

A TUDOMÁNY, AZ OKTATÁS ÉS A KÖZGYŰJTEMÉNYI KISZOLGÁLÁS ÚJ INFORMATIKAI SZINERGIÁI

**NETWORKSHOP 2026
35. Országos Informatikai Konferencia**

**2026. március 31–április 2.
Debreceni Egyetem, Debrecen**

Szerkesztette: Tick József, Kokas Károly, Holl András

**HUNGARNET Egyesület
Budapest, 2026**



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Lukács Katalin

Networkshop 2026 konferencia előadásainak közleményei

Debreceni Egyetem, Debrecen

2026. március 31–április 2.

ISBN 978-615-6792-29-7

DOI: <https://doi.org/10.31915/NWS.2026>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2026

Borítókép: [freepik.com](https://www.freepik.com)

MARC-BOTOK HARCA: A GENERATÍV MESTERSÉGES INTELLIGENCIA MODELLEK LEHETŐSÉGEI ÉS KORLÁTAI A KÖNYVTÁRI KATALOGIZÁLÁSBAN

THE BATTLE OF THE MARC BOTS: THE POTENTIAL AND LIMITATIONS OF GENERATIVE ARTIFICIAL INTELLIGENCE MODELS IN LIBRARY CATALOGING

Takács Margit

Debreceni Egyetem BTK, Könyvtár- és információtudományi Tanszék, egyetemi adjunktus

takacs.margit@arts.unideb.hu

ORCID: [0009-0004-9206-733X](https://orcid.org/0009-0004-9206-733X)

Borbély Mária

Debreceni Egyetem BTK, Könyvtár- és információtudományi Tanszék, egyetemi adjunktus

borbely.maria@arts.unideb.hu

ORCID: [0000-0002-1161-1419](https://orcid.org/0000-0002-1161-1419)

Absztrakt

A könyvtári katalogizálás technológiai fejlődése során a hagyományos katalóguscédulákat mára felváltották az online rendszerek, amelyek feldolgozó munkáját legújabbban már a mesterséges intelligencia is támogatja. A generatív GPT-modellek képesek a bibliográfiai metaadatok és MARC-rekordok automatikus előállítására, ami új távlatokat nyit, ugyanakkor komoly minőségbiztosítási kérdéseket is felvet. A tanulmány a kifejezetten könyvtári feldolgozásra fejlesztett generatív GPT-modellek által létrehozott MARC-rekordok elemzésére fókuszál. Az összehasonlító vizsgálat olyan szempontokat érvényesít, mint a szintaktikai megfelelőség, a tartalmi validitás, a feldolgozás hatékonysága és a konzisztencia. Az elemzés megmutatja az MI jelenlegi és jövőbeli helyét a katalogizálási folyamatokban.

Kulcsszavak: Mesterséges intelligencia (MI), katalogizálás, MARC21

Abstract

As library cataloging technology has evolved, traditional catalog cards have now been replaced by online systems, whose processing tasks are now supported by artificial intelligence. Generative GPT models are capable of automatically generating bibliographic metadata and MARC records, which opens up new possibilities while also raising serious quality assurance issues. This presentation focuses on the analysis of MARC records generated by generative GPT models specifically developed for library cataloging. The com-

parative study evaluates aspects such as syntactic correctness, content validity, processing efficiency, and consistency. The analysis highlights the current and future role of AI in cataloging processes.

Keywords: Artificial Intelligence (AI), cataloging, MARC21

Bevezetés

A mesterséges intelligencia (MI, Artificial Intelligence, AI), különösen a nagy nyelvi modellek (Large Language Models, LLM) megjelenése nemcsak az információszerzésre van hatással, hanem a könyvtári munka különböző területeit is befolyásolja, köztük a feldolgozómunkát. A generatív MI képességei ugyanis a szabadszöveges tartalomgyártáson túlmenően a strukturált metaadatok generálása terén is megújulási lehetőséget hordoznak. A könyvtári szakemberek számára a jelenlegi kihívást már nem annak eldöntése jelenti, hogy indokolt-e az MI-alapú eszközök alkalmazása a feldolgozási munkafolyamatokban, hanem sokkal inkább a katalogizálásra fejlesztett megoldások hatékonyságának és megbízhatóságának kiértékelése.

