

Nagy nyelvi modellek gyógyszeripari alkalmazhatóságának lehetőségei

The possible use of large language models in the pharmaceutical industry

DOI: [HTTPS://DOI.ORG/10.53793/RV.2024.1.6](https://doi.org/10.53793/RV.2024.1.6)

Absztrakt

A gyógyszerkutatás és -fejlesztés hosszú és költséges folyamat, amely számos kihívással jár. A készítmény forgalomba hozatalát követően a minőségellenőrzés és a mellékhatások monitorozásával biztosítják a gyógyszer biztonságosságát és hatásosságát. A gyógyszeripar szigorú szabályozási követelményeknek kell eleget tennie, ami nehezíti az innovációt és a hatékonyságot. A nagy nyelvi modellek egyre nagyobb teret hódítanak az életünkben, ami széles körű alkalmazhatóságuknak köszönhető. Ezek a modellek képesek nagy mennyiségű nyelvi információ feldolgozására és szöveggenerálásra. A mesterséges intelligencia alapú eszközök alkalmasnak bizonyultak hatalmas mennyiségű tudományos és orvosi információ elemzésére, ezáltal rendkívül sokoldalú segítséget nyújtanak a gyógyszerkutatás, -fejlesztés, -gyártás és minőségellenőrzés területein. Mindezek révén csökkenthetők a gyógyszerkutatás költségei, felgyorsítható az új készítmények piacra jutása, végsősoron a gyógyszeripar hatékonyabbá válhat.

KULCSSZAVAK: CHATGPT, FARMAKOVIGILANCIA, GYÓGYSZERFEJLESZTÉS, MINŐSÉGBIZTOSÍTÁS

Abstract

Drug research and development is a lengthy and costly process that comes with numerous challenges. Following the introduction of the product to the market, the safety and efficacy of the drug are ensured through quality control and by monitoring side effects. The pharmaceutical industry must comply with strict regulatory requirements, which hinder innovation and efficiency. Large language models have become an integral part of our daily lives due to their versatility. These models are capable of generating text and processing linguistic information. Artificial intelligence-based tools are suitable for analysing vast amounts of scientific and medical information, providing assistance in the fields of drug research, development, manufacturing, and quality control. By implementing these innovative tools, the costs of drug research can be reduced, the introduction of new products to the market can be accelerated, and the pharmaceutical industry can become more efficient.

KEYWORDS: CHATGPT, PHARMACOVIGILANCE, DRUG DEVELOPMENT, QUALITY CONTROL

Bevezetés

Az elmúlt években a mesterséges intelligencia (*artificial intelligence*, AI) területén látványos fejlődés tapasztalható, különösen a generatív mesterséges intelligencia területén. Ennek a fejlődésnek az egyik legkiemelkedőbb eredménye a nagy nyelvi modellek (*large language model*, LLM) robbanásszerű elterjedése és alkalmazása. Ezek az AI-alapú eszközök mély tanulási algoritmusokra épülnek, és különféle természetes nyelvi feldolgozási (*natural language processing*, NLP) feladatok

elvégzésére képesek. A legfejlettebb nyelvi modellek mély neurális hálózatokon alapulnak, például a rekurrens neurális hálózatokon (*recurrent neural network*, RNN) vagy a transzformátor architektúrákon (Kasneci et al. 2023; McCauley–Christiansen 2019).

A nagy nyelvi modellek olyan algoritmusokat használnak, amelyek a nyelvi struktúrák megértésére és előrejelzésére szolgálnak, például szóösszetételek,

szóhasználat és nyelvtani szerkezetek. A modell kimenete tehát egy becslés arra vonatkozóan, hogy mi lehet a következő szó. A modellben súlyoknak hívjuk azokat a paramétereket, amelyek meghatározzák a modell viselkedését és teljesítményét. A nyelvi modell kezdeti tanítása során a súlyok általában véletlenszerűek. Tanítás során a gépi tanulási modellt hatalmas mennyiségű szöveges adattal látják el, hogy megtanulja a nyelvet (Reynold–McDonell 2021).

A finomhangolás (*fine-tuning*) során a már előzetesen betanított nyelvi modell súlyait módosítjuk olyan módon, hogy további tanítást végzünk egy specifikus feladatra vonatkozóan. Ebben az esetben a súlyokat az új adatokra vagy feladatokra optimalizáljuk úgy, hogy a modell jobban teljesítsen az adott feladaton (Gu et al. 2022).

A nyelvi modellek alkalmazása során számos fontos szempontot kell figyelembe venni. A generált tartalmat érdemes rendszeresen ellenőrizni és validálni, hogy biztosak legyünk helyességében és megbízhatóságában. Sok más algoritmushoz hasonlóan a nagy nyelvi modellek is hajlamosak örökölni a betanítási adatokban jelenlévő torzításokat. Ez azt jelenti, hogy ha a tanító adathalmazban hibák fordulnak elő, akkor a modell is hajlamos lehet ezeket a torzításokat reprodukálni a generált szövegben. Az ilyen torzítások kezelése és csökkentése érdekében fontos a tanító adathalmazok gondos ellenőrzése (Ray 2023).

A nagy nyelvi modellek létrehozhatják az úgynevezett hallucináció jelenségét, amikor a modell olyan információt vagy tartalmat generál, ami nem felel meg a valóságnak vagy nem szerepelt az eredeti bemenetben. Ez a generált tartalom általában fiktív vagy téves információkat tartalmaz, és nem kapcsolódik szorosan a bemeneti adathoz. Hallucinációk általában akkor fordulnak elő, amikor a modell nem rendelkezik elegendő releváns információval vagy nem képes megfelelően értelmezni a kontextust. Ezért fontos, hogy a nyelvi modellek alkalmazásakor gondosan figyeljünk a generált tartalom hitelességére és pontosságára, különösen olyan területeken, ahol a pontosság kritikus szerepet játszik (Alkaissi–McFarlane 2023).

A ChatGPT nyelvi modell része a GPT (*Generative Pre-trained Transformer*) modellsorozatnak, amely az emberi nyelven történő interaktív párbeszéd létrehozására specializálódott. A ChatGPT esetében a visszajelzés alapú megerősítő tanulás (*Reinforcement Learning with Human Feedback*, RLHF) segít finomhangolni és javítani a modell teljesítményét, ami az alkalmazás hatékonysága szempontjából

kulcsfontosságú. A ChatGPT széles körben alkalmazható, ideértve a chatbotokat, kérdés-válasz rendszereket, tartalomgenerálást és fordítást (Ray 2023).

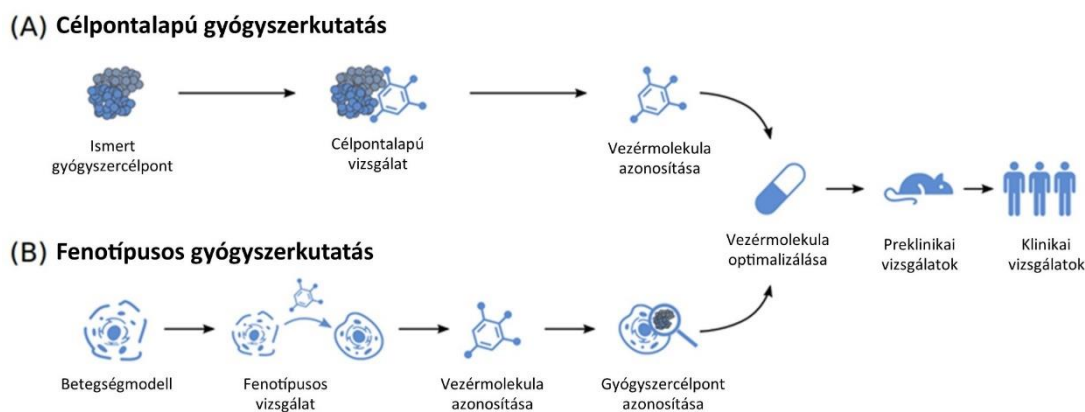
A BERT (*Bidirectional Encoder Representations from Transformers*) egy másik népszerű nagy nyelvi modell, amelyet a Google fejlesztett ki. A BERT kifejezetten azért készült, hogy a nyelvi kontextust teljes egészében figyelembe vegye. Ennek eredményeként jobban megérti a szavak és mondatok közötti kapcsolatokat, ami nagyobb pontosságot eredményez a nyelvi feladatokban. Az általános szövegfeldolgozásban alkalmazzák, például szöveg osztályozásban, összefoglalásra és keresési feladatokban (Moon–Chi–Im 2022).

Mindkét modell jelentős mértékben hozzájárult a természetes nyelvfeldolgozás területének fejlődéséhez. A ChatGPT kiválóan alkalmazható szövegenerálásra, míg a BERT kiemelkedően pontos eredményeket nyújt a szövegelemzési feladatokban. A nagy nyelvi modellek folyamatosan fejlődnek és egyre szélesebb körben alkalmazzák őket (Moon–Chi–Im 2022; Ray 2023).

Célom bemutatni a nyelvi modellek szerepét és alkalmazhatóságát a gyógyszeriparban, kiemelve azokat a területeket, ahol ezek a modellek potenciálisan segíthetnek. Említésre kerülnek a nyelvi modellek használatának potenciális előnyei és kihívásai, amelyekkel a modellek alkalmazása során szembesülhetünk.

ChatGPT a gyógyszerfejlesztés területén

A gyógyszerfejlesztés fő irányai a fenotípusos és célpontalapú kutatás (1. sz. ábra). A fenotípusos megközelítés során egy komplex betegségmodellt használva (például állatmodell) közvetlenül megvizsgáljuk a gyógyszerjelöltek hatásait, ezáltal információ szerezhető a hatóanyag várható hatékonyságát illetően. Ekkor azonban a pontos biokémiai hatások ismeretlenek és a biztonságosság nehezen jelezhető előre. A célpontalapú megközelítés során egy olyan biológiai célpontot jelölnek ki, amely közrejátszik a betegség kialakulásában és fennmaradásában. Ez a módszer lehetővé teszi a molekuláris szinten bekövetkező hatások követését és a várható kockázatok előrejelzését, azonban ezek a vezérmolekulák gyakran nem bizonyulnak elég hatékonynak (Kotschy 2021; Krentzel–Shorte–Zimmer 2023; Mohs–Greig 2017).



