

# ADAPTÍV TESZTELÉS MESTERSÉGES INTELLIGENCIA SEGÍTSÉGÉVEL

## BEVEZETŐ

Egy pedagógus munkájának talán legfontosabb része a tudás átadása, valamint az átadott tudás visszakérése, ellenőrzése és osztályozása. A kapott osztályzat egyrészt visszajelzés a diáknak, hogy mit és mennyit sikerült elsajátítani, másrészt egy visszacsatolás a pedagógusnak munkája eredményességével kapcsolatban.

A megszerzett tudás ellenőrzése igencsak nehéz folyamat. Szinte már művészet. Ahhoz, hogy pontosan meghatározzuk egy diák tényleges tudásszintjét, bonyolult és összetett tesztekre van szükség. Vagy ez már a múlt?

2026 – a mesterséges intelligencia tombolásának időszakát éljük. Mi, pedagógusok is egyre szélesebb körben kezdjük el használni ezt az eszközt, és élvezni a vele járó előnyöket. A folytatásban egy izgalmas kísérlet folyamatát szeretném megosztani.

A kísérlet címe: Adaptív tesztelés mesterséges intelligencia segítségével.

## KÍSÉRLET

Kísérletemet a következő fázisokból építettem fel:

- Probléma definiálása
- Lehetséges megoldások keresése
- Hipotézis felállítása
- Diákok tesztelése hagyományos úton
- A hipotézis igazolására szolgáló keretrendszer megépítése, tesztelése
- Diákok tesztelése adaptív módon, MI segítségével
- A kapott eredmények feldolgozása, elemzése
- Hipotézis elfogadása / elvetése

A probléma adott: hosszadalmas, bonyolult és nem minden alkalommal igazságos tesztelés. Egy lehetséges megoldás: teszteljük a diákokat adaptív módon. Ez megoldható hagyományos, papíralapú teszt segítségével is, ennek a folyamatnak viszont nagyon sok kritikus pontja van.

**Javaslat:** oldjuk meg az adaptív tesztelést digitális eszközök segítségével, MI által.

**Hipotézis:** A mesterséges intelligencia képes hatékonyabban, adaptív módon letesztelni a tanulókat.

### Milyen feltételezést tesztelünk?

Feltételezés: Az MI képes pontosabban és gyorsabban meghatározni a tanuló valódi tudásszintjét, mint egy lineáris (fix kérdéssoros) teszt azáltal, hogy valós időben elemzi a válaszok szemantikai helyességét és nehézségét.

**Miért:** A hagyományos tesztek gyakran demotiválóak (túl nehezek) vagy unalmasak (túl könnyűek). A cél a tesztelési idő lerövidítése és a mérés pontosságának növelése a „fáradási faktor” minimalizálása mellett.

**Minimális működőképességi teszt:**

Mi lenne a legkönnyebb módszer?

Egy meglévő nagy nyelvi modell (pl. Gemini, ChatGPT vagy Claude) használata egy célzott System Prompt segítségével.

A folyamat: Nem kell szoftvert fejleszteni, csak egy szabályrendszert adni az MI-nek:

- Kérdezz a tanulótól a téma kapcsán!
- Ha helyes a válasz, a következő kérdés legyen egy szinttel nehezebb!
- Ha helytelen, egyszerűsíts!
- 5–10 kérdés után becsüld meg a tudásszintjét! Ezt egy kis fókuszcsoporton (pl. 5-10 fő) azonnal lehet futtatni.

**A siker mértéke**

- Konvergencia-sebesség (Hatékonyság): Hány kérdésre volt szüksége az MI-nek ahhoz, hogy a becsült tudásszint stabilizálódjon? (Ha 5-10 kérdés alatt ugyanazt az eredményt hozza, mint egy 30 kérdéses fix teszt, a kísérlet sikeres).
- Validitási korreláció (Pontosság): Mennyire egyezik meg az MI által adott pontszám egy kontrollméréssel (pl. egy korábbi, hagyományos vizsga eredményével)? A magas korreláció ( $r > 0,7$ ) jelenti a sikert.
- Szubjektív nehézségi görbe: A tanulók visszajelzése alapján a kérdések „éppen megfelelő” nehézségűek voltak-e (ún. Flow-élmény a tesztelés alatt).

**Milyen problémát vagy kihívást szeretnének megoldani az MI-vel?**

A fő probléma a statikus tesztelés rugalmatlansága.