E kihívásra reagálva a tanulmány négy, kifejezetten könyvtári feldolgozásra létrehozott generatív MI-eszköz (CatalogerGPT, CATMELK, Cataloging Assistant, MARC-AI Cataloger) teljesítményét hasonlítja össze. A kutatás bemutatja az alkalmazott módszertant és az eszközöket, majd egy ember által készített referenciarekordhoz viszonyítva elemzi az algoritmusok kimeneteit, hogy meghatározza a mesterséges intelligencia jelenlegi és jövőbeli helyét a katalogizálási folyamatokban.

Generatív AI-alkalmazások

A generatív mesterséges intelligencia (GAI), például a GPT-technológia, betanult adatok alapján képes új tartalmakat generálni és valós idejű interakciót biztosítani. Bár az általános célú nyelvi modellek (pl. ChatGPT) felismerik a bibliográfiai struktúrát, kutatások (Taniguchi, 2024; Niveditha és mtsai., 2024) igazolják, hogy a MARC 21-rekordok készítése során gyakran pontatlanok, strukturális hibákat vétenek és hajlamosak fiktív adatok generálására (hallucinációra).

Ezen hiányosságok kiküszöbölésére a könyvtári informatikai fejlesztések két fő irányát figyelhetjük meg. Az egyik megközelítést a kereskedelmi integrált könyvtári rendszerekbe (ILS) beépített MI-alapú katalogizálástámogató eszközök jelentik, mint például az Alma AI metaadat-asszisztense vagy az OCLC új mesterséges intelligencia alapú katalogizáló eszközei. Ezek a munkafolyamatokba integrált intelligens asszisztensek szöveg vagy kép alapján generált MARC 21-rekordvázlatokkal és automatikus osztályozási szám (DDC, LCC), valamint tárgyszó (LCSH) javaslatok generálásával támogatják a feldolgozást, anélkül, hogy kivennék a döntést a könyvtáros kezéből (OCLC, 2025., Ex Libris, 2025.)

A másik fejlesztési irányt az önálló, szakterület-specifikus katalogizáló asszisztensek jelentik, melyeket szintén kifejezetten a MARC 21 formátumhoz és a különböző osztályozási rendszerekhez igazítottak. A jelen kutatás négy ilyen, konkrét intézményi szabályoktól és kereskedelmi rendszerektől független eszközt elemez: a CatalogerGPT, a CATMELK, a Cataloging Assistant és a MARC-AI Cataloger alkalmazásokat.

CatalogerGPT

A CatalogerGPT egy egyéni fejlesztés, amelyet Glen Greenly, több mint 30 éves tapasztalattal rendelkező kanadai könyvtári szakember hozott létre. Az alkalmazás kiemelkedő képessége, hogy a könyvről készült képek alapján azonosítja a bibliográfiai adatokat, kitölti a megfelelő MARC-mezőket, és képes a meglévő rekordok hibaellenőrzésére is. Továbbá tárgyszavakat, osztályozási jelzeteket (Library of Congress Classification=LCC, Dewey Decimal Classification=DDC) javasol (Toolify.ai., 2026; Greenly, é.n.).

CATMELK

A CATMELK (Cataloging using Machine Learning Knowledge) egy kutatási projekt keretében jött létre, amelyet két könyvtári szakember, Ruwan Gamage és Priyanwada Wanigasooriya fejlesztett ki Srí Lankán. A projekt fő erőssége a vizuális adatkinyerés. OCR-technológiával (optikai karakterfelismeréssel), fényképek alapján állít elő RDA¹-kompatibilis MARC 21-rekordokat. Automatikusan pótolja a hiányzó szabványos mezőket (pl. 3XX), és validálja a bibliográfiai leírásokat. Emellett a CATMELK javaslatokat tesz a megfelelő osztályozási jelzetekre (DDC², LCC³) és tárgyszavakra is (Gamage & Wanigasooriya, 2024; Gamage & Wanigasooriya, é.n.).

