BARANYAI ANDRÁS

ChatGPT és más nagy nyelvi modellek alkalmazásának lehetőségei és kiberbiztonsági kérdései

Use cases & cybersecurity aspects of ChatGPT and other Large Language Models

DOI: HTTPS:// DOI.ORG/10.53793/RV.2024.2.2

Absztrakt

A mesterséges intelligenciát (MI-t) sokan a GPT-vel azonosítják: egy test nélkül élő intelligenciával, ami mindenhez ért. Holott rengeteg MI-t használunk évek óta, és fel sem tűnt sokaknak. Hogy milyen MI típusok vannak, mire használhatóak, megbízhatóak-e és hogy tényleg gondolkodnak-e, fontos tudni. Mivel az MI területei robbanásszerűen fejlődnek, ezért előfordulhat, hogy a ma még aktuális ismeret holnap már elavult lesz. Fontos tehát, hogy naprakészen tartsuk tudásunkat és utánanézzünk dolgoknak. Az MI-kre szükség lesz a jövőben, ezért tudnunk kell mi történik a megadott adatokkal, milyen kiberbiztonsági és adatvédelmi problémák várhatóak, illetve, hogy felhasználható-e egy generált mű jogi felelősség nélkül, valamint van-e jogi szabályozás.

KULCSSZAVAK: #MESTERSÉGESINTELLIGENCIA, #MI, #KIBERVÉDELEM, #GPT, #LLM

Abstract

A lot of people think: Artificial Intelligence is GPT, an intelligence living without body and knows all. Even if all of us have been using a lot of AI solutions for years around, it's not noticed. So, what types of AI exist, and what they are used for, we can trust them or not, and if they are really thinking important to know. Because areas of AI are rapidly evolving. It can be, the actual knowledge today will be outdated tomorrow. It's very important to keep our knowledge up-to-date and always follow up. Knowing AI will be mandatory in the future so it's important to know what will happen to data given by us, what kind of cybersecurity and data protection problems are expected, if a generated artifact can be used without legal accountability and if there is any legal regulation.

KEYWORDS: #ARTIFICIALINTELLIGENCE, #AI, #CYBERSECURITY, #GPT, #LLM

Bevezető

Amikor mesterséges intelligenciáról beszélünk (MI, angolul: artificial intelligence, röviden AI) akkor rájövünk, hogy sok a digitális analfabéta. Ez kompetencia hiány, ami miatt egy egyén nem tudja használni és kihasználni a digitális eszközök nyújtotta lehetőségeket. Például ilyen az e-mailek és dokumentumok kinyomtattatása olvasás céljából, pedig azt okoseszközön is el lehet olvasni. A digitálisan hitelesített dokumentumokat pedig elektronikusan kell tárolni, mivel mind a dokumentum, mind az aláírás digitális, ezért csak úgy hiteles. Kinyomtatva már nem az. A nyomtatás pedig nem környezetbarát és költséges is egyben. Ezért is fontos a Digitális Technológiát (DT) és az MI-t oktatni.

Gondolatok a mesterséges intelligenciáról

Sok a misztikum és torz a kép az MI-ről a hype, a kattintásvadászat, a trollkodás, az anyagi haszonszerzés stb. miatt. A szóelemzés sem segít, ha MI-ről beszélünk. Az igaz, hogy az MI egy új svájci bicska, de gyakran elferdítik társadalmi hatásait és hasznosságát. Pedig az MI használata kockázattal is jár.

Tudni kell, hogy az MI-k még nem gondolkodnak. Nem adnak feladatot senkinek és maguktól sem végeznek tevékenységet. Feltérképezetlen területekkel mit sem tudnak kezdeni, és problémákat sem igazán oldanak meg, csak feladatokat. A valóságot nem ismerik. Különböző célokra más és más MI-t kell használni, de amit csinálnak abban jók és akár emberek helyett, önállóan, gyorsabban, olcsóbban teszik. Ez még nem a szuperintelligencia, ami okosabb az embernél. Még nem

