

Az AI detektorok pontosságának vizsgálata: a QuillBot és a Grammarly összehasonlítása mesterséges intelligencia által generált szövegekkel¹

Sasvári Péter²

Becsült olvasási idő: 16 perc

Link: <https://doi.org/10.13140/RG.2.2.32553.20321>

Bevezetés

Az írás és tartalomgyártás világa új korszak küszöbén áll, amelyet a mesterséges intelligencia (AI) rendszerek rohamos fejlődése jellemez. Az AI technológiák nemcsak a szövegírásban, hanem az elemzésben és ellenőrzésben is megjelentek, új kihívásokat és lehetőségeket kínálva. Az AI által generált szövegek gyors terjedésével egyre nagyobb az igény olyan eszközökre, amelyek képesek felismerni és megkülönböztetni a mesterséges és emberi írásokat. Ezt a feladatot látják el az AI detektorok, amelyek az utóbbi években egyre népszerűbbek lettek az oktatásban, tudományos kutatásokban és a tartalomgyártásban.

A jelen tanulmány célja két népszerű AI detektor, a QuillBot AI Content Detector és a Grammarly AI Detector tesztelése és összehasonlítása. Bár mindkét eszköz ígéretes megoldást kínál, a pontosságuk még mindig kérdéses. A tanulmány során végzett próbaelemzés rávilágít az AI detektorok eltérő működésére és az eredmények közötti jelentős különbségekre, ami óvatosságra int a következtetések levonásakor.

Miközben az AI detektorok egyre fejlettebbek lesznek, jelenleg még nem kínálnak teljesen megbízható megoldást, különösen a kifinomult AI modellek által generált szövegek felismerésére. Ezért a jelen tanulmány során levont következtetések is csak előzetes megállapításoknak tekinthetők, és további széleskörű kutatások szükségesek a pontosságuk és hatékonyságuk megerősítéséhez.

1. Elméleti háttér

Az **AI detektor** egy olyan eszköz vagy szoftver, amely képes azonosítani, hogy egy adott szöveget mesterséges intelligencia (AI), például szöveggeneráló modellek (mint a GPT-3 vagy GPT-4), vagy ember hozott-e létre. Az ilyen detektorok segítenek megkülönböztetni az emberi írást az AI által generált tartalomtól, és különösen hasznosak lehetnek olyan területeken, mint az oktatás, a tudományos publikációk ellenőrzése, a plágiumkeresés, valamint a tartalomgyártás.

¹ Az alábbi közlemény a Nemzeti Közszerológati Egyetem Államtudományi és Nemzetközi Tanulmányok Kar gondozásában megjelenő **Államtudományi Hírlevél** Tudományos sarok rovatában jelent meg. A korábbi hírlevelek elérhetőek az alábbi honlapon keresztül: <https://antk.uni-nke.hu/kutatas-tudomanyos-let/allamtudomanyi-hirlevel/allamtudomanyi-hirlevel-2024>.

Az információt a ChatGPT nevű mesterséges intelligencia asszisztens nyújtotta, amelyet az OpenAI fejlesztett. A megadott válaszokat és adatokat Sasvári Péter ellenőrizte és kiegészítette!

² Egyetemi docens, Nemzeti Közszerológati Egyetem, Államtudományi és Nemzetközi Tanulmányok Kar, Közszervezési és Infotechnológiai Tanszék, 1083 Budapest, Ludovika tér 2. E-mail: Sasvari.Peter@uni-nke.hu

Működése:

Az AI detektorok különböző technikákat alkalmaznak, hogy felismerjék az AI-ra jellemző szövegírási mintázatokat. Ezek közé tartozik a mondat szerkezet, a szóhasználat, a stilisztikai elemek és a nyelvtani sajátosságok elemzése. Az emberi íráshoz képest az AI által generált szövegek gyakran sablonosak, ismétlődő szerkezetűek, és bizonyos statisztikai mintákat követnek, amelyeket az AI detektorok képesek felismerni.