Cataloging Assistant alkalmazások

A ChatGPT-ben három olyan közösségi fejlesztésű Cataloging Assistant modell érhető el, amelyet a feldolgozómunka felgyorsítására hoztak létre. Ezek a megadott adatokból RDA-kompatibilis MARC-rekordokat generálnak, emellett – a választott eszköztől függően – automatikusan meghatározzák az adott könyvtári osztályozási jelzeteket (LCC vagy

1 Az RDA (Resource Description and Access = Forrásleírás és -hozzáférés) egy korszerű bibliográfiai keretrendszer, amelyet a könyvtári tartalmak leírására fejlesztettek ki. Célja olyan adatok létrehozása, amelyek a felhasználóközpontú, kapcsolattad-alapú alkalmazások fejlesztése érdekében kialakított nemzetközi modellek szerint is jól formálnak tekinthetők.

2 A DDC (Dewey Decimal Classification vagy Dewey-féle tizedes osztályozás) egy könyvtári osztályozási rendszer, amelynek célja a könyvek témák szerinti rendszerezése.

3 Az LCC (Library of Congress Classification) egy betűkből és számokból álló jelzeteket alkalmazó osztályozási rendszer, amely a dokumentumok tartalmi feltárására és visszakereshetőségének biztosítására szolgál.

DDC), valamint a polcrendi elrendezéshez szükséges Cutter-számokat (LC vagy Sanborn). (Cataloging Assistant LCC and LC Cutter, é.n.; Cataloging Assistant DDC and LC Cutter, é.n.; Cataloging Assistant LCC and Sanborn Cutter, é.n.)

MARC-AI Cataloger

A MARC-AI Cataloger egy egyéni fejlesztés, amelyet Asif Altaf hozott létre. Elsősorban az alapadatok (ISBN, cím, szerző) alapján generál RDA-kompatibilis rekordokat, emellett tárgyszavakat (LCSH), osztályozási jelzeteket (Library of Congress Classification=LCC, Dewey Decimal Classification=DDC) javasol. Kifejezetten hangsúlyozza az emberi ellenőrzés szükségességét a véglegesítés előtt (Altaf, é.n.).

Módszertan

A kutatáshoz egy referenciaként szolgáló rekordot készítettünk a kiválasztott könyv digitalizált képei alapján. A referenciarekord MARC 21 formátum (Library of Congress, 2025) alapján készült figyelembe véve az RDA leírási szabályait (RDA-HU Munkacsoport, 2026). A vizsgálat során a tartalmi feltáró mezőket (osztályozás, tárgyszavak) szándékosan kizártuk, hogy az intézményi gyakorlatok és a szubjektivitás ne befolyásolja az eredményeket. Így a vizsgálat fókuszában kizárólag a formai feltárás objektíven mérhető elemei maradtak.

A MARC 21 formátumú rekordok generálására a korábban ismertetett négy specifikus MI-alapú eszközt használtunk: CatalogerGPT, CATMELK, CatalogingAssistant és MARC-AI Cataloger. A rekordokat elsősorban a dokumentumról készített képek segítségével generáltattuk. A mesterséges intelligencia által létrehozott MARC-rekordok minőségének mérésére egy sajátos értékelési keretrendszert dolgoztunk ki, amelynek elméleti alapját Bruce és Hillmann (2004) klasszikus metaadat-minőségi dimenziói adták. A szakirodalomban és a metaadat-kezelők körében végzett felmérések alapján a legfontosabb minőségbiztosítási mutatóknak a pontosság, a teljesség és a konzisztencia bizonyultak (Park & Tosaka, 2010; Zavalin & Zavalina 2025). Ugyanakkor a nagy nyelvi modellek és a generatív MI-specifikus működési sajátosságai miatt ezt a három szempontot a könyvtári katalógizálás követelményeihez igazítva öt vizsgálati területre bővítettük:

- **Teljesség:** Azt vizsgálja, hogy a rekord tartalmazza-e az összes releváns mezőkódot, almezőt és indikátort. Ezt az LC minimumkövetelményei, illetve az összes kinyerhető adat alapján is kiszámítottuk (a felhasznált és az összes lehetséges elem hányadosaként).
- **Pontosság:** A MARC-formátum szerkezeti elemeinek (kódok, indikátorok) helyes alkalmazását méri. A nem létező, fiktív kódok hallucinálása automatikusan 0 pontot vont maga után, továbbá külön értékeltük a mezők ismételhetőségére vonatkozó szabályok betartását is.