- Időpazarlás: A hagyományos tesztek sok olyan kérdést tartalmaznak, amelyek a tanuló számára vagy túl könnyűek, vagy túl nehezek, így nem adnak érdemi információt a tudásáról.
- Mérési pontatlanság: Egy fix kérdéssor gyakran csak a felszínt kapargatja, vagy épp a tanuló egyetlen specifikus hiányossága miatt torzítva mutatja az összképet.
- Személytelenség: A standard tesztek nem képesek figyelembe venni a tanuló egyéni gondolkodásmódját vagy a válaszok mögötti logikát.

**Kiknek a számára probléma ez, és miért jelent számukra problémát? Sokakat érintő problémáról van szó?**

- Tanulók: A túl nehéz teszt szorongást és demotivációt okoz, a túl könnyű pedig unalmat. Nem kapnak valós visszajelzést a fejlődési potenciáljukról.
- Oktatók: Rengeteg időt töltenek tesztek összeállításával és javításával, mégsem kapnak tűpontos képet arról, hol tart pontosan a diák.
- Oktatási intézmények: A vizsgáztatás erőforrásigényes (idő, pénz, infrastruktúra). Minél lassabb a mérés, annál drágább a rendszer fenntartása.

## Hogyan tudna az MI segíteni ennek a problémának a megoldásában?

- Valós idejű alkalmazkodás: Minden egyes válasz után újrakalkulálja a tanuló valószínűsíthető tudásszintjét, és a következő kérdést úgy választja ki, hogy az a legtöbb új információt szolgáltatassa.
- Szemantikai elemzés: Az MI nemcsak kulcsszavakat néz, hanem érti az összefüggéseket is, így nyitott kérdésekre adott válaszokat is képes azonnal beépíteni a teszt menetébe. A kísérletben matematikai feladatokat oldottak a diákok, itt egyértelmű, hogy egy feladatnak van egy végeredménye. Az MI képes olyan tantárgyak esetében is hatékony lenni, melyeknél a diákok „saját szavaikkal” adott válaszait is megérti.
- Személyre szabott nehézség: Fenntartja a tanuló motivációját azáltal, hogy folyamatosan a „legközelebbi fejlődési zónában” tartja a feladatokat.

## Mi az, ami jól alakulhat? Milyen előnyökkel járhat?

- Radikális hatékonyság: A tesztelési idő akár 50-70%-kal is csökkenhet ugyanazon mérési pontosság mellett. A hagyományos tesztre a klasszikus értelemben vett 45 percre volt szükség. A digitális verzió esetében 20 perc állt a tanulók rendelkezésére. Hárman nem jutottak el a 10. feladatig, viszont a legjobbak (ketten) 14 perc alatt is teljesítették a feladatsort.
- Pozitív vizsgaélmény: A tanulók sikerélményt kapnak, mert a teszt „partnerként” viselkedik, nem pedig akadályként. Nagyon sokat jelentett a diákoknak a rendszertől kapott bátorítás. Papíralapú tesztelésnél sokszor a kudarcérzet szorongást, feszültséget szül. Ellenben az MI minden elrontott lépést azonnal magyaráz, felhívja a diák figyelmét a mulasztásra, és sok esetben ezen visszajelzéseknek köszönhetően gördülékenyebben vették a későbbi feladatokat és kevesebbet hibáztak.
- Mélyebb analitika: Az oktatók nemcsak egy adatot kapnak, hanem egy részletes térképet a tanuló erős és gyenge pontjairól. Ha külön beleírjuk a rendszerbe, hogy elemezze a tanuló viselkedését, akkor ezek az elemzések igazi drágakővé válnak minden pedagógusnak.

## Mi az, ami rosszul alakulhat – mi az, ami kihívást fog jelenteni?

- MI Hallucináció: Az MI tévesen ítélné meg egy helyes választ helytelennek (vagy fordítva), ha a kérdés vagy a válasz túl komplex vagy kétértelmű. A tesztelés folyamán, amíg nem hangoltuk fel elég jól a rendszert, sokszor a hibás válaszokat is elfogadta, illetve helyes válaszokat is helytelennek minősített. Ezért építettem be a párbeszéd lehetőségét: ha a diák szól, hogy szerinte helyes a válasz, akkor az MI felülírhatja, kijavíthatja előző döntését (természetesen egy újabb elemzést követően).
- Beépített elfogultság (Bias): Ha az MI olyan adatokon tanult, amelyek bizonyos megfogalmazásokat előnyben részesítenek, hátrányba hozhat bizonyos tanulói

csoportokat. Érdeemes tehát az MI általános tudásbázisára hagyatkozni; ha pedig saját tanulási alapot töltünk fel, akkor az minél széleskörűbb legyen.