Felhasználási területek:

1. **Oktatás:** A tanárok ellenőrizhetik, hogy a diákok dolgozatait mesterséges intelligencia írta-e, vagy saját munkájuk eredménye-e.
2. **Plágiumellenőrzés:** Tudományos dolgozatok, cikkek és egyéb írások ellenőrzése az AI által írt tartalmak kiszűrése érdekében.
3. **Tartalomgyártás:** Szerkesztők és tartalomgyártók ellenőrizhetik, hogy a cikkek természetesnek tűnnek-e, vagy túlságosan gépiesek, ami esetleg AI segítségével készült.
4. **Üzleti szövegek ellenőrzése:** Marketinganyagok, üzleti ajánlatok vagy egyéb szövegek eredetiségének ellenőrzése.

Korlátok:

Az AI detektorok még mindig fejlesztés alatt állnak, és nem mindig tudnak 100%-os pontossággal dolgozni. Előfordulhatnak hamis pozitív eredmények, amikor ember által írt szöveget tévesen AI által létrehozottként azonosítanak, illetve fordítva. A mesterséges intelligencia folyamatos fejlődésével az AI generálta szövegek egyre nehezebben felismerhetők, ezért az AI detektorokat is folyamatosan fejleszteni kell.

2. AI detektorok csoportosítása

Az AI detektorokat különböző szempontok alapján lehet csoportosítani, például a céljuk, működési módjuk és felhasználási területük alapján. Íme egy lehetséges csoportosítás:

1. Funkcionalitás szerinti csoportosítás

- **Parafrázis- és tartalomátfogalmazó eszközökbe integrált AI detektorok:** Ezek az eszközök elsősorban a szövegek átfogalmazására, stilizálására használhatók, de tartalmaznak egy AI detektor funkciót is, amely megállapítja, hogy az adott szöveg mesterséges intelligencia által generált-e.
 - o **Példa:** *QuillBot AI Content Detector* ide tartozik, mivel a QuillBot elsősorban parafráziseszköz, de tartalmaz AI detektor funkciót is.
- **Nyelvtani ellenőrző és szövegfinomító eszközökbe integrált AI detektorok:** Az ilyen eszközök elsődleges funkciója a nyelvtani hibák javítása és a szövegek stilisztikai javítása, de tartalmaznak AI által generált tartalom felismerő funkciókat is.
 - o **Példa:** *Grammarly AI Detector* ide tartozik, mivel a Grammarly elsődlegesen nyelvtani ellenőrző és szövegfinomító eszköz, amely integrált AI detektort kínál.
- **Plágiumellenőrző eszközökbe integrált AI detektorok:** Ezek az eszközök az AI által generált tartalmak azonosítása mellett a plágium felismerésére is szolgálnak, tehát

nemcsak az eredetiséget vizsgálják, hanem azt is, hogy a szöveg tartalmaz-e másolt részeket.

- **Példa:** Ilyen eszköz például a Turnitin, amely mind plágiumellenőrzést, mind AI által generált tartalom felismerést kínál.

2. Cél szerinti csoportosítás

- **Oktatási célú AI detektorok:** Az ilyen típusú eszközöket elsősorban oktatási intézményekben használják, például a diákok által beadott dolgozatok eredetiségének ellenőrzésére, és hogy kimutassák, mennyiben támaszkodtak mesterséges intelligenciára.
 - **Példa:** A *Grammarly AI Detector* és a *QuillBot AI Content Detector* is használható ilyen célra, mivel ezek az eszközök segítik a diákokat a nyelvi és stilisztikai finomításokban, miközben ellenőrzik az AI által generált tartalmat.
- **Tartalomgyártási célú AI detektorok:** Ezek az eszközök olyan felhasználóknak készülnek, akik professzionális írásokat, cikkeket vagy blogokat készítenek, és biztosítani akarják, hogy a tartalom természetesnek tűnjön, illetve elkerüljék az AI által generált részeket.
 - **Példa:** *QuillBot AI Content Detector* ide is sorolható, mivel a tartalomgyártók számára fontos a parafrázis és a tartalom hitelessége, de a *Grammarly AI Detector* szintén használható tartalomgyártási környezetben.
- **Kutatási és tudományos célú AI detektorok:** Az ilyen eszközöket a tudományos közegben használják a publikációk ellenőrzésére, hogy biztosítsák az eredetiséget és a kutatói munka hitelességét.
 - **Példa:** A *Grammarly AI Detector* különösen hasznos lehet kutatási és tudományos dokumentumok ellenőrzésénél, mivel nemcsak AI által generált szövegeket, hanem nyelvtani hibákat is felismer.