1. sz. ábra: A gyógyszerfejlesztés fő lépései
 Forrás: Krentzel–Shorte–Zimmer 2023

A gyógyszerkutatás több okból is egy rendkívül nehéz és összetett folyamat (Sun et al. 2022):

- 1. Biológiai komplexitás**
 Az emberi test egy rendkívül összetett rendszer és az emberben lejátszódó molekuláris folyamatok megértése különösen nehéz. A betegségeket számos tényező kiválthatja, így egy hatékony hatóanyag megtalálása kihívást jelent.
- 2. Magas kiesési kockázat**
 A klinikai vizsgálatokba belépő gyógyszerek többsége nem kap hatósági engedélyt. Ennek oka általában a nem megfelelő hatásosság vagy elfogadhatatlan mellékhatások felfedezése a klinikai vizsgálatok során.
- 3. Időigényes és költséges**
 A gyógyszerfelfedezési folyamat hosszú és költséges. Egy originális gyógyszerkészítmény kifejlesztése 15 évig is eltarthat, és akár 0,5-1 milliárd dollárba is kerülhet.
- 4. Szigorú szabályozási követelmények**
 A gyógyszerfejlesztési folyamat szigorúan szabályozott, és a gyógyszereknek preklinikai és klinikai vizsgálatok sorozatán kell átesniük.
- 5. A gyógyszerfejlesztés kiszámíthatatlan jellege**
 A hatóanyag azonosítását és kifejlesztését követően nem biztos, hogy a humán vizsgálatok során is kifejti majd a várt hatást, ugyanakkor váratlan mellékhatások is jelentkezhetnek.

6. Új célpontok hiánya

Sok ismert biológiai célpont esetében már végeztek fejlesztéseket, az új célpontok megtalálása pedig egyre nehezebbé válik.

A vezérmolekula (*lead*) felfedezése tehát kritikus lépés a gyógyszerfejlesztés folyamatában és a felsorolt kihívások ellenére a kutatók továbbra is azon dolgoznak, hogy új és hatékony gyógyszereket fejlesszenek ki. A COVID-19 járvány során egyértelművé váltak a gyógyszerfejlesztés korlátai, rámutatva a kutatási és fejlesztési folyamat felgyorsításának szükségességére (Asselah et al. 2021).

A ChatGPT ígéretesnek tűnik a gyógyszerkutatás területén, mivel alkalmas lehet az előzőekben bemutatott korlátozó tényezők enyhítésére. Számos tanulmány foglalkozik a mesterséges intelligencia alkalmazásával a gyógyszerkutatás területén (Koromina–Pandi–Patrinós 2019; Lamberti et al. 2019; Liang et al. 2020; Mak–Pichika 2019). A ChatGPT megjelenése óta pedig egyre több publikáció fektet hangsúlyt a nyelvi modellekben rejlő lehetőségek kifejtésére.

Mann tanulmányában a ChatGPT translációs medicina területén való alkalmazhatóságát tárgyalta. Az eredményei alapján elmondható, hogy a mesterséges intelligencia alapú modellek segíthetnek felgyorsítani az új gyógyszerek kifejlesztését, valamint hozzájárulnak a lehetséges mellékhatások és a gyógyszerinterakciók azonosításához is (Mann 2023).

Finomhangolva a nyelvi modellt a tudományos irodalmi adathalmazon, a ChatGPT képes a szakirodalmak és szabadalmi adatbázisok tanulmányozására annak érdekében, hogy azonosítsa a potenciális vezérmolekulákat. Az irodalom átfogó elemzése után a ChatGPT fel tudja ismerni a betegség-specifikus anyagokat, vegyületeket, géneket és egyéb

releváns információkat. Ez segítheti a kutatókat az új potenciális célpontok gyors azonosításában és az adott területen folyó kutatások aktuális állapotának felmérésében (Chakraborty–Bhattacharya–Lee 2023).

A ChatGPT használata lehetővé tette a kutatók számára, hogy hatékonyan összefoglalják, lefordítsák és átfogalmazzák a tudományos információkat. Az Insilico Medicine egy olyan biotechnológiai vállalat, amely mesterséges intelligencia segítségével potenciális gyógyszer célpontokat keres. Jelenleg azon dolgoznak, hogy a ChatGPT-t a biológiai célpont felfedező platformjukba integrálják egyfajta fejlett keresőmotorként, ezáltal megkönnyítve az adatok lekérdezést és megértését (Savage 2023).

A Synthace digitális kísérleti platform fejlesztői is elkezdtek a ChatGPT beintegrálását a platformba. A biológiai kísérletek tervezése és automatizálása nehéz és hosszadalmas folyamat lehet, a Synthace ChatGPT prototípusának célja felgyorsítani ezt a folyamatot. A tudósok szövegesen megfogalmazhatják a kísérlet leírását, amelynek feldolgozását ChatGPT-vel valósítják meg. Ezt követően pedig a Synthace platform megtervezi a kísérleti tervet (Synthace 2023).

Az irodalom átfogó elemzését követően a ChatGPT fel tudja ismerni a betegség-specifikus anyagokat, vegyületeket, géneket és egyéb releváns információkat. Képes új vezérmolekulákat tervezni a fizikai és kémiai tulajdonságaik alapján, ezzel segítve a kutatókat hatékonyabb gyógyszerfejlesztésben. Továbbá a ChatGPT alkalmas lehet egy adott vegyület farmakokinetikai, farmakodinamikai és toxicitási tulajdonságainak becslésére (Zhao–Wu 2023).

Heck megvizsgálta a ChatGPT alkalmazhatóságát a fehérje alapú gyógyszerek tervezése terén. A nyelvi modellnek feltett kérdések a 70-kDa hőstressz-fehérjékre (HSP70) vonatkoztak, pontosabban ezen fehérjéknek az egészséggel és a betegségekkel való kapcsolatára (pl. testmozgás, fáradtság, cukorbetegség, elhízás esetében), a sejtekben és szövetekben helyére, valamint azok más molekulákkal való kapcsolatára. Heck eredményei alapján elmondható, hogy a nyelvi modell által generált válaszok relevánsnak bizonyultak, ugyanakkor a válaszok javíthatóak, ha a felhasználó pontosítja a kérdéseket. Hallucináció is előfordult, mivel egyes esetekben a hamis hivatkozások és torzítások kombinációja gyanús információhoz vezetett. A szerzők arra a következtetésre jutottak, hogy jelenleg az AI nem képes teljesen helyettesíteni az emberi áttekintést (Heck 2023).

Juhi és mtsai. a ChatGPT hatékonyságát vizsgálták gyakori gyógyszer-gyógyszer interakciók előrejelzésében

és magyarázatában. Azok a betegek, akik több gyógyszert szednek, fokozott kockázatnak vannak kitéve a mellékhatások szempontjából, különösen, ha nincsenek tisztában a gyógyszereik közötti potenciális interakciókkal. A ChatGPT számára feltett két kérdés arra irányult, hogy létezik-e interakció a két gyógyszer között és ha igen, akkor ennek mi a magyarázata. Bár a 40 vizsgált interakcióból 39 esetben a válaszok helyesek bizonyultak, a helyes válaszok esetében a magyarázatok csupán 50%-a bizonyult farmakológiai szempontból helyesnek. Az átlag felhasználó szempontjából azonban a ChatGPT által adott információ helyesnek tekinthető, így a tanulmány összességében hasznosnak találta a nyelvi modellt. Azon betegek, akiknek nincs azonnali hozzáférése egészségügyi létesítményekhez, használhatják a ChatGPT-t az esetleges gyógyszerkölsönhatások kezdeti megértéséhez (Juhi et al. 2023).

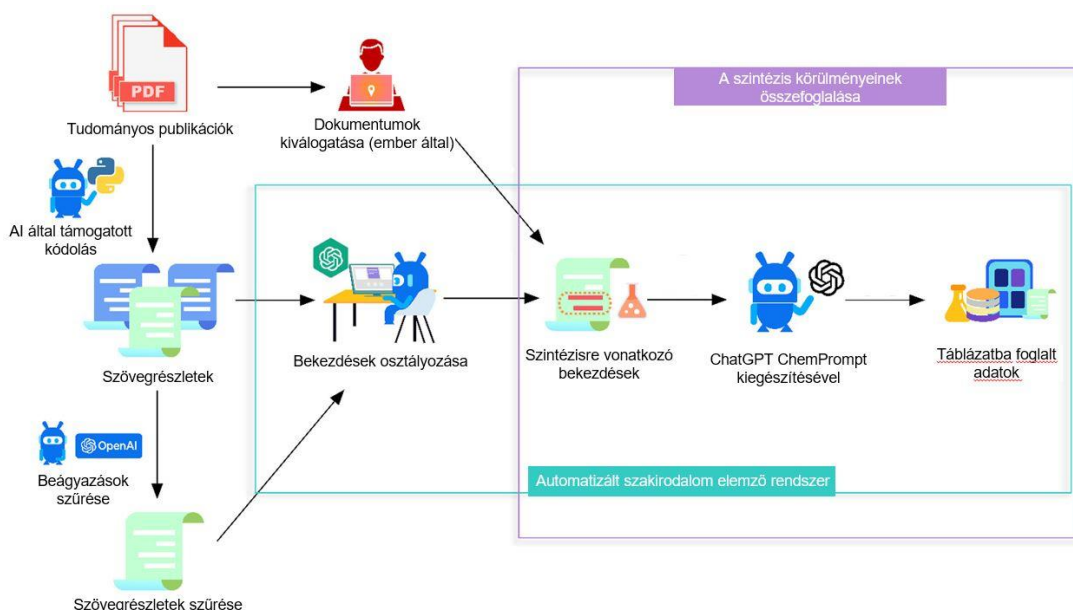
Zheng és mtsai igazolták a ChatGPT alkalmazhatóságát kémiai kutatások területén (Zheng et al. 2023). A szerzők a fémorganikus térhálók (*metal organic frameworks*, MOF) szintézisével kapcsolatos szakirodalom elemzését végezték el a modell segítségével. A ChatGPT kémiai aszisztensként való használatának folyamatábrája a 2. sz. ábrán látható. A szerzők által használt stratégia három fő szempontja a következő:

A hallucináció minimalizálása annak érdekében, hogy a ChatGPT által kitalált vagy félrevezető tartalom előidézését elkerüljék. Ez különösen fontos a kémia területén, ahol az információ pontosságának jelentős hatása lehet a kutatási eredményekre és a biztonságra.

Részletes utasítások használata, amely során a felhasználó részletes iránymutatásokat ad azért, hogy lényegre törő és pontosabb választ kapjon. Ezáltal csökkenthető a nem releváns vagy zavaró válaszok valószínűsége.

Megfelelően strukturált kimenet kérése, ami egy jól meghatározott válasz sablon beépítését jelenti (például táblázatosan). Ez megkönnyíti lényeges információk hatékony kinyerését és értelmezését.