- **Technikai instabilitás:** Az adaptív logika „beakadhat” egy szinten, vagy túl hirtelen válhat nehézséget, ami összezavarja a mérés. Nagyon precízen és egyértelműen kell meghatározni a szintek közötti átjárást a prompt tervezésénél.
- **Magyarázhatóság (Black Box):** Nehéz lehet megindokolni a tanulónak, hogy miért pontosan azt a pontszámot kapta, ha az MI pontozási logikája nem teljesen átlátható. Fontos a tanulókkal ismertetni a szabályokat a tesztelés előtt!

„Képzeld el, ha a tanulás nem egy egyenruha lenne, amit mindenki megkap, hanem egy méretre szabott öltöny, ami tökéletesen illeszkedik a viselőjére! Az adaptív tanulási módszer pontosan ezt ígéri: egy forradalmi megközelítést, ahol a tananyag és a tempó a tanuló egyéni igényeihez és tudásához igazodik, maximalizálva ezzel a megértést és a hatékonyságot. Kedves pedagógustársaim! Tartsanak velem, és fedezzük fel (újra) együtt a személyre szabott oktatás jövőjét!”

## **MI AZ ADAPTÍV TANULÁSI MÓDSZER?**

Egy olyan oktatási megközelítés, amely a tanuló egyéni képességeihez és üteméhez igazítja a tananyagot és a tanítási módszereket. Ez azt jelenti, hogy a tananyag útja nem egyenes és előre meghatározott, hanem a tanuló teljesítménye, visszajelzései és tanulási stílusa alapján folyamatosan változik. Ennek érdekében technológiai eszközök és algoritmusok segíthetnek a tananyag személyre szabásában, így a tanuló haladhat a saját tempójában, míg a rendszer optimálisan alkalmazkodik az egyéni igényekhez. Munkám során végig szem előtt tartottam azt a pedagógiai alapvetést, miszerint *minden diák fejlődése mindig a saját képességei alapján kell, hogy megvalósuljon.*

## **KÖRNYEZET**

A kísérletben szereplő diákok hatodik osztályosok, 12-13 évesek. Helyszín a péterrévei Samu Mihály Általános Iskola. A megfigyelt tantárgy a matematika, a témakör pedig: racionális számokkal végzett műveletek (összeadás, kivonás, szorzás, osztás). Az óra típusa: tudásellenőrzés. A kísérletre a digitális tanteremben került sor, így minden tanuló egyéni számítógépen keresztül kapcsolódott be a munkába. A kísérletre kifejlesztett adaptív tesztet a videotanar.com oldalról használták a diákok. Szükséges volt még egy univerzális Google-fiók a bejelentkezéshez és a prompt futtatásához.

A teszt indítása (a „start” szócskával) után egy rövid üdvözlő szöveg várta a tanulókat. Ebben elolvashatták a legfontosabb szabályokat (pl. különböző típusú számok írása, értelmezése). Tudomásul vették, hogy nem feleletválasztós tesztről van szó, hanem a feladatok megoldásával kapott végeredményt kell megadni. Azt is elolvashatták, hogy a teszt alkalmazkodik a válaszaikhoz. Mivel a diákok már előzőleg is töltöttek ki hasonló jellegű digitális felmérőt, ezek a szabályok nem okoztak nekik gondot. Elindulhatott a tesztelés.

A feladatok 5 szintre lettek felosztva. Mindenki egy 3-as szintű (abszolút középszintű) feladatot kapott elsőre. A feladatok szinte minden tanuló számára különbözőek voltak,

hiszen valós időben generálódtak (volt egy-két ismétlődő feladat, de azok véletlenszerűen jelentek meg, nem egymás mellett ülő tanulóknál).

***A feladatok nehézségi szintjeit a Bloom-féle taxonómia elvei alapján kalibráltam.***

A 3. szint (Alkalmazás) használata lehetővé tette, hogy a rendszer azonnal differenciáljon: a magabiztos tudással rendelkezők felfelé, a hiányosságokkal küzdők pedig az alapok megerősítése felé indultak el.