3. Működési mód szerinti csoportosítás

- **Algoritmus-alapú felismerés:** Ezek az eszközök bonyolult algoritmusokat használnak a szöveg struktúrájának, szóhasználatának, mondat szerkezetének elemzésére, hogy felismerjék az AI által írt szöveget.
 - **Példa:** Mind a *QuillBot AI Content Detector*, mind a *Grammarly AI Detector* ilyen algoritmusok alapján működik, elemzik a szöveget és meghatározzák annak valószínűségét, hogy AI által írt tartalommal van dolgunk.
- **Nyelvi mintázatok felismerése:** Az ilyen típusú AI detektorok az emberi írásra jellemző mintázatokat keresik, és összehasonlítják a mesterséges intelligencia által létrehozott szövegek szerkezetével.
 - **Példa:** *Grammarly AI Detector* kifejezetten figyel a stilisztikai és nyelvi hibákra, így hatékonyan azonosítja a nem emberi írásmódra jellemző mintákat.

3. QuillBot ismertetése

A **QuillBot** egy online szövegíró és parafrázis eszköz, amelyet mesterséges intelligencia hajt. Az eszközt elsősorban arra tervezték, hogy segítse a felhasználókat a szövegek átfogalmazásában,

írásának finomításában és a nyelvtani hibák javításában. A QuillBot különösen népszerű a diákok, kutatók, írók és tartalomkészítők körében, akik számára fontos a hatékony szövegalkotás, plágium elkerülése és a világos, tömör kifejezőmód.

A platform több funkciót is kínál, például:

- **Parafrázis:** A felhasználók több stíusból választhatnak, például formális vagy kreatív stíusból, hogy az átfogalmazott szöveg a kívánt hangvétellé legyen.
- **Nyelvtani ellenőrzés:** Segítséget nyújt a nyelvtani hibák, helyesírási hibák, valamint az írásjelek javításában.
- **Summarizer:** Összefoglalja a hosszabb szövegeket, kiemelve a lényeges pontokat.
- **Citation Generator:** Automatikusan idézést hoz létre különböző formátumokban, megkönnyítve ezzel a kutatók és diákok munkáját.

A QuillBot azért vált népszerűvé, mert jelentősen megkönnyíti a szövegalkotási folyamatot, és hatékony eszköz lehet azok számára, akik gyakran dolgoznak írásos tartalmakkal. A mesterséges intelligencián alapuló működése lehetővé teszi, hogy gyorsan és pontosan végezze el a szövegek átfogalmazását, mindezt úgy, hogy a felhasználói szándékot és a szöveg eredeti jelentését figyelembe vegye.

English French Spanish German

importance of scientific research and innovation to achieve sustainability and development. Specifically:

Innovation and Research Excellence: The grant proposal aims to explore the societal and administrative aspects of innovation within the public sector, while the document highlights the underperformance and internal bias of research in the Visegrad region, particularly Hungary. This directly links to the research excellence component of the grant, as it addresses the need for enhancing the quality and impact of research to foster sustainable public governance.

Policy Development and Knowledge Production: One of the main objectives of the grant is to create innovative policies in public administration, family, and community sustainability, which requires high-quality research inputs. The scientometric analysis underscores the need for better strategic interventions to improve research quality in the region. The grant's goals could benefit from insights on overcoming the internal biases and boosting international collaboration highlighted in the document.

Digitalization and Administrative Efficiency: Both documents stress the importance of digital tools in modern governance. The scientometric study addresses the lag in fields such as e-learning and human factors, both of which are critical in the fourth industrial revolution and align with the grant's focus on how digitalization impacts local governance and public services.

Regional Scientific Capacity: The scientometric report points out the lack of international recognition and internal biases in domestic publications in Hungary and the region, mirroring the challenges that the grant proposal intends to tackle in developing a more efficient and transparent public sector. The proposal's focus on knowledge dissemination and the utilization of digital

46%
of text is likely AI-generated

AI-generated	46%
AI-generated & AI-refined	0%
Human-written & AI-refined	0%
Human-written	54%

Enhance your writing in seconds
Try Paraphraser

1. ábra: Minta QuillBot használatára [1]

Az 1. ábrán a **QuillBot AI Content Detector** működését látjuk, amely egy mesterséges intelligencia alapú eszköz, amely képes azonosítani, hogy egy szöveg milyen mértékben lehet AI-generált. Az adott példa alapján a szöveg 46%-ban valószínűsíthetően mesterséges intelligencia által létrehozott tartalom, míg a fennmaradó része ember által írt szöveggént van azonosítva.