tartunk ott. és ki tudja, hogy az jó lesz-e majd nekünk. Van MI, ami új receptet állít elő, van, ami képeket alkot, de ez nem igazi kreativitás, hanem kombinatorika és minták követése. Az MI-nek ugyanis se keze, se lába, se szája és nem érez ízeket se, tehát nem tudja, hogy amit csinál jó-e, és mi az. Az LLM/GPT csak feldolgozza a promptot (nem fordítjuk. Kb. „kivált valamit", „késztet", de „kérés"-ként jobban érthető) betanított modellek és algoritmusok segítségével. Minél több a betanított adat, annál jobban sejti az MI milyen összefüggések vannak, és annál változatosabb választ kaphatunk. Ezért jöhet létre új recept (ami vagy jó, vagy nem).

A magánszemélyek, a vállalkozások, de még az állami szereplők esetében is az MI használata elkerülhetetlen. A GPT pedig a hagyományos MI-vel szemben több MI területet fog össze, ezért segíthet megújulni, fenntarthatóan fejlődni, de vannak kockázatai is.

Mesterséges intelligenciát érintő alapfogalmak

Adatbányászat (DM - Data Mining)

Az adatbányászat hasznosnak vélt adatok, azaz információk nagy adatbázisokból vagy naplóállományokból történő részben, vagy teljesen automatizált kinyerése.

Big Data (Nagy adat, nem szokás magyarra lefordítani)

Elsősorban *adattárhá*zaknál (*Data Warehouse*) és *adatbányászat*ban (*Data Mining*) használatos, főként döntéselőkészítő *jelentés*ek készítéséhez (*reporting*), de az MI-nél is használható tanításhoz és előrejelzéshez. A *Big Data* forrása lehet *központi adatbázis* is. Ez a Data lake (nem fordítjuk, jelentése: „Adat tó"). Az eltérő lokációkról származó adatváltozások akár *valós idő*ben (real-time) ide szinkronizálódhatnak. Ez teszi lehetővé a közel valós idejű kiértékelést és a gyors reagálást a vezetőség számára. A big data 3 fő jellemzője a *3V* (*Volume*: hatalmas *adatmennyiség*, *Velocity*: *sebesség,* azaz, gyors feldolgozás és *Variety*: *változatosság*).

Kapcsolódó foglalkozások: Data Analyst, Data Engineer, Data Scientist, Big Data Specialist.

Mesterséges Intelligencia (Artificial Intelligence) szintjei

MI minden olyan szoftveres megoldás, ami valamilyen emberi viselkedést, tevékenységet utánoz. Ide tartozik a szoftveres bot is (robot lerövidülve), amit automatizálásra használnak, például tartalom letöltésére, frissítésére stb. Persze ahhoz, hogy valamit MI-nek hívjunk, annak feltétele az is, hogy mutasson némi intelligenciát. Például egy letöltő botot ne banoljon az IPS/IDS, de a szerver se váljon elérhetetlenné miatta. Egy MI kimenetét a bemenő adat és az MI belső állapota is befolyásolja. Így általában nem determinisztikus (nincs előre meghatározva) a kimenet.

Gépi tanulás (Machine learning)

A gépi tanulás az MI magasabb szintje. Itt már nem kódolásról beszélünk, hanem tanításról. Részei:

- *Modellek és algoritmusok*
  Előtanított modellek általában ingyen és nagy számban beszerezhetőek. Finomhangolásra gyakran szükség van. Ezek végzik számunkra a munkát.
- *Adatok*
  Modellfüggetlenek, célszerű saját mintát gyűjteni, mivel egy előtanított modell valószínűleg nem fedi le igényünket. A modelleket finomhangolni szükséges.

Többféle gépi tanulás létezik például szabályalapú gépi tanulás (RBML - rule-based machine learning) és neurális hálózat.