- **Tartalmi validitás:** Az adatok tényszerűségét és forráshűségét ellenőrzi karakteres szinten a kritikus leíró mezőkben. Ezt a kategóriát egészítette ki a hallucináció vizsgálata, amely bináris értékelést kapott: ha az AI a forrásban nem szereplő, fiktív adatot (például kitalált oldalszámot vagy sorozatcímet) generált, az érték 0 lett, ellenkező esetben 1.
- **Szabványkövetés és logikai konzisztencia:** A formai leírási szabályok, mint a mezőkön belüli MARC21-ben előírt írásjelek és a fix hosszúságú mezők (000, 008) karakterhelyességének betartását vizsgálja. A logikai konzisztencia a mezők közötti belső ellentmondás-mentességet biztosítja (pl. a 100-as és 245-ös mező összhangját).
- **Hatékonyág és ismételhetőség:** A munkafolyamatba illeszthetőséget egyrészt a generálás sebességével (időfaktor másodpercekben), másrészt annak ellenőrzésével mértük, hogy a modell képes-e azonos bemenetből többször is pontosan ugyanazt a rekordot előállítani (determinisztikus működés).

Eredmények

A kidolgozott szempontrendszer alapján elvégeztük a négy vizsgált eszköz (Cataloger-GPT, CATMELK, Cataloging Assistant, MARC-AI Cataloger) által automatikusan generált leírások kiértékelését, a kapott eredményeket pedig egy összegző táblába helyeztük (1. táblázat).

Szempont		CatalogerGPT		CATMELK		Cataloging Assistant		MARC-AI Cataloger	
Teljesség	Mezőkódok	0,95	0,86	0,77	0,6	0,9	0,93	0,81	0,8
	Indikátorok	0,84	1	0,76	0,8	0,79	1	0,92	1
	Almezők	0,82	0,85	0,89	0,8	0,6	0,9	0,79	0,85
	Összesen	2,61	2,71	2,42	2,2	2,29	2,83	2,52	2,65
Pontosság	Mezőkódok	0		1		1		0	
	Indikátorok	0,90		0,9		1*		0,83	
	Almezőkódok	1		1		1		1	
	Ismételhetőség	1		1		1		1	
	Összesen	2,9		3,9		4		2,83	
Tartalmi validitás	Adatpontosság	0,99		0,99		0,99		0,99	
	Hallucináció	0		1		0		0	
	Összesen	0,99		1,99		0,99		0,99	

Szempont		CatalogerGPT	CATMELK	Cataloging Assistant	MARC-AI Cataloger
Szabvány- követés és logikai kon- zisztencia	MARC21 tago- lás	0,875	0,875	1	1
	Fix hosszúságú mezők	0,95	0	0,98	0,98
	logikai konzisz- tencia	1	1	1	0
	Összesen	2,575	1,625	2,605	1,605
Hatékony- ság és kon- zisztencia	Időfaktor	36 s	52 s	23 s	29 s
	Ismételhetőség	nem	nem	nem	nem

1. táblázat: A vizsgált MI-eszközök értékelésének összesített eredményei

A modellek teljességének vizsgálata rámutatott, hogy egyik mesterséges intelligencia sem tudta maradéktalanul teljesíteni a Library of Congress által megkövetelt minimumot, ráadásul a 040-es katalogizálási forrásmező generálása szinte mindegyiknek problémát okozott. Míg a kötelező adatmezők számában a Cataloging Assistant érte el a legjobb eredményt (14 mező), teljesítménye némileg ellentmondásos, mivel inkább az opcionális mezők kitöltésében volt erős, az alapkövetelmények teljesítésében már kevésbé bizonyult megbízhatónak. Összesítve a legkiegyensúlyozottabb eredményt a CatalogerGPT nyújtotta, míg a MARC-AI Cataloger és a mindössze 9 kötelező mezőt generáló CATMELK mérsékelten elmaradtak az élmezőnytől.

Bár az összes modell hibátlanul bontotta szabványos almezőkre a szöveget és maradéktalanul betartotta az ismételhetőségi szabályokat, a CatalogerGPT és a MARC-AI Cataloger érvénytelen, nem létező MARC-kódokat is hallucinált. Ezek az érvénytelen kódok valós könyvtári rendszerekben hibát okoznak, és az így feldolgozott adatok láthatatlanok maradnak az OPAC indexelésében és a találati listákban.

