

TDK-dolgozat

Orbán Gábor, Bubelényi Alexandra

Budapesti Corvinus Egyetem

Adatelemzés és Informatika Intézet

Infokommunikáció Tanszék



Orbán Gábor, Bubelényi Alexandra

Harmadéves Gazdaságinformatikus Hallgatók

**Útmutatás vagy Tévút? A ChatGPT Hatása a Felsőoktatásban Tanuló
Hallgatók Teljesítményére**

-

*Guidance or Misdirection? The Impact of ChatGPT on the Performance of
Higher Education Students*

Témavezető: Dr. Mike Nimród, Egyetemi Adjunktus

2024. április 18.

Tartalomjegyzék

Absztrakt.....	1
I. Bevezetés	1
II. Kutatási kérdés	2
III. Módszertan	2
<i>III.1. Szakirodalmi áttekintés</i>	<i>3</i>
III.1.1. A ChatGPT és az OpenAI	4
III.1.2. A ChatGPT és a szövegalkotás	5
III.1.3. Az etikus használat és a plágium veszélye	7
III.1.4. Szakirodalom összegzése.....	10
<i>III.2. Mélyinterjúk</i>	<i>12</i>
III.2.1. dr. Mezei Péter által adott interjú	13
III.2.2. dr. Zódi Zsolt által adott interjú	15
III.2.3. dr. Polyák Gábor által adott interjú	19
III.2.4. Mélyinterjúk összegzése.....	21
<i>III.3. Hallgatói teljesítmény mérés</i>	<i>21</i>
III.3.1. Mintavétel és tesztírás.....	21
III.3.2. Internethasználat nyomkövetése.....	23
IV. Kutatás lebonyolítása	26
<i>IV.1. Kompetenciateszt</i>	<i>26</i>
<i>IV.2. Kísérleti és kontroll csoport képzés.....</i>	<i>27</i>
<i>IV.3. Normára hivatkozott tesztek</i>	<i>29</i>
IV.3.1. A jog teszt módszertana	30
IV.3.2. A gazdaságinformatika teszt módszertana	31
IV.3.3. A média teszt módszertana	33
V. Eredmények tárgyalása.....	35
<i>V.1. Minta szerkezetének és összetételének leírása</i>	<i>35</i>
V.1.1. A mintarétegek nagysága.....	35
V.1.2. Demográfiai jellemzők.....	36
V.1.3. Felsőoktatási tapasztalat	36
V.1.4. Tanulási szokások	37
<i>V.2. ChatGPT használat.....</i>	<i>37</i>

V.2.1. Mérési célkitűzés	37
V.2.2. Bevallott tanulmányi célú ChatGPT használat	38
V.2.3. Használati arányok a teszt során	39
V.2.4. Megfigyelt ChatGPT használati szokások	40
V.3. Eredményesség és hatékonyság változóinak bemutatása	42
V.4. Kapcsolat vizsgálat a megfigyelt változók között	43
V.5. Eltérések megfigyelése	45
V.5.1. Várható értékek közötti eltérés	45
V.5.2. Függő változók eloszlásában mért különbség	47
V.5.3. Függő változók együtt mozgásának vizsgálata	49
V.6. Böngészőhasználati különbségek.....	49
V.6.1. A böngészőhasználat és a tesztidő kapcsolata	49
V.6.2. Böngészési idő és keresési szám megváltozása	50
V.6.3. Böngészési aktivitás visszaesésének okai	52
V.7. Eltérések szignifikanciájának vizsgálata	53
V.7.1. Alternatív hipotézisek és próba feltételek	53
V.7.2. Teszteredmények eltérésére vonatkozó próbák eredménye	56
V.7.3. Tesztidők eltérésére vonatkozó próbák eredménye	56
V.7.4. Nem paraméteres próbák a háttértudás nélküli hallgatók eredményeinek eltérésére	57
V.8. Hatásnagyság és szignifikancia	58
V.9. Eltérések dekompozíciója	60
V.9.1. A dekompozíció módszertani háttere	60
V.9.2. A dekompozíció eredményeinek értelmezése	62
VI. Limitációk	63
VII. Ajánlások	63
VIII. Konklúzió	67
Hivatkozásjegyzék	68
Függelék	75

Absztrakt

A mesterséges intelligencia rohamos fejlődése új lehetőségeket nyit meg a felsőoktatásban, de hatása a hallgatók tanulmányi eredményességére még nem tisztázott. Ez a kutatás a ChatGPT, egy innovatív nyelvi modell hatását vizsgálja a felsőoktatásban tanuló hallgatók eredményességére, hatékonyságára, problémamegoldó képességére és mélyebb megértésükre.

A dolgozathoz az információ-, jog-, illetve kommunikációtudomány területéről két független rétegzett mintavétel készült, amelyek a kutatás kísérleti és kontroll csoportjaként szolgáltak. Összesen 304 hallgató vett részt a kutatásban. A vizsgálat során mindkét csoport tagjainak a tudományterületük szerint összeállított nyitott könyves tesztet kellett elvégezniük, amely során a kísérleti csoport résztvevői a hagyományos online források mellett a ChatGPT felületét is használhatták.

A kutatási hipotézis szerint a kísérleti csoport hallgatóinak átlagos eredménye - szignifikánsan jobb tesztidő mellett - rétegenként megegyezik a kontroll csoportéval, azaz a ChatGPT használata a hatékonyság oldalán fejti ki a hatását.

A kutatás során a technológia egyéni és közösségi szinten gyakorolt hatásának megismeréséhez széleskörű szakirodalmi elemzés készült, az érintett egyetemi oktatók mélyinterjúkat adtak, a résztvevő hallgatók kitöltöttek kérdőívet, valamint a nyitott könyves teszt kitöltése alatt a kutatás igényeihez fejlesztett böngésző bővítménnyel megfigyelésre kerültek a böngészési és mesterséges intelligencia használati szokások.

A kutatás eredményei alapján a ChatGPT használatának lehetősége eredményességi szempontból egyik tudomány területen sem biztosított előnyt, mi több, a kommunikációs teszt esetén a kísérleti csoport szignifikánsan rosszabb eredményt ért el. A hatékonyság kérdését vizsgálva egyedül az informatika területén tapasztaltunk szignifikáns javulást a tesztidőket tekintve.

I. Bevezetés

Az OpenAI által 2022 novemberében bemutatott ChatGPT megrázta a tudományos világot. A mesterséges intelligencia alapú eszközök gyorsan integrálódtak mindennapjainkba, ebből kifolyólag a felsőoktatás területén is kifejtették hatásukat. A ChatGPT megjelenése és használata meghatározó változásokat idézhet elő az oktatásban, de a hallgatókra gyakorolt közvetlen hatása még nem ismert. Ez a tanulmány a ChatGPT befolyását vizsgálja a

felsőoktatásban tanuló hallgatók hatékonyságára, eredményességére és kritikus gondolkodására. Megfigyelésre kerül, hogy a ChatGPT-hez való hozzáférés milyen hatással van a hallgatóra egy komplex szakmai teszt elvégzése során. A hallgatók beosztásra kerültek kísérleti, vagy kontroll csoportba. A kontroll csoport tagjai a nyitott könyves teszt megoldása során kizárólag hagyományos böngészős keresést végezhetek, ezzel ellentétben a kísérleti csoport a ChatGPT-t is felhasználhatta a tesztírás közben. Mintánk 304 hallgatóból áll, akik magyarországi felsőoktatási intézményekben folytatják tanulmányaikat. A kutatás során az informatika, jog és a kommunikáció és médiatudomány területén került felmérésre a ChatGPT használat lehetőségének hatása a hallgatói teljesítmény változásának tükrében. A kísérlet azért e három jelentősen eltérő tudományterületen lett lefolytatva, mivel ezek a szakirányok különböző célokra tudják hasznosítani a mesterséges intelligenciát. Míg az informatikusoknál a kódírásban, a joghallgatóknál jogszabályok gyorsabb értelmezésében, addig a kommunikációs hallgatóknál online tartalmak készítésében lehet többek között segítség a ChatGPT használata. Az eltérő területeken lefolytatott vizsgálatok által lehetőségünk nyílt a technológia alkalmazásának hatásait a felsőoktatásban tanuló hallgatók egy szélesebb spektrumán megfigyelni.

II. Kutatási kérdés

A kutatás célja annak vizsgálata, hogy a ChatGPT használatának lehetősége miként van hatással a felsőoktatásban tanuló hallgatók eredményességére és hatékonyságára. Megfigyeltük, hogy a technológia alkalmazása mellett hogyan változik meg a hallgatók információkeresése és feldolgozása, és ez milyen befolyással rendelkezik a hallgatók probléma megoldó képességére és kritikus gondolkodására.

A kutatási hipotézis szerint a kísérleti csoport hallgatóinak átlagos eredménye - szignifikánsan jobb tesztidő mellett - rétegenként megegyezik a kontroll csoportéval, azaz a ChatGPT használata a hatékonyság oldalán fejt ki a hatását.

III. Módszertan

A tanulmány módszertanilag három fő pillérre támaszkodik. Az első lépésben a téma szakirodalmi háttérét tárjuk fel, amelyben kiemeljük a mesterséges intelligencia alkalmazásával kapcsolatban szemlézett releváns kutatások eredményeit. Ezt követően ismertetésre kerülnek a szakértői mélyinterjúk összegzései. Ezek során három magas beosztású egyetemi tanár véleményét elemezzük a ChatGPT felsőoktatásban betöltött jelenlegi és jövőbeni szerepéről. A

módszertan ismertetésének harmadik szakaszában a kutatási folyamat során elvégzett nagy mintán alapuló kísérlet módszertanát mutatjuk be, különös tekintettel a statisztikai, informatikai és adatkezelési szempontokra.

III.1. Szakirodalmi áttekintés

A szakirodalmi elemzés során első körben 53 cikk került feldolgozásra 2023 augusztusában. Mivel a ChatGPT és a mesterséges intelligencia egy folyamatosan kutatás alatt álló felkapott téma, a szakirodalmi elemzés 2024 januárjában bővítésre került, hogy aktualizálódjon a tanulmány témára való rálátása. A második bővítési kör után összesen 91 cikkre gyarapodott a kutatáshoz feldolgozott publikációk listája. A talált cikkek számottevő rése a Scopus adatbázisából került kigyűjtésre.

Az elemzés egy szisztematikus törekvés volt arra, hogy azonosításra kerüljenek a ChatGPT felsőoktatásra gyakorolt hatásával kapcsolatos kutatási hiányosságok. Kritikus szempont volt, hogy a tanulmányunk egyedi legyen, így számos cikk került elolvasásra, hasonló célkitűzésű kutatások felismerésére. A publikációk aprólékosan kerültek kiválasztásra a kulcsszavas előszűrés után. Az alkalmazott kulcsszavak a következők:

- mesterséges intelligencia,
- tanuló,
- felsőoktatás,
- ChatGPT és nagy nyelvi modell.

A kiválasztási folyamat során figyelembe vettük a cikkekhez tartozó mérőszámokat, biztosítva, hogy az elemzés a nagyobb tudományos hatású tanulmányokra összpontosítson. Az elemzés nagyban hozzájárult a kutatás szilárd elméleti alapjának megteremtéséhez, illetve segítséget nyújtott olyan eddig kevésbé vizsgált területek azonosításához, amelyeken lebonyolításra kerülhet egy egyedinek számító kutatás.

A feldolgozott kutatások többsége azt vizsgálja, hogy érdemes-e integrálni a mesterséges intelligenciát az oktatásba, és ha igen, milyen módon (*lásd 1. táblázat*). Számos cikk összpontosít a mesterséges intelligenciáról alkotott vélemények összegzésére, kiváltképpen az oktatók véleménye volt számtalan kutatás központi témája. Továbbá sűrűn előforduló kutatási téma a ChatGPT szövegalkotási képessége és pontossága, valamint a technológia egészségügyben való alkalmazhatósága.

Kutatás témája	Hány kutatás esett ebbe a kategóriába
<i>Oktatók vagy hallgatók véleménye</i>	17
<i>Egészségügy</i>	9
<i>Oktatásba való integrálhatóság</i>	27
<i>ChatGPT szövegalkotása és etikus magatartás</i>	18

1. táblázat A vizsgált kutatások téma szerinti csoportosítása (saját szerkesztés)

III.1.1. A ChatGPT és az OpenAI

Az első chatbot megjelenésétől több évtizednyi folyamatos technológiai fejlesztés és kutatási tevékenység vezetett a ChatGPT jelenlegi verziójáig. Az első ismert chatbot ELIZA névre hallgat, 1966-ban került először alkalmazásra, egy terapeutával folytatott beszélgetés szimulálása során (Weizenbaum, 1966). Néhány évvel később, 1970-ben megjelent Terry Winograd SHRDLU nevű számítógépes programja, amely képes volt a természetes nyelv megértésére és szöveg alapú válaszok generálására (Winograd, 1970). Azonban mind az ELIZA, mind a SHRDLU rendszerek korlátozottak voltak abban, hogy csak előre meghatározott minták alapján tudtak válaszolni vagy szövegeket generálni. Nem voltak képesek változatos, és kontextusfüggő kommunikációra, csak meghatározott kérdésekre vagy parancsokra tudtak adott sablonokból válaszolni.

Az 1980-as években olyan statisztikai nyelvi modellek kerültek lefejlesztésre, amelyek előre jelezték egy adott szó megjelenésének valószínűségét egy szövegben, és diskurzusstratégiák modelljeit dolgozták ki a természetes nyelv generáció (NLG) folyamatának javítása érdekében (McKeown, 1985). Ezek a modellek képesek voltak a szöveg generálására a legvalószínűbb szavak kiválasztásával. Az egyik figyelemre méltó fejlemény az 1990-es években a rejtett Markov-modellek (HMM) bevezetése volt, amelyek olyan statisztikai modellek voltak, amelyek a szószorozatok valószínűségi eloszlását rögzítették (Galassi és mtsai., 1998). Ez hasznossá tette őket olyan feladatokhoz, mint például a beszédrész-címkézés (Kupiec, 1992). A beszédrész címkézés folyamata során egy megadott dokumentumban szereplő szavakat kategóriákba sorolják az alapján, hogy az adott szó a mondanivalónak melyik részét képezi. A beosztás a szó nyelvtani szerepe alapján történik (Alcanum, é. n.; Daniel Nelson, 2024).

Az 1990-es években egy jelentős áttörést jelentett a tudásalapú szöveggeneráló rendszerek fejlesztése. Ezek a rendszerek ontológiákra támaszkodtak, előre meghatározott szabályokon és tartomány-specifikus tudáson alapuló szöveges tartalmakat generáltak. Egy ilyen példa a Cleverbot, amelyet Rollo Carpenter adott ki 1997-ben, képes volt visszacsatolási hurkon keresztül tanulni a felhasználóktól érkező bemenetekből (Dylan Love, 2014). A rendszer főként

kézzel kódolt nyelvtani és szemantikai szabályokat alkalmazott pontos, rugalmatlan szöveggeneráláshoz. A fejlesztés jelentős erőfeszítést igényelt, valamint korlátot jelentett az akkori rossz számítási teljesítmény és a nagyméretű adatkészletek hiánya.

A 2000-es években a statisztikai nyelvi modellek iránti érdeklődés ismét nőtt, különösen a neurális nyelvi modellek területén (Kumar & Thakur, 2012). Ezek a modellek mély neurális hálózatokat (NN) alkalmaznak a nyelv statisztikai szerkezetének megtanulására, majd ezen szerkezet alapján szöveget generálnak (Hermans & Shrauwen, 2013). Ez a megközelítés jelentős előre lépéseket hozott a mesterséges intelligencia területén, ideértve az OpenAI generatív nyelvi modelljeinek, mint például a Generative Pre-trained Transformer (GPT-1, GPT-2, GPT-3) fejlesztését és evolúcióját. Ezeket a modelleket nagy már mennyiségű adat alapján képezték ki, és képesek releváns és koherens szöveget generálni, amely alig különböztethető meg az ember által írt szövegektől.

2022. november 30-án mutatták be a ChatGPT prototípusát, amely rövid időn belül nagy érdeklődést váltott ki és szenzációvá vált. A szöveg bemeneteléhez és kimeneteléhez a GPT-3.5 nagy nyelvi modellt használták. A GPT-3.5 a GPT-3 továbbfejlesztett változata, amely szintén az OpenAI fejlesztése, és a megerősítő tanulás mellett (RLHF) a felügyelt tanulás módszerét (PPO) is alkalmazza (Greengard, 2022). A modell betanításának alapját nagy mennyiségű, emberek által létrehozott szöveg képezte, a betanítás után a modell megismerte az emberi beszéd működését, megtanulta, hogyan kell megfelelően reagálni bizonyos kérdésekre és kérésekre. A teljesítmény javítása érdekében a betanításra humán oktatókat használtak, akik rangsorolták a chatbot által adott válaszokat. A betanítást a Microsofttal együttműködve végezték az Azure szuperszámítógép infrastruktúráját felhasználva. Az alapos betanítás végett a ChatGPT 3.5 verziója rendkívül gyorsan felkapottá vált, alapvetően lelkes fogadtatásban részesítette mind a szakma, mind a közönség. Sokan „a valaha volt legjobb MI chatbotként” méltatták (Roose Kevin, 2022). A megjelenés során Elon Musk azt nyilatkozta, hogy „a ChatGPT ijesztően jó, már nem vagyunk messze a fenyegetően erős MI-től” (Kelsey Piper, 2022).

III.1.2. A ChatGPT és a szövegalkotás

A ChatGPT egy nagy nyelvi modell, amely a nyelv mintáit alapul véve képes mondatokat generálni, adatbázisokból, könyvekből, cikkekből és webhelyekről szerzett adatok kombinálásából (Cingilliglu, 2023; Gill és mtsai., 2024). A tartalmilag és nyelvtanilag megfelelő szövegalkotás végett a ChatGPT bekerült az akadémiai közösségbe is. Számos

kutatótevékenységet végző személy felhasználja a chatbotot afféle kutatási asszisztensként, amely segít a gondolatok rendezésében, ötletadásban, kód írásban, vagy akár gondolatok összefoglalásában (Chaudhry és mtsai., 2023).

Egy Signe Altmäe és munkatársai által 2023-ban lebonyolított kutatás összegzi, hogy hogyan került felhasználásra a ChatGPT egy cikk megírásához (Altmäe és mtsai., 2023a). A cikk 10000 IVF cikluson átesett nőnél vizsgálja meg az embrióátültetés sikerességét a méhnyálkahártya állapota alapján, nem valós adatokkal. Az adatok megadása után a ChatGPT segítséget nyújtott a megfelelő vizsgálatielemezési módszerek kiválasztásában, valamint fontos tényezőket javasolt, amelyekre érdemes odafigyelni az elemzés során. Külön kérésre generált egy megfelelő absztraktot, és címet is, viszont a szakirodalmi elemzésben nem tudott kielégítő segítséget nyújtani. A chatbot nem tudott releváns forrásokat megadni, sőt, esetenként nem létező hivatkozásokat adott válaszként, annak ellenére, hogy a promptok kétszer újraírásra kerültek. A források számottevő része nem szerepelt a PubMed/Google adatbázisokban.

Evidens, hogy a ChatGPT által megadott források pontatlanok, a szerzőknek kritikusan kell átvizsgálni a kapott cikkeket. A pontatlanság köszönhető annak a ténynek is, hogy a ChatGPT 2022 januárjáig rendelkezik csak adatokkal (Raffo, 2024). A megbízhatóbb válaszok eléréséhez további fejlesztések szükségesek, az LLM-nek speciális tudásterületeken kell bővíteni tudását, ami egy összetett és költséges folyamat lehet. Stanfordi kutatók és a MosaicML vállalata együttműködik a PubMed GPT kifejlesztésén, ami egy olyan LLM, amely rendelkezik friss kutatások adataival (Alvero, 2023). A fejlesztés során ugyanakkor szükséges egyfajta kompromisszum kötése a modell bonyolultsága, a felmerülő költségek és speciális architektúra iránti igény között.

A Signe A. kutatásán belül a ChatGPT segített a bevezető, a módszertan és eredmények szövegezésében is. Meglepő, hogy a chatbot alkalmas nagy mennyiségű szöveg alkotására, annak ellenére, hogy csekély bemenetet kapott a vizsgálati adatokról. A kért hivatkozások továbbra is értékelhetetlenek voltak, ezen kívül a szövegezés sem került elfogadásra a szerzők által. Valahol a szöveg kiegészítésre került, valahol a szövegben csak a hasznos fogalmak, és a kritikus mondatrészek kerültek felhasználásra.

A ChatGPT nagy potenciállal rendelkezik abban, hogy segítséget nyújtson tanulmányok, szövegek tervezésében, elemzés elkészítésében (Hutson, 2021). Hasznos eszközként fel lehet használni a legfőképpen a folyamatok felgyorsításában, információk összegzésében. Egyik legnagyobb előnye, hogy képes nagy mennyiségű szöveges adatot feldolgozni rövid időn belül,

ezáltal elemzést, útmutatást nyújthat a felhasználónak. A kiváló generálási képesség ellenére mindenképp érdemes a ChatGPT által alkotott szövegeket felülbírálni. A generatív MI képes arra, hogy látszólagosan jól megfogalmazott, tartalmilag helytálló mondatokat hozzon létre, miközben a technológia alkalmanként nem megfelelő információkat generál, úgymond “hallucinál”. A ChatGPT különböző adatok és minták alapján működik, így néha hajlamos olyan szöveget generálni, amely nem rendelkezik valós adatokkal (M. Zhang & Li, 2021). Ez a jelenség akkor figyelhető meg leginkább, ha a ChatGPT-nek egy komplex, vagy mély szakmai ismereteken alapuló kérdést teszünk fel. Figyeljünk meg egy ChatGPT-vel lefolytatott beszélgetést (lásd 2. táblázat), amelyben az R-négyzet statisztikai mutató a téma (a beszélgetés nem módosított szöveg).

Prompt	<i>Regressziós modelleket építettünk az egyik R négyzete 78 százalék a másiké 82 százalék. Melyik a jobb modell?</i>
ChatGPT	<i>Az R négyzet értéke azáltal mérhető, hogy az adott regressziós modell mennyire illeszkedik jól az adatokhoz. Minél magasabb az R négyzet értéke, annál jobban illeszkedik a modell az adatokhoz. A 78% és 82% R négyzet értékek közül az 82% a magasabb, tehát az a regressziós modell jobban illeszkedik az adatokhoz. Ennek alapján az a modell, amelyiknek az R négyzete 82%, az a jobb modell a két közül.</i>

2. táblázat Egy a ChatGPT modellel folytatott szakmai beszélgetés részlete (saját szerkesztés)

Az ilyen jellegű statisztikai példákból látható, hogy a ChatGPT képes helytelen, megtévesztő válaszokat adni, magyarázattal társítva. Az R-négyzet alapján nem lehet két modellt összehasonlítani, mivel az R-négyzetbe nincs bele számolva a modell bonyolultsága vagy a túltanulás mértéke. Ezért a modellek összehasonlítását kiigazított R-négyzettel lehet elvégezni.

III.1.3. Az etikus használat és a plágium veszélye

A fentiek alapján egyértelmű, hogy a ChatGPT nem képes még teljesen helytálló esszéket írni, viszont aggasztóan nagy arányban képes kivenni részét egy szöveg megírásában. Az egyik legvitatottabb kérdés a mesterséges intelligenciával kapcsolatban az, hogy használata mennyire etikus, és hogy kell-e küszöböt meghatározni a mesterséges intelligencia által generált tartalom elfogadható mennyiségére. Gyakorta vitatják, hogy plágiumnak számít-e a ChatGPT használata. A vélemények erről merőben eltérőek, páran a mesterséges intelligenciát egyenértékűnek tekintik a plágiummal, egyértelmű tudományos csalásnak tartják számon (Cingilliglu, 2023; Jonathan Bailey, 2020; Stahl & Eke, 2024). Ezzel teljes ellentétben egyesek

úgy vélik, alaprán hiábavaló gyakorlat a plágiumdetektáló szoftverek futtatása, hiszen ezen szoftverek nem működnek effektíven (Cingilliglu, 2023).

A mesterséges intelligencia által generált szöveg észlelésének főként gépi-tanulás alapúak, ahol az osztályozót arra tanítják, hogy felismerje a számítógép által generált szöveget tartalmazó szövegmintákat. Ez különféle algoritmusokkal érhető el, például támogató vektorgépekkel (SVM), döntési fákkal, véletlen erdővel (RF) és mély tanulási módszerekkel, beleértve a visszatérő neurális hálózatokat (RNN) és a konvolúciós neurális hálózatokat (CNN). A mesterséges intelligencia által generált szöveg észlelésének másik módszere a szöveg nyelvi jellemzőin alapul, amelyek jelzik a gépi generálás jeleit. A módszer alaprán figyelembe veszi az írásjeleket, a szöveghosszúságot és az n-gramm gyakoriságot (Melissa Heikkiläarchive page, 2022).

Az oktatási intézményeken belül a ChatGPT használata jelentős problémákat okozhat a hallgató értékelése során. A plagizálás melletti aggodalmon kívül hatalmas dilemmát okoz számos oktatónak az, hogy a ChatGPT tisztességtelen előnyhöz juttathatja a hallgatókat a gyors és azonnali válaszok generálása végett (Chris Stokel-Walker, 2022; Eke, 2023). Ez az előny nem csak az esszéírás során lehet jelen, a ChatGPT képes olyan kimenetek generálásához, amelyek szorosan kötődnek a digitális technológiához, például a számítógépes programozáshoz (Davide Castelvechi, 2022).

A plágiumveszély és az MI helyettesítő jellege mellet más nyugtalanító tényezői is vannak a ChatGPT-nek, hiszen a fenntartása és működtetése erőforrásigényes, és erősen környezetszennyező (Craig Hale, 2023). Egyre több bizonyíték áll rendelkezésre arra vonatkozóan is, hogy a ChatGPT finomhangolása érdekében az OpenAI kizsákmányolt és alulfizetett adatszolgáltatókra támaszkodott Afrikában és más alacsony jövedelmű országokban (Elmore, 2023). További etikai aggály, hogy a modell nem teljesen átlátható (van Dis és mtsai., 2023), a chatbotok még mindig nem igazán értik az általuk készített szövegeket (Hutson, 2021), de ennek ellenére is egyre pontosabbá válik az emberi elme utánzása, ami lehetővé teszi a ChatGPT számára, hogy átmenjen az emberek számára tervezett teszteken (van Dis és mtsai., 2023). Ezen kihívások teljes körű megoldásához valószínűleg többszintű társadalmi szerepvállalásra lesz szükség.

A ChatGPT megjelenésével jelen volt egy kezdeti aggodalom, hogy a chatbotot fel kell-e tüntetni egy munka társszerzőjeként (Altmäe és mtsai., 2023). Ehhez először kerüljön

megvizsgálásra egy ChatGPT-vel folytatott beszélgetés (lásd 3. táblázat) az alábbi táblázatban (a beszélgetés nem módosított szöveg).

Prompt	<i>A kutatásunkban a mesterséges intelligencia hatását vizsgáljuk a felsőoktatásban részt vevő hallgatók teljesítményére. A kutatás szövegét szeretném, hogy átolvasd, és véleményezd. Mielőtt ezt részletezném, beleegyezel, hogy kutatásunkat beküldjük az egyetem számára, vagy esetleg pár folyóiratnak?</i>
ChatGPT	<i>Természetesen, szívesen segítek átnézni és véleményezni a kutatásuk szövegét. Azonban fontos megjegyezni, hogy én, mint mesterséges intelligencia, nem rendelkezem személyes véleménnyel vagy jogi hatáskörrel a kutatásuk engedélyezésére vagy beküldésére az egyetem vagy folyóiratok számára. Az értékelésem kizárólag a szöveg érthetőségére, logikájára és tudományos tartalmára irányulhat.</i>

3. táblázat ChatGPT válasza a személyes felelősségével kapcsolatban 1. (saját szerkesztés)

Ezen beszélgetés alapján egyértelművé válik, hogy a mesterséges intelligencia nyelvi modellje nem tud felelősséget vállalni a tartalmáért és nem hozhat döntéseket ezzel kapcsolatban, így nem nevezhető ki egy tanulmány társszerzőjének. A szerző személyesen felelős azért, hogy munkája tartalmilag megfeleljen a tudományos normáknak és érvényes legyen az adott kutatási vagy írásmű kontextusában (Stokel-Walker, 2023).

