Védelem Tudomány 9. évfolyam 2. szám, 2024 április DOI: 10.61790/vt.2024.15993

A nagy nyelvi modellek alkalmazhatóságának áttekintése a katasztrófavédelmi hatósági eljárások során

Karsa Róbert
szerző
Pécsi Tudományegyetem, TTK MII tanársegéd
Email: karsar@gamma.ttk.pte.hu
ORCID: 0000-0003-0502-1508

Dr. habil. Négyesi Imre ezredes
társszerző
Nemzeti Közszolgálati Egyetem, HHK tanszékvezető
Email: negyesi.imre@uni-nke.hu
ORCID: 0000-0003-1144-1912

Absztrakt:

A mesterséges intelligencia és a gépi tanulás kutatásai az utóbbi években egyre inkább a természetes nyelvfeldolgozás (NLP) irányába mozdultak el, különösképpen a nagy nyelvi modellek (LLM) és a természetes nyelvi megértés (NLU) területén. Az NLP célja, hogy számítógépeket varázsoljunk képes különböző nyelvek megértésére és szöveges információk feldolgozására, ami forradalmi változásokat hozhat hasonlóan, mint annak idején a könyvnyomtatás vagy az internet elterjedése. A közigazgatásban jelentős előrelépések történtek a digitalizációs folyamatok terén, az online kommunikációs formák előtérbe kerülésével. Az ilyen újító technológiák, mint a GPT-3 és utódai, szövegértés, gépi fordítás és szövegek osztályozása révén javíthatják a hatósági döntéshozatalt és kommunikációt. A kutatás célja olyan módszerek kidolgozása, amelyekkel a nagy nyelvi modellek segítségével hatékonyabbá és átláthatóbbá tehetjük a közigazgatási hatósági eljárásokat és kapcsolattartási formákat. Kutatásunk során egy zárt rendszerben működő nyelvi modell létrehozására vállalkozunk, amely segítheti a katasztrófavédelmi hatósági feladatokat. Áttekintjük a nagy nyelvi modellek fejlesztését, különös figyelmet fordítva a transzformer-alapú modellek, mint a BERT és GPT alkalmazási lehetőségeire a szövegértésben és szöveggenerálásban. Bemutatjuk a közigazgatási hatósági eljárások folyamatait és azokat a pontokat, ahol a gépi tanulási módszerek hatékonyan alkalmazhatók. A kutatás során különös figyelmet fordítunk arra, hogy a nyelvi modelleket jogi szempontból is vizsgáljuk, garantálva a jogszerűség és átláthatóság megőrzését. Az eredmények alapján javaslatokat teszünk arra, hogyan lehet a nagy nyelvi modelleket hatósági eljárások keretében alkalmazni, biztosítva a hatékonyság és átláthatóság növelését a közigazgatási folyamatokban.

Kulcsszavak mesterséges intelligencia, természetes nyelvfeldolgozás, nagy nyelvi modellek, közigazgatás, ekvivalencia-elv

Abstract:

In recent years, research in artificial intelligence and machine learning has increasingly moved towards natural language processing (NLP), especially in the fields of large language models (LLM) and natural language understanding (NLU). The goal of NLP is to make computers capable of understanding different languages and processing textual information, which can bring about revolutionary changes similar to the spread of book printing or the Internet at the time. Significant progress has been made in the field of digitalization processes in public administration, with online forms of communication coming to the fore. Such innovative technologies as GPT-3 and its successors can improve official decision-making and communication through text understanding, machine translation and text classification. The aim of the research is to develop methods that can be used to make administrative official procedures and contact forms more efficient and transparent with the help of large language models. In the course of our research, we are undertaking the creation of a language model operating in a closed system, which can help the tasks of disaster protection authorities. We review the development of large language models, paying particular attention to the application possibilities of transformer-based models such as BERT and GPT in text comprehension and text generation. We present the processes of public administrative authority procedures and the points where machine learning methods can be effectively applied. During the research, we pay particular attention to examining the language models from a legal point of view, guaranteeing the preservation of legality and transparency. Based on the results, we make suggestions on how the large language models can be applied in the framework of official procedures, ensuring the increase of efficiency and transparency in public administration processes.