A tényadatok kinyerésében – ami a kiváló OCR és entitás-felismerési (név, cím, ISBN) képességeket bizonyítja – mind a négy modell szinte hibátlanul, 0,99-es pontszámmal teljesített. A hallucinációk terén azonban jelentős különbség mutatkozott: egyedül a CATMELK tudta teljesen elkerülni a fiktív adatok generálását, míg a többi mesterségesintelligencia-alapú eszköz hajlamos volt kitalált információkkal önkényesen kitölteni a hiányzó részeket.

MARC21-tagolás szempontjából a Cataloging Assistant és a MARC-AI Cataloger maradéktalanul követte a központoszási szabályokat, míg a CatalogerGPT és a CATMELK esetében az egyetlen azonosított eltérés arra korlátozódott, hogy az előírásokkal ellentétben a 300-

as mező nem ponttal zárult. A fix hosszúságú mezők területén a CATMELK teljesített a legrosszabbul, mivel ez az alkalmazás eleve nem generálta le a 008-as mezőt és a rekordfejet sem hozta létre. Bár az eszközök többsége belső ellentmondások nélkül dolgozott, a Cataloging Assistant 0 pontot kapott logikai konzisztenciára, mert a 490-es mező indikátorának beállítása ellenére nem generálta a kötelező 830-as sorozati melléktételt.

A vizsgált modellek hatékonysága (generálási ideje) eltér: a leggyorsabb a Cataloging Assistant (23 mp), míg a leglassúbb a CATMELK (52 mp) volt, ami tömeges feldolgozásnál már meghatározó különbséget jelenthet a munkafolyamatban. A legfőbb problémát azonban az ismételhetőség teljes hiánya jelenti: egyik rendszer sem volt képes azonos bemenetből kétszer pontosan ugyanazt a rekordot előállítani. Ez a kiszámíthatatlanság szigorú utólagos ellenőrzés nélkül komolyan veszélyezteti a katalógusok egységességét.

Összegzés

A kutatás összegzése szerint a generatív mesterséges intelligencia drasztikusan felgyorsítja a katalogizálást, de még nem képes teljes mértékben kiváltani az emberi szakértelmet. Bár az adatokat kiválóan nyeri ki, a szigorú szabványok betartása, az adathallucinációk és az ismételhetőség hiánya továbbra is komoly problémát jelent. Emiatt az MI jelenleg csupán kreatív asszisztensként használható, a könyvtárosi munka fókusza pedig a manuális rögzítésről a gépi adatok hitelesítésére és validálására tevődik át. A jövőben azonban, ha az MI kiküszöböli a fiktív adatok generálását, a determinisztikus működés hiányát, valamint a specifikus könyvtári szabályrendszerek alkalmazásában mutatkozó hiányosságokat, akkor a katalogizálásban betöltött szerepe szintet léphet. Azaz az MI egy megbízható, nagymértékben automatizált technológiává válhat, aminek köszönhetően a könyvtárosok a szakértelmüket és figyelmüket az adatok végső hitelesítésére és a komplex összefüggések kezelésére fordíthatják.

Felhasznált források

Altai, A. (é. n.). MARC AI Cataloger [Egyedi GPT]. OpenAI.

Elérhető: <https://chatgpt.com/g/g-NVfV7JHhH-marc-ai-cataloger> (Utolsó elérés: 2026. 02. 26.)

Bruce, T. R., Hillmann, D. I. (2004) The Continuum of Metadata Quality: Defining, Expressing, Exploiting. ALA Editions. Elérhető: <https://ecommons.cornell.edu/entities/publication/b511c75c-849f-4b52-930e-e3a9f0c043> (Utolsó elérés: 2026. 02. 26.)

Cataloging assistant: DDC and LC Cutter [Egyedi GPT]. (é. n.). OpenAI. Elérhető: <https://chatgpt.com/g/g-WtqYtttP-cataloging-assistant-ddc-and-lc-cutter> (Utolsó elérés: 2026. 02. 26.)

Cataloging assistant: LCC and LC Cutter [Egyedi GPT]. (é. n.). OpenAI.