Az így kidolgozott rendszer segítségével a szerzők kiemelkedő pontosságot értek el, F1-Pont értéke (a precizitás és fedés kombinációjából származtatott jószágmetrika) 0,90-0,99 között mozgott. Ezt követően egy gépi tanulási modellt készítettek kinyert adathalmaz segítségével, amely több mint 87%-os pontossággal becsülte meg az MOF kísérletes kristályosodási eredményeit, és azonosította az MOF kristályosodásában fontos tényezőket.



2. sz. ábra: A ChatGPT használata, mint kémiai asszisztens

Forrás: Zheng et al. 2023

A ChatGPT gyakorlati alkalmazásait figyelembe véve számos további alkalmazási lehetőség merül fel a gyógyszerkutatás és -fejlesztés területén. Az egyik ilyen terület a személyre szabott orvoslás, ahol elengedhetetlen a betegek egyéni genetikai információjának elemzése. A nyelvi modelleket nem feltétlenül kell emberi nyelvekre képezni, ugyanis akár a DNS kódok elemzésére is megtanítható. Ennek eredményeként akár olyan fehérjéket is javasolhatnak, amelyek új biológiai célpontok lehetnek (Savage 2023).

A nagy nyelvi modellek lehetővé teszik a kutatók számára, hogy hatékonyabban kommunikáljanak és összefoglalják tudományos eredményeiket. A ChatGPT egyik nagy előnye, hogy kéziratok publikálási folyamatát jelentősen megkönnyíti az angolt nem anyanyelvi szinten beszélő kutatók számára. Fatani tanulmánya alapján elmondható, hogy a ChatGPT hasznosnak bizonyult az orvosi- és fogorvostudományi szakirodalom megírásában, fordításában és összefoglalásában (Fatani 2023). Azonban ezen eszközök használata során fontos a tudományos integritás és az eredeti kutatói munka értékének tiszteletben tartása. Ennek feltétele, hogy a ChatGPT és egyéb nyelvi modellek segítségével a kutatók tisztességesen és átláthatóan használják, illetve megfeleljenek az adott tudományos közösség és a publikációs irányelvek által előírt etikai normáknak és szabályoknak.

A ChatGPT publikációs tevékenység során való alkalmazása heves vitákat váltott ki a kutatók körében. A kutatók egy része pozitívan viszonyul a mesterséges

intelligencia nyújtotta lehetőségekhez és támogatják a nyelvi modellek alkalmazását a tudományos közleményekben, míg mások határozottan ellenzik (Stahl-Eke 2024). Azonban fontos felismerni azokat a potenciális buktatókat is, amelyek a mesterséges intelligencia használatával járnak. Mivel ezek a rendszerek meglévő adathalmazokon tanulnak, előfordulhatnak torzítások és ferdtések, sőt akár felerősíthetnek bizonyos hibás feltételezéseket is. Ez súlyos következményekkel járhat, különösen az orvostudomány és a jog területén, ahol előítéleteken alapuló döntések jelentős hatással lehetnek az emberek életére. A ferdtések kockázatának csökkentése érdekében fontos, hogy a kutatók és az intézmények lépéseket tegyenek azért, hogy adataik sokszínűek és reprezentatívak legyenek. Ehhez hozzátartozik az alulreprezentált csoportokból származó adatok aktív keresése.

Felmerült ugyanakkor annak kérdése, hogy a mesterséges intelligencia lehet-e egy tanulmány társszerzője. A jelenlegi szabályok ezt nem teszik lehetővé, ami azzal magyarázható, hogy a társszerzőnek jelentős tudományos hozzájárulást kell tennie a cikkhez, illetve bele kell egyezni a társszerzővé válásba és felelősséget kell vállalni a tanulmányért. Annak ellenére, hogy az irányelv első része teljesülhet, a ChatGPT azonban mégsem tesz eleget a társszerzőségi követelményeket. Egyes szakértők ugyanakkor azt is felvetették, hogy a mesterséges intelligencia segítségével generált szövegek használata, ha ezt nem jelzik világosan, plágiumnak tekinthető (Stokel-Walker 2023).

Bár a ChatGPT hasznos eszköz lehet a gyógyszerkutatásban, fontos figyelembe venni a módszer hátrányait is:

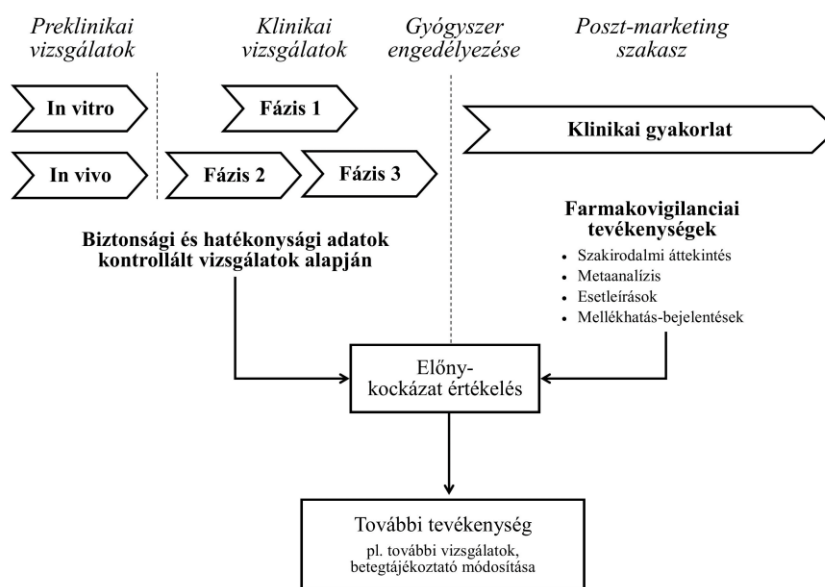
1. *Az adatok minősége és megbízhatósága*
A ChatGPT teljesítménye a tanító adatoktól függ. Ha az adatok hiányosak, elfogultak vagy pontatlanok, akkor a modell előrejelzései nem megbízhatók.
2. *Kísérletes validáció hiánya*
A ChatGPT képes előrejelzéseket és hipotéziseket generálni, de nem képes kísérleteket végezni vagy vegyületek tulajdonságait mérni. Ezért a modell által generált előrejelzéseket utólag kísérletekkel kell validálni.
3. *A biológiai háttér korlátozott megértése*
A ChatGPT modell által generált előrejelzések nem mindig tükrözik a biológiai rendszerek valódi komplexitását.
4. *Korlátozott értelmezhetőség*
A gépi tanulási modellek esetében nem egyértelmű, hogy a modell hogyan jutott az adott eredményre. A használt modell adhat olyan eredményt is, amelyre látszólag nem létezik semmilyen magyarázat. Ezt a jelenséget fekete doboznak nevezik.

Összességében tehát a ChatGPT költséghatékony lehetőséget biztosít nagy mennyiségű adat kezelésére, elemzésére és új ismeretek generálására, ezáltal segítve a kutatókat a gyógyszerkutatás és -fejlesztés felgyorsításában. Fontos azonban megjegyezni, hogy a

preklinikai és klinikai vizsgálatok nem helyettesíthetők a ChatGPT kizárólagos használatával, ugyanis a ChatGPT előrejelzéseit validálni kell. Szem előtt tartva a modellek korlátait, a ChatGPT eredményeit további módszerekkel társítva egy nagyon hasznos eszköz állhat a kutatók rendelkezésére.

A mesterséges intelligencia alkalmazhatósága a farmakovigilancia területén

A gyógyszerek mellékhatásainak jelentős részére már a kutatás-fejlesztés során fény derül és a törzskönyvezés során bekerülnek a készítmény kísérőirataiba. Egy részük azonban csak a forgalomba hozatal követő széles körű és hosszú távon való alkalmazás során válik ismertté. Az Európai Unióban minden gyógyszer szigorú ellenőrzésen megy keresztül a forgalomba hozatal előtt, melyek során preklinikai és klinikai vizsgálatok révén értesülhetünk a készítmény hatékonyságáról és biztonságosságáról (3. sz. ábra). A klinikai vizsgálatok korlátai a kis beteglétszám, a limitált betegcsoport és rövid időtartam, ezért a ritka mellékhatások nagy részére ekkor még nem derül fény. Ezért szükség van egy farmakovigilancia rendszerre, amely lehetővé teszi a készítmények előny-kockázat arányának követését, és az új adatok alapján szükség esetén intézkedéseket hoz a kockázatok csökkentése érdekében (Fittler 2012).



3. sz. ábra: A farmakovigilancia áttekintése

Forrás: Saját szerkesztés

A gyógyszerbiztonság vagy más néven farmakovigilancia a hatályos magyar jogszabály szerint: „a gyógyszerek biztonságossága érdekében a gyógyszer előny/kockázat viszonyát nyomon követő, a kockázat csökkentésére és az előnyök növelésére irányuló tevékenység” (2005. évi XCV. törvény 1. §). Az Egészségügyi Világszervezet (World Health Organization, WHO) szerinti meghatározás bővebb információt nyújt arról, hogy milyen tevékenységeket foglal magában: „a farmakovigilancia a gyógyszerek káros hatásaival, vagy gyógyszerrel kapcsolatos egyéb problémák észlelésével, értékelésével, megértésével, megelőzésével foglalkozó tevékenységek és tudományok összessége” (Fittler 2012; Hadi-Péter et al. 2010).

A farmakovigilancia fő forrásai a spontán mellékhatás-bejelentések, a beavatkozással nem járó vizsgálatok és a szakirodalom. A gyógyszerek forgalomba hozatali engedélyének jogosultjait, valamint az egészségügyi dolgozókat is törvényi előírások kötelezik arra, hogy a termékek teljes életciklusa alatt aktívan figyeljék az esetleges mellékhatásokat és nemkívánatos hatásokat (2005. évi XCV. törvény). E tevékenység résztvevői a gyógyszerügyi hatóságok, a gyógyszergyártók és forgalmazók, az orvosok, gyógyszerészek és egyéb egészségügyi szakemberek.

A spontán mellékhatás bejelentés lehetőséget nyújt a fogyasztóknak és az egészségügyi dolgozóknak arra, hogy bejelentsék a feltételezett mellékhatásokat. A szakemberek becslései alapján azonban mégis 85-94%-os a mellékhatások aluljelentési aránya, ami a bejelentések önkéntes jellegével magyarázható (Hazell–Shakir 2006). Évente 1 millió lakosonként 200-500 bejelentést tekintenek reálisnak a fejlett országokban (ahová Magyarország is tartozik), azonban hazánkban 1 millió lakosra mindössze 70-80 bejelentés van évenként. A mellékhatásokról szóló beszélgetések azonban széles körben folynak a közösségi média oldalain (Hadi-Péter et al. 2010).