Keywords: artificial intelligence, natural language processing, large language models, public administration, principle of equivalence

1. BEVEZETÉS

A mesterséges intelligencia azon belül is a gépi tanulással kapcsolatos kutatások fókusza napjainkban egyre inkább a természetes nyelvfeldolgozással (NLP[1]) kapcsolatos kutatások irányába mozdul el. Az információk tárolása szöveges formában a mai napig az egyik legáltalánosabb forma. Az így felhalmozott információmennyiség a tudás igazi kincsestárának nevezhető. A nagy nyelvi modellek (LLM[2]) igazi áttörése a természetes nyelvek megértésével (NLU[3]) kapcsolatos. Az NLU az NLP egyik részterülete, amely a szövegértésre és a szemantikai elemzésre összpontosít. A nyelv megértése azt jelenti, hogy képesek vagyunk kommunikálni a számítógéppel anyanyelvünkön, azaz egy természetes nyelven, nem csak programozási nyelveken keresztül. Ez az előrelépés olyan forradalmi változásokat hozhat el az életünkben, mint annak idején a könyvnyomtatás vagy éppen az internet elterjedése. Az a tény, hogy a számítógép bizonyos szinten már megérti a természetes nyelveket teljesen új perspektívákat ad a jövőre nézve.

Az elmúlt években a közigazgatásban jelentős előrelépések történtek a digitalizációs folyamatokban. Az ügyfél és a hatóság kapcsolati tere megváltozott, előtérbe kerültek az online kommunikációs formák. A fejlesztések során hangsúlyos a hatékonyság elve[4], amely szerint fejlett technológiák alkalmazásával, a költségek minimalizálásával a tényállás tisztázására vonatkozó követelmények sérelme nélküli eljárásokat kell lefolytatni. A közelmúltban megjelent nagy nyelvi modellek, például a GPT-3 és az azt követő verziók jelentős előrelépést hoztak a természetes nyelvfeldolgozás területén. Ezek a modellek képesek olyan feladatokra, mint a szövegek generálása, szövegértés, gépi fordítás vagy a szövegek osztályozása. Az ilyen modellek alkalmazása számos területen releváns lehet, ideértve a közigazgatási hatósági eljárások támogatását is. A kutatásunk célja az, hogy olyan módszereket keressünk, amelyek során nagy nyelvi modellek segítségével hatékonyabbá, átláthatóbbá tehetjük a hatósági döntési folyamatokat és a kapcsolattartási formákat. Ennek során több típusú nagy nyelvi modellt fogok megvizsgálni és átalakítani a közigazgatási szükségleteknek megfelelően.

A védelmi szférában nagyon sok írott szöveges formában elérhető dokumentum keletkezik. Ezek a dokumentumok nagyon jelentős tudást és nagyon sok értékes információt reprezentálnak. A dokumentumok jelentős része a védelmi és biztonsági megfontolások alapján nyilvánosan nem elérhető. Ezeknek az adatoknak a feldolgozása, tudás reprezentációk kialakítása kizárólag az adatgazda által teljesen ellenőrzött körülmények között lehetséges. Nem engedélyezhető, hogy hatósági adatok az interneten elérhető pl. Chat GPT[5], GEMINI[6] vagy más hasonló, nem a hatóság által kontrollált rendszerek részére bármi is átadásra kerüljön. Amennyiben ezt a megszorítást elfogadjuk, úgy kénytelenek vagyunk saját, a védelmi szféra egyes résztvevői számára külön-külön megoldásokat létrehozni. Kutatásom során kísérletet teszek ilyen zárt rendszerben működő nyelvi modell létrehozására, elsősorban katasztrófavédelmi hatósági feladatok támogatására.

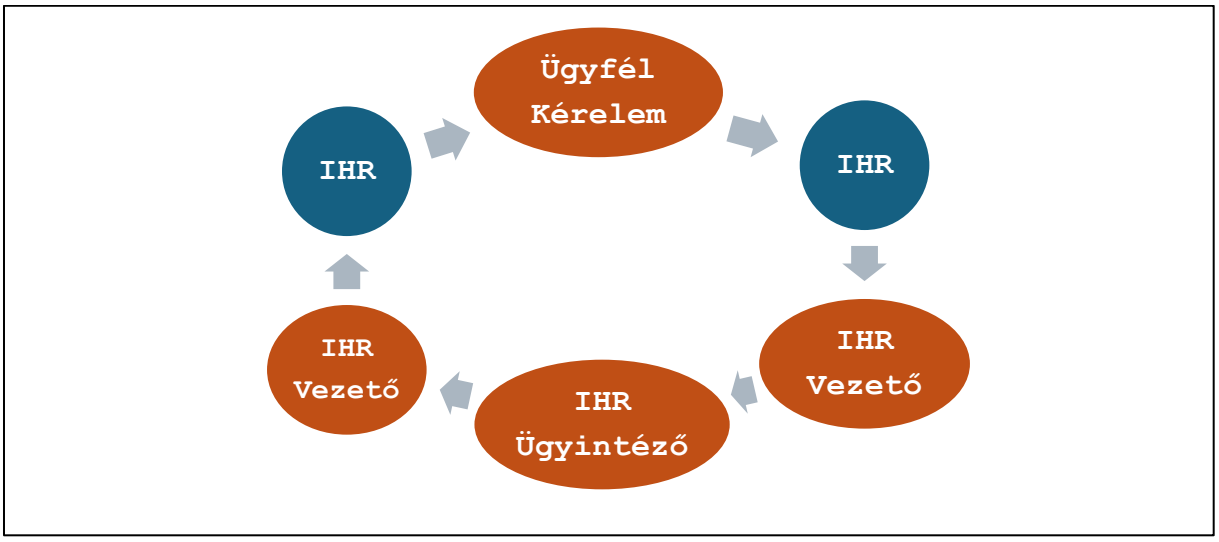
1 Natural Language Processing
2 Large Language Model
3 Natural Language Understanding
4 2016. évi CL. törvény az általános közigazgatási rendtartásról 4. §
5 https://chat.openai.com/auth/login
6 https://gemini.google.com/

