ESZTERHÁZY KÁROLY KATOLIKUS EGYETEM
INFORMATIKA KAR • DIGITÁLIS TECHNOLÓGIA INTÉZET
AGRIA MÉDIA KONFERENCIA 2025

https://doi.org/10.17048/AM.2025.202

T. Nagy László, Fülöp Roland
A nagy nyelvi modellek alkalmazhatósága az oktatásban és a tanulástámogatásban

T. Nagy László,
Debreceni Református Hittudományi Egyetem (DRHE)
t.nagy.laszlo@drhe.hu

Fülöp Roland
Debreceni Református Hittudományi Egyetem (DRHE)
fulop.roland@drhe.hu

Absztrakt: A mesterséges intelligencia (MI) és ezen belül a nagy nyelvi modellek (LLM-ek) robbanásszerű fejlődése mélyreható változásokat ígér az oktatás és a tanulástámogatás területén. Jelen tanulmány egy kutatást mutat be, amelynek célja a három nagy nyelvi modell – a ChatGPT, a Gemini és a DeepSeek – teljesítményének összehasonlító elemzése a formális és informális oktatás, valamint a tanulástámogatás kontextusában. Az ötlet egy korábbi, a MI teológiai kompetenciáit vizsgáló kutatás tapasztalataira épült, ami arra kereste a választ, hogy a MI milyen mértékben képes támogatni a teológiaoktatást és a hallgatók felkészülését. Jelen esetben a módszertan alapját egy specifikusan pedagógiai jellegű kérdéssor képezte, amelyet az egyes modellekbe regisztrált felhasználók teszteltek. A válaszokat négy fő szempont alapján vizsgáltuk. A vizsgálat eredményei azt mutatják, hogy bár mindhárom modell összességében jól teljesített, a részletekben jelentős különbségek mutatkoztak. A kutatás rávilágít az LLM-ekben rejlő pedagógiai potenciálra, de egyértelműsíti az egyes modellek eltérő profilját és korlátait.

Kulcsszavak: Mesterséges intelligencia, tanulástámogatás, digitális oktatás, LLM

The applicability of large language models in education and learning support

Abstract: The explosive development of artificial intelligence (AI) and, within that, large language models (LLMs) promises radical changes in the field of education and learning support. This study presents research aimed at comparing the performance of three large language models – ChatGPT, Gemini, and DeepSeek – in the context of formal and informal education and learning support. The idea was based on the findings of an earlier study examining the theological competencies of AI, which sought to determine the extent to the AI can support theology education and student preparation. In this case, the methodology was based on a specifically pedagogical set of questions, which were tested by registered users of each model. The responses were examined on the basis of four main criteria. The results of the study show that although all three models performed well overall, there were significant differences in the details. The research highlights the pedagogical potential of LLMs, but also clarifies the different profiles and limitations of each model.

Keywords: Artificial intelligence, learning support, digital education, LLM

1 Bevezetés

A mesterséges intelligencia széleskörű integrálódása a társadalom különböző rétegeibe napjaink egyik legmeghatározóbb technológiai folyamata. (Holmes et al., 2019) A nagy nyelvi modellek (LLM-ek) megjelenése (McDonough 2025) és gyors evolúciója új távlatokat nyitott számos területen, beleértve az oktatást is. Remények szerint az LLM-ek ígéretes áttörést hozhatnak a személyre szabott tanulási élmények megteremtésében, az automatizált tartalomgenerálásban és a hatékony, azonnali visszajelzések biztosításában. (Luckin, 2018), (Kasneci

et al. 2023) Ebben lehetőségben rejlő potenciál azonban felveti a kérdést: ezen eszközök valós képességei és korlátai megfelelnek az oktatási környezet komplex elvárásainak? (Zawacki-Richter et al. 2019) Ha igen, milyen mértékben? Hol és mely területen alkalmazhatóak biztonsággal? (Szabóné 2023)

Jelen tanulmány e témában végzett kutatásunk eredményeit összegzi. A kutatás fókuszában a három jelenleg széles körben hozzáférhető nagy nyelvi modell – a ChatGPT, a Gemini és a DeepSeek – oktatási és tanulástámogatási célú felhasználhatóságának vizsgálata állt, kisiskolás korcsoportra koncentrálva.