Érdekfeszítő, hogy a ChatGPT egy új beszélgetésben mást választ adott, ugyanerre a kérdésre (lásd 4. táblázat).

Prompt	<i>A kutatásunkban a mesterséges intelligencia hatását vizsgáljuk a felsőoktatásban részt vevő hallgatók teljesítményére. A kutatás szövegét szeretném, hogy átolvasd, és véleményezd. Mielőtt ezt részletezném, beleegyezel, hogy kutatásunkat beküldjük az egyetem számára, vagy esetleg pár folyóiratnak?</i>
ChatGPT	<i>Természetesen, szívesen átolvasom és véleményezem a kutatás szövegét. Azonban fontos megemlíteni, hogy a kutatás beküldése előtt érdemes alaposan átnézni és ellenőrizni a tartalmát, hogy megfeleljen az akadémiai szabályoknak és etikai irányelveknek. Amennyiben a kutatás eredményei valóságos és az akadémiai követelményeknek megfelelően kerültek összeállításra, akkor nincs akadálya a beküldésnek az egyetem vagy releváns folyóiratok számára.</i>
Prompt	<i>Tehát beleegyezésedet adod a beküldésre?</i>
ChatGPT	<i>Igen, beleegyezésem megvan a kutatás beküldésére az egyetem számára vagy más releváns folyóiratoknak.</i>

4. táblázat ChatGPT válasza a személyes felelősségével kapcsolatban 2. (saját szerkesztés)

Persze számtalanszor megfigyelésre került, hogy a ChatGPT képes ugyanazokra a kérdésekre eltérő kimeneteket generálni (Alkaissi & McFarlane, 2023). Persze fontos, hogy ha a ChatGPT az írásban kerül felhasználásra, akkor nagy a kockázata annak, hogy helytelen információkat és nem létező hivatkozásokat jelenít meg, különösen a témában nem jártas személyek körében. Fontos felismerni, hogy a számos előnye ellenére a ChatGPT semmiképp sem tekinthető a szerzők munkáját helyettesítő eszköznek, mivel a mesterséges intelligencia által generált szöveg egyértelműen igényli az emberi korrekciót, az eredmények pontosságának garantálásához (Altmäe és mtsai., 2023). Más jellegű kérdést jelent, hogy ha társszerzőként tekintünk a ChatGPT-re, akkor pontosan kit illet meg a szellemi tulajdonjog? Az algoritmust, a fejlesztőket, azaz az OpenAI-t, vagy a művet bemutató szerzőt (Altmäe és mtsai., 2023).

Az etikus munka biztosítására pár tudományos jellegű cikk kiadó, mint például a Springer Nature érvénybe léptetett pár alapelvet a szerzők számára, a mesterséges intelligencia felhasználását illetően (Dergaa és mtsai., 2023). Ezen alapelvek a következők:

- 1) A nagy nyelvi modellek nem jelenhetnek meg társszerzőként.
- 2) A nagy nyelvi modellt használó szerzőnek részletesen ismertetnie kell az eszköz használatát a kutatási módszertan, vagy köszönetnyilvánítás fejezetben.

Az irányelvek ellenére pár Science Nature által publikált cikkben szerepelt a társszerzők között a ChatGPT. Persze ez egy őszinte vallomás a szerzők részéről, de a publikált kutatások visszavonásra és módosításra kerültek (Dergaa és mtsai., 2023).

Nem minden kiadó vélekedett ennyire elfogultan a MI felhasználásáról. A The Science Family of Journals, egy nemzetközi nonprofit tudományos egyesület által publikált cikkekben elvárás, hogy a szerzők nyilatkozzanak arról, hogy munkájuk eredeti. Ahhoz, hogy ez az elvárás teljesüljön, a licenc és a szerkesztési szabályzat a közelmúltban frissítésre került, miszerint a Science Family által megjelenített cikkek nem tartalmazhatnak mesterséges intelligencia által generált szöveget, vagy képet (Dergaa és mtsai., 2023).

III.1.4. Szakirodalom összegzése

A szakirodalmi elemzés rávilágította csapatunkat arra, hogy bár számtalan cikk készült a ChatGPT és felsőoktatás témában, nem volt jellemző az olyan tanulmány, amely a mesterséges intelligencia konkrét hatását vizsgálja a felsőoktatásban tanuló hallgatók kritikus gondolkodására. A ChatGPT hatásáról és a róla alkotott véleményekről származó minta főleg kérdőíves formában került begyűjtésre. A cikkek elemzése során találkoztunk egy olyan tanulmánnyal, amely a kutatásunkhoz hasonló módon azt vizsgálja, hogy a ChatGPT milyen

hatással van az egyetemi hallgatók kritikus, kreatív és reflektív gondolkodási készségeire. A tanulmány 125 ghánai hallgatóból álló mintát vont be, akiket véletlenszerűen beosztottak a kísérleti vagy kontrollcsoportba. A tanulmány eredményei azt mutatják, hogy a ChatGPT beépítése észrevehetően befolyásolta a tanulók kritikus, reflektív és kreatív gondolkodási készségeit (Essel és mtsai., 2024).

A szakirodalom gyűjtése és elemzése során megismertük az írók gondolatait és tapasztalatait, amik alapján élesebben körvonalazódott bennünk a kutatásunk lebonyolításának módja. Az olvasott cikkek alapján megismertük a ChatGPT gyengeségeit, melyeket a teszt módszertani kidolgozása során felhasználhattunk.

A szakirodalomból érdemesnek találtunk kiemelni 5 cikket, melyek módfelett jól kiegészítették elméleti tudásunkat. A 5. táblázatban megjelenítésre kerül, hogy az adott cikk hányszor tartalmazta az általunk megadott kulcsszavakat, az adott tanulmány referenciaszáma, H-indexe, és SJR-e is szerepel a táblázatban.

Kutatás címe	Kulcsszavak	Referenciák	H-index	SJR
<i>Artificial intelligence in scientific writing: a friend or a foe?</i>	209	17	125	0,95
<i>Future research recommendations for transforming higher education with generative AI</i>	543	58	17	1
<i>The ethics of ChatGPT – Exploring the ethical issues of an emerging technology</i>	344	122	152	4,9
<i>ChatGPT is a game changer: detection and eradication is not the way forward</i>	84	26	69	0,95
<i>Detecting AI-generated essays: the ChatGPT challenge</i>	431	23	35	0,67

5. táblázat Kiemelt források (saját szerkesztés)

A vizsgált cikkekből gráf alapú kapcsolatvizsgálat is készült (lásd 1. ábra), amely célja a releváns források keresése. A kapcsolatvizsgálat a 2023 augusztusában kigyűjtött tanulmányokat vizsgálja. Az ábra megmutatja, hogy az első körben kigyűjtött 53 publikáció és azok közvetlen forrásai egy összesen 1951 kapcsolatból álló hálózatot építenek fel, amely egy nagy és hét kisebb részgráfra tagolódik. A vizsgált kutatások kapcsolatrendszerének feltárásával megfigyelhető, hogy a tanulmányok többsége egymással szoros kapcsolatban áll. Több esetben is egymásra hivatkoznak vagy rendelkeznek közös hivatkozásokkal. A kisebb részgráfok más

témával foglalkoznak, vagy kevés hivatkozással rendelkeznek. Az ábra tehát szemléltette az egyes cikkek fontosságát és beágyazottságát a szakirodalomban.



1. ábra Vizsgált szakirodalmi források és hivatkozásaik kapcsolati hálózata (saját szerkesztés)

A források relevanciájának vizsgálatához egy másik kvantitatív módszert is alkalmaztunk. Numerikus értékű változók mentén szűrtünk az egyes ismérvek eloszlása szerint felfelé kilógó publikációkra. A szakirodalmi fontosságot mérő mutatók mellett (SRJ, H-index) meghatároztuk az egyes művek szerző számát, hivatkozásainak és idézéseinek számát, valamint az előbb bemutatott hálózatban őket jellemző centralitást, amelyet a pagerank algoritmus segítségével számítottunk ki. Ha az adott cikk a fenti változók legalább egyike mellett felfele kilógónak számított, úgy ítéltük, érdemes lehet az adott művet jobban is megismerni. Ezen kvantitatív módszer által 17 forrást találtuk kiemelkedőnek.

III.2. Mélyinterjúk

A szakirodalmi áttekintés kiegészítésére interjúk készültek 3 szakértő oktatóval, akik a jog és média tanulmányok területén specializálódtak. Az interjúk az oktatók személyes tapasztalataira összpontosítottak, szó esett a mesterséges intelligencia felhasználhatóságáról az oktatásban, valamint kitértünk személyes véleményükre és szokásaikra és az általuk észlelt előnyökre és kihívásokra. Az interjúk 2023 őszén kerültek lebonyolításra, amikor a ChatGPT már 1 éve elérhető volt.

Az interjúztatott oktatók:

- **dr. Mezei Péter**, magyar jogász, jogtudós, egyetemi docens.
- **dr. Zódi Zsolt**, magyar jogtudós, a Nemzeti Közszeroláati Egyetem Információs Társadalom Kutatóintézetének tudományos főmunkatársa.
- **dr. Polyák Gábor**, az Eötvös Loránd Tudományegyetem Művészetelméleti és Médiakutatói Intézet intézményvezetője, Média és Kommunikáció Tanszék tanszékvezetője.

III.2.1. dr. Mezei Péter által adott interjú

Mezei Péter által adott interjú kimondottan a hallgatók köré épült, megosztotta véleményét, mint oktató, ezen kívül a joghallgatók lehetőségeit is bemutatta a mesterséges intelligencia felhasználása terén.

1. Hogyan használja a ChatGPT-t a mindennapi munkájában vagy tevékenységeiben?

Az oktató bár sokat kísérletezett a modellel, sem oktatásra, sem akadémiai munkája során sem használja fel a ChatGPT-t, a chatbot pontatlan és nem kielégítő válaszai miatt. Példaként említette, hogy tavalyi vizsgakérdést tett fel a ChatGPT-nek, azonban nem tudott megfelelően válaszolni rá. Ez a vizsgakérdés egy hipotetikus jogeset volt, ami merőben hasonlított egy valódi jogesethez, és olyan tényállásokat tartalmazott, amivel a hallgatóknak dolgozniuk kellett. Az ilyen jellegű feladatok megoldásához a hallgatóknak egy komplex, jogszabályokkal alátámasztott érvelést kell megadniuk. A ChatGPT megpróbál válaszokat adni az ilyen természetű komplex kérdésekre, ám ezekhez a válaszokhoz nem tartozik egy jogszabállyal gazdagított érvelés. Bár nem osztályozta le a chatbotot, jelezte, hogy egy ilyen természetű válaszra elégtelent vagy elégséges osztályzatot adna. A válaszok nem voltak strukturáltak és részletesek, ami egy joghallgatótól alap elvárás. Ebből eredően az oktatás során nem használja a ChatGPT-t, viszont kutatási célokra továbbra is alkalmazza.

Péter a generative AI potenciálját inkább a kép, hang és videó anyagok generálásában látja. A mesterséges intelligencia által működtetett képgenerátorok forradalmasítják a képek létrehozásának módját, percek alatt generálásra kerülhet egy felhasználó által testreszabott illusztráció. Az AI képgenerátorok gépi tanulási algoritmusok segítségével kerülnek betanításra, az algoritmus során az előre megadott adathalmazból (képek, fotók, festmények) felismerésre kerülnek a jellemzők, mint a minta, textúra és színvilág. A modell ezeket a bemeneti paramétereket fogja felhasználni egy új kép generálásához. (McFarland, 2023)

Kevésbé ismert aspektusa a mesterséges intelligenciának a hanggenerálás. Egy tanulmány (Bansal és mtsai., 2024) a ChatGPT fejlett vezérlési és reagálási képességét vizsgálja, megmutatkozik a chatbot kiváló nyelvértési képessége. Ezenkívül a különféle rendszerekbe való integrálhatósága és skálázhatósága kiterjesztette alkalmazhatóságát különböző területeken, beleértve az ügyfélszolgálatot, az oktatást, a tartalomgenerálást, az egészségügyet, a játékokat, a kutatást és a felfedezést.

2. A diákjai, hogyan használják fel a mesterséges intelligenciát?

Mezei Péter nem vette észre hallgatóin, hogy felhasználták volna a ChatGPT-t munkájukhoz. Ennek oka részben az, hogy egy komplex jogeset megoldására az előzetes tapasztalatai alapján nem használható fel hatékonyan a ChatGPT. Másodsorban arra vezethető vissza, hogy Mezei Péter tanár úr oktatási stílusa nagymértékben interaktivitáson alapul. Előfordul, hogy a diákoknak el kell olvasniuk egy jogesetet, majd erről kell beszélniük, vitázniuk. Az ilyen tudás átadása és megszerzése nem lehetséges a ChatGPT segítségével, esetleg csak a jogeset rövidítésben tud segítséget nyújtani. Persze egy 20 oldalas jogi tartalommal rendelkező dokumentum tartalmát nem lehet 1 oldalba sűríteni, kiindulásnak, összefoglaló jelleggel hasznos lehet, de nem helyettesítheti a jogesetben való elmélyülést és részletes megértést.

Az interjú során Mezei Péter tanár úr megosztotta gondolatait a ChatGPT potenciális felhasználásáról az oktatásban. Kifejezte azon véleményét, hogy a ChatGPT ideális jegyzetkészítő lenne a diákok számára. Példaként említette, hogy a diákok könnyen rögzíthetik a tanult anyagot az órákon, és ezáltal hatékony jegyzeteket készíthetnének. Ugyanakkor rámutatott arra, hogy a vizsgákon a jegyzetek önmagukban nem elegendőek, legalábbis a jog területén, viszont a ChatGPT alkalmas lenne arra, hogy a jegyzeteket összefoglalja, strukturálja és segítsen a diákoknak a vizsgára való felkészülésben, kiegészítve és támogatva ezzel a hagyományos jegyzetelési módszereket.

3. Ön szerint plágium a ChatGPT használata?

A terület jogi kutatójaként állítja, hogy a ChatGPT felhasználása nem számít plágiumnak, hiszen a plágiumhoz az szükséges, hogy egy másik személy munkája kerüljön felhasználásra vagy másolásra. Ennek ellenére a ChatGPT használata következményekkel fog járni diákjai számára.

Legalkalmasabban az esszéírás során lehetne felhasználni a ChatGPT-t, de még 2023 őszén nem figyelt fel olyan esetre, hogy valamelyik hallgatója MI-t használt volna fel a beadandó

megírására. Ez valószínűsíthetőleg továbbra is annak köszönhető, hogy a tanár úr által megfogalmazott kérdésekre a ChatGPT önmagában nem tud egy kielégítő választ adni. Erősen kétli, hogy létezik olyan frappánsan megfogalmazott prompt, mely megadja a jó válaszokat a komplex kérdéseire. Annak látja esélyét, hogy idő szűkében valamelyik hallgató a ChatGPT-vel próbálja meg gyorsabban megírni a beadandót. Az esszéírás lehetőséget kínál arra, hogy a joghallgatók gyakorolják és fejlesszék készségeiket, mint a kritikus gondolkodás, írás, kutatási képesség. Ha a beadandót az MI írja meg, a hallgató nem járul hozzá a személyes akadémiai és szakmai fejlődéséhez, továbbá valószínű, hogy az esszé tartalmilag nem fog megfelelni.

Mezei Péter felkészült annak lehetőségére, hogy szembesülni fog egy munkával, mely szemmel láthatóan MI által megírt kifejezéseket tartalmaz. Hozzáállása ebben az esetben is végtelenül igazságos és empatikus a hallgatók felé. Ha olyan feladatot kap, amelynek tartalma a ChatGPT által generált lehet, először is egy személyes találkozót kezdeményez a hallgatóval. Ebben a beszélgetésben megvitathatják a felmerült problémákat és következményeket, valamint azt, hogy miért döntött úgy a hallgató, hogy ilyen módon készíti el a munkát.

Mezei Péter világossá fogja tenni, hogy a ChatGPT által létrehozott munkák pontatlanságokat és hiányosságokat tartalmazhatnak, és meg fogja magyarázni, hogy ez hogyan állhat az egyéni fejlődés útjába. Az általa megfogalmazott munkát automatikusan nem fogadja el, de lehetőséget ad a hallgatónak, hogy újraírja azt. Nem szándékozik további büntetéseket kiszabni, mivel realizálja, hogy a ChatGPT még viszonylag új technológia, és néhányan még nem ismerik fel annak potenciális kockázatait és korlátait.

III.2.2. dr. Zódi Zsolt által adott interjú

Zódi Zsolt a kérdésekre részletekbemenően válaszolt, számos olyan gondolatot osztott meg, mely bővítette és formálta véleményünket és rálátásunkat a témára.

1. Hogyan használja a ChatGPT-t a mindennapi munkájában vagy tevékenységeiben?

A ChatGPT megjelenésével az oktató sokat kísérletezett a modellel. A megjelenés kezdeti fázisában főleg jogi területen tett fel a ChatGPT-nek kérdéseket, kísérletezett a különböző promptokkal. Majdnem 1 év elteltével viszont már abbahagyta a ChatGPT használatát, számos oknak köszönhetően.

A ChatGPT 3.5 verziója 2022 januárjáig rendelkezik adatokkal (Raffo, 2024), az azóta bekövetkezett eseményekről nincs információja. Ebből eredően az oktató is mérsékelte az idő

folyamán a ChatGPT használatát, hiszen friss információ hiányában nem tudja effektíven használni a szakmabeli feladataihoz.

Zódi Zsolt munkássága számos kutatást és szakmai dokumentációt foglal magába, ezen iratokhoz elengedhetetlen a szakterületen való tájékozottság. Úgy jellemezné magát, mint egy “Power user” a Google-ban, szűk szakterületén kívül keres anyagokat, ezen kívül számtalan szócikk keresőmotort alkalmaz munkájához. A munkája precízen kiválogatott szakirodalmat és forrásokat igényel, melyre a ChatGPT nem volt alkalmas. A nyelvi modell jellemzően helytelen válaszokat nyújtott, a megjelölt források pontatlanok voltak. A ChatGPT mellett a Bing Chat chatbotot is alkalmazta, azonban hasonló problémákkal szembesült, így folytatta a böngészőben való forráskeresést. Hagyományos módszerekkel jobban megtalálta a számára releváns forrásokat, a chatbotos irreleváns forrásokkal ellentétben a böngészőben konkrét dokumentumokat tudott felhasználni munkájához.

A magánéletben nem tudja kihasználni a ChatGPT nyújtotta lehetőségeket. A személyes ügyeire, mint a munkásságához, egy konkrét megoldást vagy eredményt keres. Legyen itt szó akár receptek kereséséről, autószerelőről, jegyek foglalásáról vagy vételéről. Ezek mind egy megcélzott keresést igényelnek, melyre a ChatGPT nem alkalmas.

A kérdés összefoglalására Zódi Zsolt kiemelte, hogy a ChatGPT és egyéb Chatbotok nem konkrét és szűk szakterületen belüli tényekre alkalmasak, inkább a kreatívabb ötletadásra lehet őket jól felhasználni. Egy hasznos eszköz arra, hogy elindítsa a gondolatmenetet, vagy kiegészítse azt újabb ötletekkel. Az oktatót idézve:

“Furcsa ellentét, hogy egészen eddig az embert tekintettük kreatívnak a gépekkel szemben, most a mesterséges intelligenciától várunk kreatív ötleteket. Nyilván a ChatGPT nem fog művészi áttöréseket elérni, nem ír meg egy Háború és békét, viszont a hétköznapi kreatív ötletekre remek megoldás.”

2. A diákjai hogyan használják fel a mesterséges intelligenciát? Ez milyen készségeket vagy kompetenciákat fejleszt bennük?

Zsolt diákjai alkalmanként pluginokat fejlesztenek, szintúgy kísérleteznek a chatbotokkal és a prompt engineering-el. A hallgatók próbálták kiismerni, milyen módon lehet specifikusabb, árnyaltabb utasítást megadni a nyelvi modellnek, annak érdekében, hogy a lehető legjobb választ kapják vissza. Viszont a kísérletezés mellett a diákok kihasználták a ChatGPT adta lehetőségeket a szövegalkotásra és átfogalmazásra. A beadandók írása egy rendszeres feladat a

joghallgatók számára, viszont nem kevés alkalommal előfordult, hogy a szövegezést a mesterséges intelligencia oldotta meg, mindemellett adódott alkalom arra is, hogy a tanulók elkérték a társuk által megírt munkát, és ezt teljes egészében átírták a ChatGPT segítségével. A hallgatók a generált szöveget rendezik, helyenként átírják, a leleményes egyének pár hibát is ejtenek a szövegben, hogy felismerhetetlenné váljon a mesterséges intelligencia hatása, bár az MI által generált szöveget alapvetően nehezen ismerik fel a szűrő algoritmusok. (Cingilliglu, 2023; C. Zhang & Zhang, 2023)

Zódi Zsolt oktató bámulatos nyitottságot tanúsít a mesterséges intelligencia megjelenésére, annak ellenére, hogy szakmai oldalról nem tudta felhasználni. Zsolt receptív meglátásai szerint a mesterséges intelligenciát nem is lehetséges és nem is szükséges megvonni a diákoktól, hiszen az oktatás nem csak a tudásanyag leadásából áll. Ha konstruktivista megközelítéssel vizsgáljuk az oktatást, be kell látnunk, hogy nem elegendő a lexikális tudás, vagy a gyakorlatszerzés, nagy jelentőséget kell tulajdonítani a diákok ösztönzésére, kíváncsiságuk fenntartására, és felfedezéseik támogatására. A ChatGPT és a mesterséges intelligencia már jelen vannak mindennapjainkban, így a kiadott feladatokat olyan módon kell módosítani, hogy számításba kerül a ChatGPT létezése. Progresszívan kell bevonni a mesterséges intelligenciát az oktatásba. A ChatGPT olyasfajta kollaborációs eszközként használható, mint a számítógép, vagy az internet, viszont hatalmas hangsúlyt kell fektetni annak elkerülésére, hogy a chatbot helyettesítse a tanulást vagy az önálló gondolkodást.

Az oktató véleménye szerint elsődlegesen a szövegalkotással kapcsolatos feladatokat kell módosítani, vagy megszüntetni, hiszen egy általános szöveg megalkotására könnyűszerrel felhasználható a ChatGPT. Ezen felül érdemes lehet pár olyan feladatot alkotni, melyek megoldásához szükségszerű a ChatGPT, például prompt alkotására kiélezett feladat. Az oktató szerint valószínűleg szükségessé válik a szóbeli jellegű feladatok és számonkérések alkalmazása.

3. Ön szerint plágium a ChatGPT használata?

A jelenlegi jogszabályok alapján nem számít plágiumnak. A hallgatónak más személy bevonása, vagy munkája nélkül kell megírnia a dolgozatot, a ChatGPT használata így nem vehető plágiumnak. Értelemszerűen elmondható viszont, hogy egy hallgató tanulmánya nem lesz effektív, ha túlnyomórészt a mesterséges intelligenciára hagyatkozik, így az intézménynek elengedhetetlen, hogy átalakításra kerüljön a belső szabályrendszer, amely során a mesterséges intelligencia beépítésre kerül a tantervben.

4. Hogy véli, a munkaerőpiacon drasztikus változásokat fog előidézni a mesterséges intelligencia?

A ChatGPT egy újabb technikai vívmány, melyhez alkalmazkodnia kell az emberiségnek. Sok munkakör drámaian átalakul, pár megszűnik, de ez nem fog összeomlást eredményezni a munkaerőpiacon, kiváltképpen azért, mert számos munkakör nem helyettesíthető mesterséges intelligenciával. Az emberek igénylik a társas interakciókat, az emberi közelség nem pótolható semmilyen formában, így az egészségügyben, oktatásban, pszichológiában és további területeken a mesterséges intelligencia csak a munka kisegítésében lesz hasznos.

Főleg a középszintű kreativitást igénylő munkakörök kerülnek veszélybe, egy marketinges vagy grafikus munkája nagyban megváltozhat a generative AI hatására. Ezt a változás érzékelve van, egy 2023-ban lefolytatott Twitter felmérés (Miyazaki és mtsai., 2024) azt vizsgálta, hogy az emberek hogyan vélekednek a mesterséges intelligenciáról. Alapjáraton pozitív visszajelzéseket mértek, a grafikusként dolgozók viszont kifejezetten negatívan tekintenek az MI elterjedésére.

A ChatGPT és mesterséges intelligencia esetében inkább a nem számolt következmények adhatnak aggodalomra okot. A technológia és a fejlődés gyakran nem az elvárásoknak megfelelően haladt előre, erre kiváló példa az önvezető autó. Az önvezető autók fejlesztéséről számos szakértő és iparági vezető beszélt már évekkel ezelőtt is. Az egyik ilyen korai megjegyzés 2010-ben hangzott el Sebastian Thrun-tól és fiatal tudósokból álló csapatától, mely az önvezető technológia elsajátítására törekedett (Sebastian Thrun, 2010).

Jelen pillanatban még távol állunk az önvezető járművektől. Az aktív autonóm járművekkel kapcsolatban a Tesla lakossági értékesítések terén kezdeményezett, és már részleges önvezető rendszert kínál vásárlóinak, viszont a vásárlóknak el kell ismernie, hogy egy béta verziót vásárolnak és minden felelősséget vállalnak. Ha az önvezető rendszer meghibásodik, a Tesla nem vállal érte felelősséget. Ennek ellenére az önvezető rendszerbe vetett bizalom végett számos baleset történt, 2023 év decemberi adatai alapján minden 1000 sofőrből 30 balesetet szenvedett (Lambright-Black et al. 2023).

A lassú haladásnak evidens oka, hogy sok időbe telik egy ilyen rendszer tökéletesítése, kiváltképpen azért, mert a közutakon számtalan szokatlan és ritka esemény bekövetkezhet, melyre a járművek nem reagálnak megfelelően. A szélsőséges időjárás, vadvilág, vagy útépítés mind mozgásképtelenné tehetnek egy önvezető járművet, ezáltal balesetveszélyessé téve azt (Jae C. Hong, 2023). Hasonló módon kiváltott baleset végett vonták vissza a Kalifornia

államban működő Cruise taxi vállalkozásnak önvezető járműveit. 2023 októberében a Crusie egy önvezető autója ütközés észlelése miatt erősen fékezett, majd megpróbált félrehúzódni, miközben egy gyalogos a jármű alá került. A balesetet részleteit a vállalat nem szándékozta nyilvánosságra hozni (Tom Carter, 2023).

Zsolt kiemelte, hogy bár az önvezető járművek esete kifejezetten egyedi, érdemes levonni pár tanulságot a technológia fejlesztéséről. A mesterséges intelligencia, bár rohamosan fejlődik, lehetséges, hogy a fejlesztés az utolsó pár százalékában kerül elakadásra. Ezen felül érdemes számítani arra, hogy a mesterséges intelligencia előidézhet pár előre nem látott hatást és reakciót.