2. A KÖZIGAZGATÁSI HATÓSÁGI ELJÁRÁSOK ÉS A NAGY NYELVI MODELLEK

1. Közigazgatási hatósági eljárások

Magyarországon a 2016. évi CL. törvény tartalmazza az általános közigazgatási rendtartás szabályait. A törvény rendkívül fontos szerepet tölt be a közigazgatási hatóságok működése és az állampolgárok jogainak védelme szempontjából. A törvény célja az állampolgárok jogainak és kötelezettségeinek, valamint a közigazgatási hatóságok tevékenységének meghatározása és szabályozása a közigazgatási eljárások során. A törvény részletesen szabályozza az eljárási rendet, például az ügyek kezdeményezését, a határidőket, az eljárási jogokat és kötelezettségeket. A törvény meghatározza az ügyfelek jogorvoslati lehetőségeit az esetleges jogsértésekkel szemben, előírja az állami hatóságok számára, hogy a közigazgatási eljárások során hozott döntéseiket és intézkedéseiket hozzák nyilvánosságra, és biztosítsák az állampolgárok jogát az információhoz. A jogszabály többek között az államigazgatás átláthatóságát és hatékonyságát kívánja biztosítani, továbbá a hivatalos kommunikációt is szabályozza, így segítve az ügyfelek és az államigazgatás közötti hatékony kapcsolatot.

Az elmúlt években jelentős digitalizációs folyamatok zajlanak a kormányzati tevékenységek terén. Ebben az új környezetben a szoftveres megoldások, mint IHR[7] váltják fel a hagyományos eljárásokat. Egy tanulmány [1] szerint a hatósági eljárásokhoz kapcsolódó követelményrendszer részeként számos olyan követelmény található, amelyek értelmezhetőek a szoftveres megoldások alkalmazása esetén. Fontos megjegyezni, hogy ezek nem csak ajánlások vagy jó gyakorlatok, hanem jogi előírások is lehetnek. A szoftveres megoldások hatása az átláthatóságra kétféle lehet: egyrészt növelheti azt a gyorsasága és interaktivitása révén, másrészt ronthatja azt az informatikai korlátok miatt. A tanulmány meghatározza a szoftveres megoldásokkal kapcsolatos legfontosabb közigazgatási követelményeket, különösen akkor, amikor ezeket kormányzati platformokon keresztül hatósági ügyekhez kapcsolódó ügyintézésre és kapcsolattartásra használják. A tanulmány továbbá új gondolatot hoz be, amelyet ekvivalencia-elvnek hív, ez akkor érvényesül, ha a humán ügyintéző feladatait teljes egészében, választási lehetőség nélkül helyettesítik szoftveres megoldással a platformon. A tanulmány eredményeit szeretném felhasználni a kutatásaim során, hiszen a közigazgatási hatósági eljárásokat nem csak informatikai, technológiai szempontból szükséges vizsgálni, hanem jogi szempontból is. Jelenleg az ekvivalencia-elv megtartása jelentős kihívás lehet a nagy nyelvi modellek részére.



1. kép: Közigazgatási hatósági eljárások folyamata (készítette a szerző)

[7] Integrált Hatósági Ügyviteli Rendszer

Az 1. képen látható a közigazgatási hatósági eljárások folyamatábrája. Az ábrán sötétkék színnel vannak jelölve azok a részfolyamatok, ahol csak szoftveres megoldások dolgoznak emberi beavatkozás nélkül. A teljes ciklusban elektronikus dokumentumok áramlanak az egyes fázisok között. Az ügyfél benyújtja a kérelmet az IHR rendszeren keresztül. Egy vezető szignálja és utasításokkal látja el az IHR-en keresztül az ügyintézőt. Az ügyintéző a tényállás tisztázása után elkészíti a kiadmányt és felajánlja azt a vezetőnek. A vezető amennyiben a döntés megfelelő, kiadmányozza azt az IHR-en keresztül. Végül az IHR megküldi a döntést az ügyfél részére. Ez a folyamatábra egy egyszerűsített változata a valóságnak, azonban alkalmas ara, hogy tanulmányozzuk az egyes szerepköröket és azok számára gépi tanulási módszerekkel történő támogatás nyújtását, esetleg folyamatok részbeni kiváltását.