A bal oldalon látható szövegrészek, amelyek ki vannak emelve, azok a mesterséges intelligencia által generált szövegrészek, vagy olyan részek, amelyeket az eszköz valószínűleg AI-írásként ismer fel. Az AI Content Detector segít a felhasználóknak felismerni az AI általi tartalom létrehozásának mértékét, ami hasznos lehet például az oktatásban, kutatásban, és olyan helyzetekben, ahol a tartalom hitelessége és eredetisége kiemelten fontos.

A QuillBot ezen funkciója lehetővé teszi, hogy felhasználók, oktatók és kutatók könnyen megállapítsák, hogy egy adott szöveg mennyire tekinthető ember által készítettnek, és segíthet elkerülni a mesterségesen generált tartalom nem kívánt használatát, különösen a plágiumellenőrzés vagy hitelességvizsgálat során. Az eszköz egy egyszerű, vizuális jelzéssel (pl. százalékos arány) mutatja meg a szöveg elemzésének eredményeit, ami jól átlátható és könnyen értelmezhető.

3. Grammarly AI Detector

A **Grammarly AI Detector** egy új funkció a Grammarly platformon, amely segít azonosítani, hogy egy szöveget mesterséges intelligencia hozott-e létre, vagy emberi írás eredménye. Ez az eszköz különösen hasznos lehet az oktatási intézményekben, kutatási folyamatokban, valamint minden olyan helyzetben, ahol fontos a szöveg hitelessége és eredetisége.

A Grammarly AI Detector funkciói:

1. **Szöveg elemzése és AI azonosítás:** Az eszköz képes elemezni egy szöveg szerkezetét, stílusát és nyelvtani szerkezetét, így meghatározza, mennyire valószínű, hogy mesterséges intelligencia generálta. A rendszer különböző algoritmusokat használ, amelyek az emberi írásra jellemző mintázatokra épülnek.
2. **Az eredmény vizuális megjelenítése:** A Grammarly egyértelmű vizuális visszajelzést ad a szöveg elemzése után. Az eredményeket százalékos formában jeleníti meg, amely azt mutatja, hogy a szöveg mekkora részét érzékeli mesterséges intelligencia által létrehozottnak. Ez az információ segíti a felhasználókat abban, hogy felismerjék, mikor lehet szükség további finomításra vagy átírásra a szöveg természetesebb megjelenéséhez.
3. **Nyelvtani és stilisztikai javítás:** Az AI detektor beépül a Grammarly által kínált szövegjavító funkciókba is, így a felhasználók nemcsak az AI által írt szövegeket azonosíthatják, de az esetleges nyelvtani hibákat, stilisztikai gyengeségeket is javíthatják.

Előnyei:

- **Plágium elleni küzdelem:** Az AI detektor különösen hasznos lehet az oktatásban és a tudományos életben, ahol fontos az eredetiség és a plágium elkerülése. Az eszköz segítségével az oktatók gyorsan ellenőrizhetik, hogy egy beadott dolgozatot mesterséges intelligencia segítségével írtak-e.
- **Szerkesztők és tartalomgyártók számára:** A tartalomgyártók és szerkesztők is kihasználhatják ezt a funkciót, így megbizonyosodjanak arról, hogy az általuk készített szöveg természetesnek és emberi írásnak tűnik.
- **Az AI által generált szöveg javítása:** A Grammarly nemcsak felismeri az AI által írt szöveget, hanem segít a felhasználóknak a szöveg átdolgozásában, hogy az jobban hasonlítson egy emberi írásra. Ez egy hasznos funkció lehet, ha valaki AI-t használ szövegíráshoz, de a végeredményt még természetesebbé szeretné tenni.

Korlátok és kihívások:

- Az AI detektorok, így a Grammarly megoldása sem mindig tökéletes, és előfordulhat, hogy bizonyos esetekben emberi írást is mesterségesnek minősít. Ahogyan az AI rendszerek fejlődnek, úgy kell fejlődniük az AI-detektoroknak is, hogy lépést tartsanak az új technológiai megoldásokkal.