Így, az első leadott válasz után a tanulók útjai szétváltak. Akik helyesen válaszolták meg a feladatot, nehezebb feladattal folytathatták. Akik rontottak, nekik a rendszer picit könnyített. Két helyes válasz után lehetett szintet lépni felfelé, két elrontott válasz után pedig lefelé. Hibás megoldás esetén a rendszer bátorította a diákokat és rámutatott a hibákra. Ha figyelmesen elolvasták az útmutatót, ez segített nekik, hogy a következő feladatoknál ne kövessék el újból ugyanazokat a hibákat.

A tanár itt kizárólag koordinátori szerepben vett részt. A tanulók a kapott feladatokat átírták a füzetekbe, megoldották, és a kapott végeredményt a kért formátumban visszaírták a rendszerbe. Ez így ment egészen addig, amíg meg nem oldották mind a 10 feladatot (vagy két sikeres, ötös fokozatú feladatot – ez automatikus befejezést jelentett).

A teszt végeztével a diákok egy értékelést kaptak. Ebben a rendszer külön dicsérettel jutalmazta a jól elvégzett területeket, és bátorított a botlások kijavítására.

A rendszer egyik külön specifikuma, hogy évszakfüggő. Attól függően, hogy a diákok mikor végzik a tesztet, az aktuális évszaknak megfelelő környezetbe helyezte őket (pl. „pontosan és magabiztosan kezeled a törteket, éppen olyan határozottan, ahogy a hólapát vágja a friss havat a kerti úton”). Ez külön motivációt ad a diákoknak a munka folytatására.

A teszt végén a rendszer összesíti a feladatokat, a leadott válaszokat, a helyes válaszokat és az adott feladat sikerességét. Így a pedagógusnak rálátása van az adott feladatokra, a diákok által elkövetett hibákra, és a hibák súlyosságára. Az utolsó szint valójában meghatározza a tanuló tudásszintjét.

A tesztelés különböző fázisaiban találkozhattunk hibás ellenőrzésekkel, téves szintlépésekkel. Ezek a hibák a rendszer tökéletesítésével és finomhangolásával elkerülhetők. A diákok a teljes tesztelést megosztották egy közös dokumentumban, így a pedagógus minden diák tesztjét, munkáját az első lépéstől az utolsóig nyomon követheti.

## **MI VAN A MOTORHÁZ ALATT? (A prompt bemutatása)**

Amint már az előzőekben említettem, a megoldás egy meglévő nagy nyelvi modell (pl. Gemini, ChatGPT vagy Claude) használata egy célzott System Prompt segítségével. Az én választásom a Gemini asszisztensre esett. Magát a promptot 6 fő részből építettem fel:

- 1. témakör: általános paraméterek:** Itt kaptak helyet a tiltott műveletek (pl. ne használjon gyökvonást és hatványozást, mert azt a diákok még nem tanulták), a szerepkör (tapasztalt matematikatanár), tulajdonságok (humoros, bátorító), stílus és a környezeti beállítások.

2. **témakör: globális matematikai szabályok:** Kötelező feladatarányok (tizedes és valódi/vegyes tört számok), negatív számok aránya, tanári megjelenítés és az elvárt formátumok.
3. **témakör: interakciós korlátozások:** Diákokra vonatkozó szabályok, reklamációs protokollok és tanári reakciók.
4. **témakör: szintek meghatározása:** Itt adtam meg, hogy az MI a különböző szinteken milyen feladatokat generálhasson. A szintek kalibrálása Bloom taxonómiája szerint történt, biztosítva, hogy minden tanuló a saját képességeihez mérten haladjon. Időbelileg ennek a résznek a megírása tartott a legtovább.
5. **témakör: működési logika:** Kezdő szint, szintugrasi szabályok, befejezési lehetőségek és egy rejtett állapotkövetés.
6. **témakör: interakciós folyamat:** Hogyan reagáljon a teszt különböző esetekben: bevezetés, feladatadás, visszajelzések, korlátozások, lezárás.

## ELEMZÉS

A kísérlet eredményeinek igazolására szükség volt egy klasszikus, többszintes feladatsorból álló papíralapú tesztelésre is. Az ott kapott eredményeket (illetve még a félévi jegyeket) hasonlítottam össze az adaptív tesztelés végén beazonosított tudásszintekkel.