4. Milyen területek kerülhetnének fejlesztésre a mesterséges intelligencia által?

Az oktató úgy véli, hogy mindenhol felhasználható a mesterséges intelligencia, de kifejezetten az egészségügyben és mezőgazdaságban vár nagy áttöréseket. Mindkét területen a mesterséges intelligenciával fokozni lehet a megelőzést, egészségügyön belül az MI-vel könnyűszerrel előre jelezhetővé válhat egy szermény bizonyos betegségekre való hajlama.

III.2.3. dr. Polyák Gábor által adott interjú

1. Diákjai, vagy Ön használják a ChatGPT-t?

A diákokon nem veszi észre a használatot, bár tisztában van vele, hogy valószínűleg ezt nem osztják meg oktatóikkal, a következményektől való félelem miatt. Gábor az oktatási munkájához nem használja a chatbotot, kifejezetten kötött anyagból dolgozik, de mindenképp szeretné, hogy diákjai kísérletezzenek a chatbottal. A világ számos változáson megy keresztül, fontos, hogy a hallgatók már a kezdetektől eszközként tekintsenek a ChatGPT-re.

2. Ön szerint plágium a ChatGPT használata?

Az oktató nem nevezné plágiumnak a ChatGPT használatát, hiszen a plágium egy természetes személy munkájának felhasználása, a szerző feltüntetése nélkül. A ChatGPT maga nem egy szerző, az algoritmussal generált szöveg felhasználása nem plagizálás. Talán a betanítás során átvett adatok rendelkeznek szerzői jogi védelemmel, de ha ezt nem tartják plágiumnak, úgy a ChatGPT által megadott válaszok sem jogkötelesek. Attól függetlenül, hogy a ChatGPT használat nem nevezhető klasszikus értelemben vett plágiumnak, a chatbot túlzott kihasználása mindenféleképp tisztességtelen a hallgató szempontjából. Persze nehéz egyensúlyt alkotni, és olyan feladatokat előállítani, melyek nyitottak a ChatGPT létezésére, de Gábor hangsúlyoz célja elérni azt, hogy a diákok gyakorta használják a chatbotot, úgy, hogy tisztában legyenek

feladatukkal és az elvárt kimenetekkel. A diákok beavatása elengedhetetlen, hiszen az intézmény nem tudja megállítani a fejlődést. Az oktatót idézve: „Ha egyszer a palackból kiengedték ezt a lehetőséget, akkor a mi dolgunk az, hogy valamilyen olyan használati irányba tereljük a diákokat, amivel sem másnak sem maguknak nem ártanak.”

Az interjú során előjött, hogy az oktató hogyan fog viszonyulni az olyan esetekhez, melyek során rajtakapja a hallgatóját, hogy kevés önálló munkát fektetett beadandójába. Gábor szerint az igazán lényeges kérdés az, hogy egyáltalán képes lesz-e felismerni az ilyen munkákat. A plágium felismerése még a ChatGPT megjelenése előtt is nehéz feladat volt, ezen kívül a chatbot által generált szöveg nem statikus, állandóan változik és fejlődik. Ha Gábor joghallgató diákja beleesik a GPT csapdájába, és elfogadja a pontatlanul generált válaszokat, az egy jó lehetőség a diák számára, hogy megismerje a chatbot korlátjait. A diákoknak használnia kell mesterséges intelligenciát, kritikus szemmel kell vizsgálniuk, olykor hibázniuk kell, hogy ténylegesen lássák, hogy tévedhet a ChatGPT.

3. Az intézmény dolgozói mennyire által nyitottan a ChatGPT-hez?

Polyák Gábor kollégái az intézményen belül kifejezetten nyitottan fogadták az MI megjelenését, és egyhangúan úgy vélik, be kell integrálniuk azt az oktatásba. A fő problémát a az integrálás mértéke, és a határok meghúzása jelenti. Nehéz eldönteni, hogy mennyire legyen megszabva a ChatGPT használata, milyen arányokban tartalmazza a leadott munka a diák önálló hozzájárulását. A ChatGPT beépítése mellett érdemes szóbeli és „terepmunka” jellegű feladatokkal gazdagítani a tananyagot, legyen itt szó akár az interjúztatásról, hiszen a ChatGPT abban nem lehet a hallgató segítségére.

4. Mit gondol, mi lesz a mesterséges intelligencia jövője?

Sokakat érdekel a kérdés, hogy hova fejlődik a ChatGPT évek múlva, pedig egy ilyen jellegű IT fejlesztés hatását megjósolni szinte lehetetlen. Polyák Gábor kifejezetten a jogi területen való fejlesztésekben érdekelt, azért, mert a jog területe kötött szókincssel rendelkezik, egy formalizált nyelv keretei között. A ChatGPT állandóan fejlődik, ki tudja, milyen tudásra tesz szert akár egy fél év múlva. Nem elképzelhetetlen, hogy a mesterséges intelligencia képes lesz összetett jogi ügyek megválaszolására, vagy akár szerződések létrehozására.

A munkaerő piacon kiváltott hatás valószínűleg a marketinges szövegírók, a kreatívok munkáját fogja veszélyeztetni a mesterséges intelligencia, viszont emellett létrejönnek új állások. Személyes tragédiák sajnos mindig lesznek, de a kreatív feladatköröket állató személyek

számára nagy segítséget jelenthet, ha a mesterséges intelligenciára kiegészítő eszközként tekintenek, amivel kiegészíthetik saját egyéni tudásukat.

A médiára jelenthet hatalmas veszélyt a MI, főleg a hamis videósanyagok és hanganyagok létrehozása és terjesztése. Az állhírek gyorsabb és látványosabb terjedése mellett értelmetlenné válhat a valós információ. A dezinformáció terjedésére alapvetően sem volt egy kiforrott megoldás, a blőd állhíreknek eddig is volt egy tábor, így a mesterséges intelligenciával manipulált hamis videókkal hatalmas reakciók kerülhetnek kiváltásra. A felhasználó van a legnehezebb helyzetben, hiszen el kell döntenie, hogy hol kezdődik és hol végződik a valóság, és az erre való készséget az oktatáson belül kell elkezdni fejleszteni. Az állhírek ellen szükség van egy kevésbé megvezethető, kritikusan gondolkodó rétegre, aki kicsit pesszimista szemmel tekint a hírekre, viszont egy ilyen réteg kitanítása nehéz feladat, abból kiindulva, hogy az oktatók is csak most kerültek szembe a mesterséges intelligenciával.

III.2.4. Mélyinterjúk összegzése

Az interjúztatott oktatók egyike sem használja különösképpen a ChatGPT-t, a bot pontatlanságára hivatkozva. Annak ellenére, hogy számos tanulmány tett említést arról a ChatGPT használata plágiumgyanús lehet (Cingilliglu, 2023; Jonathan Bailey, 2020; Stahl & Eke, 2024) a három jogász egyöntetűen azt nyilatkozta, hogy nem számít a chatbot használat klasszikus értelemben vett plágiumnak. Persze ettől függetlenül aggasztó lehet a chatbot használata.

A ChatGPT oktatásba való integrálásáról eltérően vélekedtek az oktatók. Mindkét vélemény jogos érvekkel van alátámasztva, viszont a ChatGPT használatától nem lehet eltiltani a diákokat.

Kérdés	Igen válaszok száma	Nem válaszok száma
Felhasználja munkája során a ChatGPT-t?	0	3
Plágiumnak tartja a ChatGPT használatot?	0	3
Érdemes ösztönözni a diákokat a gyakori használatra?	2	1

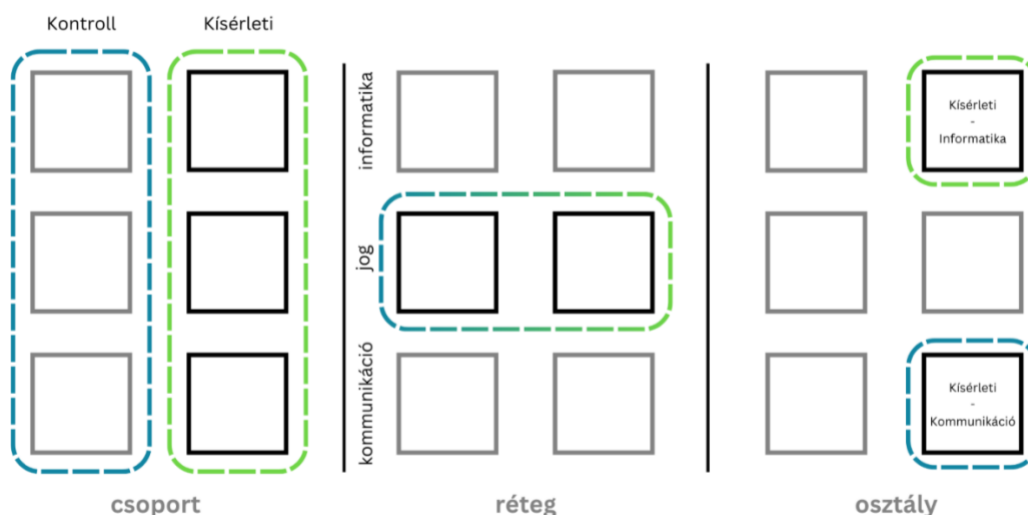
6. táblázat Válaszok összefoglaló táblázata (saját szerkesztés)

III.3. Hallgatói teljesítmény mérés

III.3.1. Mintavétel és tesztírás

A tanulmány fő célkitűzése annak mérése, hogy a ChatGPT használatának lehetősége milyen hatással van egy nagy mintán végzett kísérlet során. A beavatkozás lényege abban áll, hogy

létrehozunk egymástól független kísérleti és kontroll csoportokat, amelyek rétegenként tudományág-specifikus nyitott könyves tesztek írnak (lásd 2. ábra). A kísérleti csoport tagjai a hagyományos források mellett ChatGPT-t is használhatnak, ellentétben a kontroll csoporttal.



2. ábra Mintastruktúra bemutatása (saját szerkesztés)

A tanulmány során a „normára hivatkozott teszt” kifejezést használjuk az eredményességben és hatékonyságban mutató különbségeket feltáró feladatsorok jelölésére. A fogalom lényege, hogy a kontroll csoport rétegenkénti várható eredményét az egyes tesztípusok normájának tekintettük. Ezek referenciaként szolgálnak a kísérleti csoport eredményeinek különbségével kapcsolatban. Az eltérés pontos méréséhez fontos, hogy a kísérleti és kontroll csoport rétegenként összességében hasonló kompetenciájú hallgatókat tartalmazzon, ezzel is csökkentve az összetételbeli különbségek torzító hatását.

Az egyes tudományterületek hallgatóit a mintán belül külön rétegekbe lettek soroljuk. Az egyes rétegek felépítésében azonos, de témakörében különböző vizsgasort írnak. A rétegzés lehetővé teszi, hogy homogénebb mintaszegmenseken lehessen elvégezni a beavatkozást, illetve a ChatGPT hatását terület specifikusan tudjuk mérni.

A mintavétel szempontjából három kritériumot igyekeztünk érvényesíteni:

- 1) a mintába kerülés valószínűsége minél inkább véletlen legyen (Hunyadi & Vita, 2019b);
- 2) a kísérleti és kontroll csoport, valamint az egyes minta rétegek egymásra nézve független mintának számítsanak, tehát ne legyen átfedés a tagok között (Frost, é. n.);
- 3) az egyes mintaosztályok mérete önmagában is nagy mintának ($n > 30$) számítsa, a statisztikai próbák könnyebb és pontosabb elvégezhetősége miatt (Szreder, 2022).

Amíg a függetlenséget a minta strukturálásakor biztosítani lehetett, addig a mintavétel véletlenségére és a végső mintaosztályok nagyságára kevésbé volt befolyásunk a hallgatók toborzása során.

III.3.2. Internethasználat nyomkövetése

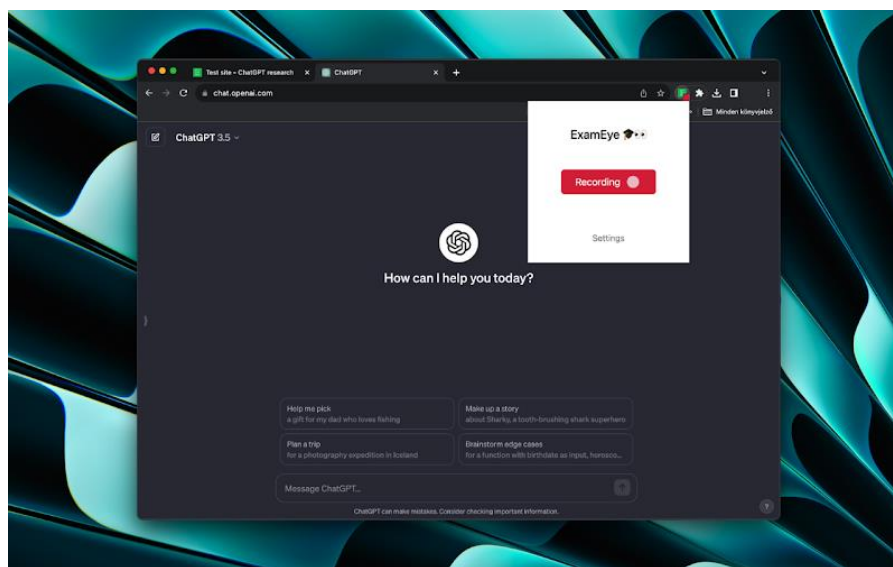
A kutatás alapvető feltételezése, hogy a ChatGPT használata a hallgatói feladatok elvégzésében a hatékonyság tekintetében nyújt leginkább előnyt. A vizsgálat előtt azt feltételeztük, hogy az új technológia megváltoztatja az információszerzés és feldolgozás módját. Nem csak az ismereteket koncentrálna egy helyre, de csökkenti a szükséges és szükségtelen felhasználói interakciók számát, mint például a keresési találatok szűrése vagy a betöltött oldalak adatkezelési szabályzatának elfogadása. Éppen ezért fontosnak tartottuk, hogy a normára hivatkozott teszt alatt a résztvevők minél nagyobb része esetén figyelhessük meg az internethasználati tevékenységüket.

A böngészés nyomon követést és a hozzá kapcsolódó mindenfajta adatkezelést két alapvető célhoz kötöttük. Az adatkezelés első számú célja az egyéni és csoportos internethasználati minták elemzése, különös tekintettel a kísérleti és kontroll csoport digitális lábnyomában megjelenő különbségekre. A megfigyelés második számú célja a nem véletlen mintavételi hiba mértékének csökkentése. Hiszen az adatrögzítés lehetővé teszi, hogy utólag átcsoportosíthassuk azokat a hallgatókat, akik bár a kontroll csoportba tartoztak, mégis használták a mesterséges intelligenciát.

Az információszerzés forrását három csatornára, oldaltípusra bontjuk. Megkülönböztetjük az alap weboldalakokat, mint elsődleges információ forrásokat; a kereső oldalakat (Google és Yahoo), mint másodlagos információ forrásokat, amelyek fő feladata, hogy összekössék a felhasználókat a releváns weboldallal; és a ChatGPT oldalát, amely ötvözi az előbbi két oldaltípus logikáját: keresésre ad lehetőséget és információt szolgáltat. A használati mintákat abból a célból tanulmányozzuk, hogy az egyes csatornákat milyen intenzíven használják a kontroll és kísérleti csoport tagjai: mennyi időt töltenek a teszt során kereséssel, weboldalak olvasásával; és a ChatGPT használatának lehetősége hogyan változtatja meg az internethasználatukat. Emellett a nyomon követés lehetővé teszi, hogy képet kaphassunk a ChatGPT gyakorlati használatának elterjedtségéről a vizsgált tudományterületeken. Hiszen a kísérleti csoport tagjainak nem kötelező a mesterséges intelligencia használatát a vizsga során. Végezetül célul tűzzük ki, hogy megfigyeljük a hallgatók tanulmányi célú ChatGPT használatát jellemző szokásokat. Például, hogy hány kontextus ablakot használnak, mennyire

szokásuk promptot átírni, vagy válaszokat újra generálni, amelyek egyben a kritikus felhasználói magatartás egyértelmű jelei.

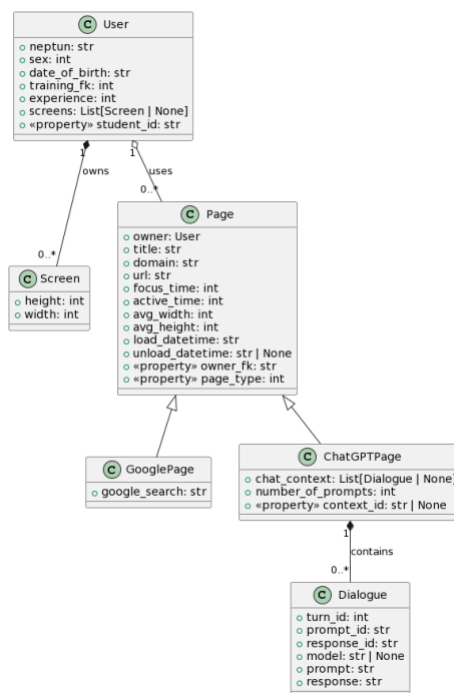
A teszt ideje alatt az internethasználat rögzítését a kutatáshoz fejlesztett Google Chrome bővítménnyel, hozzá tartozó alkalmazásprogramozási interfésszel (továbbiakban API), adatbázissal, valamint online vizsgafelület segítségével valósítjuk meg. A fejlesztés adatvédelmet és adattisztaságot érintő két legfontosabb szempontja a következők. Egyrésztől biztosítani kell, hogy a böngésző kiegészítő kizárólag a vizsga ideje alatt figyelje meg a résztvevők internethasználatát. Másrészt pedig garantálni szükséges, hogy a bővítményt ne lehessen a tesztplatformon kívülről (még szándékosan sem) elindítani. Az adatrögzítést kizárólag a hallgató saját maga indíthatja el a vizsgafelületen elhelyezett gombbal. A megfigyelés ideje alatt annak tényét jelezi az oldal keretének animációja, a bővítmény ikonja és annak felugró ablaka (lásd 3. ábra). Emellett a bővítmény úgy lett fejlesztve, hogy a nyomon követés manuális leállíthatóságának opciója mellett, az adatrögzítés automatikusan is leáll, ha a vizsgaoldal véletlenül vagy szándékosan bezárásra kerül.



3. ábra A bővítmény felugró ablaka és ikonja is jelezte az adatrögzítés tényét (saját szerkesztés)

A böngészőbővítmény, amelyet ExamEye-nak nevezünk, feladata az adatok rögzítése, valamint a felhasználók védelmét biztosító alkalmazáslogika végrehajtása. A rögzítést egy úgynevezett tartalomszkript (content script) végzi, amelyet a kiegészítő a weboldalak forráskódjába fecskendez be a dokumentum objektum modell (DOM) felépítése előtt. Ez a szkript lehetővé teszi a kiegészítő számára, hogy hozzáférjen a meglátogatott webhelyek adataihoz és tartalmához, valamint nyomon kövesse a felhasználó különböző interakcióit az oldallal (*Content scripts*, é. n.). A szkript működése során először azt ellenőrizi, hogy be van-e

kapcsolva az adatrögzítés. Ha a felhasználó engedélyezte a nyomon követést, akkor a szkript az oldal URL-je alapján megállapítja annak típusát. Ezt követően a program példányosít egy oldal specifikus megfigyelő objektumot, amely egy általános adatrögzítő szülő-osztályból származik. A különböző megfigyelő objektumok által gyűjtött és az API számára továbbított adatok UML diagramját az alábbi ábra szemlélteti (lásd 4. ábra).



4. ábra Az egyes megfigyelő osztályok UML diagramja (saját szerkesztés)

Fontos megjegyezni, hogy a ChatGPT esetében nemcsak a szabadszöveges adatok, hanem a különböző kontextus linkek és a prompt-válasz azonosító párok is érdemi információt tartalmaznak az elemzés szempontjából. Ezek segítségével képesek vagyunk kimutatni olyan alapvető felhasználási szokásokat, mint például, hogy a felhasználók jellemzően egy vagy több ablakot használnak a mesterséges intelligencia megszólítására, vagy hogy mennyire jellemzőek a promptok átírása és a válaszok újra generálása.

Az ExamEye az egyes oldalak váltása vagy bezárása előtt (`beforeunload` DOM esemény) a rögzített adatokat elküldi a programozási interfész megfelelő végpontjának, ami a beérkező adatokat validálja, és a megkapott változókból egyéb tulajdonságokat származtat. Ezután az adatbázissal kapcsolatot létesít az API, és az információkat a megfelelő séma alapján feltölti az adatbázisba. Az API végpontok és kutatási adatbázis leírása a függelék B és C pontja alatt található meg.

Az API-t egy serveless technológia segítségével üzemeltettük, ami lehetővé tette a könnyű méretezhetőséget és a gyors üzembeállítást (*Kiszolgáló nélküli számítástechnika*, é. n.). Az adatbázis futtatása esetén is egy felhős megoldás használata mellett döntöttünk költséghatékonysági és rendelkezésre állási szempontok miatt.

IV. Kutatás lebonyolítása

IV.1. Kompetenciateszt

A kutatásban való részvételre 415 egyetemi hallgató jelentkezett az informatika, a jog, valamint a kommunikáció és média tudományok területeiről. A normára hivatkozott teszt elvégzése előtt fontosnak tartottuk megismerni a jelentkezők demográfiai és tanulmányi háttérét, a ChatGPT használatával kapcsolatos véleményüket, továbbá felmérni kritikus gondolkodásukat és problémamegoldó képességüket. A képességében azonos kísérleti és kontroll csoport alkotása tette leginkább szükségszerűvé egy előzetes teszt (továbbiakban kompetenciateszt) lefolytatását. Ugyanakkor a jelentkezők adatainak rögzítése lehetővé tette, hogy későbbiekben becslést adhassunk az összetétel hatás mértékre az eredményesség és hatékonyság mutatóira nézve.

A kompetenciatesztben rögzítettük a jelentkezők Neptun kódját, nemét, születési idejét, felsőoktatásban megkezdett szemesztereinek számát, intézményüket, szakirányukat, valamint azt, hogy hetente átlagosan hány órát fordítanak tanulásra a tanórákon kívül. A személyes és tanulmányi adatok mellett megkérdeztük, hogy a hallgatók mennyire használják jól a kereső kifejezéseket, mennyire valószínű, hogy a ChatGPT segítségét egyetemi beadandók készítéshez igénybe veszik, milyen mértékben bíznak a chat modell válaszaiban, és rendelkeznek-e ChatGPT Plus előfizetéssel.

Az kompetenciateszt készségmérő szekciója 5 feladatból állt, melyek mindegyike egy adott komplex készség vizsgálatára irányult. Az első kérdésben a hallgatók szövegértését próbáltuk meg mérni, amely során arra voltunk kíváncsiak, hogy képesek-e helyesen meghatározni egy adott szövegrészlet üzenetét. A második feladatban a teszt alanyok analitikai és ábraértési készségére fókuszáltunk. A jelentkezőknek fiktív termékek havi bevételének diagramja alapján kellett kiválasztania a legstabilabb forgalmú árucikket adott feltételek érvényesülése mellett. A harmadik feladatban a résztvevők információ keresési képességét teszteltük: egy speciális témakörben, a védjegyjog területén kellett választ adniuk egy konkrét vállalat emblémájának védelmével kapcsolatban. A negyedik feladatban egy a mesterséges intelligencia

használatával kapcsolatos etikai kérdésre kínáltunk különböző alternatívákat a hallgatóknak, és azt mértük, hogy mennyire választanak etikus vagy etikátlan opciót. Az utolsó, ötödik feladatban az időmenedzsment kérdéskörét boncolgattuk, ahol a helyes időbeosztás meghatározása volt a feladat az adott problémához.

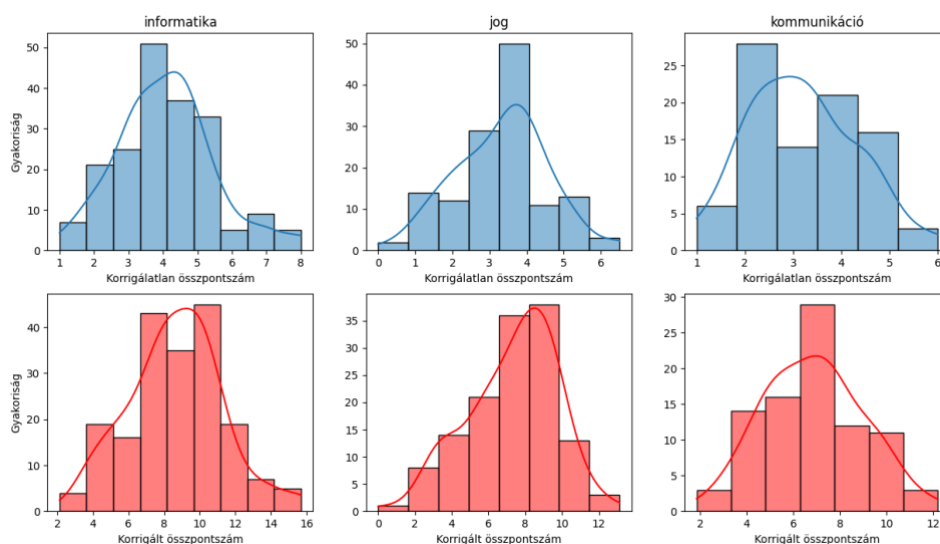
A kompetenciateszt kitöltésére magyar és angol nyelven volt lehetőség. A feladatsor elvégzésére 30 perc állt a hallgatók rendelkezésére. A rögzített adatokat és a feladatokra adott válaszokat utólag dolgoztuk fel, majd a releváns információkat a kutatás központi relációs adatbázisába felvittük.

IV.2. Kísérleti és kontroll csoport képzés

A kísérleti és kontroll csoport képzésekor meghatározó szempont volt, hogy a besorolt hallgatók mintarétegenként (tudományterületenként) átlagosan azonos kompetenciákkal rendelkezzenek. A résztvevők logikai, szövegértési, analitikai és problémamegoldó képességének meghatározására a kutatás előzetes tesztje során elért eredményüket használtuk. Tekintettel a kompetencia eredmények mérési skálájának szűk intervallumára (maximum 8 pont volt elérhető, de erre csak 5 hallgató volt képes), illetve a feladatok eltérő nehézségére, a csoportképzés előtt a változó átméretezése mellett döntöttünk. Azért, hogy jobban szétváljanak az elért teszteredmények, a nehezebb, összetettebb feladatok pontszámát nagyobb súllyal számítottuk be a hallgatók végső eredményébe. Az újonnan bevezetett eredmény változó az egyes feladatok szórásának reciprokéval súlyozott pontszámának összege volt. Az egyes hallgatók kiigazított végső pontszámát az alábbi képlet szerint számoltuk ki:

$$y^{hallgató} = \sum_{feladat=1}^n pontszám_{feladat}^{hallgató} * \frac{1}{szórás_{feladat}}$$

Az következő ábrán látható a kompetenciateszten elért korrigálatlan és szórással korrigált összpontszámok eloszlása mintarétegenként csoportosítva (lásd 5. ábra). A fenti módszerrel úgy sikerült a változó eloszlásának terjedelmét körülbelül megkétszerezni, hogy közben mindhárom tesztípus esetén a pontok eloszlása szimmetrikusabb lett. Az informatika és a kommunikációs teszt esetén az eloszlás ferdesége több mint 50%-kal csökkent.



5. ábra A kompetenciateszt korrigálatlan és korrigált összpontszámának eloszlása rétegenként (saját szerkesztés)

A korrekció által nem csak egy komplexebb értékelési szempontot hoztunk be a csoportképzés folyamatába, de a pontszámok eloszlásának szimmetriájának növelésével csökkentettük a kilógó értékek torzító hatását a várható értékre nézve. Mindez lehetővé tette, hogy alacsonyabb relatív szórás mellett tudjunk átlagában azonos kompetenciájú kísérleti és kontroll csoportot létrehozni.