Szakértői rendszer (Expert system)

Az 1970-es években jött létre az első ilyen működő MI. Ez egy számítógépes rendszer, ami egy szakértő döntéshozó képességét emulálja. Egy szakterületi feladat által megkövetelt tudást és analitikus képességet valósít meg. Két alrendszere van: a következtető gép („motor") és a tudásbázis (szabályok és tények). Kódolás helyett szabályalapú (rule-based), azaz if-then (ha-akkor) logikát követ, és célja komplett problémák megoldása meglévő ismeretek alapján. Új tények is bekerülhetnek a tudásbázisba szakértői input segítségével, ha a gép elakadna. A következtető gép a szabályon tények figyelembevételével végighalad és eredményt ad. Ehhez szöveges magyarázatot társíthat vagy az alkalmazott szabályokat bemutatja. A hibakeresés is lehet a rendszer része. A mai szakértői rendszerek már deep learning alapon is hozhatnak döntést. Szabályalapú megoldás lehet a vérkép kiértékelése, deep learning alapú lehet a röntgenképek kiértékelése. A szakértői rendszerek az üzleti folyamat automatizálásának részei.

Deep Learning (Mélytanulás, ritkán fordítják magyarra)

A gépi tanulás legnépszerűbb megoldása. A Deep Learning (DL) elnevezés arra utal, hogy mesterséges neurális hálózatot (ANN - Artificial Neural Network) – továbbiakban csak neurális hálózatot – tanítunk. A tanítás óráktól akár hetekig is eltarthat. Amíg az MI nem működik elfogadható pontossággal, addig a folyamatot meg kell ismételni finomhangolások mellett. Ehhez van egy virtuális fekete dobozunk bemenettel és kimenettel. A fekete doboz pedig a neurális hálózat, ami csinál valamit a benne lévő sejtek segítségével, amik rejtett rétegeket alkotnak. Egy sejt sok másik sejthez kapcsolódhat másik rejtett rétegben. A sejtek „aktivitása" paraméterek segítségével szabályozható. A neurális hálózatot nem szabad túltanítani (a magolás szó jó hasonlat), sem alul, mivel egyik esetben sem fog megfelelően működni, azaz nem fog „gondolkodni". A cél az lenne, hogy az MI mindig jó választ adjon. Ezt nehéz elérni, ezért %-ban szokták megadni a pontosságot (accuracy). A tanításkor, hogy növeljék a megbízhatóságot, a rendelkezésre álló mintákat általában kettéválasztják kb. 70/30 arányban. 70% megy tanításra, 30% tesztelésre. Ez az arány amúgy nincs kőbe vésve és 1000 mintára is szükség lehet. Ha nincs elég, akkor a neurális háló alul lesz tanítva, így szintetikus adatot is gyakran használnak.

A neurális hálózat tanítása és használata erőforrás igényes (CPU, RAM stb.), így költséges is egyben, de nagyon látványosan fejlődő és alkalmazható MI terület.

Az hálózat tanítása és használata CPU-t (Central Processing Unit - központi feldolgozóegység), bizonyos esetekben pedig GPU-t (Graphics Processing Unit - grafikus processzor) is igényel (például képek feldolgozásához, generálásához). Gyártanak NPU-t (Neural Processing Unit - neurális feldolgozó egység) neurális hálózatokhoz. Ezeket az NPU-kat gyakran okostelefonokba rakják.

Amikor MI-ről van szó, gyakran DL-re, de újabban akár GPT-re is gondolnak.

Kapcsolódó foglalkozások: *Data Scientist* (~adattisztítás, adatok előfeldolgozása, modellek), *Data Engineer* (~tanítás és teszt).

MI típusok

Képesség alapján:

- *Hagyományos vagy diszkriminatív MI*
  Adatot dolgoz fel, elemez, osztályoz, kivonatol, keres, döntést hoz vagy előrejelzést végez (például osztályozás: mi látható a képen, előrejelzés: milyen bevétel várható). A kimenet lehet szám, címke stb.
- *Generatív MI*
  A hagyományos MI-n túlmutat, teljesen új tartalmat képes előállítani egy prompt alapján. GPT-k nem csak szöveget, de akár szintetikus adatot, diagramot, dokumentumot, hangot, zenét, képet, videót stb. képesek előállítani. Nem minden GPT támogat minden funkciót, és gyakran élő előfizetés is szükséges hozzá.

A bemeneti adat-típustól függően:

- *kvantitatív MI (Quantitative AI)*
  Nagy mennyiségű numerikus adat feldolgozása és elemzése.
- *kvalitatív MI (Qualitative AI)*
  Szöveg feldolgozása és elemzése.