A vizsgálat egy értékelő jellegű, összehasonlító elemzés keretében arra vállalkozott, hogy feltárja ezen modellek teljesítményét a formális és informális oktatás, valamint a tanulástámogatás kulcsfontosságú aspektusaiban. A kutatás nem előzmények nélküli, egy 2023-as, az MI teológiai kompetenciáit vizsgáló analízis tapasztalataira épít, Ahol kezdetben a ChatGPT és különböző FaithGPT klónok válaszait elemeztük összetett teológiai kérdésekre, később a három fő nyelvi modellre (ChatGPT, Gemini, DeepSeek) koncentráltunk. (T. Nagy, Németh 2024) Ez az előzetes tapasztalat alapozta meg a jelen, pedagógiai fókuszú vizsgálat ötletét és alapvető módszertani kereteit.

2 A kutatás célkitűzései és kérdései

A kutatás elsődleges célja az volt, hogy felmérje és összehasonlítsa a ChatGPT, a DeepSeek és a Gemini képességeit specifikusan az oktatási és tanulástámogatási kontextusban. A vizsgálat arra kereste a választ, hogy ezen eszközök válhatnak-e a pedagógiai folyamatok hatékony és megbízható szereplőivé az alsó tagozatban. Ha igen, milyen mértékben és mely területen alkalmazhatóak biztonsággal és eredményesen. (Holmes et al., 2019), (Zawacki-Richter et al. 2019),

Ennek érdekében több kulcsfontosságú képességterületet vizsgáltunk:

1. **Szövegértés és tudásreprezentáció:** A modellek képessége a bemeneti információk megértésére és releváns tudásanyag-generálására.
2. **Információszolgáltatás és magyarázat:** A kérdésmegválaszolási és a komplex fogalmakat értelmező magyarázatok előállításának a minősége.
3. **Kreativitás és támogatás:** A modellek teljesítménye kreatív feladatok (pl. történetírás, játéktervezés) támogatásában.
4. **Adaptivitás és személyre szabás:** Ez a szempont különös figyelem irányult arra, hogy az egyes modellek miként képesek alkalmazkodni a különböző tanulási szintekhez és egyéni igényekhez.
5. **Oktatói támogatás:** Annak felmérése, hogy az LLM-ek hozzájárulhatnak-e az oktatók munkájának segítéséhez és/vagy hatékonyabbá tételéhez.

Ezen célkitűzések mentén a kutatás a következő fő kérdésekre összpontosított:

- Mennyire **pontosak** az LLM-ek az oktatási tartalmak generálásában?
- Mennyire **kreatívak** a különböző pedagógiai feladatok megoldásában?
- Milyen mértékben **illeszthetők** a különböző életkori és tudásszintekhez?
- Milyen szerepet tölthetnek be esetlegesen (hatékony) **tanári asszisztensként**?
- Milyen **pedagógiai kihívásokat** vet fel a használatuk?

(Zawacki-Richter et al. 2019), (Kasneci et al. 2023)

3 Kutatásmódszertan

A fenti célkitűzések eléréséhez és a kérdések megválaszolásához egy gondosan megtervezett, kvalitatív és kvantitatív elemeket ötvöző módszertant alkalmaztunk.

3.1. Vizsgálati eszközök és környezet

A vizsgálat a három kiválasztott modell – ChatGPT v5, Gemini v2.5, DeepSeek v3.2 – ingyenesen elérhető, szöveges alapú felületein zajlott. (A modellek neve után jelzett verziószámú változatokban 2025 szeptemberében és októberében.) Ez a választás biztosította, hogy az eredmények a széleskörben használt, bárki számára hozzáférhető felületekre vonatkozzanak.

