifejezéseket ott, ahol a „levéltár” leszűkítő vagy félrevezető lenne.

Az archívumok esetében érdemes megkülönböztetnünk a dokumentumok gyűjteményét és a dokumentumok megőrzésével megbízott intézményt. Az Európa Tanács 2000-ben született ajánlása szerint az „archívum” szó a következő jelentéssel bír: (1) amikor kis „a”-val írják: az iratok összessége, függetlenül a keletkezéstől, formátumtól vagy adathordozótól, amelyeket bármely magán- vagy jogi személy tevékenysége során készített vagy kapott, és a levéltárnak tartós megőrzésre átad; (2) amikor nagy „A”-val írják: az archívumok megőrzésével megbízott (köz)intézmények. Ugyanakkor nem mindig egyértelmű, hogy a dokumentumok összességéről vagy az archívumi intézményről, esetleg mindkettőről van-e szó; mindenesetre diszkusszióink mind a két archívumfogalomra kiterjed.

A legtöbb MI által kreált vagy módosított dokumentum esetében a kezdeményező szándéka és a kiinduló anyag forrása kideríthető, gyakran éppen MI segítségével. Más esetekben közvetett módon vethető össze a dokumentum és a valóság, illetve feltételezhető a humán közreműködés, szándék. A legszembetűnőbbek az MI által generált vagy módosított képek és videók, de egyre elterjedtebbek az MI által íratott dolgozatok, publikációk, amelyek nem létező forrásokra hivatkoznak, értelmes vagy értelmetlen gondolatmeneteket tartalmaznak, MI által íratott jogi állásfoglalások, amelyek nem létező vagy hamis jogeseteken, érveléseken alapulnak, de a humán olvasó számára – ellenőrzés nélkül – hihetőnek tűnnek. Ezekben az esetekben hibrid intelligenciával, hibrid dokumentumokkal állunk szemben, ahol a humán közreműködés vagy kezdeményezés mértéke változó. Ehhez járulnak az MI hibái, tévedései, amit eufémisztikusan „hallucinációnak” neveznek, és a hibák feldúsulását, köztudatba kerülését eredményezhetik.

Tegyük fel, hogy képesek vagyunk megállapítani vagy legalább is feltételezni az ilyen dokumentumok szintetikus voltát. Megbízhatunk-e humán intelligenciánkban, vagy a mesterséges intelligenciában, hogy kérdésünkre kideríti, van-e kapcsolata az adott dokumentumnak a valósággal, vagy teljes egészében szintetikus generált?

Az archívumok és más emlékezetőrző intézmények – akár pozitivista megközelítéssel a dokumentumok semleges őrzőinek tekintjük ezeket, akár az intézmény és munkatársai ágenciáját tartjuk meghatározónak – a múlt letéteményesei, amelyeket esetleges szelektív gyűjtőkörük, esetleges restriktív hozzáférési politikájuk ellenére általában megbízható forrásoknak tartunk. Más szóval – a személyes történelem forrásain túlmenően – ezen intézmények dokumentumaiból alakul ki a múltról alkotott képünk, vagy legalább is ezekhez fordulunk referenciaként.

Ha feltételezhető a dokumentum szintetikus volta, akkor felvethető a kérdés, hogy az archívumoknak és más emlékezetőrző intézményeknek egyáltalán meg kell-e őrizniük ezeket a dokumentumokat? Erre a kérdésre eddigi beszélgetéseink résztvevői egyöntetűen igenlően válaszoltak, mondván, hogy ezeknek az intézményeknek az a feladata, hogy a mai létező világ lenyomatát és dokumentumhagyatékát megőrizzék, a szintetikus és hibrid dokumentumokkal együtt – ennek módja azonban nem magától értetődő.

További kérdésként vethető fel, hogy ha az ilyen dokumentumok az archívumok állományának sztenderd részévé válnak, és az archívumok felhasználói – akik lehetnek humán és MI entitások egyaránt – ezeket veszik alapul kutatómunkájukhoz, és ezekből újabb (szintetikus vagy humán közreműködéssel készült) dokumentumokat, történeti forrásokat kreálnak, újabb nézeteket alakítanak ki és terjesztenek, noha a kiinduló szintetikus dokumentumok valóságtartalma nem ítéhető meg, akkor vajon elképzelhető-e, hogy egymástól és a valóságtól független vagy nagyon eltérő, fragmentált világképek alakulnak ki a társadalomban?

Eddig az archívumok és más emlékezetőrző intézmények az MI vonatkozásában szinte kizárólag arra koncentráltak, hogy az MI-eszközök miként használhatók az adott intézmény munkájának segítésére, gyorsítására, hatékonyságának növelésére. Azt gondolom, hogy eljött az idő, amikor ezeknek az intézményeknek más szempontból is foglalkozniuk kell az MI hatásával. Igaz, hogy a legtöbb mai archívumban, különösen a hagyományos közlevéltárakban, feltehetőleg nincsenek ilyen szintetikus dokumentumok, de már most sem kizárható a létük adatgyűjteményekben vagy az Internet Archive-ban. Az archívumoknak, még ha a múlttal foglalkoznak is, kell gondolkodniuk a jövőről is, és legalább gondolat kísérlet szintjén ki kell alakítaniuk politikájukat a szintetikus dokumentumok vonatkozásában.

2025. október 20-án és 21-én – a résztvevők ráérése függvényében két csoportra bontva, de azonos expozícióval és illusztrációkkal – személyes beszélgetést tartottunk erről a témáról a Blinken OSA Archívumban. A beszélgetésnek jelen vitarovatunk szerzői is résztvevői voltak, avagy előzetesen írásban fejtették ki véleményüket. A kétszer másfél órás beszélgetésről hangfelvétel készült, amelyet először nyers gépi leírással konvertáltunk szöveggé, majd aprólékos manuális korrigálást követően az összesen 45 oldalnyi, 170.000 karakter terjedelmű szöveget megküldtük a résztvevőknek emlékeztetőként, illetve a további érvelések nyersanyagául.

A jelen vitaindítóban exponált, illetőleg a személyes beszélgetéseken elhangzott vélemények alapján azt kértem a vita résztvevőitől, hogy írásban fejtsek ki álláspontjukat az alábbi kérdésekben, illetve az ezekből következő további kapcsolódó kérdésekben:

- 
- Az emlékezetőrző intézmények képesek lesznek-e (képesnek kell-e lenniük) a szintetikus dokumentumok felismerésére?
  - Az archívumoknak, ha felismerik, meg kell-e őrizniük a szintetikus dokumentumokat? Ha igen, hogyan?
  - A szintetikus dokumentumok elszaporodása megváltoztatja-e majd az archívumok „hiteleshely” mivoltát a közfelfogásban?
  - Milyen társadalmi következményekkel járhat a szintetikus dokumentumok tömegessé válása és mesterséges intelligencia általi felhasználása?

---

**Így hivatkozzon erre a cikkre:**

*<https://dx.doi.org/10.22503/inftars.XXVI.2026.1.1>*