Az eljárások során az ügyfelek által elektronikusan benyújtott dokumentumok elemzését, például a hiánypótlás szükségességének meghatározását át kell alakítani egy gépi tanulási problémára, például egy dokumentum osztályozási feladatra, majd arra megoldást kell keresni elsősorban a nagy nyelvi modelleket felhasználva.

A kutatásunk során ezt a körfolyamatot elemezzük és keressük azokat a pontokat, ahol a meglévő feladatokat át lehet alakítani gépi tanulási problémára, és meg is lehet oldani a rendelkezésre álló nagy nyelvi modellek segítségével.

A hatósági eljárások során nagyon sok írott szöveges formában elérhető dokumentum keletkezik. Ezek a dokumentumok nagyon jelentős tudást és nagyon sok értékes információt reprezentálnak. A dokumentumok jelentős része adatvédelmi és biztonsági megfontolások alapján nyilvánosan nem elérhető. Ezeknek az adatoknak a feldolgozása, tudás reprezentációk kialakítása kizárólag az adatgazda által teljesen ellenőrzött körülmények között lehetséges csak. Nem engedélyezhető, hogy ezen információkból nem a hatóság által kontrollált rendszerek részére, bármi is átadásra kerüljön. Amennyiben ezt a megszorítást elfogadjuk, úgy kénytelenek vagyunk saját, a hatóságok számára külön-külön megoldásokat létrehozni.

2. Nagy nyelvi modellek

Amikor számítógépek segítségével próbálunk meg leírni egy folyamatot, akkor modelleket készítünk, amelyek reményeink szerint egy elvárt viselkedést mutatnak. A nyelvi modell egy olyan valószínűségi eloszlás a szavak sorozatai között, ahol a modell minden egyes szóhoz valószínűségi értéket rendel egy szekvenciában, azaz a szövegben a következő szót kell előre jeleznie az előtte meglévő szavak alapján [2]. A nagy nyelvi modell már olyan típusú nyelvi modell, amelynek hátterében egy neurális hálózat van. Ezek a neurális hálózatok az információkat nagyszámú paramétereikben (számok) tárolják. A nagy nyelvi modellek esetén ezeknek a paramétereknek a száma több milliárdos nagyságrendet is elérhet.

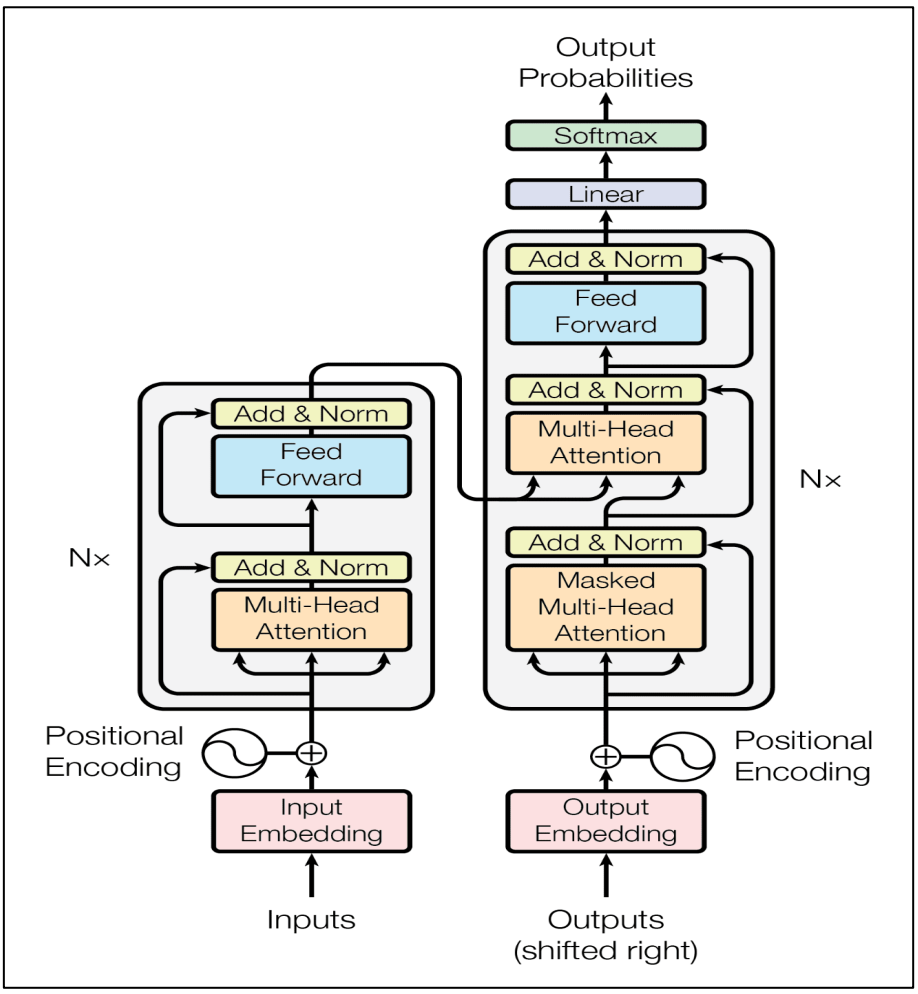
A nyelvi modellek fejlődésében a transzformer architektúra megjelenése igazi mérföldkőnek tekinthető, hiszen a ma ismert szinte összes nagy nyelvi modell ezen a technológián alapul [3]. Az ilyen típusú modellek képzéséhez nagy mennyiségű szöveges adatra van szükség.