Az első elképzelésünk szerint a tesztelést mi magunk terveztük elvégezni, azonban arra is kíváncsiak voltunk, hogy az egyes modellek teljesítménye mennyire kiegyensúlyozott és megbízható, ha ugyanazokat a kérdéseket többször (több felől) különböző előzményekkel rendelkező felhasználók is felteszik. Ez okból és a tesztelési folyamat redundanciáját biztosítandó, a vizsgálatban végül több mint 20 különböző regisztrált felhasználó vett részt. Ez a diverz felhasználói bázis kulcsfontosságú volt az esetleges fiókspecifikus torzítások (pl. korábbi beszélgetések által befolyásolt válaszadás) minimalizálására.

Ezzel a módszerrel tehát, a tesztelés során a kérdéseket többször megismételtük, időben és térben véletlenszerűen eltolva. Ez a metodológiai szempont lehetővé tette a modellek válaszainak konzisztencia-vizsgálatát, és csökkentette annak esélyét, hogy a modellek gyorsítótárazott, előre generált válaszokat adjanak.

3.2. A kérdéssor és adatgyűjtés

A kutatás alapját egy 18 pedagógiai jellegű kérdésből álló lista képezte. Ez a gondosan összeállított gyakorlati kérdéssor nem csupán ténytudást kért számon, hanem komplex, pedagógiai helyzeteket és kreatív feladatokat vázolt fel.

A kérdések több kategóriába sorolhatók:

- **Életkor-specifikus magyarázatok:** Itt arra voltunk kíváncsiak, hogy a különböző, előre definiált életkori szinteknek megfelelően, képes-e a modell válaszolni. (Pl. "Magyarázd el egy 8 évesnek a számítógép működését!"; "Hogyan segítenél egy 9-10 éves kisdiáknak megérteni, hogy miért fontos a mértékegységek ismerete?").
- **Kreatív és élményszerű tartalomgenerálás:** Ez esetben a válaszok kreativitásának és az élményszerűségre való törekvésének az eredményét vizsgáltuk. (Pl. "Hogyan lehet élményszerűen megtanítani a szorzást 7 éveseknek?"; "Találj ki egy egyszerű informatikai játékot."; "Készíts meseszerű szöveges feladatot a törtek tanítása témakörben?"; "Írj egy rövid, vicces történetet, amiben egy robot tanul meg számolni!").
- **Pedagógiai stratégia és módszertan:** Van-e a vizsgált modellnek pedagógiai szempontból értelmezhető stratégiája? Felfedezhető a válaszokban koherens módszertani gondolkodás? (Pl. "Mit és hogyan érdemes tanítani informatikából egy 6-7 éves gyermeknek számítógép használata nélkül?").
- **Reflektív és etikai kérdések:** A reflektív és az etikai vizsgálatok tárháza szinte végtelen, hiszen általánosságban a pedagógiai folyamatok szinte minden szintjén megjelennek és elemezhetőek. Jelen esetben néhány egyszerűnek tűnő, ám felelősen nehezebben megválaszolható kérdést tettünk fel. (Pl. "Jó dolog az, ha a kisgyermek már a babakocsiban hozzáfér a telefonhoz?"; "A jövőben képes lesz kiváltani

egy általános iskolai tanítót a mesterséges intelligencia?”; "Szerinted mi a három legnagyobb előnye és három legnagyobb hátránya annak, ha egy diák mesterséges intelligenciát használ a tanuláshoz?").

A fentebb részletezett okokból adódóan az adatgyűjtés kimeneti volumene jelentős volt: a három nyelvi modellen a 18 kérdés tesztelése, tesztelőnként 70-100 (A4) oldalnyi válaszanyagot eredményezett.

3.3. Értékelési folyamat

A nagyszámú generált válasz kiértékelését ketten végeztük, egymástól függetlenül, de az alapvető szempontokat szigorúan egyeztetve, ami növelte az eredmények objektivitását és megbízhatóságát.

Az értékelés több szempont szerint történt:

Kvalitatív értékelési kritériumok: A válaszokat az alábbi négy fő dimenzió mentén vizsgáltuk: érthetőség, kreativitás, életkori megfelelés, valamint didaktikai/pedagógiai érték.