Az elmúlt évtizedben egyre növekedett az érdeklődés a közösségi média farmakovigilancia területén való felhasználása iránt. A hivatalos felmérések során hónapok vagy évek is eltelhetnek míg az eredmények elérhetővé válhatnak, a közösségi médián megosztott adatokat szinte azonnal le lehet kérdezni. Powell és munkatársai megállapították, hogy bizonyos esetekben nyilvános Facebook és Twitter bejegyzésekben a mellékhatásra vonatkozó említést tettek, még mielőtt bejelentés érkezett volna a gyógyszerügyi hatóságoknak (Powell et al. 2016). A különböző platformok által kínált álnevesítés ugyanakkor megkönnyíti a nyitottságot a betegek részéről.

A közösségi média alapú farmakovigilancia először 2010-ben jelent meg (Leaman et al. 2010), azóta pedig egyre nagyobb népszerűsége lett. A legtöbb

tanulmány a Twitter, Facebook és Reddit közösségi média platformokat használja olyan feladatokhoz, mint a káros gyógyszerhatások felderítése és az off-label gyógyszeralkalmazás elemzése. A közösségi oldalakon megosztott bejegyzések alapján történő információgyűjtés a következő nehézségekbe ütközik:

- A gyógyszereket kereskedelmi névvel, a hatóanyag nevével és köznyelvi kifejezésekkel is leírhatják.
- A mellékhatásokra olyan kifejezésekkel is utalhatnak, amelyek nem szerepelnek orvosi szótárakban.
- A közösségi média informális jellege miatt a bejegyzések helyesírási és nyelvtani hibákat, rövidítéseket és szlengeket is tartalmazhatnak.
- A bejegyzésből nem mindig egyértelmű, hogy mi a beteget érintő konkrét mellékhatás.
- A közösségi médiában megjelenő tartalmak csak kis része vonatkozik a mellékhatásokra.

A 4. sz. ábrán láthatóak a mellékhatásokra vonatkozó információk közösségi média oldalakról való kinyerésének általános lépései. A gyógyszerrel kapcsolatos bejegyzések az adott közösségi oldal alkalmazásprogramozási felületének (application programming interface, API) segítségével nyerhetőek ki. A gyógyszernevek lehetséges elírásait is érdemes kulcsszóként használni annak érdekében, hogy minél nagyobb adatkészlet álljon rendelkezésre. Ehhez egy fonetikai helyesírási szűrő (phonetic spelling filter) segítségével gyakori helyesírási hibákat lehet generálni (Pimpalkhute et al. 2014). Roche és munkatársai manuálisan hoztak létre egy szótárt a leggyakoribb szavakkal és szinonimáikkal. Így például a „levothyrox” szinonimái a „levo”, a „levothyro” és a „levotyrox” szavak voltak. A feldolgozás során az összes ilyen alakot lecserélték a közös szinonimára, tehát a „levothyrox”-ra (Roche–Robert–Salam 2023). Carpenter és munkatársai a GPT-3 nagy nyelvi modell segítségével szinonimákat generáltak egyes gyógyszerek és kábítószer neveihez. Az így létrehozott lexikon hamis pozitív példákat is tartalmazott, így szükség volt az adatok szűrésére. Például a heroin szinonimái között szerepelt az oxikodon is, annak ellenére, hogy teljesen más hatóanyagról van szó. A modell azonban számos olyan szinonimát generált, amelyek nem szerepeltek a korábban használt kábítószer lexikonokban (Carpenter–Altman 2023).

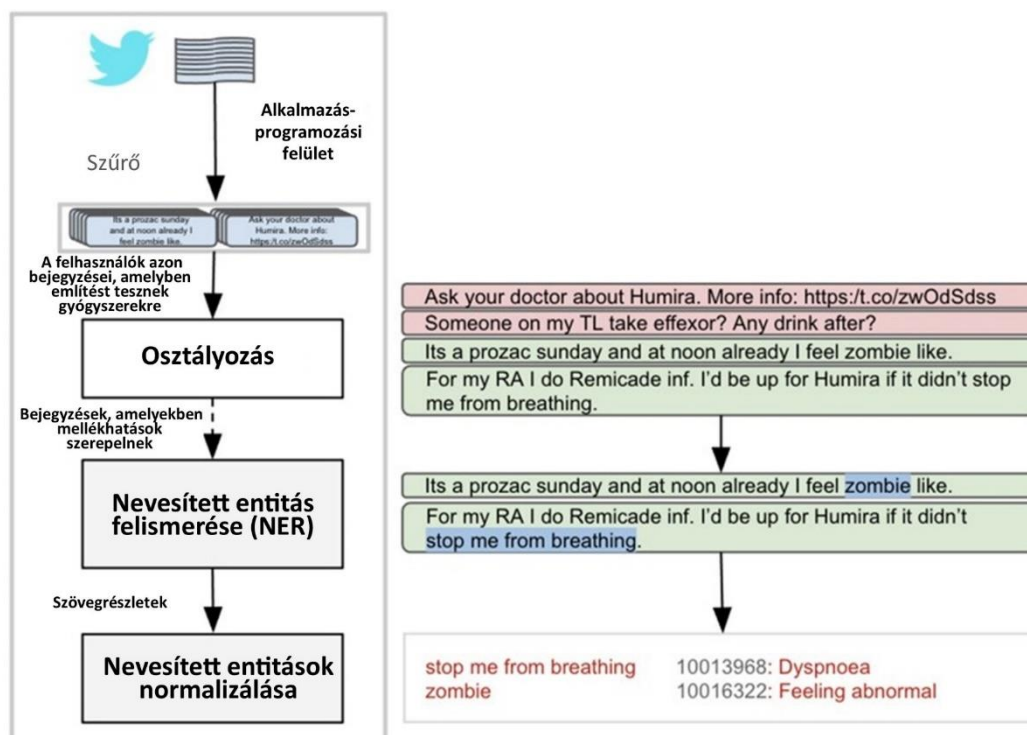
Érdemes ugyanakkor a helyesírási hibákat generálni a billentyűzet elrendezése alapján is. A módszer során az egyes kifejezések betűit rendre a legközelebbi betűre cserélik le. Az elírások jelentőségét igazolja, hogy a lehetséges helyesírási hibák figyelembevételével

Tekumalla és munkatársai 15%-kal több adatot gyűjtöttek össze a COVID-19 kezelésével kapcsolatban a Twitter felületén (Tekumalla–Banda 2020).

A folyamat következő lépése az adatok szűrése, mely során a hirdetéseket, az automatikus válaszokat és a kérdéseket eltávolítják az adathalmazból. A bejegyzésekben levő információk kinyerése névelmefelismeréssel (named entity recognition, NER) valósítható meg. Kihívást jelent az indikáció szerinti alkalmazás („Szörnyű fejfájásom van, és most bevettem egy ibuprofent”) és a mellékhatás („Mióta elkezdtem

szedni a szertralint, olyan rosszul érzem magam”) elkülönítése. Ennek a feladatnak a megoldására alkalmasnak bizonyultak a nagy nyelvi modellek.

Az információk kinyerését követően a bejegyzéseket csoportosítani kell egy általános, nemzetközileg is elfogadott terminológia szerint. A gyógyszerek esetében általánosan a MedDRA (Medical Dictionary for Regulatory Activities) vagy az egységes orvosi nyelvrendszer (Unified Medical Language System, UMLS) orvosi terminológiája szerint csoportosítják a mellékhatásokra vonatkozó adatokat.



4. sz. ábra: Mellékhatások kinyerési folyamata a Twitteren

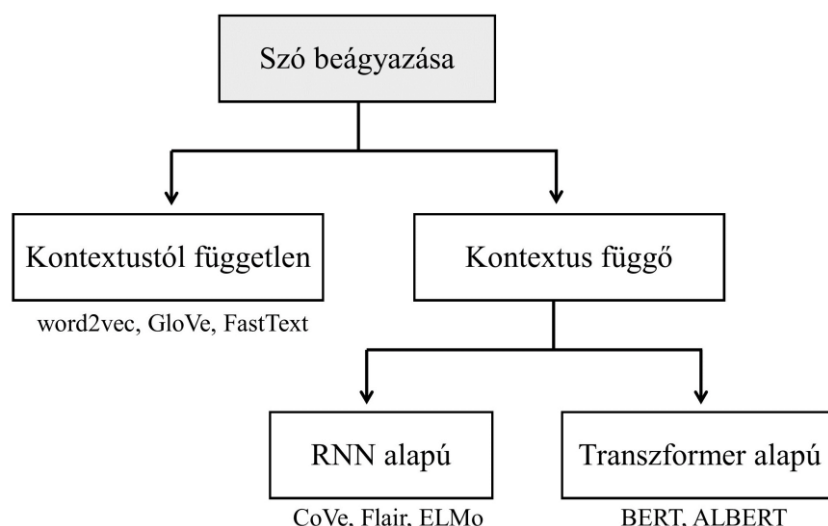
Forrás: Magge et al. 2021

A mellékhatásokra vonatkozó információk kinyerését tárgyaló tanulmányok összefoglalása a 1. sz. táblázatban látható. A szakirodalomban adatforrásként leggyakrabban használt közösségi oldal a Twitter, annak ellenére, hogy a felhasználók csak rövid, 280 karakternél

rövidebb bejegyzéseket tudnak közzétenni. A tanulmányokban használt természetes nyelvfeldolgozási módszerek csoportosítása az 5. sz. ábrán látható. A BERT alapú nyelvi modellek bizonyultak a legnépszerűbbnek.