A Grammarly AI Detector tehát egy hasznos eszköz a modern írási folyamatok támogatásában, amely segíti a felhasználókat a szövegek eredetiségének megőrzésében és a mesterséges intelligencia által létrehozott tartalmak megfelelő azonosításában.

AI Detector by Grammarly

Navigate responsible AI use with our AI checker, trained to identify AI-generated text. A clear score shows how much of your work appears to be written with AI so you can submit it with peace of mind.

The document titled "On the Periphery of the European Social Sciences—A Scientometric Analysis of Publication Performance, Excellence, and Internal Bias in Social Sciences in the Visegrad Countries" by Sasvári and Lendvai examines the publication performance of social sciences in the Visegrad countries (Czech Republic, Hungary, Poland, and Slovakia) from 2019 to 2023. This study analyzes regional publication trends, disciplinary backlogs, journal biases, and the excellence of published work. It highlights the challenges these countries face in achieving scientific excellence, international collaboration, and the visibility of their publications in the global scientific community.

How the Document Relates to the Grant Proposal:
The analysis in this document complements the Tématerületi Kiválósági Program on "Fenntartható államkormányzás és innovatív közszolgáltatások" because both documents emphasize the importance of scientific research and innovation to achieve sustainability and development. Specifically:

Innovation and Research Excellence: The grant proposal aims to explore the societal and administrative aspects of innovation within the public sector, while the document highlights the underperformance and internal bias of research in

0% of this text appears to be AI-generated ⓘ

Go beyond AI detection

- ✓ Cite AI sources instantly
- ✓ Format citations in MLA, APA, or Chicago style in a click
- ✓ Catch accidental plagiarism

Check text Clear 431/2000 words Get Grammarly

Acknowledge AI usage with instant citations formatted in MLA, APA, or Chicago style. Get Grammarly It's free →

2. ábra: Grammarly AI Detector működés közben [2]

A képen a **Grammarly AI Detector** működését láthatjuk, amely elemzi a feltöltött szöveget annak megállapítására, hogy a tartalmat mesterséges intelligencia generálta-e vagy emberi írás eredménye. A példában a Grammarly kijelzi, hogy a feltöltött szöveg 0%-ban mesterséges intelligencia által írt, vagyis az eszköz nem észlel AI-generált tartalmat.

Az eszköz használata különösen hasznos lehet, ha biztosítani szeretnénk a szöveg hitelességét és eredetiségét, például kutatási munkák, tanulmányok vagy más dokumentumok esetében. A felhasználók egyszerűen feltölthetik vagy beilleszthetik a szöveget, és az AI Detector gyorsan elemzi, majd megmutatja az eredményt százalékos arányban. Ez segíthet elkerülni az AI általi tartalom automatikus felhasználását olyan helyzetekben, ahol fontos az emberi írás.

Továbbá, az eszköz része a Grammarly szélesebb körű szövegjavító és nyelvtani ellenőrzési funkcióinak, így a felhasználók nemcsak az AI által írt szöveget ismerhetik fel, hanem további stilisztikai és nyelvtani segítséget is kaphatnak a szövegek finomításához.

4. Összefoglalás

Az AI rendszerek és az AI detektorok megjelenése új korszak kezdetét jelzi az írás, tartalomgyártás és tudományos publikációk világában. Ahogyan a mesterséges intelligencia egyre kifinomultabbá válik, úgy nő az igény arra, hogy felismerjük és különválasszuk az emberi és AI által generált szövegeket. Az AI detektorok, mint a QuillBot AI Content Detector és a Grammarly AI Detector, fontos szerepet játszanak ebben a folyamatban.

Ez a tanulmány egy próbaelemzésen alapul, ahol egy teljes mértékben AI által generált szöveget használtunk tesztelésre. Az eredmények eltérései - 46% QuillBot és 0% Grammarly esetében - rávilágítanak arra, hogy az AI detektorok működése nem feltétlenül egyenletes, és óvatosan kell kezelni a kapott eredményeket. Bár ezek az eszközök ígéretesek a mesterséges tartalom felismerésében, jelenlegi formájukban még nem kínálnak teljesen megbízható megoldást, különösen a szofisztikált AI modellek által generált szövegek esetében.