A modern nyelvi modellek egyik közös tulajdonsága a szóbeágyazások használata, ahol szavakat vagy csak szó részleteket sokdimenziós matematikai vektorokká alakítunk át úgy, hogy a vektortérben egymáshoz közel eső szavak jelentése hasonló legyen. A szóbeágyazásnak egyik ismert megvalósítása a Word2Vec eljárás [4]. A Word2Vec mögött meghúzódó alapötlet az, hogy a szöveg korpuszban minden szót egyetlen vektor reprezentál, amely az összes kontextus (a közeli szavak) alapján kerül kiszámításra, amelyben a szó előfordul. Ennek a megközelítésnek az intuíciója az, hogy a hasonló jelentésű szavak hasonló kontextusban fordulnak elő.

A szavakat reprezentáló beágyazott vektorok nem csak a szavak jelentését, hanem a szót körülvevő szavak kontextusának a jelentését is megragadják. Amikor dokumentum osztályozást végzünk tulajdonképpen ezt a beágyazási technikát használjuk fel. A mondatokat vagy akár egész dokumentumokat alakítunk át sokdimenziós vektorokká és azokat osztályozzuk. Ez az osztályozási technika gyakran alkalmazott a hagyományos gépi tanulási eljárások során, azonban korlátot jelenthet a szavak többértelműsége és kontextus függőségük miatt.

3. A transzformer modell

2017-ben egy igazi áttörés történt a nyelvi modellek kutatása terén, megjelent az 'Attention Is All You Need' [3] a figyelem minden, amire szükség van című kutatás, és elhozta a transzformer architektúrát, amely egy teljesen új megközelítést alkalmazott. Az architektúra két részből áll: kódolóból és dekódolóból. Ez a felépítés kizárólag a figyelemmechanizmuson alapul, és mellőzi a rekurziót és a konvolúciót. Az eredményeket két gépi fordítási feladattal végzett kísérlettel támasztották alá. Az eredmények jobb minőségűek voltak, a képzési folyamat párhuzamosítható volt és jelentősen kevesebbet időt vett igénybe. Ekkor még talán a szerzők sem gondolták, hogy valódi forradalom kezdődött a nyelvi modellek terén. A figyelem minden, amire szükséged van cikk jelentősége abban áll, hogy bevezette és elterjedtté tette a figyelemalapú neurális hálózatokat először gépi fordításban, majd a dokumentumok feldolgozásában.



2. kép: Transzformer architektúra (Forrás: ld.[3])

Az eredeti transzformer architektúra képe látható a 2. képen, az egyes blokkokban különböző típusú neurális hálózati elemek találhatóak.

A kép baloldalán találjuk a transzformer kódoló modulját, a jobb oldalon pedig a dekódolóját. A kódoló modul feladata a bemeneti szöveg átalakítása kódolt vektorokká, amit a dekódoló modul visszaalakít ismét szöveggé. Egy angolról francia nyelvre történő fordítási feladat során a transzformer bemeneti, kódoló része dolgozza fel a bemeneti angol nyelvű szöveget, majd a kódoló által előállított vektorok segítségével a dekódoló egység azt feldolgozza és francia nyelvű szöveget állít elő a kimenetén.