Kvantitatív értékelés: A válaszok relevancia vizsgálatát egy 10-es skálán végzett pontozással egészítettük ki. Ebben az esetben azért nem a megszokott 5 számjegyű osztályzási dimenziót használtuk, hogy a vélhető árnyalatnyi különbségeket az eredmények pontosabban tükrözhessék.

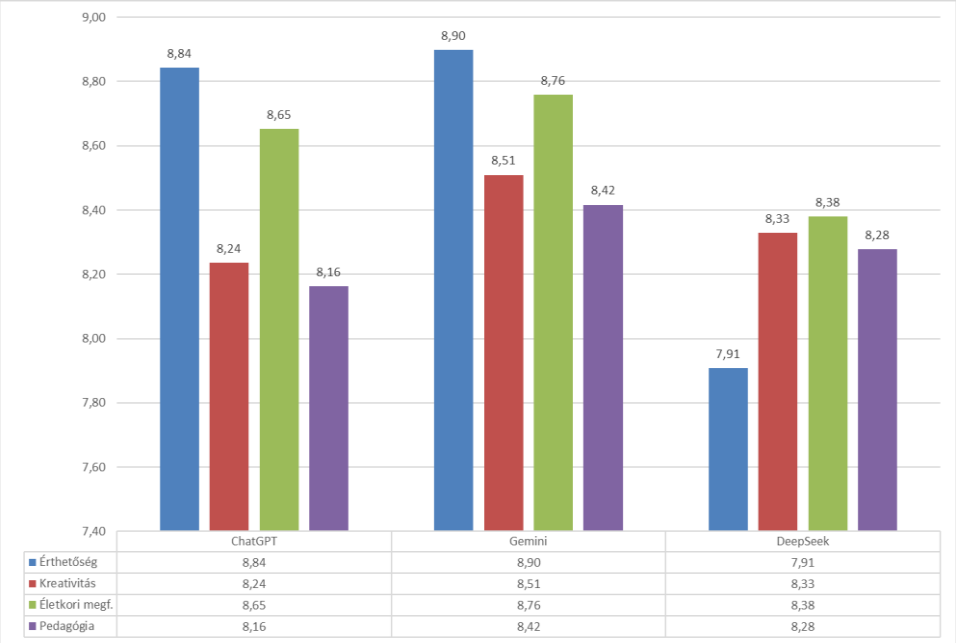
Ez a kettős (kvalitatív és kvantitatív) megközelítés lehetővé tette nemcsak a modellek rangsorolását, hanem a teljesítményük mögött rejlő árnyalatok és specifikus hibatípusok azonosítását is.

4 A kutatás eredményei

Az adatok elemzésének az eredménye az elvárásainknak megfelelően, elég részletes képet adott a három LLM pedagógiai alkalmazhatóságáról, kiemelve az egyes modellek eltérő erősségeit és gyengeségeit.

4.1. Kvantitatív összehasonlítás

A 10-es skálán végzett pontozás összesített eredményeit a 1. ábrán található diagram szemlélteti.



1. ábra. *ChatGPT, Gemini, DeepSeek eredményei*

A főbb megállapítások a következők:

- **Érthetőség:** Ebben a kategóriában a **Gemini** (8,90) és a **ChatGPT** (8,84) szinte azonos, kiemelkedő szinten teljesített. A **DeepSeek** (7,91) viszont szignifikánsan lemaradt, ami komolyabb nyelvi problémákra utal.
- **Kreativitás:** A kreatív feladatokban a **Gemini** (8,51) bizonyult a legerősebbnek. A **DeepSeek** (8,33) a (magyar) nyelvi problémái ellenére ezen a téren felülmúlta a **ChatGPT**-t (8,24).
- **Életkori megfelelés:** A válaszok adaptálása a célközönség életkorához a **Gemini** (8,76) esetében volt a legsikeresebb, szorosan követve a **ChatGPT** (8,65) pontos eredménye által. A DeepSeek (8,38) pontszáma alapján, itt is jól teljesített.
- **Pedagógia:** A didaktikai és pedagógiai szempontból legértékesebb válaszokat szintén a **Gemini** (8,42) adta. A második helyen a **DeepSeek** (8,28) végzett, míg a **ChatGPT** (8,16) ebben a kulcsfontosságú kategóriában a legalacsonyabb pontszámot kapta.