TANULÓ	FÉLÉVI OSZTÁLYZAT	HAGYOMÁNYOS TESZT	MI TESZT	ELTÉRÉS: FÉLÉV - MI	ELTÉRÉS: HAGY. TESZT - MI
Tanuló 1	3	3	3	0	0
Tanuló 2	4	4	4	0	0
Tanuló 3	2	2	1	1	1
Tanuló 4	3	2	3	0	1
Tanuló 5	1	2	1	0	1
Tanuló 6	4	3	4	0	1
Tanuló 7	4	4	4	0	0
Tanuló 8	3	4	4	1	0
Tanuló 9	2	1	1	1	0
Tanuló 10	1	1	2	1	1
Tanuló 11	1	1	1	0	0
Tanuló 12	1	1	1	0	0
Tanuló 13	2	2	2	0	0
Tanuló 14	2	2	1	1	1
Tanuló 15	4	5	4	0	1
Tanuló 16	4	4	2	2	2
Tanuló 17	2	1	1	1	0
Tanuló 18	3	2	3	0	1
Tanuló 19	1	1	1	0	0
Tanuló 20	3	3	1	2	2

(A táblázat adatai alapján végzett elemzés:)

**Megfigyelés:** Az MI-alapú tesztelés eredményei átlagosan alacsonyabbak (2,20), mint a hagyományos teszté (2,40) vagy a félévi jegyeké (2,50). Ez arra utalhat, hogy az adaptív algoritmus szigorúbb értékelési görbét alkalmaz, vagy hatékonyabban szűri ki a véletlenszerűen eltalált válaszokat.

**Hagyományos teszt – MI teszt** Az összefüggés mértéke:  $r \approx 0,76$ . Ez erős pozitív korrelációt jelez. A különbség (eltérés) 11 tanulónál jelentkezett, ebből két esetben (Tanuló 16 és 20) kiugróan nagy, 2 egységnyi eltérés tapasztalható.

**Félévi osztályzat – MI teszt** Az összefüggés mértéke:  $r \approx 0,78$ . Meglepő módon az MI-teszt szorosabb korrelációt mutat a félévi átlaggal, mint a hagyományos, egyszeri teszttel. Ez azt sugallja, hogy az adaptív tesztelés jobban tükrözi a tanuló tartós tudásszintjét (amelyet a félévi jegy reprezentál), mint egy statikus, hagyományos feladatsor.

**Megjegyzések:** Amennyiben az MI-teszt adaptív jellege miatt „lefelé” kalibrálta a tanulót, az jelezheti a tudásban lévő strukturális hiányosságokat, amelyeket a hagyományos teszt (esetlegesen szerencsés tippelés vagy begyakorolt sémák miatt) nem tudott felszínre hozni. Ez a jelenség tipikusan a „szigetszerű tudásra” utal. A hagyományos teszteken a begyakorolt sémák mentén a tanuló jól teljesít, de az MI adaptív kérdései gyorsan rátapintanak a fundamentális hiányosságokra. Az egyetlen példa (Tanuló 10), ahol az MI-teszt jobb eredményt hozott (2), rejtett potenciálra vagy szorongásmentesebb teljesítményre utalhat.

*Megjegyzés: az elemzés a mesterséges intelligencia által generált javaslatok alapján készült*

## ÖSSZEGZÉS

### Mit tanultunk az MI-alapú mérésből?

A vizsgálat eredményei azt mutatják, hogy az MI-alapú adaptív tesztelés nemcsak kiváló alternatíva, hanem hatékony kiegészítője is a hagyományos módszereknek. A félévi osztályzatok és az MI-eredmények közötti szoros kapcsolat igazolja, hogy a rendszer pontosan méri a diákok tudását, miközben a tesztelés idejét több mint felére (45 percről 20 percre) csökkenti.

A hagyományos tesztekkel szemben az MI folyamatosan új és ismeretlen feladatokat generál, ami rákényszeríti a tanulókat a valódi gondolkodásra. Fontos azonban látni a korlátokat is. Az automatizált visszacsatolások pedagógiai haszna csak akkor válik valóra, ha a diákok valóban elolvassák és átgondolják a rendszer által adott személyre szabott javaslatokat.

**Végszóként:** Az adaptív tesztelés rengeteg időt és energiát spórol meg a tanárnak és diáknak egyaránt. Ugyanakkor hiba lenne kizárólag erre építeni. A pedagógiai munka akkor hiteles, ha változatos: az MI-technológiát érdemes a hagyományos módszerek mellett, a tanulók egyéni fejlesztésének támogatására használni.