A figyelem mechanizmust 2015-ben dokumentálták [5]. A cikk bemutatta, hogy a figyelemalapú neurális hálózatok hogyan lehetnek hatékonyak a beszédfelismerésben. A figyelem mechanizmus azonban nem csak a beszédfelismerésben, hanem a szövegek feldolgozásában is hasznos. A figyelem mechanizmusának lényege a hosszú távú (a szekvenciában egymástól távol eső) kapcsolatok kezelése és rugalmasabb feldolgozása egy szekvenciában. Míg a korábbi neurális hálózatoknak nehézséget okozott a hosszabb szekvenciák összefüggéseinek kezelése, a figyelem segítségével a modell képes figyelmet fordítani a releváns információkra és könnyebben átugorhatja a nem releváns részeket. A közigazgatási hatósági eljárásokban jellemzően hosszú szövegeket kell feldolgozni, ezért fontos szempont, hogy a modell képes legyen kezelni ezeket a dokumentumokat. A hatósági munkában megjelenő szövegek értelmezése annak jogi jellege miatt nehéz feladat, de a transzformerek képesek kezelni tetszőleges szöveget, amennyiben megfelelő mennyiségű tanítóadat érhető el. [6]

3. Kódoló típusú modellek

A 2. kép bal oldalán látható a kódoló egység, amely tulajdonképpen szövegből számokat, pontosabban vektorokat készít, amelyek reprezentálják a szöveg jelentését. Ezeknél a modelleknél az eredeti transzformer architektúrából csak a kódoló részt használják fel. A kódoló típusú nagy nyelvi modellek olyan gépi tanulási modellek, amelyek szövegkódolási feladatokra összpontosítanak, azaz szöveg-bemenetet alakítanak át egy kódolt reprezentációvá, amely a szöveg jelentését és tartalmát hordozza. Ezek a modellek többek között használatosak az automatikus szövegértelmezésben, a szöveg osztályozásban, a gépi fordításban. A modellek előnye, hogy a kódolt reprezentációk szemantikai és kontextuális információt is hordoznak, vagyis egy szöveg jelentését képesek matematikailag reprezentálni. A legismertebb és legelső ilyen architektúra a BERT [7], egy olyan típusú nagy nyelvi modell, amelyet a Google kutatói fejlesztettek ki. A BERT forradalmi előrelépést jelentett a természetes nyelvfeldolgozás terén.

A nyelvi modelleket az előzetes tanítás során hatalmas mennyiségű nyelvi adattal tréningezik. Ez a folyamat arra kényszeríti a modellt, hogy a szövegrészletekben rejlő összefüggéseket és kontextusokat megértse. A BERT modell a képzés során úgy tanul, hogy a szövegben a kihagyott (maszkolt) szavakat megpróbálja helyesen visszaállítani. Ez azt jelenti, hogy a tanulás során a „megjósolandó" szó előtti és az azt követő szavakat is látja, innen ered a kétirányú megnevezés. A modell így gazdag reprezentációkat tanulhat a szöveg különböző szintjein.

A BERT előnye, hogy kontextuális reprezentációkat készít a szavak számára, azaz a szó jelentése és reprezentációja attól függ, hogy az milyen szövegkörnyezetben található. 2019-től ezt a modellt beépítették a Google keresőjébe is. Ezt a modelltípust szöveg osztályozási, entitásfelismerési [8], szöveg összefoglalási feladatokra [9] lehet felhasználni. A közigazgatási eljárásban egy benyújtott dokumentumról a kezdeti lépésben több vizsgálatot kell elvégezni, ilyenek a hatáskör, illetékesség, teljeskörűség. Ezek a feladatok tulajdonképpen osztályozási feladatok, amely során döntéseket kell meghozni. Ezeket a döntéseket tudnánk segíteni, részben automatizálni a nagy nyelvi modellekkel.

Az entitások kinyerése alatt többek között az eljárásban résztvevő ügyfelek megállapítását is érthetjük.

4. Dekódoló típusú modellek

A 2. kép jobb oldalán látható a dekódoló egység, amely a transzformer architektúrában a kódolt reprezentációkból és a saját kimenetén megjelenő szavak, szótöredékek visszacsatolásából a saját bemenetére, állítja elő a következő szót, szótöredéket. A dekódoló típusú nagy nyelvi modellek olyan gépi tanulási modellek, amelyek kizárólag az eredeti transzformer architektúra dekódolóját használják, tehát nem rendelkeznek a kódoló által előállított információkkal, így kizárólag a saját bemenetükön megjelenő szöveget használják fel a szöveg folytatásához.

A legismertebb ilyen modell a GPT[8] nevezetű nagy nyelvi modell, amelyet az OpenAI fejlesztett ki [10]. A GPT modell tanítása a kódoló típusú modellekhez hasonlóan szintén hatalmas mennyiségű szöveges adattal történik. A képzés során a modellnek mindig a következő szót kell kitalálnia, tehát a képzés alatt a modellnek nincs információja a kitalálandó szó utáni szavakról. A modell megtanulja a nyelv mintázatait, és reprezentációkat fejleszt ki a szavak és a szövegkörnyezet közötti kapcsolatokra. A GPT megérti a kontextust és képes előrejelzéseket tenni a szövegben következő szavakra vagy mondatokra vonatkozóan.