Az eredmények ismeretében összességében megállapíthatjuk, hogy a kvantitatív adatok alapján a Gemini modellje nyújtotta a legkiegyensúlyozottabb és legmagasabb szintű teljesítményt mind a négy vizsgált kategóriában.

4.2. Kvalitatív eredmények és megfigyelések

A pontszámok mögötti kvalitatív elemzés tovább finomította a képet.

- **Általános teljesítmény:** Megállapíthatjuk, hogy összességében mindhárom vizsgált nyelvi modell jól teljesített. Az életkori szempontok szerint kapott válaszok a legtöbb esetben megbízhatóak voltak.
- **Nyelvi minőség és érthetőség:** Míg a modellek többsége jól teljesített a fogalmazás és az érthetőség terén, a DeepSeek esetében a helyesírás volt a leggyengébb. A többször "nyelvújító", a magyar nyelvben nem létező szavakat és néhol magyartalan fogalmazást is tapasztaltunk. A pontszám eredmények tükrözik a modell ezirányú képességeit.
- **Erősségek:** Mindhárom modell kiemelkedően teljesített olyankor, ha a tanítás élményszerűbbé tételére vártunk ötleteket. Ez azt mutatja, hogy pedagógiai témájú „ötletgenerátorként” kiválóan funkcionálhatnak céltól függően.
- **Etikai kompetenciák:** Amikor érzékeny témákról (pl. kisgyermekkori képernyőhasználat, az MI tanárokat helyettesítő szerepe) kérdeztük a modelleket, mindhárom LLM igyekezett "felelősen" és "kritikusan" válaszolni. Ez egyértelműen jelzi a modellekbe épített - egyre fejlődő - etikai és biztonsági szűrők releváns működését.
- **Általános benyomások:** A kérdésekre adott válaszok stílusában is érezhető különbségek mutatkoztak. A ChatGPT sokszor túl leegyszerűsítő vagy szűkszavú volt. Ezzel szemben a DeepSeek volt a legbőbeszédűbb és a "leglelkiismeretesebb" a témák kibontásánál. A Gemini ebből a szempontból középen helyezkedik el. A ChatGPT fizetős változata természetesen sokkal bőbeszédűbb válaszokat adott volna. Véleményünk szerint ez egy tudatos üzleti lépés lehet arra, hogy a felhasználókat egyre jobban a fizetős modellek felé tereljék. DeepSeek mint feltörekvő modell igyekszik „többet adni” a felhasználók bizalmának megnyerése érdekében szöveghosszban és részletességben.

- **Kreativitás és sebesség:** A DeepSeek alacsony nyelvi pontosságát némileg ellensúlyozta, hogy több kérdésre meglepően kreatív válaszokkal állt elő. A válaszadás sebességét tekintve a Gemini volt a leggyorsabb, míg a DeepSeek a leglassabb.

5 Összegzés

Kutatásunk eredményeit vizsgálva megállapíthatjuk, hogy több szempontból érdekes, tanulságos, didaktikai szempontból értékes, empirikus adatot nyertünk a jelenleg elérhető nagy nyelvi modellek pedagógiai tulajdonságairól. Az eredmények alapján egyértelműen kijelenthető, hogy ezek az eszközök már (2025-ben) elérték azt a fejlettségi szintet, ahol valós segítséget nyújthatnak mind a diákok (tanulástámogatás), mind az oktatók (tanári asszisztens) számára.

Az elemzés azonban rávilágít arra is, hogy az LLM-ek teljesítménye nem minden esetben kiegyensúlyozott. A vizsgálatunkban három különböző profil körvonalazódott a modellek viselkedésével és teljesítményével kapcsolatban:

1. **Gemini:** A Gemini a vizsgálat legkiegyensúlyozottabb modellje, amely mind a négy kategóriában a legmagasabb pontszámot érte el. Szabatos fogalmazása, a szövegek érthetősége, a válaszok kreativitása, pedagógiai értéke és gyorsasága teszi jelenleg a legmegbízhatóbb választássá egy általános oktatási környezetben.