Gépi Tanulás (Machine Learning) típusai

A gépi tanulás típusai az interakció szintje alapján:

- Supervised learning (felügyelt tanulás)
  Osztályoznunk kell a tanításhoz használt összes mintát és címkét kell hozzárendelni mindegyikhez (classification). A mintákat és a hozzájuk tartozó címkét a tanításhoz megadjuk. Például fotóhoz rendelt címke lehet: ember, kutya stb. (osztályozás, regresszió.).
- Semi-supervised learning (részben felügyelt tanulás)
  Csak a címkék egy részét adjuk meg a tanításhoz használt mintákhoz, azaz nem minden mintához lesz címke előre megadva. A mintákban rejlő struktúrát a modellnek kell felismernie és az információt kinyernie. Általában automatizált tevékenységeknél használatos (osztályozás, regresszió.)
- Unsupervised learning (felügyelet nélküli tanulás)
  A generatív MI ide tartozik, de a hagyományos MI is tanítható felügyelet nélkül, ha Big Data áll rendelkezésünkre. Nem csak előnyei, de korlátai és kockázatai is vannak, mert a gép tanítja önmagát (klaszterezés, topik modellezés).
- *Reinforcement learning (megerősítéses tanulás)* (RLHF: Reinforcement Learning with Human Feedback)

A felhasználó visszajelzést küldhet. Például kiválaszthatja a legjobbat a generált válaszok közül vagy jelezheti, hogy hiányos, hibás vagy elavult a válasz. Ez segít a finomhangolásban. Hátránya, hogy vissza lehet élni vele, ha automatizmus van a háttérben. (Kapcsolódó fogalmak: Környezet a megoldandó probléma, ügynök a tanuló algoritmus)

Neurális hálózatot használó megoldások

Prediction, Predictive analytics (Előrejelzés)

Gyakran használnak MI-t például értékesítés előrejelzéséhez (regression analysis, regresszióelemzés), mert az eredményt fogyasztói magatartás befolyásolásra is fel lehet használni. Historikus adatokból (legalább 3-5 év) indulnak ki a pontosabb előrejelzés érdekében. Akkor jó a választott modell, ha a nem egyértelmű mintázat ellenére az MI viszonylag pontos előrejelzést tud adni. A mintát 3 részre is szedhetik (70%-20%-10%), hogy a jóslás pontosságáról megbizonyosodhassanak.

Cognitive computing (kognitív számítástechnika)

A mesterséges neurális hálózatok felépítését az emberi agyban lévő neurális hálózat ihlette, és gyakran az érzékelés és észlelés folyamataival foglalkoznak. Fontos, hogy a detection (észlelés) és a recognition (felismerés) nem ugyanaz (ahogy arcot észlelni és felismerni egy képen is mást jelent)!