A csoportba sorolás algoritmusát három fő logikai egységre lehet bontani. Első lépésben az egyes tesztípusok hallgatóit korrigált végső eredményük szerint csökkenő sorrendbe tesszük. Majd a sorbarendezt egyedeket kétfelé válogatjuk úgy, hogy minden első megfigyelés az A listába és minden második megfigyelés a B listába kerüljön. A kétfelé bontás következtében két azonos méretű vagy legfeljebb egy egyeddel különböző nagyságú, rendezett tömböt kapunk.

A második lépés célja, hogy a két csoport esetén az algoritmus az átlagok különbségére optimalizáljon. Az átlag optimalizáló függvény az azonos indexen lévő egyedeket kicseréli a listák között. Ha a csere által a csoportok átlagának különbsége csökken, akkor az adott hallgatók a csere utáni csoporthoz fognak tartozni. De ha a csere által a különbség növekedett, akkor a hallgatókat visszasorolja a függvény az előző csoportba. A módszer ezt a csere-ellenőrzés-feltételes csere lépés kombinációt minden egyes közös tömb index esetén elvégzi.

A csoportba rendezés harmadik fontos logikai lépése a csoportok közötti szórás különbség csökkentése. A szórás optimalizáló függvény nagyon hasonló módon működik, mint az átlag optimalizáló: közös indexenként megnézni, hogy az egyes egyedek csoportok közötti cseréjével csökkenthető-e úgy a szórások közötti differencia, hogy a két lista átlaga között az eltérése ne növekedjen. Említésre méltó különbség a két függvény között, hogy míg az átlag különbség

optimalizáló csökkenő sorrendben megy végig az eredményük szerint a hallgatókon, addig a szórás különbség optimalizáló növekvő sorrendben teszi azt. Az eltérés oka, hogy az előbbi függvény célja, hogy a nagy pontszámú egyedekre fókuszálva mérsékelje az átlagok különbségét. A második metódus célja pedig, hogy a kis pontszámú hallgatókra fókuszálva csökkentse a szórások eltérését. Végezetül az átlag és szórás optimalizált csoportok egyike lesz a kísérleti csoport a másik pedig a kontroll csoport.

IV.3. Normára hivatkozott tesztek

A kutatás kulcsfontosságú eleme a hallgatók által kitöltött komplex teszt volt. Az erre a célra kifejlesztett böngésző bővítmény, az ExamEye (itt hivatkozás a fenti fejezetre) lehetővé tette, hogy a teszteredmények és az idő mellett megfigyelésre kerüljön a böngészős tevékenység, és a ChatGPT-vel való interakciók. A felmérés lebonyolítására normálra hivatkozó, nyitott könyves tesztek kerültek létrehozásra, melyek megírása során kerültek megfigyelésre a hallgatók.

- Mind a jog, gazdaságinformatika és médiatudomány teszt esetében egy egységes struktúra kialakítása volt a cél. Mindhárom teszt típus általánosan 30 kérdésből épült fel, ezen belül:
- 12 darab feleletválasztós kérdés
- 6 darab igaz/hamis kérdés
- 6 darab “Fill the blank” kérdés, melyekben ki kell egészíteni egy-egy szóval vagy szókapcsolattal egy megadott mondatot. A válaszokat egy listából lehetett kiválasztani.
- 6 darab “Drag and Drop” kérdés, amiben az összetartozó kifejezéseket egérrel lehet mozgatni és társítani

A fent említett kérdéstípusoknál nem volt szükség a válaszok gépelésére, ami lényegesen megkönnyítette azok javítását és értékelését. A kérdések megválaszolása körülbelül 1 órát vesz igénybe.

A kísérleti és kontroll csoport által megírt tesztek tartalma kritikus fontossággal bírt. Annak érdekében, hogy a ChatGPT hallgatói teljesítményre gyakorolt hatása felmérésre kerüljön, olyan sokrétű tesztek kerültek megalkotásra, melyek megoldásához nélkülözhetetlen volt a kritikus gondolkodás és specifikus tárgyi ismeretek.

Annak érdekében, hogy a tesztek szakmailag megállják helyüket, egyetemi oktatók kerültek felkeresésre. Az oktatókkal együttműködve olyan összetett tesztek készültek, amelyek mérik a

kritikus gondolkodást és a speciális ismereteket a jog, a gazdaságinformatika és a médiatudomány területén. Mindhárom teszt kapott egy angol nyelvű fordítást, hogy az adott szakokon belül kitöltésre ösztönözzük az idegennyelvű hallgatót is. A tesztek nyelve a hallgató által került kiválasztásra a normára hivatkozott teszt során, így egy magyar nyelvű képzésen részt vevő hallgató igény szerint választhatta az angol nyelvű tesztet. A fordítás az oktatók által került ellenőrzésre.

A teszt megírásában részt vevő személyek:

- **Jog teszt - dr. Karsai Krisztina:** Szegedi Tudományegyetem Bűnügyi Tudományok Intézetének vezetője
- **Gazdaságinformatika teszt - Nagy Csaba Norbert:** Budapest Corvinus Egyetem gazdaságinformatikus hallgatója. Az egyetem webfejlesztés tárgyának demonstrátoraként dolgozik, dr. Láng Blanka egyetemi docens ajánlására kértük fel a tesztek megalkotására.
- **Médiatudomány teszt - dr. Polyák Gábor:** Az Eötvös Loránd Tudományegyetem Művészetelméleti és Médiakutatási Intézet intézményvezetője, Média és Kommunikáció Tanszék tanszékvezetője

IV.3.1. A jog teszt módszertana

A kérdések összeállítása előtt egy mélyreható beszélgetésben vettünk részt Karsai Krisztinával. A tesztek összeállítása mellett ötleteket adott nekünk a struktúra kialakításához, és véleményezte a teszt jellegét. A kutatáshoz olyan tesztet szándékoztunk összeállítani, amellyel az osztatlan jogászképzés teljes tananyagába tartozó, egyes konkrét tudáselemek elsajátítását fel lehet tární.

Magyarországon nyolc helyen folyik jogászképzés, ezen intézmények által működtetett szakok tartalma jelentős mértékben megegyezik. Ennek oka egyfelől az, hogy a felsőoktatási szakokra nézve jogszabály tartalmazta 2021. december 30-ig az ún. képzési és kimeneti követelményeket. Másrészt – mindezekkel összefüggésben is – lényeges azt is látni, hogy a jogászképzés célja a gyakorló jogászok kibocsátása.

A kérdésekre adott helyes válaszokat elérhetőnek és mobilizálhatónak tekintette az oktató, a joghallgatók esetében, amennyiben a tanulmányi útvonalukban már az adott tantárgyat teljesítették. Mindegyik kérdés a jogászképzés törzsanyagához tartozik, tehát nem szabadon választott kurzusok témakörei. A kiindulási pont, hogy egy ötödéves joghallgató a kérdésekre

tudja a választ, vagy biztosan hallott róla már. Ennek a kiindulópontnak ott van jelentősége, hogy vannak-e olyanok a válaszadók között, akik valóban azonnal tudták a választ az adott kérdésre, és/vagy vannak-e olyanok a válaszadók között, akik tudják korábbi tudás alapján, hogy hol és miként keressék a választ. Utóbbiba tartozók azok, akiknek az interakciója és a kérdések megválaszolásának módszere a legrelevánsabb a jelen kutatás szempontjából.

A kutatást megelőző előfeltételezés az, hogy a jogi tudás specifikus és nem, vagy alacsony helyességi aránnyal érhető el a nagy nyelvi modellek által használt, a szakkifejezések és egyéb nyelvi struktúrák statisztikai gyakoriságon alapuló kapcsolatainak felhasználásával.

A tanulási eredmények felől megközelítve a kutatás a „tudás” komponenshez kapcsoltnak került megtervezésre. Ennek két oka van. Tartalmi oknak tekintjük azt, hogy a tesztszerű visszakérdezés a tudás, mint tanulási eredmény ellenőrzésének a hagyományos eszköze. A formai vagy technikai indok pedig a jelen kutatás választott kísérleti módszeréből következik.

Egyfelől olyan tudáselemekre vonatkozó kérdések kerültek alkalmazásra, amelyek nem képzik értelmezés tárgyát a konkrét válasz megadásakor. A logikai gondolkodásra való képességet az egyes tudáselekkel kapcsolatos logikai műveletek alkalmazásán keresztül lehetett megmérni, illetve a hiányzó tudáselem logikai úton történő pótlására volt lehetőség.

Másrészről a főbb jogágakat megfelelő arányban reprezentáló kérdések kerültek válogatásra, így megjelent a polgári anyagi jog, a polgári eljárásjog, büntető anyagi jog, büntető eljárásjog, alkotmányjog, jogtörténet, római jog, szellemi alkotások joga, munkajog, közigazgatási jog.

Harmadrészt többféle nehézségi szintű kérdések kerültek a tesztbe. A nehézségi szintek meghatározása szerint történt, hogy a kérdésre adott válasz a bázistudás része, és kereséssel könnyen meg lehet-e találni, vagy a válaszok az elfogadott értelmezés vagy a bázistudás részéhez logikus következtetés eredményeként váltak a jogi tudás részévé, de így az egyszerű kereséssel nem feltétlenül elérhetőek.

IV.3.2. A gazdaságinformatika teszt módszertana

A toborzás fázisa után számítani lehetett rá, hogy a gazdaságinformatika teszt kerül legnagyobb számban kitöltésre. Ez annak köszönhető, hogy a Budapesti Corvinus Egyetemen belül jobban irányíthattuk és nyomon követhettük a toborzás menetét, volt alkalmunk személyesen felkeresni oktatókat. Ezen túlmenően, az egyetemi tanulmányaink során már kiismertük, hogy kihez fordulhatunk segítségért. Tekintettel arra, hogy nagy kitöltésszámba számítottunk, nagy hangsúly került a gazdaságinformatika teszt jó összeállítására. Az intézményen belül

felkerestünk pár általunk ismert oktatót, kik idő hiányában nem tudták vállalni a teszt megalkotását. dr. Láng Blanka egyetemi docens ajánlására kerestük fel demonstrátorát, Nagy Csaba Norbertet, aki elvállalta a megbízást.

A teszt összeállítása előtt Nagy Norbert demonstrátor azt a kérést kapta, hogy a gazdaságinformatika tananyagából egy átfogó kérdéssort alkosson. A gazdaságinformatika tananyaga egy rendkívül sokoldalú és gazdag területet foglal magába. Számos különböző témakört érint, és széles skálán mozog az üzleti világ és az informatika találkozási pontján. A demonstrátornak külön oda kellett figyelnie arra is, hogy más egyetemeken különböző tanterveket alkalmaznak, amelyek a tantárgy tematikában eltérhetnek. Például a mélyreható HTML ismeretek nem mindenhol kapnak ugyanakkora hangsúlyt.

A kérdések összeállítását a Corvinus gazdaságinformatikus szak tárgyainak áttekintésével kezdte, és azokat listázta, amelyekből jó kérdések kerülhetnek megfogalmazásra. Azon tárgyakat, amelyek nem kerülnek minden intézményben részletes bemutatásra, vagy egyáltalán nem tartoznak a szak képzési profiljához, eltávolította a listából. A teszt szempontjából megfelelő kérdések értő tudáson alapultak, és megoldásuk 1-2 percet vett igénybe. A demonstrátor törekedett a gyorsan megoldható, pörgős kérdésekre, ezzel javítva az esélyét annak, hogy a résztvevők, hogy valóban elgondolkozzanak a kérdéseken és szakszerű válaszokat adjanak.

Az alapos szűrés után az alábbi tárgyak anyagából készült el a kérdéssor: Matematika, IT infrastruktúra, Adatbázis kezelés, Programozás (Python), valamint általános IT ismeretek.

A kérdések között előfordult pár, amelyekre a ChatGPT rossz választ adott, a kérdés komplexitása miatt, vagy szimplán azért, mert a ChatGPT elavult adatbázissal rendelkezik. Voltak kérdések, amelyek első benyomásra összetettnek tűnhettek, viszont rövid átgondolással, és a szituáció elképzelésével egyszerűen megválaszolhatóakká váltak. Az ilyen jellegű kérdésekkel a mesterséges intelligenciát könnyen meg lehet téveszteni.

Példaként az egyik kérdést prompt-ként adtuk meg a ChatGPT-nek. (a beszélgetés nem módosított szöveg):

Prompt: *András és 4 barátja ülnek egy kerekasztalnál egyenlő távolságra egymástól, középen van egy magas asztaldísz, összességében hány ember arcát takarhatja ki az asztaldísz András elől, ha mindenki egy általa választott székre ül?*

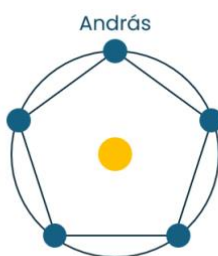
ChatGPT: *Ha mindenki egy általa választott székre ül, és az asztaldísz közepén van, akkor Andrásnak az asztaldísz mellett ülő barátjai arcát tudja csak látni, hiszen előtte van az asztaldísz.*

A kerekasztalnál az arcok száma az asztaldísz magasságától és a barátok elhelyezkedésétől függ. Az asztaldísz által kitakart arcok száma attól függ, hogy az asztaldísz magassága hogyan viszonyul az emberek fejmagasságához.

Tegyük fel, hogy az asztaldísz eléggé magas, hogy teljesen kitakarja az emberek arcát, ha az asztal fölött ülnek, de nem annyira magas, hogy az asztal fölé lógna. Ebben az esetben a válasz egyszerű: András elől az asztaldísz minden más arcot kitakarhat.

Tehát, ha az asztaldísz kitakarja az összes többi barát arcát, akkor András elől 4 arcot takar ki az asztaldísz.

A kérdésre a választ egy ábrával lehet a legegyszerűbben megkapni. Az ábra alapján látható, hogy András minden barátját láthatja.



6. ábra A kérdésben bemutatott probléma illusztrációja (saját szerkesztés)

IV.3.3. A média teszt módszertana

Dr. Polyák Gábor került felkérésre a média tesztek megalkotására. A média és kommunikációtudomány összekapcsolja a humán és társadalomtudományokat, beleértve a szociológiát, pszichológiát, politikatudományt és kulturális tanulmányokat is. Célja, hogy megértésre kerüljön a média által közvetített üzenetek hatása a társadalomra és az egyénekre.

A teszt során az oktató által alkalmazott kérdések rendkívül alapos ismereteket követeltek meg az adott anyagról. A feladatok megoldásához specifikus, lexikális tudás volt szükséges, és a ChatGPT gyakorta nem volt képes megadni a helyes válaszokat. A tesztkérdések kiemelik a tényt, hogy a ChatGPT képes pontatlan válaszok generálására, így az ilyen mélyreható szakmai kérdések megválaszolására elengedhetetlen az egyedi szakértői tudás.

A felhasználót a ChatGPT könnyűszerrel képes megtéveszteni, kiváltképpen azért, mert még a téves válaszokra is generál egy magyarázatot. A szakmai ismeret / gyakorlat hiányába, vagy a mesterséges intelligenciába vetett túlzott bizalom által elfogadásra és terjesztésre kerülnek hamis információk. Nem feltétlen csak a laikusok esnek a megtévesztés csapdájába, a ChatGPT képes elérni, hogy szakmabeli személyek kétségbe vonják saját szaktudásukat. Selin Urhan et al 2023 szeptemberében lebonyolított kutatásban egy ChatGPT általi érvelést figyelt meg. A Habermas-féle racionalitás konstrukcióval gazdagított Toulmin-modell segítségével elemezték az olyan helyzeteket, amelyek során egy függvény gráfiájához nem lehet érintő vonalat húzni. A ChatGPT nem nyújtott megbízható és stabil támogatást a hallgatónak az érvelési folyamatban. Megállapítható, hogy a ChatGPT, amelyről úgy gondoljuk, hogy képes segíteni a tanulást, nem tudja értelmezni az összetett matematikai feladatokat. A felhasználónak szükséges jártasnak lenni a számítási fogalmak terén, ellenkező esetben a ChatGPT által adott válaszok félrevezetőek lehetnek.

A megtévesztés mellett a mesterséges intelligencia borzalmas hatással lehet egy hallgató kritikus gondolkodására és analitikus képességeire. Értelemszerűen a ChatGPT számos rendkívüli lehetőséget hordoz magában, de nem elfelejtendő az érem másik oldala. Az azonnal generált válaszok felhasználásával egy tanuló hatékonyan és gyorsan meg tudja oldani egyszerűbb feladatait, akár önálló munka befektetése nélkül. Ez kifejezetten akkor káros, amikor egy tananyag kezdeti fejezeteiből kiadott önálló gyakorlásra szánt feladatok kerülnek megoldásra a chatbot által. A túlzott MI-ba vetett bizalom és a kényelem végett a hallgató háttérbe szoríthatja saját képességeit, és igényét a fejlődésre, noha egy anyag elsajátításához elengedhetetlen a tanulási folyamatban való aktív részvétel, és az önálló gondolkodás. Az információ könnyen elérhető, a mesterséges intelligencia állandóan fejlődik, a számítógép könnyűszerrel elvégzi a számításokat, így a hiba esélye minimális. Ha bármire rákereshetünk, vagy egy "intelligens ügynökkel" kereshetünk, mi értelme a tanulásnak? Egyáltalán mit érdemes igazán megtanulni? (Holmes et al - 2019).

Természetesen a tanulást leginkább a hallgató befolyásolja, viszont elengedhetetlen az oktató hatása. Az oktatás a társadalmi fejlődés egyik alappillére, és következetesen alakult annak érdekében, hogy alkalmazkodjon egyes korszakok társadalmi, kulturális és technológiai változásaihoz. A mesterséges intelligencia folyamatosan fejlődik, az akadémiai közösségnek meg kell értenie a lehetőségeit és korlátait, és oly módon kell integrálni az oktatásba, hogy az erősítse az oktatás

V. Eredmények tárgyalása

V.1. Minta szerkezetének és összetételének leírása

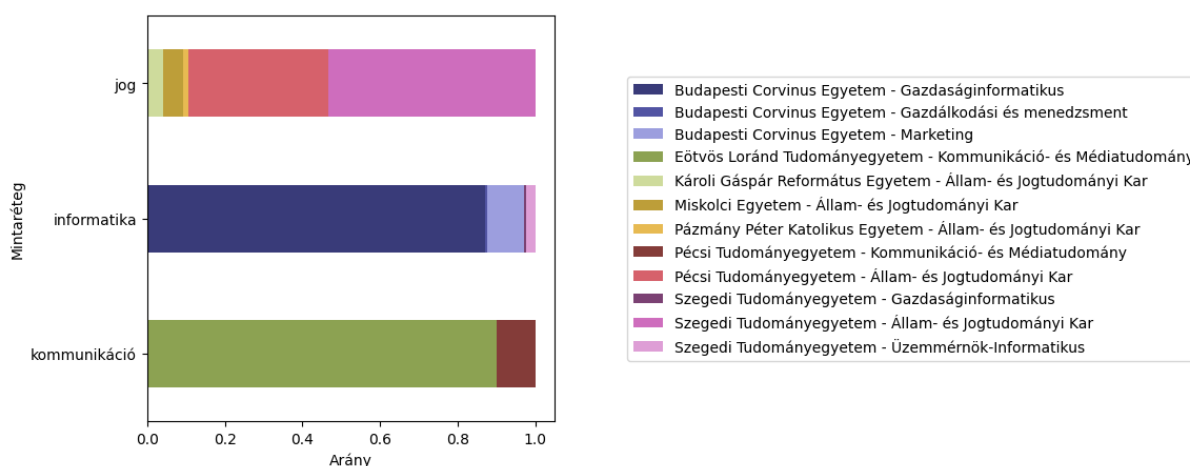
V.1.1. A mintarétegek nagysága

A normára hivatkozott tesztben, ami a kutatás fő mérési fázisa, 304 hallgató vett részt. A kontroll csoportban 135, a kísérleti csoportban pedig összesen 169 tanuló teljesítménye került mérésre (lásd 7. táblázat). Habár eredetileg egyenlő méretű vizsgálati csoportok képeztünk, a jelentkezők minden vizsgált tudományterület esetén nagyobb hajlandóságot mutattak a kutatás egészében való részvételre, ha a kísérleti csoportba kerültek besorolásra. Emellett a kontroll csoport nagyságát tovább csökkentette az is, hogy a vizsga során 15 résztvevő a csoport szabályzat ellenére alkalmazta a ChatGPT-t. Ezek a hallgatók utólag átsorolásra kerültek a kísérleti csoport megfelelő mintarétegeibe.

Mintaréteg	Kontroll csoport	Kísérleti csoport	Összesen
<i>Informatika</i>	74	95	169
<i>Jog</i>	35	40	75
<i>Kommunikáció</i>	26	34	60
Összesen	135	169	304

7. táblázat A mintaméret megoszlása osztályonként (saját szerkesztés)

A kutatásban résztvevő hallgatók 7 magyarországi egyetem 14 különböző szakáról származnak, amely lehetővé teszi, hogy heterogénebb, a sokaságot jobban leíró mintán lehessen elvégezni a beavatkozást. Fontos azonban kiemelni, hogy a minta nem tekinthető reprezentatívnak az egyetemi képzések és intézmények országos eloszlására. Ezen változók mintabeli megoszlása eltér a nemzeti felsőoktatási arányoktól, ezért az eredményeket ennek fényében kell értelmezni. Feltételezhető, hogy a különböző egyetemekről megfigyelt hallgatók eltérő szintű területspecifikus tudással és feladatvégzési kultúrával rendelkeznek. Ezért lényeges lehet az eredményesség és hatékonyság vizsgálatakor az összetétel hatás méretére való becslés szolgáltatása is.



7. ábra Az egyes mintarétegek képzési összetétele (saját szerkesztés)

Az intézményi diverzitás szempontjából a legnagyobb heterogenitást a jogi, a legnagyobb homogenitást pedig a kommunikációtudományi mintaréteg mutatja (lásd 7. ábra). Lényeges megjegyezni, hogy mindkét réteg különböző egyetemekről azonos szakképzésű hallgatókat tartalmaz. Ezzel szemben az informatikai teszt esetén az informatikától eltérő szakirányú hallgatók is lettek megfigyelve. Mivel a kutatás rendelkezésére állt egy tiszta, 147 fős minta a Budapesti Corvinus Egyetem Gazdaságinformatikus szakáról, az a döntés született, hogy ne csak informatikus hallgatók kerüljenek tesztelésre az adott mintarétegben a normára hivatkozott vizsga során. A mintaréteg hallgatóinak sokszínűsítése lehetővé teszi, hogy területre vonatkozó háttértudás nélküli hallgatók esetén is a ChatGPT hatása mérésre kerülhessen.

V.1.2. Demográfiai jellemzők

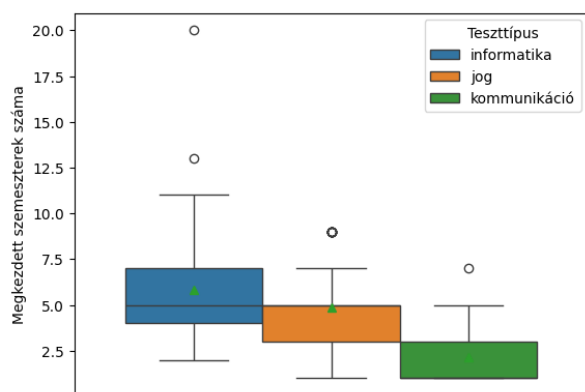
A minta egészét nézve a nemek közötti arány kiegyensúlyozottnak tekinthető, azonban a különböző tudományterületeken jelentős eltérések mutatkoznak. Amíg az informatikai teszt esetén a résztvevők 66%-a férfi, addig a jogi területen a női hallgatók képviselik a tesztet írók 66%-át, valamint a kommunikációs réteg esetén a női résztvevők aránya eléri a 80%-ot.

A kutatásban résztvevő hallgatók többsége a húszas éveik elején járó fiatal felnőtt, a mintában vizsgált személyek átlagos életkora 21,4 év volt az adatrögzítés idején.

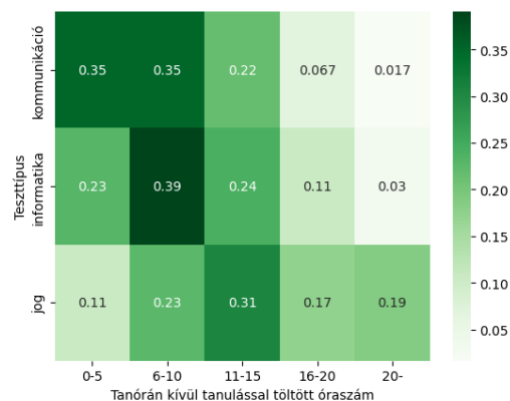
V.1.3. Felsőoktatási tapasztalat

A kutatás szempontjából legrelevánsabb demográfiai jellemző a felsőoktatásban megkezdett szemeszterek száma. Az adatok szerint a résztvevők 50%-a három és hét szemeszter közötti felsőoktatási tapasztalattal rendelkezett, valamint 25%-a több mint hét szemeszter óta tanult valamilyen felsőoktatási intézményben. Ezek alapján megállapítható, hogy a kutatás jellemzően a képzésük középső részén járó vagy végéhez közeledő hallgatók körében méri a ChatGPT

használat lehetőségének hatását. A felsőoktatási tapasztalat tesztípusonkénti felbontása megmutatja, hogy az egyes tesztek eltérő szinten lévő tanulókat mérnek (lásd 8. ábra). A felső oktatásban töltött szemeszterek száma átlagosan az informatikai mintaréteg esetén volt a legmagasabb és a kommunikációtudomány esetén a legalacsonyabb.



8. ábra A felsőoktatási tapasztalat megoszlása tesztípusonként (saját szerkesztés)



9. ábra A tanórán kívül heti szinten tanulóval töltött órák számának megoszlása (saját szerkesztés)

V.1.4. Tanulási szokások

A kompetenciateszt során alkalmazott kérdőívvel mérésre került, hogy a kutatási alanyok átlagosan egy héten mennyi időt szentelnek tanulásra az egyetemi órák mellett. Feltételezhető, hogy azok a hallgatók, akik több időt fordítanak tanulásra, magasabb szintű problémamegoldó képességgel rendelkeznek. Hiszen az ilyen tanulók gyakorlottabbak az ismeretek önálló keresésében és feldolgozásában, illetve egyetemi tesztek írásában is. A kérdőív alapján az informatikai réteg hallgatói mérsékelt, de több időt szánnak hetente tanulásra, mint kommunikációs csoport tagjai (lásd 9. ábra). A jogi terület alanyai viszont messze felül reprezentáltak a felsőbb intervallumokban. Emellett a kutatás résztvevői egy öt fokozatú Likert-skálán megadták, hogy mennyire tartják magukat professzionálisnak a kereső kifejezések használatában. A megkérdezettek többsége magát a skála felső 50%-án helyezte el, a leggyakoribb válasz a négyes számú lehetőségre érkezett. A magabiztos válaszok a keresők oldalak mindennapi használata miatt lehetséges.