Alkalmazott eszköz	Közösségi média oldal	Forrás
BERT	Facebook Twitter	(Hussain et al. 2022) (Biseda–Mo 2020; Hussain et al. 2021; Hussain et al. 2022; Klein et al. 2020; Prakash Babu–Eswari 2020; Tekumalla–Banda 2020; Zhu et al. 2020)
BLSTM-RNN	Twitter	(Cocos–Fiks–Masino 2017; Tekumalla–Banda 2020)
word2vec	Naver Twitter	(Lee et al. 2023) (Magge et al. 2021; Sarker–Gonzalez 2017)
FastText	Twitter	(Magge et al. 2021)
CNN	Doctissimo Twitter	(Roche–Robert–Salam 2023) (Tekumalla–Banda 2020)

1. sz. táblázat: A mellékhatásokra vonatkozó információk kinyerését tárgyaló tanulmányok összefoglalása
Forrás: Saját szerkesztés



5. sz. ábra: A természetes nyelvfeldolgozási (Natural Language Processing, NLP) módszerek csoportosítása
Forrás: Saját szerkesztés

A Magge és munkatársai által kifejlesztett DeepADEMiner ígéretesnek tűnik a mellékhatásokkal kapcsolatos adatok kinyerésére (Magge et al. 2021). A kidolgozott módszer egyik korlátozó tényezője azonban, hogy a szerzők azt feltételezték, hogy azok a hatások, amelyek a gyógyszernevekkel együtt szerepelnek a bejegyzésben, a gyógyszer káros hatásai. Azonban ez nem feltétlenül igaz azokban az esetekben, ahol több ugyanis kihívást jelent a beteg adatainak védelme és a gyógyszerbiztonság követése közötti egyensúly megteremtése. Az új technológiák által nyújtott lehetőségeket oly módon kell hasznosítani, hogy az etikus legyen, megfelelően a szabályozásoknak,

gyógyszer vagy káros hatás is szerepel. A mellékhatás ugyanis lehet egy gyógyszerinterakció eredménye, vagy akár a másik gyógyszer mellékhatása is. A kinyert adathalmaz bővíthető a gyógyszer és mellékhatás közötti kapcsolatokkal, ezáltal javítva a kidolgozott modellt.

A közösségi oldalakon történő információgyűjtés során azonban felmerül az adatvédelem kérdésköre, tiszteletben tartva az adatvédelmet (Dreyer et al. 2015). Ahogy nő a közösségi oldalak felhasználóinak száma és a közösségi oldalak adataival foglalkozó szereplők száma, úgy nő az esetleges adatvédelmi jogsértések kockázata is. A közösségi oldalakon megosztott adatok

személyes adatoknak minősülnek mindaddig, amíg egy azonosított vagy azonosítható személyhez kapcsolódnak. Bár a közösségi médiáról származó adatok nyilvánosak, ez a tény nem vonja meg az adatvédelmi jog által nyújtott védelmet (Greenleaf 2012). Az ilyen adatok feldolgozása továbbra is tisztességes és törvényes keretek között kell, hogy megtörténjen. A gyors technológiai fejlődés arra ösztönzi a jogalkotó szerveket, hogy szigorúbb adatvédelmi törvényeket hagyjanak jóvá vagy módosítsák a meglévő szabályozást. Ezek az új adatvédelmi szabályozások a kutatókat kihívások elé állítják (Lengsavath et al. 2017; Pappa–Stergioulas 2019).

A közösségi oldalak bejegyzései gyakran olyan diagnózisokra utalhatnak, amelyek egy internetes kutatás eredményén, nem pedig egészségügyi szakemberek által megerősített diagnózisokon alapulnak. Éppen ezért megkérdőjelezhető az információk megbízhatósága. Eltérés mutatkozik abban is, hogy a betegek és az egészségügyi dolgozók mely tüneteket jelentik. Topaz és munkatársai összehasonlították a betegek közösségi médiában megosztott aggodalmait az elektronikus egészségügyi nyilvántartásokban található orvosi jelentésekkel, és szignifikáns korrelációkat fedeztek fel (Topaz et al. 2015). Azonban a hagyományos spontán mellékhatás bejelentéshez viszonyítva a súlyos mellékhatások alulreprezentáltak a közösségi médiában, ugyanis a betegek elsősorban a kevésbé súlyos (nem életveszélyes vagy nem kórházi kezelést igénylő) eseteket osztják meg (Golder–Norman–Loke 2015).

Hussain és munkatársai felmérték a COVID-19 védőoltást követő nemkívánatos események (*Adverse Event Following Immunisation*, AEFI) említésének gyakoriságát és jellegét a közösségi médiában (Twitter és Facebook) az Egyesült Királyságban. Az adatokat 2020. december 8. (az Egyesült Királyság COVID-19 oltási kampányának kezdete) és 2021. április 30. közötti angol nyelvű Facebook és Twitter bejegyzések képezték. A kutatók azt tapasztalták, hogy a közösségi médiát használók nagy többsége nem reprezentatív az Egyesült Királyság lakosságára nézve (fiatalabbak, gazdagabbak, magasabb végzettséggel). Az életkor ebben az esetben különösen fontos, mivel a vizsgált időszakban az Egyesült Királyságban a COVID-19 oltások nagy részét az idősebb korosztályok kapták meg. A kiválasztási torzítás (*selection bias*) csökkenthető, ha a közösségi oldalokról nyert információkat kiegészítő adatforrásként használják a hagyományos források mellett. A szerzők ugyanakkor arra is felhívják a figyelmet, hogy a közösségi médiát használók egy része elvégzése, azok dokumentálása, a minőséget befolyásoló tényezővel szemben támasztott követelmények azonosítása és betartása. A GMP elengedhetetlen része

sokkal aktívabb, ami ugyancsak torzíthatja az eredményeket. Éppen ezért fontos kontextusba helyezni a felhasználók bejegyzéseit leíró statisztikák és hálózatelemzés segítségével (Hussain et al. 2022).

A WebMD egészségügyi információs webhely adatait felhasználva Park és munkatársai a betegek hormonpótló terápiával kapcsolatos véleményét vizsgálták pajzsmirigy-alulműködés esetén. A különböző nemű és életkorú betegek eltérő tapasztalatokról és aggodalmakról számoltak be a levotiroxinnal (szintetikus pajzsmirigyhormon) kapcsolatban. A nők elsősorban a külső megjelenésre gyakorolt hatások miatt szorongtak, míg a férfiak inkább az esetlegesen előforduló fájdalommal kapcsolatban fejezték ki az aggodalmukat (Park–Hong 2018).

A farmakovigilancia-tevékenység részét képezi ugyanakkor a nagy mennyiségű adat kiértékelése. Wang és munkatársai tanulmánya alapján a ChatGPT sikeresnek bizonyult a gyógyszerek mellékhatásainak azonosításában. A szerzők az egészségügyi dolgozók mellékhatás bejelentéseit használták bemeneti adatként és a ChatGPT eredményeit összevetették a szakemberek által ugyanabban az adathalmazban kijelölt gyógyszer-mellékhatás kapcsolatokkal. A ChatGPT sikeresen azonosította az összes mellékhatás-gyógyszer párt, emellett olyan szövegrészeket is kiemelt, amelyek felett a szakértők elsiklottak (Wang–Ding–Luo 2023).

ChatGPT a Helyes Gyógyszergyártási Gyakorlat (GMP) területén

A Helyes Gyógyszergyártási Gyakorlat (*Good Manufacturing Practice*, GMP) biztosítja, hogy a gyógyszerkészítményt mindig olyan minőségi követelmények szerint gyártsák és ellenőrizzék, amelyek biztosítják, hogy a termékek a forgalomba hozatali engedély követelményeinek eleget tesznek és alkalmasak arra a célra, amire szánták. Az Európai Unió területén a rendszeres GMP vizsgálatokat a nemzeti gyógyszerügyi hatóságok végzik, emellett az Európai Gyógyszerügynökség (*European Medicines Agency*, EMA) is felügyeli a vizsgálatokat annak érdekében, hogy a GMP szabványok betartásra kerüljenek (Abou-El-Enein et al. 2013).

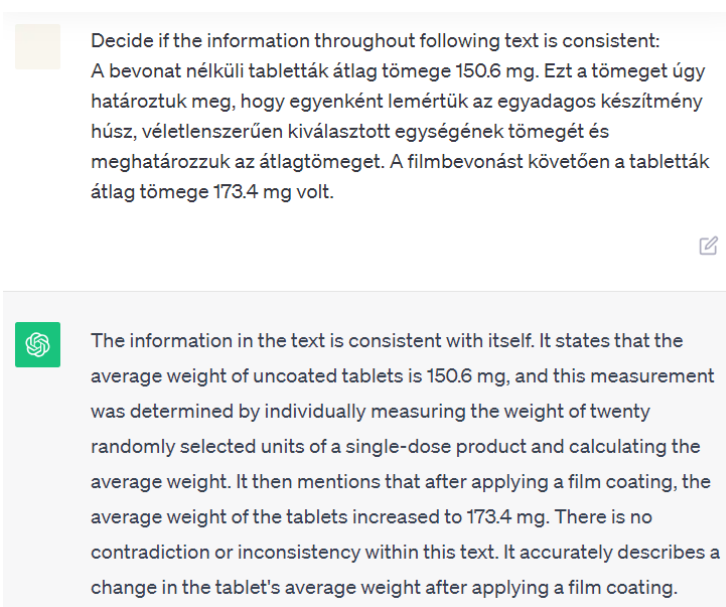
A GMP a gyógyszeripari termelés minden területét lefedi, a nyersanyagoktól, az infrastruktúrán és a berendezéseken, a dolgozók képzésén át a személyes higiénijáig. A GMP részét képezik a szükséges mérések tehát a dokumentáció, ez ugyanis lehetővé teszi a gyártási tétel nyomon követését és az utólagos átvizsgálását. A gyógyszergyártás során nagy

mennyiségű adat keletkezik, amelynek pontosnak és ellentmondásmentesnek kell lennie. Ezeket az információkat pedig rögzíteni, tárolni kell úgy, hogy később bármikor visszakereshetők legyenek. A gyártás, illetve vizsgálat során keletkezett nyersadatok helyességét kiegészítő ellenőrzéssel is igazolni kell, különösen, ha ezeket az adatokat kézzel viszik be a rendszerbe (például a kimért anyagok tömegét, a gyártási számát).

A gyógyszerbiztonság szempontjából tehát elengedhetetlen a gyártás és minőségellenőrzés során készített dokumentációban szereplő lehetséges ellentmondások és hibák felismerése. Luo és munkatársai több nagy nyelvi modell esetében is megvizsgálták, hogy mennyire alkalmasak az ellentmondások és összefüggések azonosítására. A vizsgált modellek közül a ChatGPT bizonyult a legmegfelelőbbnek, azonban fontos kiemelni, hogy a ChatGPT esetében téves következtetéseket és

elfogultságot is érzékeltek. Mindezek ellenére a ChatGPT felgyorsíthatja a gyártási és minőségbiztosítási dokumentáció ellenőrzését és segíthet az ellentmondások felderítésében (Luo–Xie–Ananiadou 2023).