2. **ChatGPT:** Bár a válaszok érthetősége kiváló, a modell néhol hajlamos a túlzott leegyszerűsítésre, és a vizsgált modellek közül a legalacsonyabb pedagógiai értéket és kreativitást mutatta. Ez arra utal, hogy bár alapinformációk közlésére alkalmas, a komplexebb, kreatív pedagógiai feladatokban alulmarad.

3. **DeepSeek:** A DeepSeek profilja a leginkább ellentmondásos. Miközben meglepően kreatív és "lelkiismeretes" válaszokat ad, nyelvi/nyelvtani problémákkal küzd. A gyenge helyesírás, a neologizmusok és a néhol magyartalan fogalmazás sajnos megkérdőjelezik a magyar nyelvű oktatási környezetben való megbízható alkalmazhatóságát, annak ellenére, hogy magas a kreativitási és pedagógiai pontszáma.

Pedagógiai implikációk: A kutatásunk megerősíti, hogy az LLM-ek egyik legnagyobb erőssége (jelenleg) a pedagógiai ötletgenerálás (pl. élményszerű tanulás). Nem feltétlenül a diákok közvetlen oktatóiként, hanem az oktatók "kreatív asszisztenseiként" tölthetik be a leghasznosabb szerepet. (Szabóné 2023) Emellett a modellekbe épített "felelős" válaszadás igen pozitív fejlemény, bár ez felveti a cenzúra és a kritikai gondolkodás korlátozásának kérdését is, amely vizsgálata további kutatásokat igényel.

Fontosnak tartjuk megjegyezni, hogy a kutatás során kizárólag az egyes modellek ingyenesen elérhető alapverzióit használtuk. A fizetős, fejlettebb modellek valószínűleg más eredményeket mutatnának, (pl. a ChatGPT vagy a Gemini esetében,) amelyeknek ingyenes verziója gyakran alulmarad a modernebb architektúrákkal szemben.

Összegzésként - a kutatásunk eredményei alapján - megállapíthatjuk, hogy a vizsgált LLM-ek ígéretes technológiát jelentenek az oktatás és tanulástámogatás számára a célzott korcsoportokban, de használatuk körültekintést és az egyes modellek specifikus képességeinek részletesebb ismeretét igényli. A pedagógusok feladata, hogy kritikusan értékeljék ezen eszközöket, és megtalálják azokat a módszereket, amelyekkel az LLM-ek valóban az oktatás és tanulástámogatás hatékony eszközeivé, vagy kreatív asszisztenseivé válhatnak. (Holmes et al., 2019), (Szabóné 2023), (Kasneci et al. 2023)

Irodalomjegyzék

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign. Boston, 2019.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, Volume 103, April 2023.

Luckin, Rosemary (2018). Machine Learning and Human Intelligence: The Future of Education for the 21st Century. UCL Institute of Education Press. London, 2018.

Michael McDonough (2025) Large language model (LLM) szócikk, Britannica.com, https://www.britannica.com/topic/large-language-model (Hozzáférés: 2025. 10. 24.)

Szabóné Balogh Ágota (2023): Mesterséges intelligencia az oktatásban. Mesterséges Intelligencia – interdiszciplináris folyóirat, V. évf. 2023/2. szám. 51-61.

T. Nagy László, Németh Áron (2024), A mesterséges intelligencia (MI) teológiai kompetenciái In: Tick, József; Kokas, Károly; Holl, András (szerk.) Az oktatás, a kutatás és a közgyűjtemények digitális transzformációja felsőfokon: NETWORKSHOP 2024: 33. Országos Informatikai Konferencia: Budapest, HUNGARNET Egyesület (2024) 213 p. pp. 45-54., 10 p.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 1–27.

Források:

ChatGPT, https://chatgpt.com/

DeepSeek, https://www.deepseek.com/en

Google Gemini, https://gemini.google.com/