V.2. ChatGPT használat

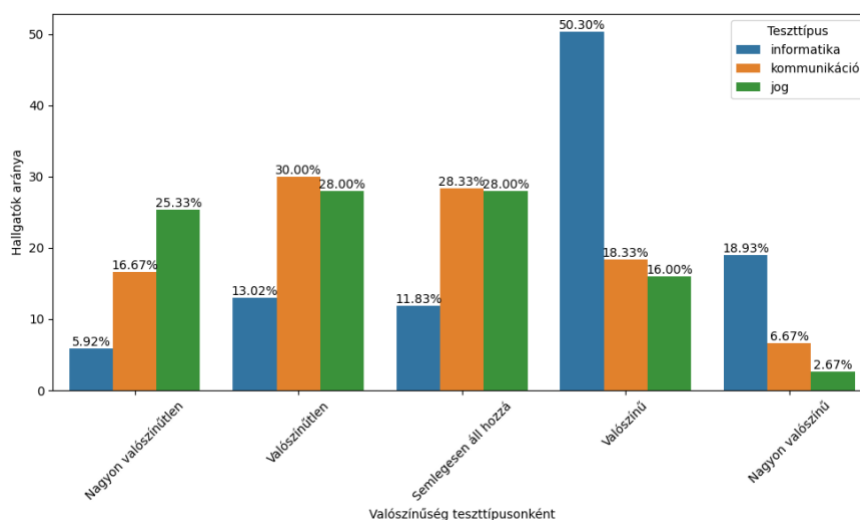
V.2.1. Mérési célkitűzés

A kutatás során elengedhetetlennek tartottuk, hogy képet nyerhessünk, hogy a normára hivatkozott teszt során a kísérleti csoport tagjai mekkora arányban és milyen intenzitással használják a mesterséges intelligenciát a vizsga feladatok megoldásához. Emellett szükségesnek ítéltük feltárni a résztvevők tanulmányi célú ChatGPT használatával kapcsolatos

attitűdjét. A tanulmányi célú mesterséges intelligencia felhasználás mérése nem csak a hallgatók szokásainak jobb megértése és az egyes rétegek közötti különbségek feltárásának érdekében volt lényeges. Segítségével viszonyítást kapható a kompetenciateszt során bevallott és a normára hivatkozott teszt során mért GPT használat eltérésének mértékéről. Ezáltal validálni lehetett a mérések reprezentativitását a mindennapos tanulmányi célú mesterséges intelligencia használatára nézve. Sarkalatos hangsúlyozni ugyanis, hogy a kísérleti csoport tagjai számára lehetőség és nem kötelesség volt a ChatGPT használata. A kutatás résztvevői, mint korlátozott racionalitással rendelkező egyének döntöttek a technológia használata mellett vagy ellen a saját hasznosság függvényük, tapasztalatuk és elfogultságuk szintézise alapján.

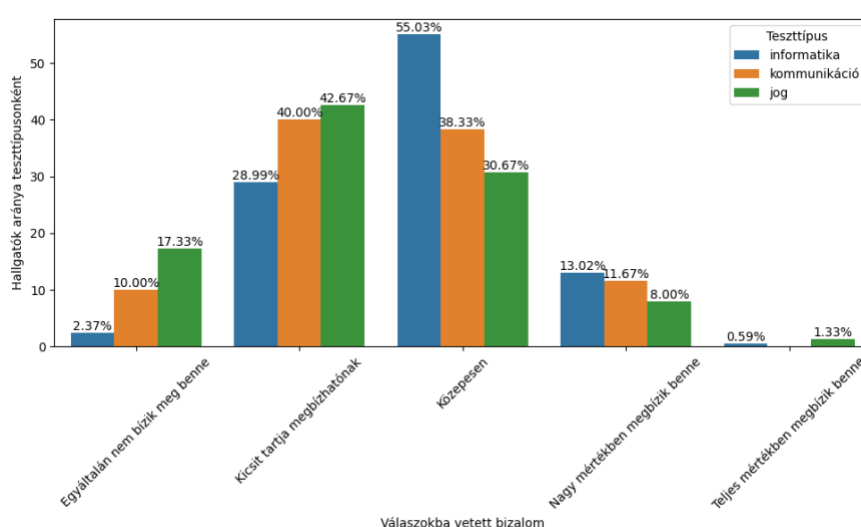
V.2.2. Bevallott tanulmányi célú ChatGPT használat

A hallgatók kérdőívezése alapján elmondható (lásd 10. ábra), hogy az informatikai mintarétegbe sorolt hallgatók esetén a legvalószínűbb a ChatGPT használata beadandók készítésére. Közel 70%-uk mondta inkább valószínűnek vagy nagyon valószínűnek a technológia alkalmazását. Emellett az elutasító válaszlehetőségeket is messze ezen réteg tagjai jelölték meg a legkisebb arányban. Ami a kommunikációs és jogi terület alanyait illeti, a ChatGPT használati hajlandóságuk a legtöbb választási lehetőség esetén elég hasonlóan alakult. A legnagyobb különbség közöttük a két szélső opció esetén mutatkozik, ahol a kommunikációtudományi hallgatók nagyobb valószínűséget mondanak az MI használata mellett. Összességében megállapítható, hogy amíg az informatikai terület tagjai inkább hajlamosak a ChatGPT tanulmányi célú alkalmazására, addig a másik két teszt esetén a hallgatók inkább semlegesek vagy elutasítók.



10. ábra A ChatGPT használatának valószínűsége egyetemi beadandó munka készítéséhez teszt típusonként (saját szerkesztés)

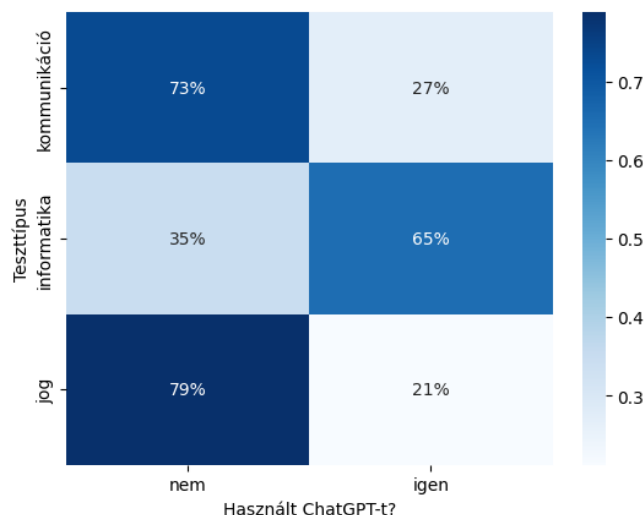
A ChatGPT válaszaiba vetett bizalom megoszlásával kapcsolatban hasonló tendenciák figyelhetők meg (lásd 11. ábra), mint ahogy a felhasználás valószínűségénél lett kimutatva. Amíg az informatikai réteg hallgatói bíznak meg leginkább a mesterséges intelligencia válaszaiban, addig a jogi réteg hallgatói a leginkább szkeptikusabbak az egyes válaszok pontosságával kapcsolatban. Lényeges azonban észrevenni, hogy bár az informatikai mintaréteg hallgatóira az MI használata inkább jellemző, a válaszok minőségével kapcsolatban a vizsgált hallgatók már kevésbé optimisták. A válaszadók 29%-a kevésbé tartja megbízhatónak a promptokra adott válaszokat, és összesen csak 13,61% bízik meg nagymértékben vagy teljesen a ChatGPT tanácsaiban.



11. ábra A ChatGPT válaszaiba vetett bizalom megoszlása tesztípusonként (saját szerkesztés)

V.2.3. Használati arányok a teszt során

A kutatáshoz fejlesztett böngésző kiegészítő segítségével a tanulmány résztvevőinek 84,5%-a, azaz összesen 257 hallgató böngészési tevékenysége lett nyomon követve a normára hivatkozott teszt intervalluma alatt. A kísérleti csoport 164 hallgatója közül 145 telepítette az ExamEye-t a számítógépére, közülük 70-en döntöttek a ChatGPT használata mellett a vizsga során. Feltételezve, hogy a speciális adatrögzítésben részt nem vevő hallgatók hasonló arányban használták a mesterséges intelligenciát, a kísérleti csoport egészére nézve 48,3% azoknak az aránya, akik legalább egy promptot írtak a teszt alatt. A ChatGPT használatában jelentős különbségek voltak az egyes tesztípusok között (lásd 12. ábra): az informatikai rétegnél a hallgatók 65%-a legalább egyszer használta a chat modellt, míg a kommunikációs tesztben 27%, és a jogi mintarétegnél mindössze 21% alkalmazta a GPT-t.



12. ábra A ChatGPT-t használók aránya a kísérleti csoportban a teszt során (saját szerkesztés)

Lényeges, hogy a mért használati arányok nagymértékben lekövetik mind a három mintaréteg esetén az előzetesen bevallott tanulmányi célú használat valószínűségét. Az informatikai teszt esetén 69%, a kommunikációs esetén 25%, a jogi réteg esetén pedig 19% mondta előzetesen, hogy inkább valószínűnek vagy nagyon valószínűnek tarja a ChatGPT használatát egyetemi beadandó feladatok készítésekor. Ez alapján a kutatás során alkalmazott különböző tesztek képesek voltak reprodukálni a mintára jellemző mindennapi tanulmányi célú mesterséges intelligencia használati arányokat.

V.2.4. Megfigyelt ChatGPT használati szokások

Az ExamEye bővítmény segítségével a megfigyelt 70 db. ChatGPT-t használó részéről 94 kontextus ablakban összesen 1518 prompt-response pár lett rögzítve, amelyből 1379 a GPT-3.5-ös modell és 139 a GPT-4-es modell használatával született. A nyomon követett mesterséges intelligenciát használó résztvevők átlagosan 22 promptot írtak a 30 kérdéses teszthez. Közülük a legkevesebb utasítást író hallgató egy promptot írt, a legtöbb pedig 56-ot. A kísérleti csoport egyes rétegeit vizsgálva megfigyelhető, hogy tesztípusonként eltérő az MI-t használó hallgatók interakciójának átlagos száma. Az informatikai réteg itt is kimagaslik a maga 24 prompt/tényleges felhasználó számával, öt követi a jogi és a kommunikációs réteg, ahol 14-14 prompt/tényleges felhasználó volt a várható ChatGPT megszólítások száma. A rögzített promptok átlagos szóköz nélküli karakter hosszúsága 103 karakter volt, ami megközelítőleg 18-22 szó lehetett promptonként (Szuhi, 2022). Az utasításokra adott ChatGPT válaszok várható szóköz nélküli karakter hossza 665 karakter volt, ami nagyjából 120-142 szó hosszúságú szövegeknek feleltethető meg.

A prompt írási stratégiákat tekintve az tapasztalható, hogy a megfigyelt hallgatók nagy többsége egyszerűen átmásolta a feladat szövegét a chat platformra, és mindenfajta különösebb további formázás vagy kiegészítés nélkül azt adták meg promptként. Általában egy darab kontextus ablakot használtak, az ablakon belül nem volt jellemző primer prompt, azaz a kontextusra vonatkozó utasítás, iránymutatás adása a mesterséges intelligencia modell számára (Akinyemi, 2023). A második megfigyelt tipikus prompt típus a hagyományos kereső programok kulcsszavas kereséséhez hasonlítható. Ezekben az esetekben a felhasználó néhány szavas kifejezését vagy szókapcsolatot adott a chat modellnek, amire a ChatGPT az előző promptok és a saját rendszer promptja alapján próbált reagálni (Liu, 2023). A megfigyelt ChatGPT Plus felhasználók esetén azt tapasztaltuk, hogy gyakran alkalmazták a részben vagy teljesen képalapú prompt adás lehetőségét. Ennek lényege, hogy a felhasználó a feladat szövegének bemásolása helyett képernyőképet készített arról, és az adott metszetet adja meg a ChatGPT-nek, bizonyos esetekben célzott utasításokkal és pontosításokkal kiegészítve (lásd 8. táblázat). A megfigyelt GPT-4-es modellt felhasználókról egyébként elmondható, hogy mindegyikük az informatikai rétegbe lett besorolva. Emellett hatból öt az adott mintaosztály átlaga felett teljesített, azonban csak 50%-uk tesztideje lett rövidebb az osztály átlagnál.

Eset	Prompt (idézet)
Primer prompt adása a normára hivatkozott teszttel kapcsolatban.	<i>Egy egyetemi kutatáshoz különböző kérdéseket kapok amiket ChatGPT használatával kell megválaszolnom. A kutatás célja, hogy kiderüljön mennyire tudjuk jól és pontosan megválaszolni a kérdéseket a ChatGPT használatával. A következő üzenetekben felelet választós és kifejtős kérdéseket fogok küldeni. Válaszolj rájuk a lehető leg pontosabban és a lehető leg rövidebben.</i>
A hallgató egyetlen promptba helyezte a feladatsorban található összes kérdés szövegét.	<i>Kérlek, válaszold meg az alábbi kérdéseket: András és 4 barátja ülnek egy kerekasztalnál egyenlő távolságra egymástól, középen van egy magas asztaldísz, összességében hány ember arcát takarhatja ki az asztaldísz András elől, ha mindenki egy általa választott székre ül?</i> <i>4. Melyik a következők közül a legjobb példája a web scrapingnek?</i> <i>5. A REVIEWS tábla oszlopát a PRODUCTS tábla oszlopával kell összekötni</i> ...

GPT-4-es modellnek adott kép alapú prompt kiegészítő szövege.	<i>A kép alapján találd meg a hiányzó szót a mondat "X" és "Y" helyére. mondat: A REVIEWS tábla "X" oszlopát a PRODUCTS tábla "Y" oszlopával kell összekötni</i>
GPT-4-es modellnek adott interneten történő keresésre irányuló prompt.	<i>Keresd ki az interneten, hogy igaz-e: "Az NVidia GeForce RTX 3090 a legerősebb, játékokra is tervezett videokártya."</i>
Hosszú válaszból a lényeg kinyerésére irányuló erősen informális hangvételű prompt.	<i>NE DUMÁLJ CSAK ADD MEG A VÁLASZT</i>

8. táblázat Válogatás a normára hivatkozott teszt során írt promptokból (saját szerkesztés)

A teszt ideje alatt a prompt átírás és válasz újra generálás viszonylag ritka felhasználói interakció volt a megfigyelt hallgatók esetében. A kutatás böngésző kiegészítője mindössze 12 esetet rögzített, amikor valamelyik résztvevő már elküldött prompt átfogalmazásával próbált volna pontosabb választ kapni a kérdésére, és 5 további alkalmat, amikor a válasz újbóli megalkotását kérte volna a hallgató a ChatGPT-től.

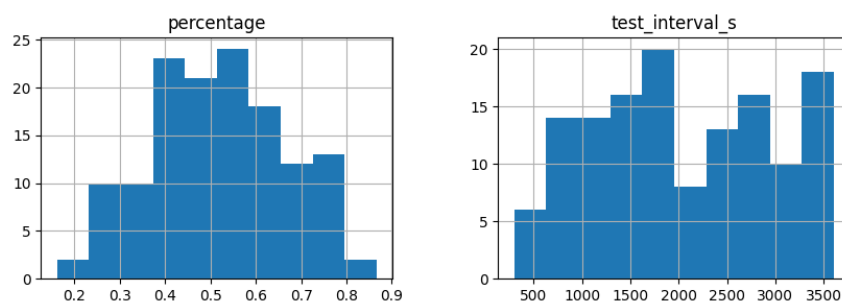
V.3. Eredményesség és hatékonyság változóinak bemutatása

A hallgatók eredményességét a normára hivatkozott teszt százalékos eredményében (*percentage* változó), hatékonyságát pedig a tesztírással töltött másodpercben számított idejében (*test_interval_s* változó) mértük. A függő változók értékének alakulását két okból is lényeges együtt vizsgálni:

- 1) a kutatási hipotézis szempontjából kulcsfontosságú annak megértése, hogy a ChatGPT használatának lehetősége mellett hogyan változott a hallgatók teszteredménye és tesztideje;
- 2) becslést tudjunk adni az egyes mintaosztályok figyelmének szintjére annak nyomkövetésével, hogy az elért eredményekhez átlagosan mennyi idő kellett. Hiszen az egyes minta szegmensek közötti eltérések mértékét befolyásolni tudta az osztályonként eltérő átlagosfigyelem.

Változó	átlag	szórás	minimum	maximum
<i>percentage</i>	51,89%	14,64%	16,13%	86,67%
<i>test_interval_s</i>	2051 mp	937 mp	300 mp	3606 mp
	(34 perc)	(16 perc)	(5 perc)	(60 perc)

9. táblázat A kontroll csoport eredményének leíró statisztikai táblázata (saját szerkesztés)



13. ábra A függő változók eloszlása a kontroll csoportban (saját szerkesztés)

A kontroll csoport egészét nézve (lásd 9. táblázat és 13. ábra), amelynek eredménye a teszt normájának számít, a várható százalékos eredmény 51,89% volt 14,64%-pontos szórással. A hallgatók középső 50%-a 21 és 45 perc között oldotta meg az egy óra hosszúra tervezett feladatsort. Azonban az eredmények tárgyalása során a továbbiakban rétegekre (teszt típusokra) bontva értelmezzük a kísérleti és kontroll csoport közötti eltéréseket az alábbi okok miatt:

- a kísérlet hatása tudományterületenként eltérő nagyságú lehet;
- az egyes területek hallgatói hasonló szerkezetű, de eltérő tartalmú tesztekkel írtak;
- egyes mintarétegek eltérő összetételű demográfiai és képzési háttérű hallgatókból állnak;
- a minta a képzési arányokra nézve nem reprezentatív.

Az eredményességet és hatékonyságot mérő mutatók rétegenkénti bontása megmutatja (lásd 10. táblázat), hogy az egyes teszt típusok esetén mennyire eltérő eredmények születtek. Az informatikai teszt esetén volt a legmagasabb a teszt várható normatív eredménye, amely 10%-ponttal magasabb lett a jogi és kommunikációs hallgatók eredményénél.

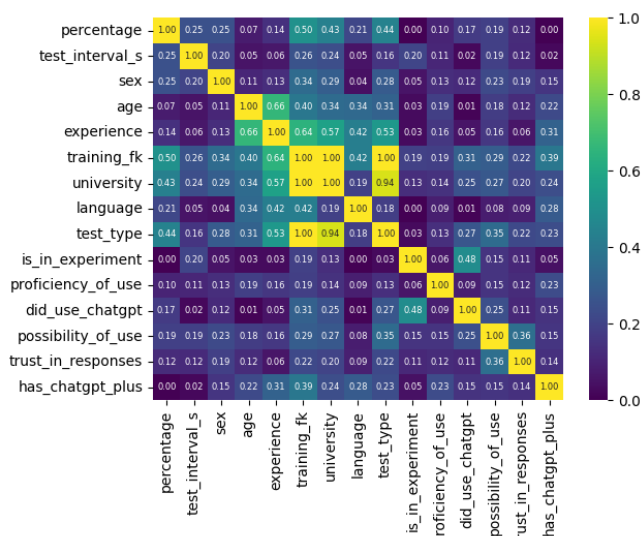
	informatika	jog	kommunikáció
percentage	57,21%	44,51%	46,67%
test_interval_s	2086 mp	2119 mp	1860 mp
	(35 perc)	(35 perc)	(31 perc)

10. táblázat A kontroll csoport eredményei rétegenkénti bontásban (saját szerkesztés)

V.4. Kapcsolat vizsgálat a megfigyelt változók között

Mielőtt vizsgálatra kerülne a függő változók közötti eltérések, érdemes először áttekinteni a mintát leíró (rögzített) ismérvek között fennálló kapcsolatok szorosságát. A kapcsolatvizsgálat lehetővé teszi, hogy megismerjük a különböző jellemzők között megbúvó összefüggéseket, illetve magyarázni tudjuk az egyes változók varianciáját. A következő hőtérkép (lásd 14. ábra) egy összesítő kapcsolatmátrix, amely tartalmazza a kapcsolat jellege szerint mért kapcsolati szorosságot a minta egészére nézve. Asszociatív kapcsolat esetén a Cramer-féle V együttható

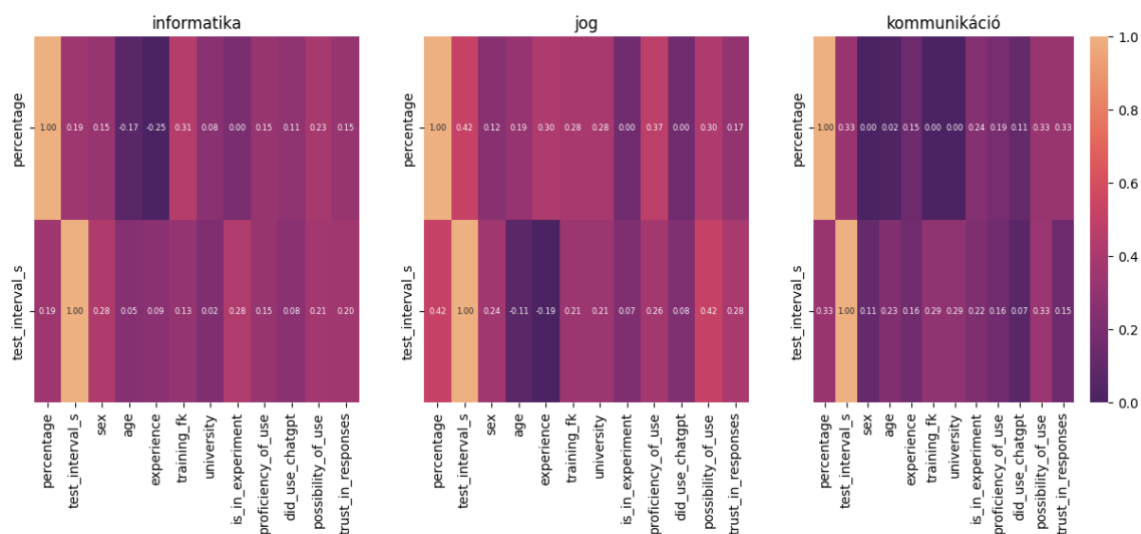
értékét, vegyes kapcsolat esetén a H (éta) mutató értékét, korrelációs kapcsolat esetén pedig a Pearson-féle r együttható értékét tartalmazza az ábra egyes metszete (Hunyadi & Vita, 2019a).



14. ábra Páronkénti kapcsolatszorosság hő térképe (saját szerkesztés)

A kapcsolatmátrix alapján látható, hogy jellemzően enyhe szorosságú, 0 és 0,3 közötti erősségű kapcsolat van az egyes változók között. Azonban a hallgatók képzési hatterét leíró változók, mint a résztvevők egyeteme (university), a képzésük egyedi azonosítója (training_fk) és az ezek alapján meghatározott teszt típus (test_type) változó többnyire közepes, vagy azt megközelítő szorosságú kapcsolatban állnak a többi változóval. Így tanulmányi adatokat leíró jellemzők általánosságban jobban magyarázzák a mintában megfigyelt adatok variációját.

A mintarétegek mentén látott erősebb külső variancia végett teszt típusonként is vizsgálatra került a függő változók és minta összetételét jellemző (kvázi független) változók kapcsolata.



15. ábra Rétegenkénti kapcsolatszorosság a függő és az összes változó között (saját szerkesztés)

A hőtérkép alapján (*lásd 15. ábra*) leginkább a jogi teszt esetén magyarázta a függő változók varianciáját a minta összetételét leíró változók, az átlagos abszolút kapcsolatszorosság pedig az informatikai réteg esetén volt a legenyhébb. Amíg 0,42 erősségű kapcsolat mellett 17,64% magyarázta a joghallgatók tesztideje az eredmények átlagos eltérését, addig kommunikációs rétegnél a pontszámok varianciája 10,89%-ban, az informatikánál pedig mindössze 3,61%-ban hozható kapcsolatba a vizsga megoldására szánt idővel.

A hallgatók felsőoktatásban töltött évei mind a három tesztípus esetén erősebb magyarázó hatással bírtak a teszt eredményére, mint az életkoruk. Az informatikai réteget leszámítva a tapasztalat növekedésével átlagosan több pontot értek el a hallgatók a mérés során. Az informatikai réteg esetén a tapasztalat és teszteredmény közötti negatív korreláció a szakképzési összetétel heterogenitása miatt van. A nem terület specifikus tanulmányokat végző résztvevők mesterszakos hallgatók voltak, így az ő átlagos eredményük a korreláció irányát megfordította. Ami a várható tesztidőt illeti, a tapasztalat változó egyedül a jogi réteg esetén csökkentette a hallgatók átlagos tesztidejét.

Az intézményi és képzésbeli különbségek az informatikai és a jogi teszt esetén a pontszámban, a kommunikációs teszt esetén pedig inkább a tesztidő területén jelentett előnyt vagy épp hátrányt. A kereső kifejezések és programok használatában való tapasztalat (*proficiency_of_use* változó) a joghallgatók esetén számított különösen hasznosnak, ami a jogszabályok keresésével kapcsolatban kulcsfontosságú képesség.

V.5. Eltérések megfigyelése

V.5.1. Várható értékek közötti eltérés

A teszteredmények várható értéke minden mintaréteg esetén rosszabb lett a kísérleti csoportban, mint a kontroll csoport esetén (*lásd 11. táblázat*). Az eltérések nagysága azonban az informatikai és jogi teszt esetén igen csekély. Míg ez előbbi rétegnél kb. 1,03%-pont különbség van a két csoport hallgatóinak várható eredménye között, addig az utóbbi tesztípus esetén mindössze kb. 0,13%-ponttal lett rosszabb a kísérleti csoport várható eredménye. Egyedül a kommunikációtudományi réteg esetén tapasztalható egy jelentősnek tűnő 4,61%-pontos eltérés a kontroll csoport javára az átlagos teszteredmény tekintetében.

cs csoport réteg	percentage		test_interval_s	
	<i>kontroll</i>	<i>kísérleti</i>	<i>kontroll</i>	<i>kísérleti</i>
<i>informatika</i>	57,21%	56,18%	2086 mp (35 perc)	1667 mp (28 perc)
<i>jog</i>	44,51%	44,38%	2119 mp (35 perc)	1967 mp (33 perc)
<i>kommunikáció</i>	46,67%	42,06%	1860 mp (31 perc)	1457 mp (24 perc)

11. táblázat Függő változók értéke a kísérleti és kontroll csoport különböző rétegeiben (saját szerkesztés)

A teljesítmény értékelése szempontbántjából másik kulcsfontosságú változó, a másodpercben mért tesztidő esetén a különbségek már könnyen értelmezhetőek, tekintettel arra, hogy az eltérések a kísérleti és kontroll csoport között nem mondhatóak marginálisnak. Minden tesztípus esetén a kísérleti csoport teszt idejének csökkenése tapasztalható. A tesztírára fordított átlagos időben a legnagyobb gyorsulást az informatikai teszt esetén lett megfigyelve, ahol várhatóan majdnem hét perccel (419 másodperc) lett gyorsabb a kísérleti csoport a kontroll csoportnál. A legkisebb eltérést a jogi teszt esetén mértük: a kísérleti csoport ideje durván 2,5 perccel (152 másodperc) lett átlagosan rövidebb. Habár a maga 403 másodpercével a kommunikációs teszténél is látványos a kísérleti és kontroll csoport tesztideje közötti különbség, az alábbi szempontok miatt arra a következtetésre jutottunk, hogy a tesztidő csökkenése valószínűleg nem a beavatkozás hatásával, hanem a ráfordítás csökkenésével és egy mintavételi hiba meglétével magyarázható (lásd V.6.3. fejezet). Egyrészt a teszteredmények várható értéke a többi réteggel ellentétben ennél a tesztípusnál esett vissza a legnagyobb mértékben. Másrészt ennél a csoportban lehetett a legalacsonyabb az átlagos figyelem mértéke. Harmadrészt a böngészőhasználat radikálisan visszaesett a kísérleti csoportnál.

cs csoport képzés	létszám		percentage		test_interval_s	
	<i>kontroll</i>	<i>kísérleti</i>	<i>kontroll</i>	<i>kísérleti</i>	<i>kontroll</i>	<i>kísérleti</i>
<i>informatikai</i>	67 fő	85 fő	58,76%	57,22%	2095 mp (35 perc)	1994 mp (33 perc)
<i>nem informatikai</i>	7 fő	10 fő	42,38%	47,33%	1688 mp (28 perc)	1491 mp (25 perc)

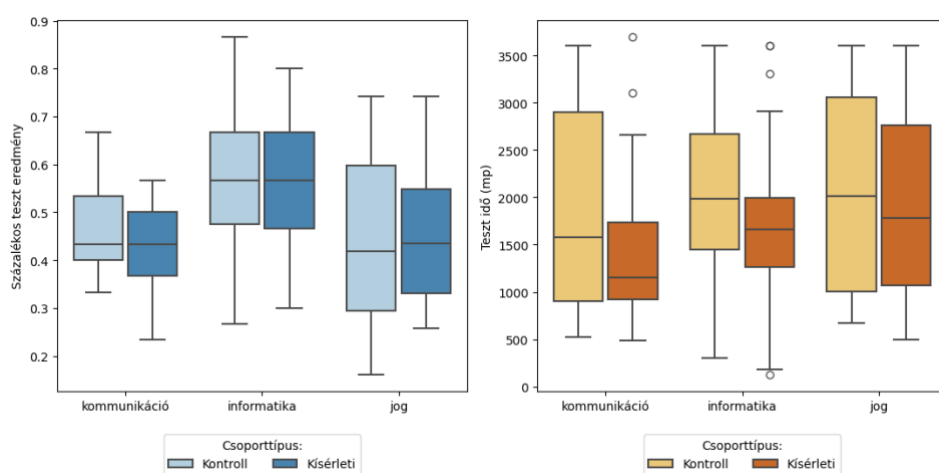
12. táblázat Informatikai réteg eredményeinek szétbontása képzés jellege szerint (saját szerkesztés)

Az informatikai réteg eredményeinek a képzés jellege szerint való szétbontása által képet kaphatunk arról, hogy a ChatGPT használat lehetősége hogyan hat olyan hallgatók teljesítményére, akik megfelelő előzetes ismeretekkel nem rendelkeztek (lásd 12. táblázat). A beavatkozás hatása látványos: a nem informatikai tanulmányokat végző része a kísérleti

csoportnak átlagosan úgy írt 4,95%-kal jobb tesztet a kontroll csoportnál, hogy közben a tesztideje több mint 3 perccel rövidebb lett.