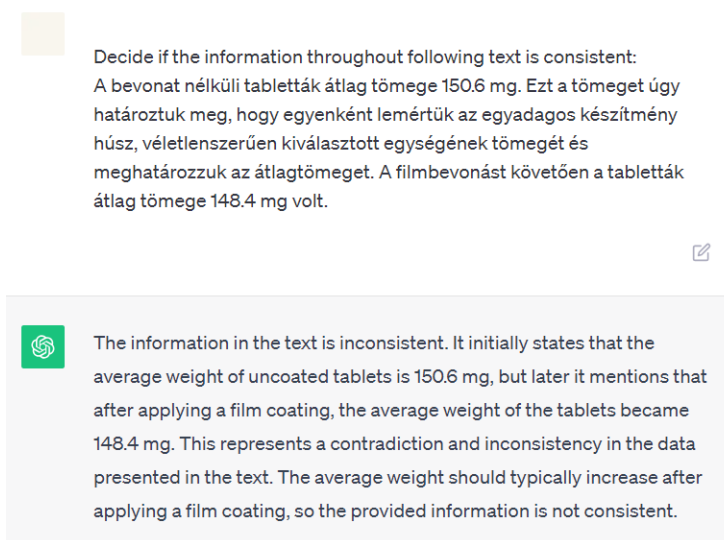
Szemléltetésként a ChatGPT számára azt a feladatot adtam, hogy állapítsa meg, hogy a mellékelt rövid szövegek konzisztensek-e. A szövegek két szám adatot tartalmaztak arra vonatkozóan, hogy mennyi a bevonat nélküli, illetve a filmtabletták átlag tömege. A 6. sz. ábrán látható, hogy abban az esetben, amikor a bevonat nélküli tabletták tömege kisebb volt, mint a filmbevonatos tabletták átlag tömege, a ChatGPT nem talált ellentmondást a szövegben. A filmbevonás során ugyanis a tabletták tömege valóban nő, így a két számérték valóban nincs ellentmondásban egymással. Ennek ellenére a válaszban fölöslegesen arra vonatkozóan is szerepel információ, hogy hány tablettát kellett ehhez megvizsgálni.



6. sz. ábra: Példa, amikor a ChatGPT nem talált ellentmondást a szövegben
 Forrás: Saját példa és szerkesztés

Abban az esetben, amikor a filmbevonatos tabletták tömege kisebb volt, mint a bevonat nélküli tablettáké, a ChatGPT helyes magyarázatot adott arra vonatkozóan, hogy miért van ellentmondás a szövegben (7. sz. ábra), a

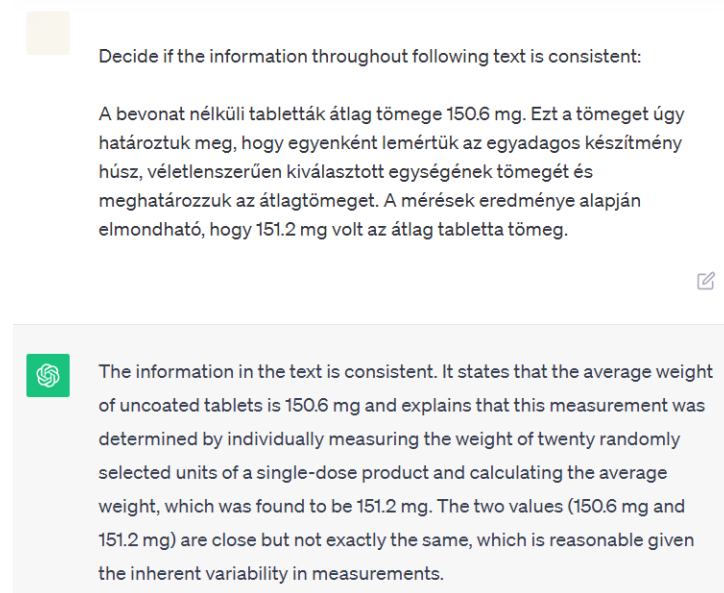
filmbevonatos tabletták tömege ugyanis nem lehet kisebb a bevonat nélküli tablettáknál.



7. sz. ábra: Példa a ChatGPT alkalmazására az ellentmondások felderítésében
Forrás: Saját példa és szerkesztés

Ezt követően olyan szöveg esetében teszteltem a ChatGPT-t, amelyben ugyanannak a mintának az átlag tömege volt eltérő (egyszer 150,6 mg, majd 151,2 mg). A válasz ebben az esetben is tartalmazott főleg információt, ugyanis megismételte a mérési eljárást leírását, mely szerint húsz darab véletlenszerűen kiválasztott tablettá tömegét kell lemérni (8. sz. ábra). A ChatGPT nem talált ellentmondást a bejegyzésben

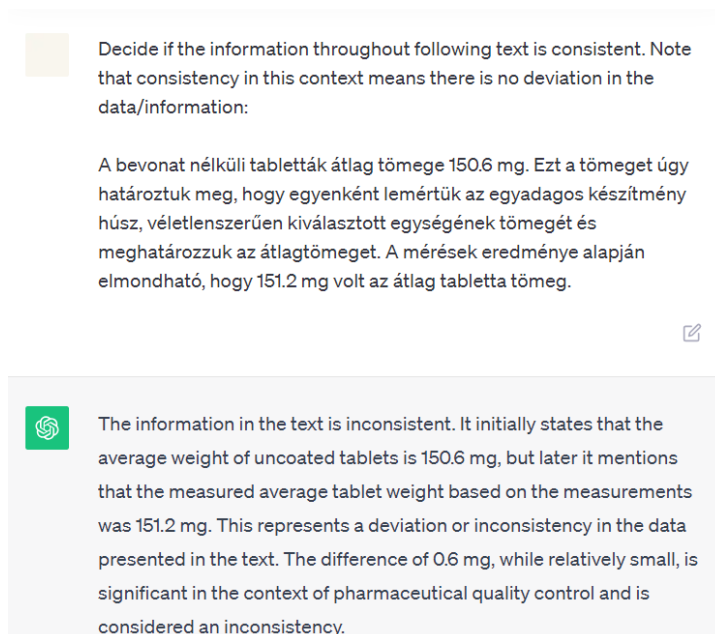
annak ellenére, hogy a két érték közel áll egymáshoz, de nem teljesen azonosak. Ezt azonban elfogadhatónak találta, tekintettel a mérésekkel járó változékonyságra. Mindez nem helyes, tekintettel arra, hogy nem felel meg a gyógyszeripar elvárásainak, ha ugyanazon minta tömege ekkora mértékben eltérhet. Az ilyen jellegű hibák felismerése elengedhetetlen abban az esetben, ha a ChatGPT-t vagy más nyelvi modellt a gyárak dokumentációjának ellenőrzése céljából használnák.



8. sz. ábra: Példa a ChatGPT által adott téves következtetésre
Forrás: Saját példa és szerkesztés

Pontosítva a felhívószöveget és az elvárásokat, a ChatGPT a 9. sz. ábrán látható választ generálta. Annak ellenére, hogy sikeresen azonosította az ellentmondást, a 0,6 mg-os tömegbeli eltérést viszonylag kevésnek találta. Ahogy korábban is említettem, a

gyógyszeriparban ez jelentős eltérést jelent, amire a ChatGPT is felhívja a figyelmet. A válasz nem tartalmaz olyan fölösleges információkat, mint a 6. sz. és a 8. sz. ábrákon bemutatott esetekben.



9. sz. ábra: Példa a ChatGPT által adott válaszra a pontosítást követően

Forrás: Saját példa és szerkesztés

Az OpenAI lehetővé tette a nyelvi modellek, köztük a GPT-3.5 Turbo és a GPT-4 finomhangolását, így a fejlesztők a modelleket az egyedi igényekhez igazíthatják. A nyelvi modellek finomhangolásával lehetőség van arra, hogy jobban teljesítsen célfeladatok esetében. A továbbfejlesztett irányíthatóság biztosítja, hogy a modellek pontosabban kövessék az utasításokat, és a megbízható kimeneti formázás garantálja a következetes AI-válaszokat. Ezáltal csökkenthetőek vagy akár elkerülhetőek a ChatGPT által adott téves következtetések és a fölösleges információk (Badini et al. 2023).

A ChatGPT és egyéb nyelvi modellek gyógyszeriparban való alkalmazhatósága azonban további nehézségekbe ütközik. Gyógyszeripari vállalatok ugyanis szigorú követelményeket támasztanak a nyers adatokkal szemben, különösen az elektronikus formában tárolt adatok esetében. Az ilyen információk mentése, archiválása és más rendszerekbe történő átvitele során annyira szigorú feltételeknek kell eleget tenni, hogy emiatt a vállalatok nem kapcsolják össze a meglévő számítógépes rendszereiket a mérőműszerekkel. Ezért a hatóság által is rendkívül

szigorúan szabályozott gyógyszergyártás digitalizációja csak lassú ütemben, a teljes kutatási és gyártási szemlélet átforgalmazásával valósítható meg.

A digitalizáció jelentősen javíthatja a gyógyszeripari folyamatok minőségét, elsősorban a dokumentáció és gyártás terén, biztonságosabbá, hatékonyabbá és következetesebbé téve azokat. A digitális platformok használata számos módon javíthatja a folyamatokat, mivel lehetővé teszi a gyors adatgyűjtést, a kísérleti eredmények valós idejű megosztását és a termelés különböző szempontjainak nyomon követését. Ezáltal a digitalizáció csökkentheti a termelési költségeket, és javíthatja a termékminőséget. Sok gyógyszergyártó azonban mégis tartózkodik digitális gyártási eljárások bevezetésétől, mivel úgy látják, hogy rendszereik, adataik és embereik nincsenek felkészülve. A nemzetközi gyógyszerügyi hatóságok ösztönzésére azonban az utóbbi években több gyógyszergyár is elkezdett kísérletezni a digitalizációval és azon belül a mesterséges intelligencia lehetséges alkalmazásaival a gyártás és minőség-ellenőrzés területén (Hole–Hole–McFalone–Shaw 2021). A digitalizáció során a vállalatnak szüksége van kiberbiztonsági, mesterséges intelligencia és egyéb stratégiák kidolgozására is, ami

kezdetben jelentős költségeket jelent majd. A vállalat versenyképességének fenntartása érdekében azonban elengedhetetlen, hogy a gyár elfogadja és alkalmazza a digitalizáció eszközeit. A digitalizáció elterjedésével a mesterséges intelligencia és a nyelvi modellek egyre nagyobb teret kapnak majd a gyógyszergyártás és minőség-ellenőrzés területén is.

Összefoglalás

A gyógyszeripar területén a nagy nyelvi modellek alkalmazása kiterjed a gyógyszerkutatás, a farmakovigilancia és a GMP témaköreire egyaránt.

A szakirodalmak és szabadalmi adatbázisok tanulmányozásával a nagy nyelvi modellek azonosíthatják a lehetséges vezérmolekulákat és a biológiai célpontokat, ezáltal csökkentve a gyógyszerkutatás költségeit és felgyorsítva a folyamatot. A ChatGPT képes új vezérmolekulákat tervezni fizikai és kémiai tulajdonságaik alapján, ezzel segítve a kutatókat hatékonyabb gyógyszerfejlesztésben. Továbbá a ChatGPT alkalmas lehet egy adott vegyület farmakokinetikai, farmakodinamikai és toxicitási tulajdonságainak becslésére is.