A GPT modell generatív jellege azt jelenti, hogy képes új szöveget generálni, amely a tanítóadatokból tanult nyelvi szerkezeteket és jellemzőket követi. A GPT modellek sokoldalúak, és széles körben alkalmazhatók szöveggenerálásra, ember és gép közötti párbeszéd megvalósítására. A GPT előnye, hogy kontextuális reprezentációkat készít az előző szövegrészletek figyelembevételével és így képes értelmes és koherens szöveg generálására.

A 2018-ban megjelent GPT típusú modellek egyre nagyobb méretűek és jobbak lettek, ma a legjelentősebb és legfejlettebb nyelvi modellek közé tartoznak, és a természetes nyelvfeldolgozás (NLP) alkalmazások széles körben használják őket.

Ezt a modelltípust generatív jellege miatt elsősorban kérdések megválaszolására, hatósági döntés tervezetek generálására lehet felhasználni. A hatósági eljárásokban a kapcsolattartás az ügyfelekkel többféle módon valósulhat meg. A nagy nyelvi modellek lehetővé teszik, hogy a gépek mint virtuális ügysegédként jelenjenek meg a közigazgatásban. Ezek az ügysegédek a nap bármely szakaszában képesek lehetnek információt nyújtani egy adott üggyel kapcsolatosan. Az információ nyújtása alatt olyan készségeket értek, amelyeket a jelenlegi eljárásokban csak az ügy intézői tudnak megtenni, tehát emberszerű viselkedéssel kommunikálni természetes nyelvünkön.

5. Transzformer alapú modellek képzése

A nagy nyelvi modellek képzése rendkívül költségigényes feladat, ezért nagyon kevés cég képes arra, hogy saját modellt fejlesszen. A fejlesztési költségek a modell növekedésével (paraméter számának növekedésével) skálázódnak. A BLOOM[9] nevezetű 176 milliárd paraméterrel rendelkező nyelvi modell képzése nagyjából 118 napig tartott és 433196 kWh áramfogyasztással járt. A képzést 384 NVIDIA A100 GPU-val 48 számítási egység segítségével végezték el. [11] Magyarországon a Komondor szuperszámítógép képes csak hasonló feladatok elvégzésére. A Komondor specifikációja szerint annak GPU particiója 232 db NVIDIA A100-as GPU-val rendelkezik és a fenti BLOOM modell képzése számításaim szerint több mint 190 napig tartana, amennyiben a szuperszámítógép csak a nagy nyelvi modell képzésével foglalkozna.

[8] Generative Pre-trained Transformer
[9] BigScience Large Open-science Open-access Multilingual Language Model (BLOOM)

A nagy nyelvi modellek képzése jelenleg csak felhő környezetben valósítható meg. Egy ilyen képzést mutat be Feldmann Ádám cikke, amelyben munkatársaival a HILBERT modellt késztik el. [12]

A nagy nyelvi modellek közül több kezeli a magyar nyelvet is, azonban a magyar nyelvű szövegek aránya a teljes képzési korpuszon belül nagyon kicsi, például a Llama-2-7B modell esetén 0,003%. [13]

Hazánkban jelenleg a legnagyobb nyilvánosan elérhető és magyar nyelvű adatokon képzett modell a Nyelvtudományi Kutatóközpont által készített PULI modell család. Ezek közül az első volt a 6,7 milliárd paraméteres PULI-GPT-3SX. Ez a modell kódoló típusú. Kifejlesztésre került egy dekódoló típusú modell is PULI BERT-Large néven. A modelleket 32,4 milliárd szavas korpuszon tanították be. [14] Többek között ezeket az alap modelleket fogjuk megvizsgálni a kutatásaink során összehasonlítva, más nem magyar készítésű nyelvi modellel. Arra számítunk, hogy a magyar nyelvű modellek jóval koherensebb, nyelvtanilag jobb szövegeket fognak generálni köszönhetően a több magyar nyelvű képzési adatnak.

Meg kell említeni, hogy az OpenAI[10] által létrehozott Chat GPT[11] az egyik legfejlettebb nyelvi modell, jelenleg több változata is létezik, az első verzió a 3.5-ös verziónevet kapta és a GPT3-ra épülő finomhangolt verzió volt, 2022 decemberében adták ki. Jelenleg már elérhető a 4.5-ös modell is, amely elődjénél fejlettebb képességekkel bír, azonban az OpenAI nyilvánosan nem adott ki információkat a modell hátteréről.