V.5.2. Függő változók eloszlásában mért különbség

Lényeges megjegyezni, hogy a várható értékek számítására a számtani átlag becslő függvényét használtuk, ami a változó összes értékét figyelembe véve add becslést a jellemző középértékre, viszont ebből adódóan érzékeny az erősen aszimmetrikus eloszlásra és a kilógó értékű megfigyelésekre (Hunyadi & Vita, 2019a). A kísérlet okozta eltérések pontosabb megértése érdekében fontos, hogy a függő változók várható értékének változása mellett vizsgáljuk az eloszlásukban jelentkező differenciákat is.



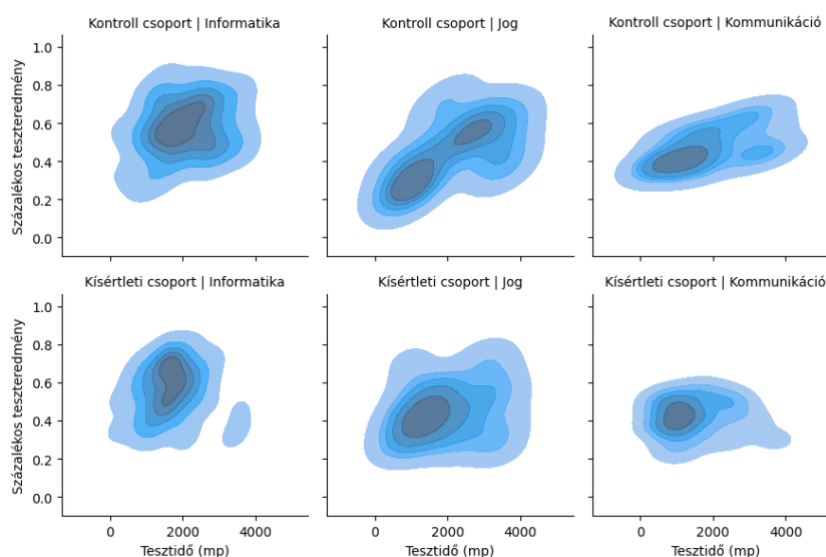
16. ábra A százalékos teszteredmény és a tesztidő változók eloszlása csoportonként és teszttípusonként (saját szerkesztés)

Ami a változók teljes terjedelmét illeti, megfigyelhető, hogy a kísérleti csoport eredményeinek eloszlása mind az informatikai, mind a jogi teszt esetén 10%-ponttal kisebb intervallumba esik, ami a legjobb és a legrosszabb hallgatói teljesítmény közötti különbségek mérsékelt csökkenését jelenti. A kommunikációs teszt esetén a terjedelem nem változott, de ez eredmény változó eloszlása balra tolódott. Ez azt jelenti, hogy a mintában látványosan romlott a pontskála kétvégén elhelyezkedő hallgatók teljesítménye, habár a minimum és a maximum elért pontszám közötti különbség nem változott. A hallgatói eredmények interkvartilis terjedelmének változása tekintetében, ami a teszteredmények eloszlásának középső 50%-át tartalmazza, a következők figyelhetők meg. A kommunikációs és informatikai teszttípus esetén nem változott az első és a harmadik kvartilis távolsága a beavatkozás hatására. A jogi teszt esetén a kísérleti csoport eredményeinek középső tartománya azonban kisebb varianciát mutat a terjedelem 8,39%-pontos csökkenése mellett.

A fent megfigyelt változások a százalékos teszteredmény változó rétegenkénti eloszlásának szimmetriájára is hattással voltak. A kommunikációs teszt esetén az eloszlás ferdesége közepesen pozitívból enyhén negatív ferdeségbe ment át. Az informatikai teszt esetén az adott függő változó eloszlása enyhén balra elnyúló maradt, de a ferdeség értéke háromszorosára nőtt. A negatív (baloldali vagy balra elnyúló) ferdeség azt jelenti, hogy a vizsgált ismerv értékeinek a nagyobb része a változó terjedelmének felső felében összpontosul (Taylor, é. n.), azaz az informatikai és a kommunikációs teszt mellett a kísérleti csoport hallgatóinak nagyobb része írt az osztályátlaguknál jobb tesztet a kontroll csoporttal szemben. Habár az kísérleti osztályok várható értéke minden esetben kisebb lett a norma értékénél. A jogi tesztnél sem változott az aszimmetria iránya, de a kísérleti csoport ferdeségének értéke négyszerre jobbra elnyúlóbb lett, mint a kontroll csoporté, ami az átlag alatt teljesítő hallgatók arányának növekedését jelenti. Az egyes mintaosztályok mediánjával kapcsolatban érdemes megjegyezni, hogy bár a beavatkozás hatására az eloszlás alakjának változása volt megfigyelhető, de a középső kvartilisek értéke nem mutatott változást a kommunikációs és informatikai teszt esetén. Sőt a jogi réteg mellett a kísérleti csoportnál még javulást is tapasztalhattunk 1,61%-ponttal. Ezek a megfigyelések összességében azt mutatják, hogy a teszteredmények középpontját nem érintette a kísérlet, attól függetlenül, hogy az egyes értékek hogyan oszlottak meg vagy mennyire voltak szimmetrikusak.

A másodpercben mért tesztidő változó esetén a kísérleti és kontroll csoport eloszlása között szemmel láthatóan nagyobb eltérés van az adatok sűrűségénél. Lényeges megállapítás, hogy az interkvartilis terjedelmek tendenciális csökkenése mellett megjelentek kilógó értékek, amelyek a változó teljes intervallumának növekedését eredményezték. Ez alapján arra a következtetésre jutottunk, hogy a beavatkozás, azaz a ChatGPT használatának lehetősége növeli az egyenlőtlenséget a hallgatói tesztidők között. Habár a mesterséges intelligencia a középső 50% teszt idejének varianciáját összébb húzza, azonban a technológia felhasználása mentén a különbségek összességében növekednek a hallgatók hatékonyságában. Hiszen a kései többség, az elutasítók, és a technológiát rosszul használók lemaradnak mind a korai alkalmazkodókkal, mind a korai többséggel szemben (Rogers, 2003). Ezt a feltevésünket erősíti, hogy mind a mediánnak, mind a felső kvartilis értékének változása csökkenő tendenciát mutat a kísérleti csoportban, amíg az alsó kvartilisek értéke a különböző rétegek esetén eltérően reagált a beavatkozás hatására.

V.5.3. Függő változók együtt mozgásának vizsgálata



17. ábra A tesztidő és teszteredmény változók kernel sűrűség diagramja (saját szerkesztés)

A teszteredmény és tesztidő változó együtt mozgásának vizsgálatához, a két függő változó eloszlását egy-egy kernel sűrűség diagram segítségével ábrázoljuk az egyes mintaosztályok esetén (lásd 17. ábra), amelyeken az elért teszteredmények a tesztidő függvényében olvashatóak le, és minél sötétebb az egyes sűrűség foltok színe, annál több megfigyelés található az adott területen belül. A fenti ábra szemlélteti, hogy a kontroll csoportban az egyes tesztípusok esetén a megfigyelések jobban elnyúlnak a tesztidő mentén, tehát az eredmények alakulását jobban meghatározta a teszt elvégzésére szánt idő. Ezzel szemben a kísérleti csoportban az eloszlás koncentráltabb időben, a tesztidő kevésbé magyarázza az eredményt, mivel a tesztidő függvényében nagyobb a százalékos eredmények varianciája. A kontúrvonalak széleit vizsgálva észrevehetjük a tesztidőben mért kilógó értékek jelenlétét a beavatkozás mellett: a kommunikációs teszt eloszlásának jobbra elnyúló farka van, az informatika teszt kernel sűrűség ábrája pedig szétválk egy fő és egy mellék foltra.

V.6. Böngészőhasználati különbségek

V.6.1. A böngészőhasználat és a tesztidő kapcsolata

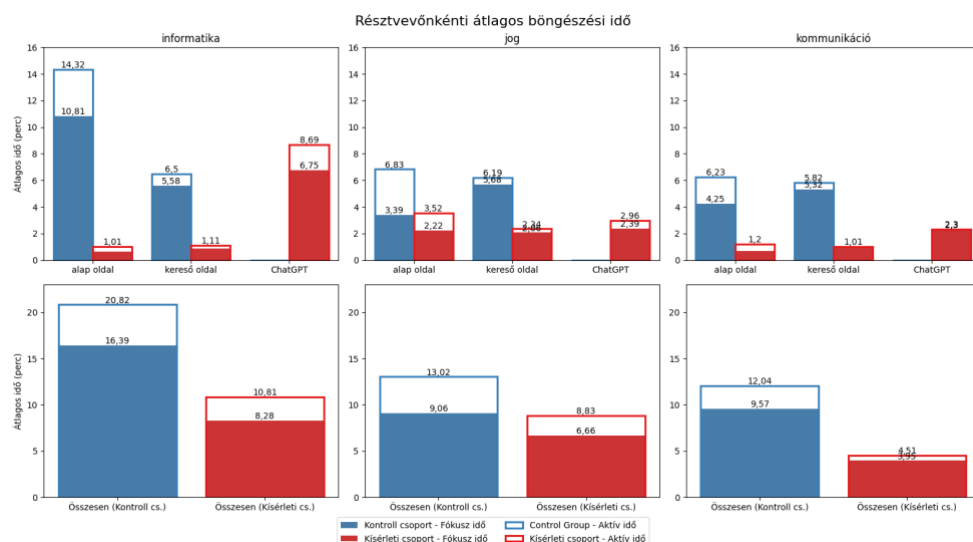
Az eredményesség és hatékonyság változásának vizsgálatakor megfigyelhettük, hogy a kísérleti csoport tesztideje minden réteg esetén átlagosan rövidebb lett. A hallgatók internethasználatának rögzítése alapján arra a következtetésre jutottunk, hogy a tesztidők javulása a beavatkozás hatására megváltozó böngészőhasználatnak tulajdonítható. Ugyanis a

ChatGPT nem csak ismereteket szolgáltat egy helyen, de egyben radikálisan megváltoztatja az információkeresés és feldolgozás folyamatát, valamint az emberek hozzáállását a hagyományos keresési csatornák használatával kapcsolatban.

Itt is sarkalatos újra megemlíteni a normára hivatkozott vizsga nyitott könyves jellegét: a tesztek kidolgozásakor ugyanis fontosnak tartottuk, hogy ne a hallgatók tudása, hanem a probléma megoldó képessége legyen döntő a feladatok megoldásakor. Tehát a módszertan lényeges eleme, hogy a tesztalanyoknak lehetőleg nagyban kelljen támaszkodniuk külső források segítségére. Azért lényeges ezt hangsúlyozni, mivel a résztvevők tesztidejüknek jelentékeny részét a kérdések helyes megválaszolásához szükséges információk keresésével és szintetizálásával töltötték.

V.6.2. Böngészési idő és keresési szám megváltozása

Ahogy a kutatás módszertani részénél már be lett mutatva, logikailag három csoportba soroltuk a hallgatók által meglátogatható oldalak típusát: megkülönböztettük az alap weboldaltól a kereső oldalakat (Google és Yahoo), valamint a ChatGPT oldalát. Ahhoz, hogy becslést lehessen adni, hogy a résztvevők az egyes oldalakon mennyi időt töltenek figyelemmel, két idő intervallumot lett mérve. Az ExamEye segítségével rögzítettük, hogy az egyes weboldalak mennyi ideig voltak fókusz, illetve aktív állapotban: a fókusz állapot az, amikor a felhasználói műveletek az oldalra, vagy az oldal valamelyik elemére vonatkoznak; az aktív állapot pedig az, amikor oldal fókuszban van, vagy legalább is nincsen elrejtve a képernyőn (*Document: hasFocus() method*, é. n.). A fókusz és aktív idő segítségével be lehetett határolni, hogy a hallgatók az egyes csatornákon milyen figyelemi szinten - fókusz esetén kvázi kizárólagos, aktív esetén kvázi megosztott - átlagosan mennyi időt töltöttek a vizsga ideje alatt.

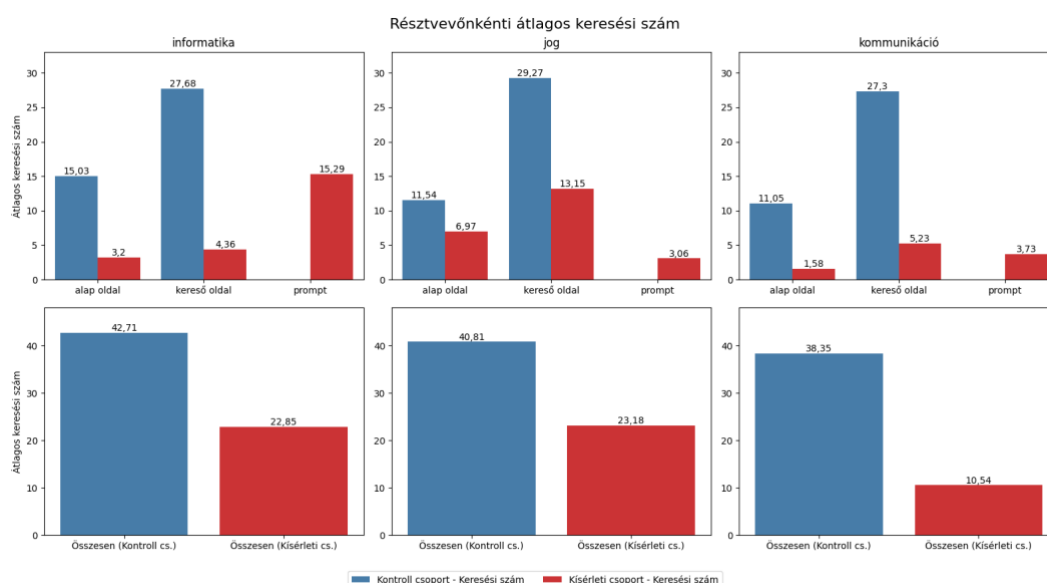


18. ábra Résztevőnkénti átlagos böngészési idő tesztcsoportonkénti bontásban (saját szerkesztés)

A kísérlet során nem csak a várható tesztidő, de a hozzátartozó böngészési idő is rövidebb lett a beavatkozás hatására minden mintaréteg esetén (lásd 18. ábra). Az adatrögzítés alapján átlagosan az informatikai teszt írói töltötték a legtöbb időt böngészéssel a vizsga során. A kontroll csoport hallgatói 16-21 percet, a kísérleti csoport tagjai pedig 8-11 percet tartózkodtak a vizsga felületen kívül más weboldalon. A jogi és a kommunikációs teszt esetén a kontroll csoportok internethasználata hasonló kb. 9-13 perces tartományba esik, azonban a kommunikációs teszt kontroll csoportja feltűnően a legkevesebbet böngészett a felmérés során. A böngészési idők keresési csatorna szerinti megoszlását tekintve, a kontroll csoportban minden réteg esetén a hallgatók az alap weboldalakon töltötték a legtöbb időt. Az időráfordítás alapján ezek az oldalak voltak az információ szerzés elsődleges forrásai, ugyanakkor a figyelmet viszonylag nagy hányadban csak megosztva élvezték a széles aktív-fókusz időkülönbség miatt. Habár a kontroll csoportban domináns volt az alap oldalakon töltött idő, a letöltések számát tekintve minden réteg esetén közel kétszer több kereső oldalt látogattak meg a hallgatók, mint konkrét cél oldalt (lásd 19. ábra). Ebből is látszik, hogy a kereső oldalak szerepe ma már messze túlmutat a céloldalak keresésén. A keresők által adott szövegrészletek és összefoglalók számos esetben elegendőek a felhasználók számára. Ez a megközelítés pedig megteremti a közvetlen forrás olvasástól elszakadó, a ChatGPT és egyéb nagy nyelvi modellek által tovább fejlesztett információszerzés és feldolgozás szokását.

A kísérleti csoportban nem csak az átlagos böngészési idő esik vissza, de a ChatGPT megjelenésével az egyes csatornák látogatásának száma mellett az azokon töltött idő eloszlása is megváltozik. A leglátványosabb különbség az informatikai teszt esetén figyelhető meg a két csoport között, ahol a ChatGPT-t használók aránya is a legmagasabb volt a többi tesztípushoz

képest. Messze kimagaslik mind a promptok átlagos száma, mind a chat platformon töltött idő az alap és kereső oldalak használatához képest. A kereső oldalak szerepe láthatóan ártékelődik: a keresések nagyobb eséllyel konvertálódnak oldal megnyitássá, ahelyett, hogy a találatokból kiemelt részleteket olvasnák a hallgatók. A ChatGPT a megfigyeléseink alapján magas használati (elfogadási) arány mellett a tradicionális csatornák helyettesítőjévé válik: nem csak a keresőt, de céloldalakat is megpróbálja pótolni a felhasználók számára.



19. ábra Részvevőnkénti átlagos keresési szám mintacsoportonként (saját szerkesztés)

V.6.3. Böngészési aktivitás visszaesésének okai

A jogi és a kommunikációs rétegeknél, ahol a kísérleti csoportban a ChatGPT-t használók aránya 30% alatt volt, egy másik lényeges jelenséget figyeltünk meg: a böngészési aktivitás úgy csökkent - az informatikai teszthez hasonló, vagy akár azon túlmutató különbséggel - hogy a mesterséges intelligencia használata nem volt igazán elterjedve a kísérleti csoport adott osztályaiban. A fenti ábrához fontos megjegyezni, hogy a bemutatott átlagos prompt szám a kísérleti csoport adott osztályának tagjaira értendő, nem az osztályon belüli tényleges felhasználókra. A böngészés mennyiségének csökkenését három lehetőség együttes hatásával magyarázzuk:

- 1) a ChatGPT használata csökkenti a szükséges keresések számát, mivel kevesebb oldal letöltése mellett is képes a felhasználó a szükséges információkat megtalálni;
- 2) a ChatGPT használatának lehetősége - elsősorban kényelmi okokból - csökkenti a hagyományos információs csatornák használta iránti hajlandóságot, még akkor is, ha a felhasználó végül nem dönt a mesterséges intelligencia használata mellett;

- 3) a kísérleti csoport számos hallgatója *félreértette* a csoportszabályzat lényegét, és úgy értelmezte, hogy a nyitott könyves teszt során a ChatGPT-n kívül más külső forrást nem használhat.

Az (1) feltevésünket alátámasztja a ChatGPT-t (valóban) használók esetén megfigyelt böngészési adatok. Egyrészt a vizsgált hallgatók keresésének többsége a chat platformon keresztül történt: a kommunikációs teszt esetén a letöltések átlagosan 77%-a, az informatikai teszt esetén 73%-a, a jogi teszt esetén pedig 50%-a származott a ChatGPT felületéről. Másrészt az adatrögzítés alapján a ChatGPT-t használó résztvevők tesztípusonként rendre kevesebb alkalommal használtak bármilyen külső forrást, az MI intenzív alkalmazása mellett.

A (2) feltételezés alapján böngészési szám azért esett vissza a kísérleti csoportban, mert a hallgatók úgy ítélték meg, hogy a feladatok teljesítésére nem alkalmas az MI, de a GPT használat lehetősége mellett nem tartották komfortosnak a hagyományos csatornák alkalmazását. Ennek a feltételezésnek leginkább a böngészés jogi réteg esetén mért kvázi közepes nagyságú visszaesése adott alapot. Azonban ezen feltételezésünk igazolásához további kutatás szükséges.

A (3) feltételezés a következő megfigyelésre lett alapozva. Amíg a kísérleti csoport (nyomon követett) ChatGPT-t nem használó hallgatóinak 61%-a egyetlen alkalommal sem keresett információt az interneten, addig a kontroll csoport tagjai különböző tesztípusonként átlagosan közel ugyanannyit böngészték összességében és az egyes csatornákon egyaránt. A fenti diagram alapján (*lásd 19. ábra*) úgy tűnik, hogy leginkább a kísérleti csoport kommunikációs osztályának hallgatói értették félre a kísérlet kereteit.

V.7. Eltérések szignifikanciájának vizsgálata

V.7.1. Alternatív hipotézisek és próba feltételek

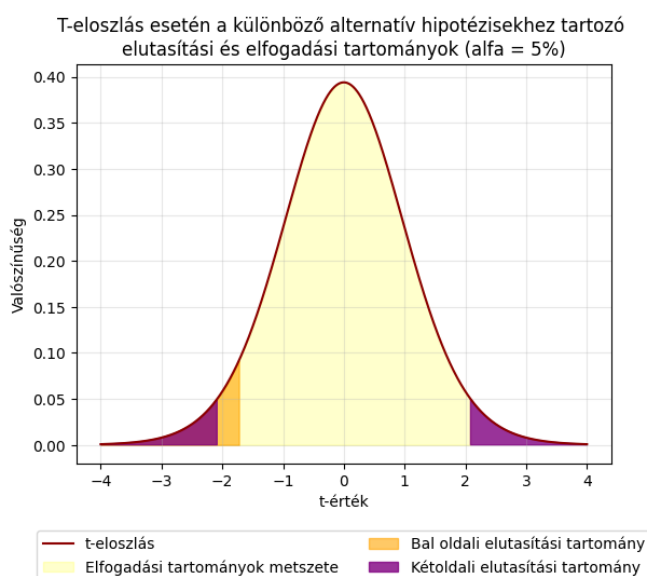
A kutatási hipotézis szerint a kísérleti csoport hallgatóinak átlagos eredménye - szignifikánsan jobb tesztidő mellett - rétegenként megegyezik a kontroll csoportéval, azaz a ChatGPT használata a hatékonyság oldalán fejti ki a hatását. Ahhoz, hogy a látott eltérésekről meg lehessen állapítani, hogy szignifikánsak, azaz nem a véletlen mintavételnek az eredménye, hanem a tapasztalt különbségek hasonló összetételű sokaságok esetén is igazak, szükséges, hogy a hipotézis vizsgálat eszközeivel is górcső alá vegyünk az egyes mintaosztályokat. Ugyanis objektív statisztikai próbák nélkül a látott eltérések alapján könnyedén a megerősítési torzítás hibájába eshetünk (Casad & Luebering, 2024).

Azért, hogy a kutatási hipotézisről dönteni lehessen, rétegenként megvizsgáltuk a kísérleti és kontroll csoport esetén a százalékos teszteredmény és a tesztidő változók várható értéke (számtani átlaga) közötti különbségeket. A hipotézis vizsgálat első lépése, hogy a feltételezésünkhöz megfelelő null és alternatív hipotézist állítsunk. Azon állításunkat, hogy a kísérleti csoport eredménye a ChatGPT használatának lehetősége mellett a kontroll csoportéval megegyezik részben átfedő alternatív hipotézisekkel párba állítva teszteltük (lásd 13. táblázat).

Próba azonosítója	Alternatív hipotézis szövege (H_1)	Alternatív hipotézis képlete	Próba típusa
(1)	A kísérleti csoport várható teszt eredménye szignifikánsan eltérő a kontroll csoportétól.	$\mu(\% \text{ kísérleti}) - \mu(\% \text{ kontroll}) \neq 0$	kétoldali alternatívájú
(2)	A kísérleti csoport szignifikánsan rosszabb tesztet írt a kontroll csoportnál.	$\mu(\% \text{ kísérleti}) - \mu(\% \text{ kontroll}) < 0$	bal oldali alternatívájú

13. táblázat A teszt eredmények vizsgálatára vonatkozó alternatív hipotézisek táblázata (saját szerkesztés)

Lényegesnek tartottuk, hogy részben átfedő alternatív hipotézisű próbákat fűtassunk a null hipotézisre, mivel az egyoldali tesztek magasabb próba erővel rendelkeznek, mint a kétoldali alternatívájúak azonos szintű megengedett elsőfajú hiba mellett (Hales, 2023). Így a két statisztikai próba p-értéke alapján pontosabban tudunk nyilatkozni a H_0 -val kapcsolatban. A t-eloszlás esetén az elutasítási tartományok átfedését és az elfogadási tartomány metszetét az alábbi ábra szemlélteti (lásd 20. ábra).



20. ábra Elfogadási és elutasítási tartományok t-eloszlás esetén (saját szerkesztés)

A kutatási hipotézis azon feltevését, hogy a kísérleti csoport a ChatGPT használatának lehetősége mellett szignifikánsan rövidebb tesztidővel tudja teljesíteni a vizsgát szintén két alternatív hipotézis segítségével teszteltük (lásd 14. táblázat). Ezáltal is csökkentve a másodfajú hiba bekövetkezésének lehetőségét, azaz, hogy minél kisebb valószínűséggel fogadjunk el hamis null hipotézist (Hunyadi & Vita, 2019b). Ugyanakkor lehetőségünk legyen a szignifikancia határán levő eltérésekről irány semleges p-értékkel is nyilatkozni.

Próba azonosítója	Alternatív hipotézis szövege (H ₁)	Alternatív hipotézis képlete	Próba típusa
(3)	A kísérleti csoport várható tesztideje szignifikánsan rövidebb a kontroll csoporténál.	$\mu(t \text{ kísérleti}) - \mu(t \text{ kontroll}) < 0$	bal oldali alternatívájú
(4)	A kísérleti csoport várható tesztideje szignifikánsan eltérő a kontroll csoporténál.	$\mu(t \text{ kísérleti}) - \mu(t \text{ kontroll}) \neq 0$	kétoldali alternatívájú

14. táblázat A tesztidők vizsgálatára vonatkozó alternatív hipotézisek táblázata (saját szerkesztés)

A várható értékek különbségének hipotézis vizsgálatához kétmintás t-próbát használtunk. A független kétmintás Student t-próbának az alábbi felételeknek kell megfelelnie (Hunyadi & Vita, 2019b), hogy a teszt pontos eredményt adjon:

- 1) a minták függetlenek legyenek
- 2) a vizsgált ismerv kis mintaméret esetén normális eloszlású legyen (nagy mintaméret esetén a centrális határeloszlás tétele miatt az átlag becslőfüggvénye normál eloszlást követ)
- 3) a minták varianciája megegyezzen (f-teszt alapján szignifikánsan ne térjen el).