A nagy nyelvi modellek lehetővé teszik a kutatók számára, hogy hatékonyabban kommunikáljanak és összefoglalják tudományos eredményeiket. Azonban ezen eszközök használata során fontos a tudományos integritás és az eredeti kutatói munka értékének tiszteletben tartása. A nyelvi modellek egy további lehetséges alkalmazási területe a személyre szabott orvoslás, ahol elengedhetetlen a betegek egyéni genetikai információjának elemzése. A szintetikus szerves kémia területén nagy jelentőséggel bír, hogy a kémiai kísérletek eredményei és a kísérleteket befolyásoló tényezők becsülhetőek egy gépi tanulási modellel, amelyet a nyelvi modell segítségével kinyert adathalmazzal tanítottak.

A farmakovigilancia területén a legtöbb tanulmány a közösségi média oldalain történő információgyűjtés kérdéskörével foglalkozik. Az így gyűjtött adatok elemzésével fontos gyógyszer-mellékhatásokra vonatkozó információkra derülhet fény. A farmakovigilancia-tevékenység részét képezi ugyanakkor a nagy mennyiségű adat kiértékelése, amelyet jelenleg szakemberek végeznek. A nyelvi modellek segítségével azonban gyorsan és hatékonyan meghatározhatóak a gyógyszer-mellékhatás kapcsolatok.

Két vállalat (Insilico Medicine és Synthace) már elkezdte beépíteni a ChatGPT-t a rendszerükbe. A

gyógyszergyárakban való alkalmazás azonban még várat magára, ami az ipart jellemző szigorú szabályozásokkal magyarázható. A gyógyszergyártó vállalatok sokáig tartózkodtak digitális gyártási eljárások bevezetésétől, mivel úgy látták, hogy rendszereik, adataik és embereik nincsenek erre felkészülve. A nemzetközi gyógyszerügyi hatóságok ösztönzésére azonban az utóbbi években több gyógyszergyár is elkezdett kísérletezni a digitalizációval, azon belül a mesterséges intelligencia lehetséges alkalmazásaival a gyártás és minőség-ellenőrzés területén.

Összességében tehát a ChatGPT költséghatékony lehetőséget biztosít nagy mennyiségű adat kezelésére, elemzésére és új ismeretek generálására, ezáltal segítve a kutatóknak a gyógyszerkutatás és -fejlesztés felgyorsításában. Fontos azonban kiemelni, hogy a nyelvi modellek alkalmazhatóságát validálni kell más módszerekkel, mivel a modelleknek vannak korlátai.

A nyelvi modellek használata számos etikai szempontot is felvet, mint például hitelesség, plágium, adatvédelem és torzítások. Bizonyos esetekben megkérdőjelezhető az adatok minősége és megbízhatósága, akár hallucináció is felléphet. Emellett gondot jelenthet a jelenség tudományos háttérének korlátozott megértése is, mivel a nyelvi modell által generált válaszok nem mindig tükrözik a rendszerek valódi komplexitását.

A hibás válaszok előfordulása bizonyos mértékben javítható a modell finomhangolásával és részletesebb utasítások megadásával. Figyelembe véve ezeket a korlátokat és a modellek eredményeit, más kutatási eszközökkel együttesen alkalmazva a nagy nyelvi modellek rendkívül hasznos eszközei lehetnek a gyógyszeriparnak.

Köszönetnyilvánítás

A Doktoranduszi Kiválósági Ösztöndíj Program (DKÖP) által támogatott projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatallal kötött támogatási szerződés alapján valósult meg.

Rövidítésjegyzék

Rövidítés	Angol kifejezés	Magyar kifejezés
AEFI	<i>Adverse Event Following Immunisation</i>	Oltást követő nemkívánatos események
AEFI	<i>Adverse Event Following Immunisation</i>	Nemkívánatos események
AI	<i>Artificial Intelligence</i>	Mesterséges intelligencia
API	<i>Application Programming Interface</i>	Alkalmazásprogramozási felület
BERT	<i>Bidirectional Encoder Representations from Transformers</i>	
(BLSTM) RNN	<i>Bidirectional Long Short-Term Memory</i>	
EMA	<i>European Medicines Agency</i>	Európai Unióban az Európai Gyógyszerügynökség
FDA	<i>Food and Drug Administration</i>	Egyesült Államok Élelmiszer- és Gyógyszerügyi Hivatala
GMP	<i>Good Manufacturing Practice</i>	Helyes Gyógyszergyártási Gyakorlat
GPT	<i>Generative Pre-trained Transformer</i>	
LLM	<i>Large Language Model</i>	Nagy nyelvi modell
MedDRA	<i>Medical Dictionary for Regulatory Activities</i>	
MOF	<i>Metal Organic Frameworks</i>	Fémorganikus térhálók
NLP	<i>Natural Language Processing</i>	Természetes nyelvi feldolgozás
OGYÉI		Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet
PSUR	<i>Periodic Safety Update Reports</i>	Időközönként időszakos biztonságossági jelentés
QPPV	<i>Qualified Person Responsible for Pharmacovigilance</i>	Farmakovigilanciáért felelős személy
RLHF	<i>Reinforcement Learning with Human Feedback</i>	Visszajelzés alapú megerősítő tanulás
RNN	<i>Recurrent Neural Network</i>	Rekurrens neurális hálózat
UMLS	<i>Unified Medical Language System</i>	Egységes orvosi nyelvrendszer

WHO	World Health Organization	Egészségügyi Világszervezet
-----	---------------------------	-----------------------------

2. sz. táblázat: Rövidítésjegyzék
 Forrás: Saját szerkesztés

Irodalomjegyzék

- Abou-El-Enein, M.–Römhild, A.–Kaiser, D.–Beier, C.–Bauer, G.–Volk, H.-D.–Reinke, P. (2013) Good Manufacturing Practices (GMP) manufacturing of advanced therapy medicinal products: a novel tailored model for optimizing performance and estimating costs. *Cytotherapy*, Vol. 15. No. 3. pp. 362–383. doi: 10.1016/j.jcyt.2012.09.006
- Alkaissi, H.–McFarlane, S. I. (2023) Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, Vol. 15. No. 3. e35179. doi: 10.7759/cureus.35179
- Asselah, T.–Durantel, D.–Pasmant, E.–Lau, G.–Schinazi, R. F. (2021) COVID-19: Discovery, diagnostics and drug development. *Journal of Hepatology*, Vol. 71. No. 1. pp. 168–184. doi: 10.1016/j.jhep.2020.09.031
- Badini, S.–Regondi, S.–Frontoni, E.–Pugliese, R. (2023) Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting. *Advanced Industrial and Engineering Polymer Research*, Vol. 6. No. 3. pp. 278–287. doi: 10.1016/j.aiepr.2023.03.003
- Biseda, B.–Mo, K. (2020) *Enhancing Pharmacovigilance with Drug Reviews and Social Media*. doi: 10.48550/arXiv.2004.08731
- Carpenter, K.A.–Altman, R.B. (2023) Using GPT-3 to Build a Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance. *Biomolecules*, Vol. 13. No. 2. 387. doi: 10.3390/biom13020387
- Chakraborty, C.–Bhattacharya, M.–Lee, S. S. (2023) Artificial intelligence enabled ChatGPT and large language models in drug target discovery, drug discovery, and development. *Molecular Therapy - Nucleic Acids*, Vol. 33. pp. 866–868. doi: 10.1016/j.omtn.2023.08.009
- Cocos, A.–Fiks, A. G.–Masino, A. J. (2017) Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *Journal of the American Medical Informatics Association*, Vol. 24. No. 4. pp. 813–821. doi: 10.1093/jamia/ocw180
- Dreyer, N. A.–Blackburn, S.–Hliva, V.–Mt-Isa, S.–Richardson, J.–Jamry-Dziurla, A. et al. (2015) Balancing the Interests of Patient Data Protection and Medication Safety Monitoring in a Public-Private Partnership. *JMIR Medical Informatics*, Vol. 3. No. 2. doi: 10.2196/medinform.3937
- Fatani, B. (2023) ChatGPT for Future Medical and Dental Research. *Cureus*, Vol. 14. No. 4. e37285. doi: 10.7759/cureus.37285
- Fittler, A. (2012) Mit tehetnek a gyakorló gyógyszerészek a biztonságos gyógyszeralkalmazás érdekében? A farmakovigilancia és a gyógyszermellékhatás bejelentések szabályozásának változásairól. *Gyógyyszerészet*, Vol. 56. No. 11. pp. 650–655.
- Golder, S.–Norman, G.–Loke, Y. K. (2015) Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *British Journal of Clinical Pharmacology*, Vol. 80. No. 4. pp. 878–888. doi: 10.1111/bcp.12746
- Greenleaf, G. (2012) Global data privacy laws: 89 countries, and accelerating. In: *Privacy Laws & Business International Report*, Vol. 115. pp. 1–14.
- Gu, Y.–Tinn, R.–Cheng, H.–Lucas, M.–Usuyama, N.–Liu, X. et al. (2022) Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Transactions on Computing for Healthcare*, Vol. 3. No. 1. pp. 1–23. doi: 10.1145/3458754
- Hadi-Péter, E. S.–Hadi, K.–Hadi, F.–Szentirmai, B.–Hankó, B. (2010) Farmakovigilancia – mellékhatás bejelentés. *Gyógyyszerészet*, Vol. 54. No. 5. pp. 259–264.
- Hazell, L.–Shakir, S. A. W. (2006) Under-Reporting of Adverse Drug Reactions. *Drug Safety*, Vol. 29. No. 5. pp. 385–396. doi: 10.2165/00002018-200629050-00003
- Heck, T. G. (2023) What artificial intelligence knows about 70 kDa heat shock proteins, and how we will face this ChatGPT era. *Cell Stress and Chaperones*, Vol. 28. No. 3. pp. 225–229. doi: 10.1007/s12192-023-01340-1
- Hole, G.–Hole, A. S.–McFalone-Shaw, I. (2021) Digitalization in pharmaceutical industry: What to focus on under the digital implementation process? *International Journal of Pharmaceutics: X*, Vol. 3. 100095. doi: 10.1016/j.ijpx.2021.100095
- Hussain, S.–Afzal, H.–Saeed, R.–Iltaf, N.–Umair, M. Y. (2021) Pharmacovigilance with Transformers: A