Elérhető: <https://chatgpt.com/g/g-abwV39yWz-cataloging-assistant-lcc-and-lc-cutter> (Utolsó elérés: 2026. 02. 26.)

Cataloging assistant: LCC and Sanborn Cutter [Egyedi GPT]. (é. n.). OpenAI. Elérhető: <https://chatgpt.com/g/g-qO3cUTaDi-cataloging-assistant-lcc-and-sanborn-cutter> (Utolsó elérés: 2026. 02. 26.)

Ex Libris. (2025.). Shaping tomorrow: Introducing Alma's AI vision - Chapter 1: AI in Alma's next-generation platform & the evolving user experience. Elérhető: <https://exlibrisgroup.com/blog/shaping-tomorrow-introducing-almas-ai-vision-chapter-1-ai-in-almas-next-generation-platform-the-evolving-user-experience/> (Utolsó elérés: 2026. 06. 08.)

Gamage, R., Wanigasooriya, P. (2024) Using Generative AI for Bibliographic Description: A Study with ChatGPT 4, Journal of the University Librarians Association of Sri Lanka, 27(2), p. 257–284. <https://doi.org/10.4038/jula.v27i2.8083>

Gamage, R., Wanigasooriya, P. (é. n.). CATMELK [Egyedi GPT]. OpenAI. Elérhető: <https://chatgpt.com/g/g-YQm6gByUe-catmelk>

Greenly, G. (é. n.). CatalogerGPT [Egyedi GPT]. OpenAI. Elérhető: <https://chatgpt.com/g/g-8ymg1Ftwo-catalogergpt> (Utolsó elérés: 2026. 02. 26.)

Library of Congress (2025) MARC 21 format for bibliographic data. Elérhető: <https://www.loc.gov/marc/bibliographic/> (Utolsó elérés: 2026. 02. 26.)

Niveditha, B., Harinarayana, N. S., Balachandran, C. (2024) Error Analysis in ChatGPT's MARC21 Records: A Study of RDA Conformity. Journal of Information and Knowledge, 61(4), p. 187–195. <https://doi.org/10.17821/srels/2024/v61i4/171481>

OCLC (2025.). OCLC introduces new AI tools to make cataloging faster and smarter. Elérhető: <https://www.oclc.org/en/news/releases/2025/20251208-ai-recordmanager-connexion.html> (Utolsó elérés: 2026. 06. 08.)

- Park, J., R., & Tosaka, Y. (2010) Metadata quality control in digital repositories and collections: Criteria, semantics, and mechanisms. *Cataloging & Classification Quarterly*, 48(8), 696–715. Elérhető: <https://cci.drexel.edu/faculty/jpark/articles/Metadata%20Quality%20Control%20in%20Digital%20Repositories%20and%20Collections.pdf> (Utolsó elérés: 2026. 02. 26.)
- RDA-HU Munkacsoport (2026, január 12) Javaslat az RDA magyarországi bevezetésének első szakaszára: 1. Monografikus források (nyomtatott és elektronikus könyvek). Országos Széchényi Könyvtár Könyvtári Intézet. Elérhető: https://ki.oszk.hu/sites/default/files/dokumentumtar/javaslat_az_rda_magyarorszagi_bevezetesenek_első_szakaszara_20260112.pdf (Utolsó elérés: 2026. 02. 26.)
- Taniguchi, S. (2024) Creating and Evaluating MARC 21 Bibliographic Records Using ChatGPT. *Cataloging & Classification Quarterly*, 62(5), p. 527–546. <https://doi.org/10.1080/01639374.2024.2394513>
- Toolify.ai. (2026). CatalogerGPT – Features and capabilities.
Elérhető: <https://www.toolify.ai/gpts/g-8ymgiFtwo> (Utolsó elérés: 2026. 02. 26.)
- Zavalin, V., Zavalina, O. L. (2025). Are we there yet? Evaluation of AI-generated metadata for online information resources. *Information Research an International Electronic Journal*, 30(iConf), p. 732–740. <https://doi.org/10.47989/ir30iConf47215>. Elérhető: <https://publicera.kb.se/ir/article/view/47215/37087> (Utolsó elérés: 2026. 02. 26.)