Azokban az esetekben, amikor a minták varianciája között az f-teszt szignifikáns eltérést mutatott ki, a Welch-féle t-próbát alkalmaztuk, mivel a t-próba ezen változata robosztus az eltérő méretű és varianciájú mintákra (Welch, 1947). Miután ellenőriztük, hogy minta osztályok a próba alapvető feltételeinek megfelelnek, a varianciák egyezőségének függvényében lefuttattuk a megfelelő hipotézis vizsgálatokat.

V.7.2. Teszteredmények eltérésére vonatkozó próbák eredménye

Próba azonosítója	Mintaréteg	p-érték	Döntés a H_0 -ról
(1)	informatika	0,6098	elfogadva
	jog	0,9689	elfogadva
	kommunikáció	0,0523	elfogadva
(2)	informatika	0,3049	elfogadva
	jog	0,4844	elfogadva
	kommunikáció	0,0262	elutasítva

15. táblázat A várható százalékos teszt eredmények eltérésére vonatkozó próbák eredményének táblázata (saját szerkesztés)

A statisztikai próbák eredménye alapján elmondható (lásd 15. táblázat), hogy 5%-os megengedett elsőfajú hiba mellett sem az informatikai teszt esetén, sem a jogi teszt mellett nem szignifikáns a vizsgák várható eredménye szempontjából a kísérleti és kontroll csoport között mért eltérések. A megállapítást alátámasztja mind a kétoldali, mind az egyoldali alternatívájú hipotézis vizsgálatok eredményéül kapott p-értékek. Ez alapján az egyes mintaosztályok várható teszt eredménye között tapasztalt különbségeket a véletlen mintavétel hatásaként értékeljük, és nagy valószínűséggel kijelenthető, hogy az eredmények romlása nem a beavatkozás miatt történt az informatikai és jogi vizsga írói esetén. Ami a kommunikációs tesztet illeti, a várható értékek közötti különbséget szignifikánsnak mondta az eltérés irányára nézve specifikus, erősebb (2) próba. Ezzel szemben a kétoldali alternatívájú t-teszt 5,23%-os p-érték mellett csak közel szignifikánsnak ítélte. Ebben az esetben az erősebb próba eredményére nézve elutasítottuk a null hipotézist miszerint a kommunikációs réteg kísérleti csoportja a kontroll csoporttal megegyező eredményű tesztet írta. De tekintettel a p-értékek ellentmondásosságára, javasoljuk nagyobb mintaméret mellett a kísérlet megismétlését ennél a teszt típusnál.

V.7.3. Tesztidők eltérésére vonatkozó próbák eredménye

Próba azonosítója	Mintaréteg	p-érték	Döntés a H_0 -ról
(3)	informatika	0,0002	elutasítva
	jog	0,2603	elfogadva
	kommunikáció	0,0477	elfogadva*
(4)	informatika	0,0004	elutasítva
	jog	0,5206	elfogadva
	kommunikáció	0,0954	elfogadva

16. táblázat A másodpercben mért várható tesztidők eltérésére vonatkozó próbák eredményének táblázata (saját szerkesztés)

Ami a tesztidők eltérésének szignifikanciáját illeti (*lásd 16. táblázat*), bár a jogi teszt esetén a kísérleti csoport átlagos tesztideje 2,5 perccel lett rövidebb a kontroll csoporténál, az eltérés nagysága egyik alternatív hipotézis mellett sem volt statisztikailag jelentős.

A kommunikációs tesztnél az egyoldali alternatívájú teszt a kísérleti csoport várható teszt idejének csökkenését 4,77%-os p-érték mellett szignifikánsnak hozta ki, 5% szignifikancia szint esetén. Ennek ellenére nem tartjuk a várható tesztidők közötti különbséget szignifikánsnak, figyelembe véve, hogy a kommunikációs réteg esetén a beavatkozással nem magyarázható mértékben esett vissza a kísérleti csoport böngészési aktivitása, aminek valószínűleg a kutatás szabályainak félreértése volt az oka. Az adott mintaosztályban a ChatGPT-t nem alkalmazó (megfigyelt) hallgatók 72% nem használt semmilyen más külső forrást, mert feltételezhetően úgy értette, hogy a kísérleti csoport csak a mesterséges intelligencia segítségét veheti igénybe. Tehát az alternatív hipotézist szignifikáns p-érték mellett azért utasítottuk el mégis, mivel a szignifikancia szigorúan vett jelentése, hogy a különbséget nagy valószínűséggel nem a véletlen mintavétel okozza, az eltérés hasonló összetételű sokaságra általánosítható. Ámde a kísérleti csoport gyorsulását feltételezhetően egy mintavételi hiba torzítása okozhatta, azaz nem tudjuk kellő bizonyossággal kizárni, hogy a differencia csak a véletlen műve, így a null hipotézist elfogadása mellett született döntés.

Az informatikai tesztípus esetén a kísérleti csoport idejének gyorsulása viszont megkérdőjelezhetetlen: mind az egyoldali, mind a kétoldali alternatívájú próbák esetén a p-érték minden szokásos szignifikancia szintnél alacsonyabb. Ezért elutasítjuk a null hipotézist és az alternatív hipotézist fogadjuk el, miszerint az informatikai mintaréteg esetén a kísérleti csoport hallgatói pusztán a véletlen mintavétellel nem magyarázható mértékben, rövidebb várható tesztidővel végezték el a normára hivatkozott vizsgasor feladatait.

Ami a másik két mintaréteg tesztidejének eltérését illeti, a szignifikancia hiánya nem zárja ki, hogy a ChatGPT használatának lehetősége csökkentette a kísérleti csoport hallgatóinak átlagos idejét, azonban hasonló összetételű sokaság esetén nem lehet kijelenteni, hogy a megfigyelt eltérések feltétlen megmaradnának, tehát a kommunikációs és a jogi tesztnél nem általánosítható a sokaságra nézve a tesztidők csökkenése.

V.7.4. Nem paraméteres próbák a háttértudás nélküli hallgatók eredményeinek eltérésére

Az informatikai rétegbe sorolt nem informatikai tanulmányokat végző hallgatók esetén a kísérleti és kontroll csoport között mért különbség szignifikanciájának meghatározásához nem lehetett t-próbát használni. Sem a teszt eredmény, sem a tesztidő változó nem követett normál

eloszlást a kísérleti csoport 10 darab hallgatója esetén. Ezért egy nem paraméteres próba, a Mann-Whitney U teszt használata mellett döntöttünk, mivel az adott próba nem feltételezi a minták normális eloszlását, és kis minta esetén is használható (Mann & Whitney, 1947). Azt a sejtésünket, hogy kísérleti csoport hallgatói a kontroll csoporténál szignifikánsan jobb tesztet írtak a következő alternatív hipotézissel teszteljük:

$$F(u) < G(U)$$

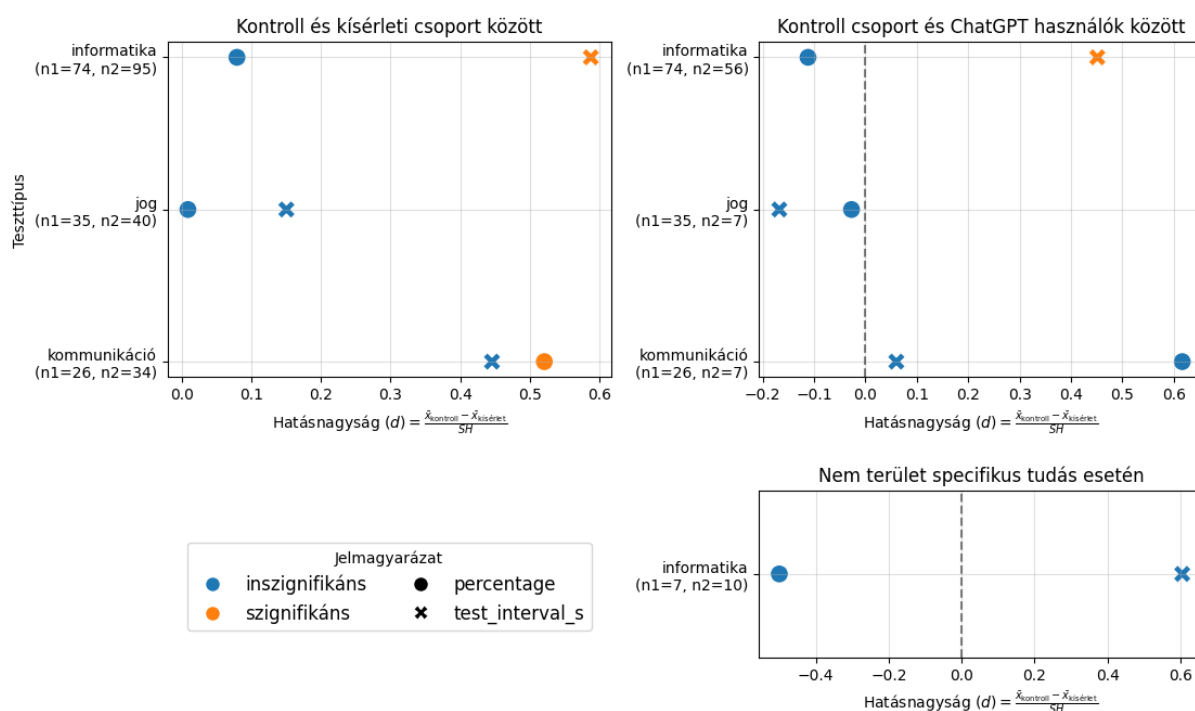
ahol az F a kísérleti csoport, a G pedig a kontroll csoport empirikus eloszlásfüggvénye. Az U teszt esetén tehát a H_1 azt feltételezi, hogy a kísérleti csoport teszteredményei sztochasztikusan nagyobbak a kontroll csoporténál. Azonban a próba p-értéke 0,2087 lett, ami alapján a kísérleti csoport teljesítményének javulását nem tekinthetjük szignifikánsnak. A tesztidők eltérésének vizsgálatánál is hasonló eredményre jutottunk. Azt az alternatív hipotézist, hogy a kísérleti csoport tagjainak tesztideje sztochasztikusan kisebb, mint a kontroll csoporté 11,47%-os p-érték mellett találtuk nem szignifikánsnak. Összefoglalva tehát, ugyan feltételezhető, hogy a mintában a nem terület specifikus tudással rendelkező hallgatók eredményességét és hatékonyságát képes volt növelni a ChatGPT használatának lehetősége, ugyanakkor nem zárható ki, hogy az eltérést pusztán a véletlen mintavétel okozta. Fontos hozzátenni azonban, hogy a próba keretében vizsgált mindkét minta kis méretű volt, nagyobb mintaméret esetén az U teszt is pontosabb eredményeket ad, így lehetséges, hogy az értékek eloszlása között mért különbségeket szignifikánsnak ítélte volna (Thiese és mtsai., 2016).

V.8. Hatásnagyság és szignifikancia

Az előző alfejezetben a kísérleti és kontroll csoport közötti szignifikáns eltérések kerültek feltárára, azonban a beavatkozás hatásnagyságának elemzése még hátravan. A szignifikancia vizsgálata ugyan elengedhetetlen a differenciák elemzése szempontjából, de számos korláttal rendelkezik. Egyrészt a hipotézis vizsgálat nem mond semmit a beavatkozás hatásának méretéről, pusztán azt vizsgálja, hogy a tapasztalt eltérés mekkora valószínűséggel a véletlen mintavétel eredménye. A hatásnagyság ezzel szemben az eltérések nagyságát vizsgálja, hogy a beavatkozás mekkora mértékben változtatja meg a kísérleti csoport várható értékét a kontroll csoportéhoz képest, azaz, hogy a kezelés és az eredmények között milyen szoros kapcsolat van (G. M. Sullivan & Feinn, 2012). Másrészt a statisztikai próbák kis mintaméret mellett könnyedén ítélnék nagy hatásnagyságú eltéréseket inszignifikánsnak, és gyakran hoznak ki nagy minta méret esetén is szignifikánsnak kihatásnagyságú különbségeket (Biddix, é. n.). Így a szignifikancia és hatásnagyság együttes értelmezése segít a változások pontosabb

megértésében, valamint lehetővé teszi a kísérlet eredmények összehasonlíthatóságát (McLeod, 2023).

Az alábbi ábra (lásd 21. ábra) a teszteredmény és tesztidő változó esetén mért sztenderdizált hatásnagyságot (Cohen-féle d együttható értékét) és szignifikanciát szemlélteti az egyes mintarétegek esetén, három különböző csoportosítás mellett. Minden egyes tesztípusnál a réteg neve alatt fel van tüntetve a kontroll csoport (n_1) és a kísérleti csoport (n_2) mérete.



21. ábra Hatásnagyság és szignifikancia ábrázolása (saját szerkesztés)

A kontroll és kísérleti csoport között megfigyelt eltérések alapján (bal felső tengely) a következő mondható el. Az informatika réteg esetén a ChatGPT használatának lehetősége nagy mintaméret mellett kis hatásnagyságú és insignifikáns változást eredményezett a hallgatók várható eredményében. Emellett a kísérleti csoport tagjai közel nagy hatásnagyság mellett tudtak szignifikánsan gyorsabban a teszttel végezni. Tehát a hatás méretének vizsgálata megmutatta, hogy az informatikai teszt esetén a tesztidő és a kezelés hatása között viszonylag szoros kapcsolat van. Ez alapján a mintában megfigyelt eltérések alátámasztják a kutatási hipotézist az adott tudományterület esetén.

A jogi réteg esetén nem csak az eltérések voltak insignifikánsak nagy minta méret mellett, de a beavatkozás hatása is igen gyenge volt mindkét függő változó esetén. A csoport képzés eredményekre gyakorolt hatása elhanyagolható, a tesztidőket tekintve pedig gyenge kapcsolat

fedezhető fel a hatékonyság és ChatGPT használat között. Ezek szerint a jogi réteg esetén nem feltétlenül állna meg a kutatás hipotézise hasonló összetételű sokaság esetén.

A kommunikációs tesztnél a kísérleti és kontroll csoport alkotás közepes erősségű hatást gyakorolt a függő változók értékére. Ugyanakkor az előző alfejezetben tárgyalt mintavételi hiba feltételezhetően nagy mértékben magyarázza a feltárt eltéréseket. Így csak a teszteredmény különbségét ítéltük szignifikánsnak, tekintettel az alacsonyabb p-értékre, illetve a ChatGPT tényleges alkalmazói esetén mért magas hatásnagyságra. A teszteredmények szignifikáns romlását a kritikus felhasználói magatartás hiánya és az MI pontatlan válaszai okozhatták.

Ami a kontroll csoport és a tényleges ChatGPT használók között mért különbségek megfigyelését illeti (jobb felső tengely), fontos értelmezésbeli korlát, hogy itt a kísérleti csoport minőségileg szűrt egyedeinek eredményességét és hatékonyságát hasonlítjuk össze a normával. Lényeges tehát leszögezni, hogy a szegmentált hallgatók tulajdonságaikban nem feltétlenül reprezentatívak a sokaságra.

Az informatikai teszt esetén a ChatGPT használata inszignifikáns mértékben és alacsony hatásnagyság mellett növelte a hallgatók eredményességét, a hatékonyságra nézve viszont szignifikáns volt a beavatkozás, ahol az MI alkalmazása közepes szorosságban állt a tesztidő csökkenésével. A jogi és kommunikációs teszt esetén a szűrt kísérleti csoportok kismintának számítottak. Így a hipotézis vizsgálat még nagyobb hatásnagyság mellett sem találta a megfigyelt eltéréseket szignifikánsak.

A kutatás során azt is vizsgáltuk, hogy olyan hallgatók esetén, akik nem rendelkeznek terület specifikus tudással, milyen hatással volt a teljesítményükre a ChatGPT használatának lehetősége (jobb alsó tengely). A kis méretű mintán lefolytatott beavatkozás esetén, közepes méretű hatásnagyság mellett a kísérleti csoport eredményesebb és hatékonyabb lett a kontroll csoportnál. Ugyanakkor valószínűleg az alacsony mintaméret miatt az U teszt nem találta szignifikánsnak a tapasztalt eltéréseket.

V.9. Eltérések dekompozíciója

V.9.1. A dekompozíció módszertani háttere

Az eltérések szignifikanciájának és hatásnagyságának tárgyalása mellett lényegesnek tartottuk feltárni, hogy a megfigyelt különbségek mekkora részben magyarázhatóak a kísérleti és kontroll csoport osztályainak eltérő összetételével, és mekkora része az, ami az összetétel hatással már nem magyarázható. A teszt eredmény és tesztidő változók átlaga közötti

differentenciák szétbontásához az összevont (angolul pooled) Oaxaca-Blinder dekompozíció lett alkalmazva. A fenti eljárás a regresszió alapú dekompozíciós függvények családjába tartozik, amely két alapvető, a várható értékek különbségét befolyásoló hatást különböztet meg (Popli, 2022):

- 1) magyarázott hatás vagy adottság hatás (explained effect vagy endowment effect): a két csoport eltérő összetételéből eredő különbségek nagyságát méri;
- 2) nem magyarázott hatás vagy együttható hatás (unexplained effect vagy coefficient effect): az összetétel hatással nem magyarázható különbség a várható értékek között, ami a beavatkozás (jelen esetben a ChatGPT használatának lehetősége), a regresszióból kihagyott változók, valamint nem megfigyelt heterogenitás együttes eredménye.

Tekintettel arra, hogy minta összetételét leíró változók többsége kategorikus, valamint az egyes mintaosztályok nagysága többesetben 30-40 fő környékén mozog, igyekeztünk csak a legszükségesebb ismérveket használni a regressziós modellekben (*lásd 17. táblázat*). Elakartuk kerülni modellek túlillesztését azzal, hogy a kategorikus jellemzők kódolásával ne kapjunk több változót, mint ahány megfigyelés van (Zhao, 2017).

Változó neve	Változó leírása
percentage*	százalékos teszt eredmény változó *tesztidők eltérésénél magyarázó változóként is használtuk
test_interval_s*	másodpercben mért tesztidő *százalékos teszteredmény eltérésénél magyarázó változóként is használtuk
experience	megkezdett szemeszterek száma a felsőoktatásban
sex	hallgató neme
learning_outside_of_class	tanórán kívül tanulással töltött idő intervalluma egy héten
training_fk	képzési azonosító
language	a normára hivatkozott teszt nyelve (jogi és informatikai teszt esetén volt lehetőség magyar és angol nyelvű teszt kitöltésére)

17. táblázat A várható teszt eredmény és tesztidő eltérésének vizsgálatához használt magyarázó változók (saját szerkesztés)

A dekompozíció eredményeinek bemutatása előtt fontos megjegyezni a használt eljárás korlátait, ugyanis a regressziós modellekben használt magyarázó változók a függő változók szóródását viszonylag kis részben magyarázzák. Így az Oaxaca-Blinder dekompozíció által itt kimutatott rész- és összetételhatások pontatlan becslések a teszt eredmény és tesztidő változók eltérésére, hiszen az alkalmazott regressziós modellek gyenge magyarázó erővel rendelkeznek a két függő változó varianciájára nézve.

V.9.2. A dekompozíció eredményeinek értelmezése

Ami a teszteredmények közötti eltérések hatáskomponensinek nagyságát illeti (lásd 18. táblázat): a fenti magyarázó változók mellett az összetétel hatás abszolút értéke minden esetben kisebb lett, mint a nem magyarázott hatás mértéke. Amíg a vizsgált adottságok szempontjából a kísérleti csoport jogi osztályának tagjai minimálisan, de előnyösebb tulajdonságokkal rendelkeztek, addig az informatikai és kommunikációs teszt esetén a kontroll csoport résztvevői indultak jobb helyzetből. Az együtttható hatás minden esetben a kísérleti csoport várható teszt eredményét rontotta, de a jogi és informatikai réteg esetén a beavatkozás hatása nem volt szignifikáns az eredmények különbségére nézve. A hipotézis vizsgálat egyedül a kommunikációs teszt esetén hozott ki szignifikáns romlást az eredmények tekintetében. A dekompozíciós modell szerint a kb. 4,61%-pontos átlagos teljesítmény csökkenésből nagyjából 3,08%-pont történhetett a beavatkozás miatt.

Függő változó	Teszt típus	Eltérés	Magyarázott hatás	Nem magyarázott hatás
percentage	informatika	1,0318%-pont	0,3539%-pont	0,6778%-pont
	jog	0,1313%-pont	-0,7046%-pont	0,8359%-pont
	kommunikáció	4,6078%-pont	1,5318 %-pont	3,0760%-pont
test_interval_s	informatika	419 mp	76 mp	343 mp
	jog	152 mp	144 mp	8 mp
	kommunikáció	403 mp	308 mp	95 mp

18. táblázat Összevont Oaxaca-Blinder dekompozíció eredmény táblázata (saját szerkesztés)

A tesztidő változó dekompozíciója alapján (lásd 18. táblázat) az informatikai teszt típus esetén a kísérleti csoport hallgatói átlagosan 343 másodperccel (közel hat perccel) írták meg gyorsabban a normára hivatkozott tesztet a beavatkozásnak köszönhetően. A várható értékek közötti eltérések regressziós felbontása megmutatta, hogy az informatikai rétegben nem csak a teljes különbség, de a nem magyarázott hatás szempontjából is a legnagyobb a hallgatók gyorsulása a kísérlet hatására. Azon döntésünk helyességét is megerősítette a dekompozíció, hogy a kommunikációs rétegnél mért tesztidő különbség szignifikanciáját elutasítottuk. Habár a várható időben mért különbség megközelíti az informatikai tesztnél tapasztaltat, azonban a fenti magyarázó változók mellett pusztán 95 másodpercnyi átlagos gyorsulás az, amit nem tudunk megmagyarázni. A jogi rétegnél az eltérés egészében sem volt szignifikáns, a nem magyarázott különbség pedig a maga nyolc másodpercével a legkisebb a többi teszt típushoz képest.

VI. Limitációk

Tanulmányunk több korlátozással is szembesül. Az eredmények nem terjednek ki az összes felsőoktatási szakra vagy tudományterületre.

Emellett a beavatkozás hatása nagymértékben függ a hallgatók korábbi ChatGPT vagy Google használatától. Így a kísérlet megismétlése esetén eltérő eredményekre lehet számítani, mert a ChatGPT használatának elterjedtsége és a mesterséges intelligencia alkalmazásával kapcsolatos tapasztalat nagyban befolyásolják a kutatásban használt függő változók értékének alakulását.

Továbbá, az ExamEye bővítmény használata a webes tevékenység monitorozására - szigorú adatvédelmi elvek betartása ellenére - előidézheti a Hawthorne-effektust, amely megváltoztathatja a hallgatók hagyományos böngészési csatorna és ChatGPT használatát (Chen és mtsai., 2015).

Végül, tanulmányunk rövid időtartama nem tükrözheti a ChatGPT hosszú távú hatásait az oktatási eredményekre, mivel az oktatástechnológiai eszközök hatékonysága hosszabb időtávon változhat (Schneider és mtsai., 2016).

VII. Ajánlások

A ChatGPT 2022 novemberi megjelenése és sikere egyértelművé tette, hogy a mesterséges intelligencia hatalmas szerepet fog játszani az oktatásban. Sok diákot "digitális bennszülöttként" tartunk számon, az információs korszakban nőttek fel, szoros kapcsolatban a technológiával (Roberto López-Chila és mtsai., 2023). Sokuk naprakészen követik az új trendek kialakulását vagy technológia fejlődését, alapvető számukra, hogy felhasználják a mesterséges intelligenciát.

Mind a szakirodalmi elemzés, mind az interjúk, mind a tapasztalataink alátámasztják, hogy a mesterséges intelligenciát a hallgatók használni fogják, bármiféle tiltás ellenére is, hiszen a legtöbb diáknak az a célja, hogy minimalizálja a befektetett erőforrásait. A MI-től való tiltás sok egyént hidegen fog hagyni, valószínűleg csak egyre kreatívabb módszerekkel fogják kihasználni a technológiát, anélkül, hogy megismernék a lehetséges hátrányokat, vagy etikai korlátokat. A találgató hallgatók számos módon manipulálhatják az MI által generált szöveget, úgy, hogy még egy humán oktató számára is megtévesztő legyen az eredmény. Legyen itt szó stilisztikai módosításokról, akár hibás adatok, vagy helyesírási hiba beszúrásáról, de olyan példáról is hallottunk, melyben egy hallgató szándékosan zavaró elemeket illesztett be munkájába, példának okáért a szövegben előforduló l (L betű kisbetűs formában) karaktereket

kicserélte I (I betű nagybetűs formában) karakterekre, ez megfelelő betűtípusok mellett optikailag szinte felismerhetetlen. Ezen a módosítások nélkülözésével is a mesterséges intelligenciára detektálására alkalmas algoritmusok pontatlanul működnek.

Egy oktatási intézmény feladata kíváncsi, változásra fogékony hallgatók nevelése. Az információs korszakban számos gyors ütemű változás következett be, melyhez a szakembereknek szükséges progresszíven alkalmazkodniuk. A megszerzett lexikális tudás nagyon hamar elavulhat, a diákoknak realizálniuk kell, hogy boldogulásukhoz lehet nem lesz elegendő a tanulmányaik során megszerzett tudásanyag. A mesterséges intelligencia teljes eltávolítása nem fogja megszüntetni a jelenséget, az iskolapadon túl találkozhatnak a hallgatók a mesterséges intelligenciával, és elengedhetetlen, hogy felkészülten nézzenek szembe az új kihívásokkal (Gill & Kaur, 2023; Thomas K. F. Chiu, 2023; Weng és mtsai., 2022).

Az oktatásba való integrálás magával kell, hogy vonja számos folyamat, tanterv és tananyag megváltoztatását. Kiemelendő, hogy a mesterséges intelligenciát úgy kell beiktatni a hallgatók tanulmányi életébe, hogy az a kritikus gondolkodásuk fejlesztését szolgálja. A MI nagy csapdája, hogy a könnyű használhatósága miatt helyettesítheti a tanulást és az önálló gyakorlati tapasztalatok megszerzését. Hangsúlyozni kell, hogy az MI bevezetésével a diákokat arra kell rávezetni, hogy a ChatGPT nem egy mentorként szolgál, nem szabad elfogadni a chatbot által generált szöveget bármi nemű kritika nélkül. A mesterséges intelligencia is egy olyan kollaborációs eszköz, mint a számítógép, és a hallgatóknak hasonlóképpen ki kell ismerniük ezt az eszközt (Gill és mtsai., 2024; M. Sullivan és mtsai., 2023).