- Framework to Detect Adverse Drug Reactions Using BERT Fine-Tuned with FARM. *Computational and Mathematical Methods in Medicine*. Vol. 2021. 5589829. doi: 10.1155/2021/5589829
- Hussain, Z.–Sheikh, Z.–Tahir, A.–Dashtipour, K.–Gogate, M.–Sheikh, A.–Hussain, A. (2022) Artificial Intelligence–Enabled Social Media Analysis for Pharmacovigilance of COVID-19 Vaccinations in the United Kingdom: Observational Study. *JMIR Public Health and Surveillance*, Vol. 8. No. 5. e32543. doi: 10.2196/32543
- Juhi, A.–Pipil, N.–Santra, S.–Mondal, S.–Behera, J. K.–Mondal, H. (2023) The Capability of ChatGPT in Predicting and Explaining Common Drug–Drug Interactions. *Cureus*, Vol. 15. No. 3. e36272. doi: 10.7759/cureus.36272
- Kasneji, E.–Sessler, K.–Küchemann, S.–Bannert, M.–Dementieva, D.–Fischer, F., et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, Vol. 103. 102274. doi: 10.1016/j.lindif.2023.102274
- Klein, A. Z.–Cai, H.–Weissenbacher, D.–Levine, L. D.–Gonzalez-Hernandez, G. (2020) A natural language processing pipeline to advance the use of Twitter data for digital epidemiology of adverse pregnancy outcomes. *Journal of Biomedical Informatics: X*, Vol. 112S. 100076. doi: 10.1016/j.jybinx.2020.100076
- Koromina, M.–Pandi, M. T.–Patrinós, G. P. (2019) Rethinking Drug Repositioning and Development with Artificial Intelligence, Machine Learning, and Omics. *OMICS A Journal of Integrative Biology*, Vol. 23. No. 11. pp. 539–548. doi: 10.1089/omi.2019.0151
- Kotschy, A. (2021) A gyógyszerkutatás új irányzatai: hatékonyság és biztonságosság. *Scientia et Securitas*, Vol. 2. No. 2. pp. 177–183. doi: 10.1556/112.2021.00028
- Krentzel, D.–Shorte, S. L.–Zimmer, C. (2023) Deep learning in image-based phenotypic drug discovery. *Trends in Cell Biology*, Vol. 33. No. 7. pp. 538–554. doi: 10.1016/j.tcb.2022.11.011
- Lamberti, M. J.–Wilkinson, M.–Donzanti, B. A.–Wohlhieter, G. E.–Parikh, S.–Wilkins, R. G.–Getz, K. (2019) A Study on the Application and Use of Artificial Intelligence to Support Drug Development. *Clinical Therapeutics*, Vol. 41. No. 8. pp. 1414–1426. doi: 10.1016/j.clinthera.2019.05.018
- Leaman, R.–Wojtulewicz, L.–Sullivan, R.–Skariah, A.–Yang, J.–Gonzalez, G. (2010) Extracting Adverse Drug Reactions from User Posts in Health-Related Social Networks. In: *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pp. 117–125. Uppsala, Sweden. Association for Computational Linguistics.
- Lee, S.–Woo, H.–Lee, C. C.–Kim, G.–Kim, J. Y.–Lee, S. (2023) Drug_SNSMiner: standard pharmacovigilance pipeline for detection of adverse drug reaction using SNS data. *Scientific Reports*, Vol. 13. No. 1. 3779. doi: 10.1038/s41598-023-28912-6
- Lengsavath, M.–Dal Pra, A.–de Ferran, A. M.–Brosch, S.–Härmark, L.–Newbould, V.–Goncalves, S. (2017) Social Media Monitoring and Adverse Drug Reaction Reporting in Pharmacovigilance: An Overview of the Regulatory Landscape. *Therapeutic Innovation and Regulatory Science*, Vol. 51. No. 1. pp. 125–131. doi: 10.1177/2168479016663264
- Liang, G.–Fan, W.–Luo, H.–Zhu, X. (2020) The emerging roles of artificial intelligence in cancer drug development and precision therapy. *Biomedicine and Pharmacotherapy*, Vol. 128. pp. 110255. doi: 10.1016/j.biopha.2020.110255
- Luo, Z.–Xie, Q.–Ananiadou, S. (2023) ChatGPT as a Factual Inconsistency Evaluator for Text Summarization. *Arxiv*, 2303.15621. doi: 10.48550/arXiv.2303.15621
- Magge, A.–Tutubalina, E.–Miftahutdinov, Z.–Alimova, I.–Dirkson, A.–Verberne, S. et al. (2021) DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. *Journal of the American Medical Informatics Association*, Vol. 28. No. 10. pp. 2184–2192. doi: 10.1093/jamia/ocab114
- Mak, K. K.–Pichika, M. R. (2019) Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, Vol. 24. No. 3. pp. 773–780. doi: 10.1016/j.drudis.2018.11.014
- Mann, D. L. (2023) Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine: A JACC: Basic to Translational Science Interview With ChatGPT. *JACC: Basic to Translational Science*, Vol. 8. No. 2. pp. 221–223. doi: 10.1016/j.jacbts.2023.01.001
- McCauley, S. M.–Christiansen, M. H. (2019) Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, Vol. 126. No. 1. pp. 1–51. doi: 10.1037/rev0000126
- Mohs, R. C.–Greig, N. H. (2017) Drug discovery and development: Role of basic biological research. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, Vol. 3. No. 4. pp. 651–657. doi: 10.1016/j.trci.2017.10.005
- Moon, S.–Chi, S.–Im, S. B. (2022) Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT).

- Automation in Construction*, Vol. 142. 104465. doi: 10.1016/j.autcon.2022.104465
- Pappa, D.–Stergioulas, L. K. (2019) Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. *International Journal of Data Science and Analytics*, Vol. 8. No. 2. pp. 113–135. doi: 10.1007/s41060-019-00175-3
- Park, S. H.–Hong, S. H. (2018) Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy From Online Patient Medication Reviews: Text Mining of Social Network Data. *Journal of Medical Internet Research*, Vol. 20. No. 10. pp. e11085. doi: 10.2196/11085
- Pimpalkhute, P. –Patki, A.–Nikfarjam, A.–Gonzalez G. (2014) Phonetic Spelling Filter for Keyword Selection in Drug Mention Mining from Social Media. *AMIA Joint Summits on Translational Science proceedings*, Vol. 2014. pp. 90–95.
- Powell, G. E.–Seifert, H. A.–Reblin, T.–Burstein, P. J.–Blowers, J.–Menius, J. A. et al. (2016) Social Media Listening for Routine Post-Marketing Safety Surveillance. *Drug Safety*, Vol. 39. No. 5. pp. 443–454. doi: 10.1007/s40264-015-0385-6
- Prakash Babu, Y.–Eswari, R. (2020) Identification of Medication Tweets Using Domain-specific Pre-trained Language Models. In: *Proceedings of the Fifth Social Media Mining for Health Applications Workshop*, pp. 128–130.
- Ray, P. P. (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, Vol. 3. pp. 121–154. doi: 10.1016/j.iotcps.2023.04.003
- Reynolds, L.–McDonell, K. (2021) Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–7. New York, NY, USA. doi: 10.1145/3411763.3451760
- Roche, V.–Robert, J. P.–Salam, H. (2023) A holistic AI-based approach for pharmacovigilance optimization from patients behavior on social media. *Artificial Intelligence in Medicine*, Vol. 144. 102638. doi: 10.1016/j.artmed.2023.102638
- Sarker, A.–Gonzalez, G. (2017) A corpus for mining drug-related knowledge from Twitter chatter: Language models and their utilities. *Data in Brief*, Vol. 10. pp. 122–131. doi: 10.1016/j.dib.2016.11.056
- Savage, N. (2023) Drug discovery companies are customizing ChatGPT: here's how. *Nature Biotechnology*, Vol. 41. pp. 585–586. doi: 10.1038/s41587-023-01788-7
- Stahl, B. C.–Eke, D. (2024) The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, Vol. 74. 102700. doi: 10.1016/j.ijinfomgt.2023.102700
- Stokel-Walker, C. (2023) ChatGPT listed as author on research papers: many scientists disapprove. *Nature*, Vol. 613. No. 7945. pp. 620–621. doi: 10.1038/d41586-023-00107-z
- Sun, D.–Gao, W.–Hu, H.–Zhou, S. (2022) Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, Vol. 12. No. 7. pp. 3049–3062. doi: 10.1016/j.apsb.2022.02.002
- Synthace. (2023) <https://www.synthace.com/news/synthace-platform-integrates-with-chatgpt> [Letöltve: 2023.09.28.].
- Tekumalla, R.–Banda, J. M. (2020) Characterizing drug mentions in COVID-19 Twitter Chatter. In: *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, doi: 10.18653/v1/2020.nlpCOVID19-2.25
- Topaz, M.–Lai, K.–Dhopeswarkar, N.–Seeger, D. L.–Sa'Adon, R.–Goss, F. et al. (2015) Clinicians' reports in electronic health records versus patients' concerns in social media: A pilot study of adverse drug reactions of aspirin and atorvastatin. *Drug Safety*, Vol. 39. No. 3. pp. 243–250. doi: 10.1007/s40264-015-0381-x
- Wang, H.–Ding, Y. J.–Luo, Y. (2023) Future of ChatGPT in Pharmacovigilance. *Drug Safety*, Vol. 46. No. 8. pp. 711–713. doi: 10.1007/s40264-023-01315-2
- Zhao, A.–Wu, Y. (2023) Future implications of ChatGPT in pharmaceutical industry: drug discovery and development. *Frontiers in Pharmacology*, Vol. 41. pp. 17–19. doi: 10.3389/fphar.2023.1194216
- Zheng, Z.–Zhang, O.–Borgs, C.–Chayes, J. T.–Yaghi, O. M. (2023) ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *Journal of the American Chemical Society*, Vol. 145. No. 32. pp. 18048–18062. doi: 10.1021/jacs.3c05819
- Zhu, M.–Song, Y.–Jin, G.–Jiang, K. (2020) Identifying personal experience tweets of medication effects using pre-trained RoBERTa language model and its updating. In: *EMNLP 2020 - 11th International Workshop on Health Text Mining and Information Analysis, LOUHI 2020, Proceedings of the Workshop*, pp. 127–137. doi: 10.18653/v1/2020.louhi-1.14
- 15/2012. (VIII. 22.) EMMI rendelet az emberi alkalmazásra kerülő gyógyszerek farmakovigilanciájáról. 2005. évi XCV. törvény Az emberi alkalmazásra kerülő gyógyszerekről és egyéb, a gyógyszerpiacot szabályozó törvények módosításáról.