3. KÖVETKEZTETÉS

Jelen áttekintésben felvázoltuk, hogy a közigazgatásban a digitalizációs folyamatok előretörése tapasztalható, és hangsúlyos a hatékonyság elve. Az online kommunikáció kiemelkedő fontosságúvá vált, és a nagy nyelvi modellek, például a GPT-3 és utódai, fontos szerepet játszanak ennek a megvalósításában. Bemutattuk a közigazgatási hatósági eljárások egyszerűsített folyamatát, a nagy nyelvi modellek fő típusait és felhasználási lehetőségeiket a hatósági feladatellátásban. A kutatás célja, hogy a hatósági folyamatok során felmerülő problémákat kezelhető osztályozási, entitásfelismerési, szöveggenerálási vagy egyéb gépi tanulási feladatokká alakítsunk át azért, hogy a közigazgatási hatósági eljárásaink hatékonyabbá átláthatóbbá váljanak. Az IHR rendszer gépi tanulási képességekkel kiegészítve elérhető lenne, hogy az ügyfél által benyújtott dokumentumokat azonnal feldolgozva és kiértékelve, még a benyújtáskor megállapításra kerüljenek a hatáskör, illetékesség, esetlegesen hiányzó dokumentációval kapcsolatos problémák. Ezeket azonnal vissza tudnánk csatolni a bejelentő felé, így segítve az ügyintézést. Azonban az elért eredmények visszaellenőrzése is feladatunk az ekvivalencia-elv mentén. Megállapítható, hogy a nagy nyelvi modellek fejlesztése terén nagyon sok szakanyag található, hiszen ennek a területnek éppen a robbanásszerű fejlődését láthatjuk. Azonban a közigazgatásban a mesterséges intelligencia használata még jórészt feltáratlan terület, a fellelhető tudás anyag nagy része inkább a bírósági döntéshozatalhoz kapcsolható. A téma kutatásában nagyon sok lehetőség rejlik és értékes, valóban használható eredmények érhetők el.

[10] https://openai.com/
[11] https://openai.com/chatgpt

5. IRODALOMJEGYZÉK

[1] B. Hohmann, „Chatbotok a kormányzati platformok szolgálatában”, *BELÜGYI SZEMLE: A BELÜGYMINISZTÉRIUM SZAKMAI TUDOMÁNYOS FOLYÓIRATA (2010-) 71 : 4,* pp. 691-709., 2023.

[2] D. Jurafsky és J. H. Martin, „Speech and Language Processing (3rd ed. draft)” 2023. [Online]. Elérhetőség: https://web.stanford.edu/~jurafsky/slp3/3.pdf (2023.11.01.)

[3] A. Vaswani, „Attention Is All You Need,” *Advances in Neural Information Processing Systems,* pp. p./pp. 5998--6008, 2017.

[4] T. Mikolov, „Efficient Estimation of Word Representations in Vector Space.,” 2013. [Online]. Elérhetőség: https://arxiv.org/pdf/1301.3781.

[5] J. Chorowski, „Attention-based models for speech recognition.,” *In Neural Information Processing Systems,* p. pp. 577–585, 2015.

[6] K. Hornik, M. Stinchcombe és W. Halbert, „Multilayer Feedforward Networks are Universal Approximators.,” *Neural Networks. Vol. 2. Pergamon Press.,* p. pp. 359–366., 1989.

[7] J. Devlin, „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.,” *arXiv preprint,* p. arXiv:1810.04805, 2018.

[8] D. Nemeskei, „Értsük meg a magyar entitásfelismerő rendszerek viselkedését!,” *XVII. Magyar Számítógépes Nyelvészeti Konferencia,,* p. pp. 409–418. , 2021.

[9] Z. G. Yang, „Automatikus összefoglaló generálás magyar nyelvre BERT modellel.,” *XVI. Magyar Számítógépes Nyelvészeti Konferencia,,* p. pp. 319–329., 2020.

[10] A. N. K. S. T. &. S. I. Radford, „Improving language understanding by generative pre-training.,” 2018.

[11] A. Luccioni, „Estimating the carbon footprint of BLOOM, a 176B parameter language model.,” *arXiv (Cornell University).,* 2022.

[12] Á. Feldmann, „HILBERT, magyar nyelvű BERT-large modell tanítása,” *XVII. Magyar Számítógépes Nyelvészeti Konferencia,* pp. pp. 29-36., 2021.

[13] H. Touvron, „Llama 2: Open foundation and Fine-Tuned chat models,” *arXiv.org,* 2023b.

[14] Z. G. a. D. Yang, „Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre,” *XIX. Hungarian Computational Linguistics Conference,* pp. 247--262, 2023.