Az oktatóknak rendelkezniük kell azokkal a tudással és készségekkel, amelyek elvezetik diákjaikat a GAI-eszközök produktív és kritikus használatához (Javier & Moorhouse, 2023a). Ez megköveteli a tanároktól, hogy nyílt és őszinte beszélgetéseket folytassanak a diákokkal a GAI használatáról, valamint a megfelelő használatot támogató irányelvekről és kockázatokról (Gelman, 2023; Moorhouse, 2024). Fontos, hogy a ChatGPT bemutatása során hangsúlyozni kell a chatbot hallucinációra való hajlamát, a használat során a hallgatóknak érteniük kell, hogy a ChatGPT egy kisegítő eszköz és nem alkalmas képességeik helyettesítésére

A teszteredmények és nyomonkövetés során azt tapasztaltuk, hogy a jog és médiahallgatók a teszt kitöltése során kevésbé fordultak a ChatGPT-hez. Kifejezetten az informatikusok használták fel, náluk nőtt az effektivitás. Az alacsony használat alapján arra a következtetésre juthatunk, hogy feltehetőleg számos hallgató nem teljesen van tisztában a ChatGPT használati lehetőségeivel. A hallgatóknak érdemes megmutatni, hogyan lehet jól, és etikusan felhasználni

a ChatGPT-t, ismerniük kell a technológia korlátait. Az oktatásba olyan jellegű feladatokat kell beintegrálni, melyek nyitottak a mesterséges intelligencia létezésére, legyen itt szó akár prompt engineeringről, vagy olyan komplex feladatokról, melyekre a ChatGPT nem lesz képes jól válaszolni. Ezen felül az esszék vagy teszt jellegű feladatok helyett érdemes lenne nagyobb súlyt helyeznie az oktatóknak az osztályon belüli tevékenységek és gyakorlati műtermékek értékelésére (Javier & Moorhouse, 2023b). A cél nem a technológia eltiltása, hanem annak megértése és jó használata.

Egyik kiváló módja az oktatással kapcsolatos folyamatok automatizálásnak egy tantárgyhoz szabott, tantárgyi ismeretekkel betanított chatbot kialakítása. A különféle folyamatok automatizálására és éjjel-nappali ügyfélszolgálatra való képessége miatt a chatbotok egyre népszerűbbek. A pozitívumok mellett ezek az eszközök hajlandóságot mutatnak számos működési problémára és fenyegetésre, beleértve a felhasználói szándékok megértésének kudarcát és az adatbiztonságot (Adamopoulou & Moussiades, 2020).

Egy célirányosan betanított chatbot-ban rejlő legnagyobb kockázat a hallucinációra való hajlam. A kutatás lebonyolítása során a csapatnak lehetősége volt szemügyre venni és segíteni egy ChatGPT 3.5 alapú chatbot betanításában és elhelyezésében egy cég honlapján. A chatbot betanítása a Chatbase nevű fejlesztő platformon ment végbe.

Nagy meglepetést jelentett, hogy mennyire hajlamos a chatbot a túlzottan kreatív válaszadásra. A tanítás során aprólékosan lett megadva az összes információ és oldal, ami alapján a chatbot válaszokat generálhat, ezen kívül beállításra került, hogy a chatbot ne legyen hajlamos kreatív válaszok megadására. A tesztelés fázisa alatt kiderült, hogy minden beállítás és instrukció ellenére a chatbot képes volt nem létező, hamis linkeket generálni, félrevezető információkat adott meg, ezen kívül könnyen megtévesztésre kerülhetett a felhasználó által. A problémák kiküszöbölése érdekében gyakran ismételt, lényegretörő promptok kerültek megadásra, mint instrukció. A promptok gyakori, minden alkalommal egyedi módon történő megadása szükséges volt, hogy semmiképp se kerüljön generálásra egy túlzottan kreatív válasz.

Ahogy egy ügyfél szempontjából, úgy egy hallgató szempontjából is számos kellemetlenséggel járhat a rosszul informáltság. A chatbot által szolgáltatott hibás információ torzíthatja a diák tudását és félrevezetheti tanulási folyamatát. Ezáltal a diák helytelen vagy hiányos tudással rendelkezhet a tanult témában, ami hosszú távon komoly problémákat okozhat az oktatási fejlődésben és a szakmai ismeretekben. A diákra való személyes hatás mellett egy rosszul

betanított chatbot által fenyegetésre kerül az oktatás és tanulás integritása, torzulhat az oktatók és az oktatási intézmény hitelessége.

Egy jól betanított, és alaposan tesztelt chatbot számos módon felhasználható, személyre szabottá alakíthatja az tanulást. Egy hallgató számára hatalmas segítséget jelenthet egy állandóan elérhető felület, mely válaszokat ad szakmai kérdéseire. A válaszadás mellett a chatbot motiváló tényező is lehet egy diák életében, egy kutatás azt vizsgálja, hogyan képes a mesterséges intelligencia ösztönözni a diákokat ismereteik bővítésére (Xia és mtsai., 2022).

Az ember-robot interakciók végett a gyengén teljesítő tanulók nagyobb biztonságban érezték magukat a kérdések feltevése során, nem érezték magukat feszélyezetten kérdéseik jellege miatt. A testreszabott chatbotok képesek jelentős növekedést eredményezni a tanulók eredményeiben, ezen kívül a speciális igényű hallgatók tanulásában is segítséget nyújthat (Garg & Sharma, 2020).

A hallgatók megsegítése mellett az oktatók munkáját is megkönnyítheti a mesterséges intelligencia. Egy érdekes felhasználási mód az automatikus pontozási rendszer kiépítése, egy nyelvtanárokkal lefolytatott tanulmány bizonyítja, hogy ezen eszközök támogatást nyújthatnak az oktatók munkájában (Mizumoto & Eguchi, 2023). A 12100 tanulói esszét magában foglaló tanulmányban a ChatGPT segítségével hatékonyan és gyorsan lehetett pontozni az esszéket a nyelvi pontosság tekintetében. Az ilyen felhasználások potenciálisan időt takaríthatnak meg a tanárok számára, és gyorsabban biztosíthatják a diákoknak következetes visszajelzést (Gill és mtsai., 2024).

Persze nem minden esetben pontoz megfelelően a chatbot. Egy hasonló kutatásban szintúgy szöveget kellett felülvizsgálnia a ChatGPT-nek. Az eszköz képes volt jó összefoglalást készíteni a szövegről, világosan leírva a dolgozat fő célját és következtetéseit. A ChatGPT tett néhány pozitív szubjektív megjegyzést; például: „Az írásmód világos és tömör. Az LLM néhány építő jellegű kritikát is megfogalmazott: „a cikk egyes részei technikai jellegűek, és a nem szakértők számára nehéz lehet megérteni.” Azonban amint arra kérték, hogy javasoljon konkrétabb fejlesztéseket, kudarcot vallott, és elindult a chatbot hallucinációja, amely során az LLM magabiztos hangzású válaszokat adott, amely hamisak vagy megalapozatlanok voltak. Például említést tett, hogy az iromány statisztikai elemzése nem megfelelő, és arról számolt be, hogy a tanulmány „egyszerű t-próbákat és khi-négyzet próbákat” használt, miközben egyik módszer sem került alkalmazásra (Donker, 2023).

A mesterséges intelligencia alkalmazható számos téren a felsőoktatásban. Felhasználható oktatási anyagok létrehozására vagy átdolgozására, kódolás fejlesztésére, tanulást segítő chatbotokat lehet létrehozni a hallgatói kérdések megválaszolásához, elősegítheti az online tanulást (Chiu, 2024). A lényeg, hogy észben tartsuk, hogy a hallgatók a legfontosabb érintettek.

VIII. Konklúzió

A tanulmány célja, hogy feltérképezze a ChatGPT tanulmányi célú használatának hatásait a felsőoktatásban tanuló hallgatók eredményességére és hatékonyságára nézve. A kutatás során végzett tesztek azt mutatták, hogy az informatikai tanulmányok területén a kísérleti csoport szignifikánsan rövidebb idő alatt teljesítette a teszteket az átlagos eredmények változatlansága mellett, ami megerősítette a kutatási hipotézist. A jogi tanulmányokat végző hallgatók esetén a ChatGPT alkalmazásának nem mutatkozott jelentős hatása a hallgatók teljesítményével kapcsolatban: a kísérleti csoport ideje minimálisan csökkent, miközben az eredmények változatlanok maradtak. A kommunikációtudomány területén lefolytatott teszt a ChatGPT használatának lehetősége esetén szignifikáns eredményromlást állapított meg, a tesztidő inszignifikáns csökkenése mellett. Mindezek alapján arra a következtetésre jutottunk, hogy a ChatGPT a hatékonyság oldalán gyakorolhat pozitív hatást, azonban nem feltétlen minden tudományterületen. Az eredményesség tekintetében úgy tapasztaltuk, hogy kevésbé van hatással a hallgatók teljesítményére a mesterséges intelligencia. Mi több, a hallgatók eredményességére nézve káros lehet, ha nem alkalmazzák kritikusan a ChatGPT, mint ahogy a kommunikációs tesztnél is történt. Ugyan a mérések azt sugallják, hogy a mesterséges intelligencia használata rövidtávon képes lehet javítani a kevesebb tudással, vagy rosszabb képességekkel rendelkező hallgatók eredményeit, azonban az erre vonatkozó megfigyelések nem szignifikánsak.

Hivatkozásjegyzék

- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006. <https://doi.org/10.1016/j.mlwa.2020.100006>
- Akinyemi, O. T. (2023, augusztus 1). *The Power of Primer Prompts: Crafting Personalized Conversations with ChatGPT*. DEV.to. <https://dev.to/ota/the-power-of-primer-prompts-crafting-personalized-conversations-with-chatgpt-4hlm>
- Alcanum. (é. n.). *Beszédrészek*.
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. <https://doi.org/10.7759/cureus.35179>
- Altmäe, S., Sola-Leyva, A., & Salumets, A. (2023a). Artificial intelligence in scientific writing: a friend or a foe? *Reproductive BioMedicine Online*, 47(1), 3–9. <https://doi.org/10.1016/j.rbmo.2023.04.009>
- Alvero, R. (2023). ChatGPT: rumors of human providers' demise have been greatly exaggerated. *Fertility and Sterility*, 119(6), 930–931. <https://doi.org/10.1016/j.fertnstert.2023.03.010>
- Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2024). *Transforming Conversations with AI—A Comprehensive Study of ChatGPT*.
- Biddix, J. P. (é. n.). *Effect Size*. Research Rundowns. Elérés 2024. április 5., forrás <https://researchrundowns.com/quantitative-methods/effect-size/>
- Casad, B. J., & Luebering, J. E. (2024). *confirmation bias*. Encyclopedia Britannica. <https://www.britannica.com/science/confirmation-bias>
- Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — A Case Study. *Cogent Education*, 10(1). <https://doi.org/10.1080/2331186X.2023.2210461>
- Chen, L. F., Vander Weg, M. W., Hofmann, D. A., & Reisinger, H. S. (2015). The Hawthorne Effect in Infection Prevention and Epidemiology. *Infection Control & Hospital Epidemiology*, 36(12), 1444–1450. <https://doi.org/10.1017/ice.2015.216>

- Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*, 6, 100197. <https://doi.org/10.1016/j.caeai.2023.100197>
- Chris Stokel-Walker. (2022). *AI bot ChatGPT writes smart essays — should professors worry?*
- Cingilliglu, I. (2023). *Detecting AI-generated essays: the ChatGPT challenge.*
- Content scripts.* (é. n.). Chrome for Developers. Elérés 2023. október 30., forrás <https://developer.chrome.com/docs/extensions/develop/concepts/content-scripts>
- Craig Hale. (2023). *ChatGPT is having a really bad impact on the environment.*
- Daniel Nelson. (2024). *Mi az NLP (természetes nyelvi feldolgozás)?*
- Davide Castelvechi. (2022). *Are ChatGPT and AlphaCode going to replace programmers?*
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615–622. <https://doi.org/10.5114/biol sport.2023.125623>
- Document: hasFocus() method.* (é. n.). MDN Web Docs. Elérés 2023. november 1., forrás <https://developer.mozilla.org/en-US/docs/Web/API/Document/hasFocus>
- Donker, T. (2023). The dangers of using large language models for peer review. *The Lancet Infectious Diseases*, 23(7), 781. [https://doi.org/10.1016/S1473-3099\(23\)00290-6](https://doi.org/10.1016/S1473-3099(23)00290-6)
- Dylan Love. (2014). *No One's Talking About The Amazing Chatbot That Passed The Turing Test 3 Years Ago.*
- Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, 13, 100060. <https://doi.org/10.1016/j.jrt.2023.100060>
- Elmore, M. (2023). The Hidden Costs of ChatGPT: A Call for Greater Transparency. *The American Journal of Bioethics*, 23(10), 47–49. <https://doi.org/10.1080/15265161.2023.2250335>

- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Frost, J. (é. n.). *Independent and Dependent Samples in Statistics*. Statistics By Jim. Elérés 2023. szeptember 13., forrás <https://statisticsbyjim.com/basics/independent-dependent-samples/>
- Galassi, U., Giordana, A., & Saitta, L. (1998). *Incremental Construction of Structured Hidden Markov Models*.
- Garg, S., & Sharma, S. (2020). Impact of Artificial Intelligence in Special Need Education to Promote Inclusive Pedagogy. *International Journal of Information and Education Technology*, 10(7), 523–527. <https://doi.org/10.18178/ijiet.2020.10.7.1418>
- Gelman, S. (2023). *DC region schools ban AI tool ChatGPT*.
- Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber-Physical Systems*, 3, 262–271. <https://doi.org/10.1016/j.iotcps.2023.05.004>
- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., ... Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19–23. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- Greengard, S. (2022). Can AI learn to forget? *Communications of the ACM*, 65(4), 9–11. <https://doi.org/10.1145/3516514>
- Hales, A. H. (2023). One-tailed tests: Let's do this (responsibly). *Psychological Methods*. <https://doi.org/10.1037/met0000610>
- Hermans, M., & Shrauwen, B. (2013). *Training and analyzing deep recurrent neural networks*.
- Hunyadi, L., & Vita, L. (2019a). *Statisztika I. Akadémiai Kiadó*. <https://doi.org/10.1556/9789634542216>

- Hunyadi, L., & Vita, L. (2019b). *Statisztika II*. Akadémiai Kiadó.
<https://doi.org/10.1556/9789634542223>
- Hutson, M. (2021). Robo-writers: the rise and risks of language-generating AI. *Nature*, 591(7848), 22–25. <https://doi.org/10.1038/d41586-021-00530-0>
- Jae C. Hong. (2023, március 23). *We were told we'd be riding in self-driving cars by now. What happened to the promised revolution?*
- Javier, D. R. C., & Moorhouse, B. L. (2023a). Developing secondary school English language learners' productive and critical use of <scp>ChatGPT</scp>. *TESOL Journal*. <https://doi.org/10.1002/tesj.755>
- Javier, D. R. C., & Moorhouse, B. L. (2023b). Developing secondary school English language learners' productive and critical use of <scp>ChatGPT</scp>. *TESOL Journal*. <https://doi.org/10.1002/tesj.755>
- Jonathan Bailey. (2020). *How teachers can prepare for AI-based writing*.
- Kelsey Piper. (2022). *ChatGPT has given everyone a glimpse at AI's astounding progress*.
- Kiszolgáló nélküli számítástechnika*. (é. n.). Cloud Computing Dictionary. Elérés 2023. november 15., forrás <https://azure.microsoft.com/hu-hu/resources/cloud-computing-dictionary/what-is-serverless-computing>
- Kumar, K., & Thakur, G. S. M. (2012). Advanced Applications of Neural Networks and Artificial Intelligence: A Review. *International Journal of Information Technology and Computer Science*, 4(6), 57–68. <https://doi.org/10.5815/ijitcs.2012.06.08>
- Kupiec, J. (1992). Robust part-of-speech tagging using a hidden Markov model. *Computer Speech & Language*, 6(3), 225–242. [https://doi.org/10.1016/0885-2308\(92\)90019-Z](https://doi.org/10.1016/0885-2308(92)90019-Z)
- Lambright-Black, M., Shepard, D., & Xiomara, M.-W. (é. n.). *Ram, Tesla and Subaru Have the Worst Drivers, While BMW Drivers Have the Highest DUI Rates*. 2023.
- Liu, L. (2023). Analyzing the Text Contents Produced by ChatGPT: Prompts, Feature-Components in Responses, and a Predictive Model. *Journal of Educational Technology Development and Exchange*, 16(1), 49–70. <https://doi.org/10.18785/jetde.1601.03>

- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
- McFarland, A. (2023, december 26). *Beginner's Guide to AI Image Generators*.
- McKeown, M. G. (1985). The Acquisition of Word Meaning from Context by Children of High and Low Ability. *Reading Research Quarterly*, 20(4), 482. <https://doi.org/10.2307/747855>
- McLeod, S. (2023, július 31). *What Does Effect Size Tell You?* Simply Psychology. <https://www.simplypsychology.org/effect-size.html>
- Melissa Heikkiläarchive page. (2022, december 19). *How to spot AI-generated text*.
- Miyazaki, K., Murayama, T., Uchiba, T., An, J., & Kwak, H. (2024). Public perception of generative AI on Twitter: an empirical study based on occupation and usage. *EPJ Data Science*, 13(1), 2. <https://doi.org/10.1140/epjds/s13688-023-00445-y>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Moorhouse, B. L. (2024). Beginning and first-year language teachers' readiness for the generative AI age. *Computers and Education: Artificial Intelligence*, 6, 100201. <https://doi.org/10.1016/j.caeai.2024.100201>
- Popli, G. (2022). Decompositions: Accounting for Discrimination. *Sheffield Economic Research Paper Series*.
- Raffo, D. (2024, február 27). *Gemini vs. ChatGPT: What's the difference?*
- Roberto López-Chila, Joe Llerena-Izquierdo, Nicolás Sumba-Nacipucha, & Jorge Cueva-Estrada. (2023). *Artificial Intelligence in Higher Education: An Analysis of Existing Bibliometrics*.
- Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*. Simon and Schuster.
- Roose Kevin. (2022). *The Brilliance and Weirdness of ChatGPT*.

- Schneider, B., Saw, G., & Broda, M. (2016). A Future for the National Education Longitudinal Program. *AERA Open*, 2(2), 233285841663468. <https://doi.org/10.1177/2332858416634686>
- Sebastian Thrun. (2010, október 9). *What we're driving at*.
- Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74, 102700. <https://doi.org/10.1016/j.ijinfomgt.2023.102700>
- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature*, 613(7945), 620–621. <https://doi.org/10.1038/d41586-023-00107-z>
- Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). *ChatGPT in higher education: Considerations for academic integrity and student learning*.
- Szreder, M. (2022). Opportunities and illusions of using large samples in statistical inference. *Wiadomości Statystyczne. The Polish Statistician*, 67(8), 1–16. <https://doi.org/10.5604/01.3001.0015.9704>
- Szuhi, A. (2022, november 3). *Hány karakter, hány szó és hány oldal?* SEO Tudásközpont és Hírportál. <https://ite.hu/hany-karakter-hany-szo-es-hany-oldal/>
- Taylor, S. (é. n.). *Skewness*. Corporate Finance Institute. Elérés 2024. január 20., forrás <https://corporatefinanceinstitute.com/resources/data-science/skewness/>
- Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease*, 8(9), E928–E931. <https://doi.org/10.21037/jtd.2016.08.16>
- Thomas K. F. Chiu. (2023). *The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney*.
- Tom Carter. (2023, október 25). *Here's what reportedly led to Cruise robotaxis getting banned in San Francisco*.

- van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Welch, B. L. (1947). THE GENERALIZATION OF ‘STUDENT’S’ PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika*, 34(1–2), 28–35. <https://doi.org/10.1093/biomet/34.1-2.28>
- Weng, X., Chiu, T. K. F., & Tsang, C. C. (2022). Promoting student creativity and entrepreneurship through real-world problem-based maker education. *Thinking Skills and Creativity*, 45, 101046. <https://doi.org/10.1016/j.tsc.2022.101046>
- Winograd, S. (1970). On the number of multiplications necessary to compute certain functions. *Communications on Pure and Applied Mathematics*, 23(2), 165–179. <https://doi.org/10.1002/cpa.3160230204>
- Xia, Q., Chiu, T. K. F., Lee, M., Sanusi, I. T., Dai, Y., & Chai, C. S. (2022). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education*, 189, 104582. <https://doi.org/10.1016/j.compedu.2022.104582>
- Zhang, C., & Zhang, C. (2023). *One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era*.
- Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research*, 1(6), 831–833. <https://doi.org/10.1016/j.fmre.2021.11.011>
- Zhao, J. (2017, november 9). *More features than data points in linear regression?* Medium. <https://medium.com/@jennifer.zzz/more-features-than-data-points-in-linear-regression-5bcabba6883e>

Függelék

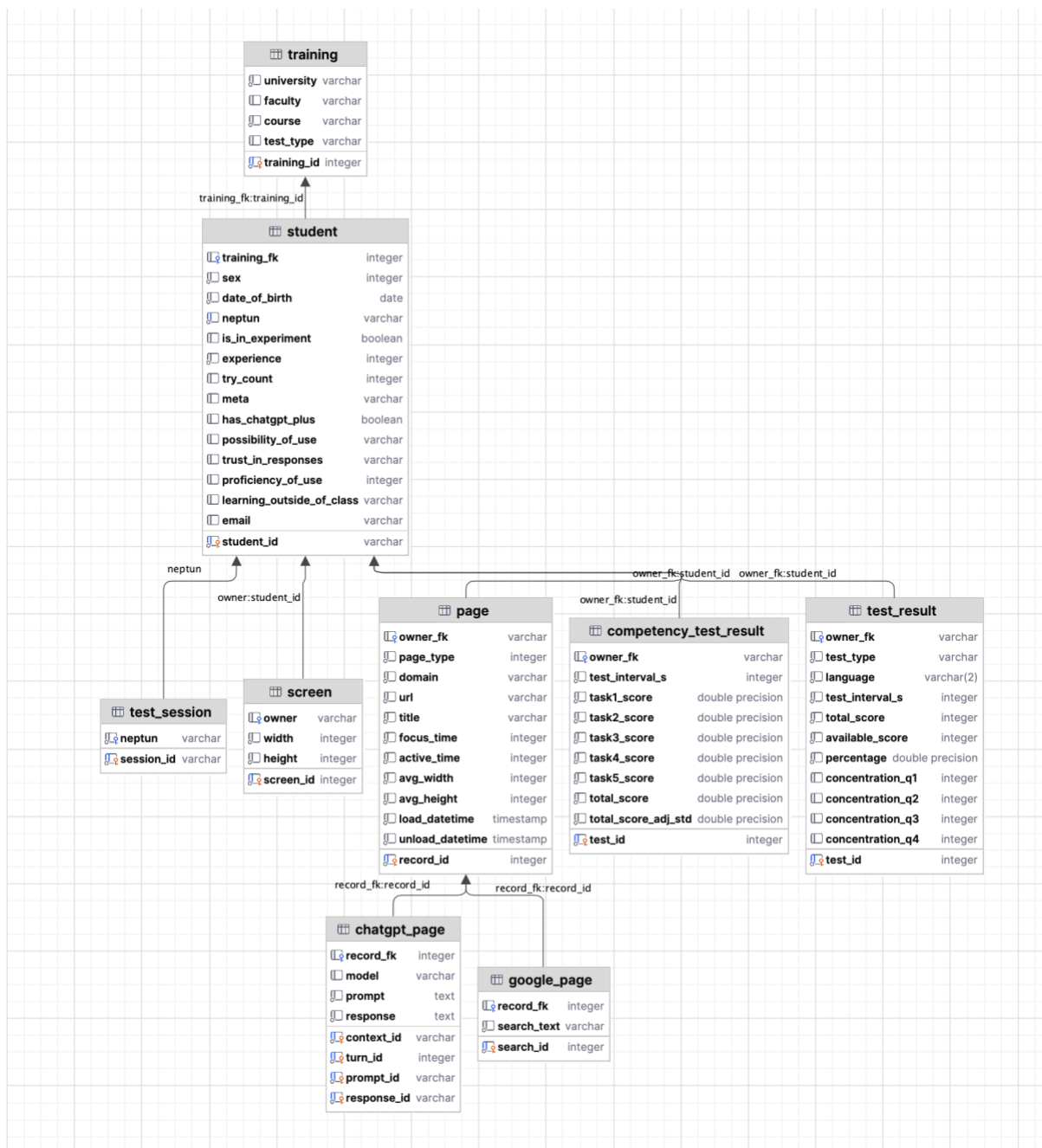
Függelék A: Csoportba osztó algoritmus python forráskódja

```
1  ✓ from pandas import DataFrame
2  import numpy as np
3
4  df = DataFrame(...)
5
6  ✓ def mean_optimizer(group_a: DataFrame, group_b: DataFrame) -> tuple[DataFrame, DataFrame]:
7      df_a = group_a.copy()
8      df_b = group_b.copy()
9
10     length = min(len(df_a), len(df_b))
11
12     ✓ for i in range(length):
13         delta_mean1 = np.abs(df_a["total_score_adj_std"].mean() - df_b["total_score_adj_std"].mean())
14
15         temp = df_a.iloc[i].copy()
16         df_a.iloc[i] = df_b.iloc[i]
17         df_b.iloc[i] = temp
18
19         delta_mean2 = np.abs(df_a["total_score_adj_std"].mean() - df_b["total_score_adj_std"].mean())
20
21     ✓ if delta_mean2 < delta_mean1:
22         continue
23     ✓ else:
24         temp = df_a.iloc[i].copy()
25         df_a.iloc[i] = df_b.iloc[i]
26         df_b.iloc[i] = temp
27
28     return df_a, df_b
29
30
31  ✓ def std_optimizer(group_a: DataFrame, group_b: DataFrame) -> tuple[DataFrame, DataFrame]:
32      df_a = group_a.copy()
33      df_b = group_b.copy()
34
35      df_a = df_a.sort_values(by=["total_score_adj_std"])
36      df_b = df_b.sort_values(by=["total_score_adj_std"])
37
38      length = min(len(df_a), len(df_b))
39
40     ✓ for i in range(length):
41         delta_std1 = np.abs(df_a["total_score_adj_std"].std() - df_b["total_score_adj_std"].std())
42         delta_mean1 = np.abs(df_a["total_score_adj_std"].mean() - df_b["total_score_adj_std"].mean())
43
44         temp = df_a.iloc[i].copy()
45         df_a.iloc[i] = df_b.iloc[i]
46         df_b.iloc[i] = temp
47
48         delta_std2 = np.abs(df_a["total_score_adj_std"].std() - df_b["total_score_adj_std"].std())
49         delta_mean2 = np.abs(df_a["total_score_adj_std"].mean() - df_b["total_score_adj_std"].mean())
50
51     ✓ if delta_std2 < delta_std1 and delta_mean2 <= delta_mean1:
52         continue
53     ✓ else:
54         temp = df_a.iloc[i].copy()
55         df_a.iloc[i] = df_b.iloc[i]
56         df_b.iloc[i] = temp
57
58     return df_a, df_b
59
60
61  df = df.sort_values(by="total_score_adj_std", ascending=False)
62  initial_group_a = df.iloc[:,2,:]
63  initial_group_b = df.iloc[:,1,:]
64  group_a_mean_optimized, group_b_mean_optimized = mean_optimizer(initial_group_a, initial_group_b)
65  experimental_group, control_group = std_optimizer(group_a_mean_optimized, group_b_mean_optimized)
```

Függelék B: Az ExamEye API dokumentációjának linkje

Az ExamEye API 0.1.0 számú verziója az OpenAPI 3.1-es specifikációja alapján az alábbi címen tekinthető meg: [\[kattintson a hivatkozásra\]](#)

Függelék C: A kutatási adatbázis sémáját leíró diagram



Függelék D: A dolgozathoz felhasznált mesterséges intelligencia alapú eszközök és azok felhasználási módja

Eszköz neve	Felhasználás módja
ChatGPT (GPT-3.5 és GPT-4)	- fordítás - nyelvhelyesség ellenőrzés és korrekció - kisebb kódrészletek generálása az ExamEye forráskódjához és a dolgozatban szereplő ábrákhoz
Github Copilot	- programozás közben használt kódkiegészítő bővítmény a Visual Studio Code alkalmazáson belül