Ezért a tanulmány végkövetkeztetéseit is fenntartással kell kezelni, hiszen ez csak egy próba volt, és az AI detektorok pontosságát további szélesebb körű kutatások során kell megerősíteni. Az eszközök fejlesztés alatt állnak, és folyamatos finomítást igényelnek, hogy lépést tudjanak tartani a mesterséges intelligencia gyors fejlődésével.

A mellékletben szereplő Nature hír [3] alátámasztja, hogy a szakterületi AI detektorok – amelyek egy adott típusú szövegre fókuszálnak – sokkal jobb eredményeket mutatnak, mint az általános célú detektorok. Az elemzés során megfigyeltük, hogy a kifejezetten tudományos cikkekhez kifejlesztett detektorok, mint például a ChatGPT-detektor, kimagasló pontossággal képesek felismerni az AI-generált tartalmat. Ez különösen fontos lehet az akadémiai publikációk világában, ahol az eredetiség és a hitelesség kulcsfontosságú.

Ez a megfigyelés megerősíti, hogy az AI felismerésére szolgáló eszközöket célszerű specifikus szakterületekre fejleszteni, mivel az általános célú detektorok csak korlátozott pontosságot nyújtanak. A mellékletben bemutatott illusztráció jól szemlélteti az általános és a szakterületi detektorok közötti különbséget, rávilágítva az utóbbiak jelentős előnyeire.

Felhasznált irodalom

- [1.] QuillBot AI Content Detector, <https://quillbot.com/ai-content-detector>
- [2.] Grammarly AI Detector, <https://www.grammarly.com/ai-detector>
- [3.] McKenzie Prillaman (2023): 'ChatGPT detector' catches AI-generated papers with unprecedented accuracy, <https://doi.org/10.1038/d41586-023-03479-4>
- [4.] Heather Desaire, Aleesa E. Chua, Min-Gyu Kim, David Hua (2023): Accurately detecting AI text when ChatGPT is told to write like a chemist, Cell Reports Physical Science, Volume 4, Issue 11101672, <https://doi.org/10.1016/j.xcrp.2023.101672>
- [5.] [2.] Heather Desaire, Aleesa E. Chua, Madeline Isom, Romana Jarosova, David Hua (2023): Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools, Cell Reports Physical Science, Volume 4, Issue 6101426, <https://doi.org/10.1016/j.xcrp.2023.101426>

További AI detektorok

1. AI Detector By Copyleaks, <https://copyleaks.com/ai-content-detector>
2. Trusted GPT-4, ChatGPT and AI Detector tool by ZeroGPT, <https://www.zerogpt.com/>

Melléklet

‘ChatGPT detektor’ példátlan pontossággal azonosítja az AI által generált cikkeket [3]

A gépi tanuláson alapuló eszköz az írási stílus jellemzőit használja fel annak megkülönböztetésére, hogy egy szöveget ember vagy mesterséges intelligencia készített-e.

Egy 2023-ban megjelent tanulmány szerint egy gépi tanuláson alapuló eszköz könnyedén felismeri, ha kémiai tárgyú cikkeket a ChatGPT nevű chatbot írta. A tanulmányt 2023. november 6-án publikálták a *Cell Reports Physical Science* folyóiratban. [4] Ez a speciális osztályozó, amely felülmúlta két másik mesterséges intelligencia (AI) detektor teljesítményét, segíthet az akadémiai kiadóknak az AI-szöveggenerátorok által készített cikkek azonosításában.

„A szövegelemzés területének nagy része általános detektort szeretne, amely bármire működik,„ mondja Heather Desaire, a Kansasi Egyetem kémikusa és a tanulmány társszerzője. *„De azzal, hogy egy adott típusú cikkekre koncentráltunk, igazán a pontosságra törekedtünk.”*

A kutatási eredmények szerint az AI-detektorok fejlesztését elősegítheti, ha a szoftvert specifikus írástípusokra szabják, mondja Desaire. *„Ha gyorsan és egyszerűen lehet valamit felépíteni, akkor nem nehéz különböző területekre is alkalmazkodni.”*

A stíluslemek

Desaire és munkatársai először 2023. júniusban írták le a ChatGPT-detektorukat, amikor a *Science* folyóirat Perspektíva cikkein alkalmazták. [5] A gépi tanulás segítségével a detektor az írási stílus 20 jellemzőjét vizsgálja, beleértve a mondat hosszúságok változatosságát, bizonyos szavak és írásjelek gyakoriságát, hogy megállapítsa, egy tudós vagy a ChatGPT írta-e a szöveget. Az eredmények azt mutatják, hogy *„egy kis funkcióhalmaz segítségével magas pontosság érhető el,„* mondja Desaire.

A 2023-as tanulmányban [5] a detektort tíz kémiai folyóirat bevezető szakaszainak elemzésére képezték ki, amelyeket az Amerikai Kémiai Társaság (ACS) publikált. A kutatócsoport azért választotta a bevezetőt, mert ez a szakasz könnyen megírható a ChatGPT számára, ha hozzáfér a háttérodalomhoz, mondja Desaire. A kutatók 100 publikált bevezetőt használtak fel ember által írt szöveggént, majd megkérték a ChatGPT-3.5 modellt, hogy írjon 200 bevezetőt az ACS folyóirat stílusában. Ebből 100 esetben a címeket adták meg a modellnek, a másik 100 esetben pedig az absztraktokat.

Amikor az eszközt emberek által írt és AI által generált bevezetőkön tesztelték ugyanazon folyóiratokban, a detektor 100%-os pontossággal azonosította a címek alapján írt ChatGPT-3.5 bevezetőket. Az absztraktok alapján készült ChatGPT-generált bevezetőik esetében a pontosság kissé alacsonyabb volt, 98%. Az eszköz ugyanolyan jól működött a ChatGPT-4 által írt szövegekkel is, az OpenAI által fejlesztett más AI-detektorokhoz képest. A ZeroGPT AI-detektor³ például 35-65% közötti pontosságot ért el, míg az OpenAI saját osztályozója csak 10-55%-os pontosságot mutatott.

³ A **ZeroGPT** egy mesterséges intelligencia (AI) detektor, amelyet kifejezetten az AI által generált szövegek felismerésére fejlesztettek ki. A célja, hogy segítsen megkülönböztetni az ember által írt tartalmat a különböző AI modellek, például a GPT (Generative Pretrained Transformer) által készített szövegektől.

Az új ChatGPT detektor még olyan bevezetőknél is jól teljesített, amelyeket nem az adott folyóiratokból képeztek ki, és még különböző promptok alapján generált AI szövegeknél is sikeres volt, beleértve azokat is, amelyek célja az AI-detektorok összezavarása volt. Azonban a rendszer rendkívül specializált tudományos folyóiratok cikkeire, és amikor valódi egyetemi újságcikkekkel találkozott, nem ismerte fel őket ember által írottként.

Szélesebb problémák

„A szerzők munkája valami lenyűgöző dolog,” mondja Debora Weber-Wulff, a berlini HTW Alkalmazott Tudományok Egyetem számítástechnikai szakértője, aki az akadémiai plágiumvizsgálattal foglalkozik. Számos meglévő eszköz megpróbálja meghatározni a szerzői jogot AI-generált szövegeknél, ahelyett hogy az írásstílus jellemzőire összpontosítana, mondja. „Sosem gondoltam volna, hogy a stilometriát⁴ alkalmaznám a ChatGPT-re.”

Weber-Wulff azonban rámutat arra is, hogy vannak más tényezők, amelyek az AI használatát hajtják az akadémiai világban. Sok kutató nyomás alatt áll, hogy gyorsan publikáljon, vagy nem tekinti a cikkírást a tudományos munka fontos részének. Az AI-felismerő eszközök nem oldják meg ezeket a problémákat, és nem tekinthetők „mágikus szoftveres megoldásnak egy társadalmi problémára”.

⁴ A **stilometria** az írott szövegek stílusának és nyelvi jellemzőinek matematikai és statisztikai elemzése. A stilometria célja, hogy a szövegek alapján azonosítsa szerzőket, illetve megállapítsa, hogy egy adott szöveget milyen stílusjegyek jellemezzék. Ezt a módszert gyakran használják szerzőazonosításra, plágiumellenőrzésre, vagy arra, hogy megkülönböztessék az emberi és az AI által írt szövegeket.