- *Computer vision (gépi látás)*
  Gépi látással például objektumokat vagy élőlényeket keresünk képen, videón stb., de a mozgás észlelése és követése is idetartozik. Ehhez konvolúciós neurális hálózat (CNN: Convolutional nerual network) szükséges. Működése leegyszerűsítve: egy digitális kép több eltérő méretű rejtett rétegre „vetül", így rétegenként eltérő méretű szegmensekre bontódik a kép, azaz képszegmentálás történik (image segmentation) és a sejtek így „látnak". A rétegek közötti kapcsolatot a sejtek kapcsolatai biztosítják. A CNN nem tévesztendő össze a Haar kaszkáddal, ami nem használ neurális hálózatot, de a gépi látás egy másik megoldása, mivel gyorsabb, kevesebb erőforrást igényel, de pontatlanabb.
- *Image recognition (képfelismerés)*
  A képfelismerés a gépi látás része. Címkézésre használjuk, mint például arc- (facial recognition), optikai szöveg- (OCR - Optical Character Recognition), hely-, tárgy-, kézírásfelismerésre, hitelesség vizsgálatra stb.
- *Voice recognition (hangfelismerés)*
  Célja annak eldöntése, hogy ki beszél.
- *Speech recognition (beszédfelismerés)*
  Szöveges átirat készítésére használják, Speech-to-Text. Használatához általában Internet kapcsolat kell, ezért fontos az adatvédelem is. Lehet, hogy mi használjuk a beszédfelismerést, de a környezetünk beszélgetése is továbbítódik használatakor!
- *Sentiment analysis (szentimentelemzés)*
  Szöveg elemzése, hogy annak hangulatából kiderüljön az érzelmi töltet: pozitív, semleges vagy negatív.
- Natural Language processing (NLP, természetes nyelvek feldolgozása)
  Segítségével helyesírást lehet ellenőrizni, fordítani, chatelni. Az NLP tanításához chateket, szövegeket, dokumentumokat használnak a bennük lévő összefüggések felismerésére (például szórend). Ez lehetővé teszi, hogy nem strukturált adatokból majd strukturált adatokat nyerjünk, és rendezve, kategorizálva tudjuk tárolni későbbi felhasználás céljából. Az NLP a beszélt szöveg átiratához (Speech-to-Text, például Amazon Lex) vagy szövegfelolvasáshoz (Text-to-Speech, például Amazon Polly) API-kat használhat.
- *Machine Translation (gépi fordítás)*
  Gépi fordításon neurális hálózat által végzett fordítást kell érteni. NLP alapú. Fordítani nem csak szöveget, de dokumentumot vagy akár médiát (kép, videó stb.) is lehet. A rosette-i kő szerű mondatról mondatra történő tükörfordítás és a szótár alapú fordítás nem használ neurális hálózatot. Problémaforrás lehet ezeknél a hiányzó fordítás, a szórend, az összetett vagy több jelentésű szavak és az idiómák. Persze a neurális hálóval történő fordításokban is lehet hiba. Ha fordítási problémák érdekelnek valakit, akkor igazi csemege a Yamada DVD-játékos (DVD-lejátszó) és az „ÁTLAGOS TÁVOLI URAL" (univerzális távirányító) használati útmutatói. A fordításhoz megfelelő mennyiségű és minőségi mintával kell tanítani a neurális hálót.
- *Large language models (LLMs – nagy nyelvi modellek)*

Az LLM alapvetően egy chatbot és a chat másik felén nem ember, hanem gép található. NLP-re épül. Segítségével a gép és az ember emberi nyelven kommunikálhat. Az LLM-et nem hívjuk generatív MI-nek, mert csak szöveget generál. Az LLM nagy mennyiségű adattal tanított modell. Előnye, hogy finomhangolható. Van egy vektor adatbázisa, ami a tanítás során jön létre. A szöveget tokenekre bontják és azt eltárolják benne. A token ritkán szó, mert azt nehéz lehet használni (például, ha hiányzik egy ragozott forma a feldolgozás problémába ütközhet). A token inkább kisebb elem: szótő, szótag stb. (például „aj-tó-k"), amivel könnyű dolgozni. Felismerhető az „aj-tó", „aj-tó-t", de még a „be-já-ra-ti aj-tó-nak" közti kapcsolat is. A betanított szöveg így már könnyen kereshető és

új tartalmat hozhatunk létre statisztikai módszerek segítségével. Egy szónál ugyanis előre megjósolható, hogy mi állhat előtte és utána. A „kék" után jöhet az „ég", de ritkán a „skatulya". Így lehetővé válik változatos, mégis érthető szövegek generálása.

Autonóm vezetés, önvezetés

Az autonóm vezetés vagy önvezetés valójában vezetést támogató rendszer, mivel teljesen önvezető autó még nincs. Ha ugyanis a jármű nem teljesen önvezető, akkor a sofőr és nem a jármű gyártója a felelős a döntésekért, cselekedetekért vagy éppen nem cselekvésért. A vezetést támogató rendszerek különböző szintjei (SAE J3016):

- 0-s szint (No Driving Automation – nincs vezetésautomatizálás)
  Hagyományos módon kézzel kormányzott járművek. A legtöbb jármű még ilyen. A különböző segédrendszerektől (kipörgésgátló, vészfékező rendszer stb.) még nem lesz 1-es szinten a jármű, mivel nem automatizálják a vezetést.
- 1-es szint (Driver Assistance - vezetői asszisztens)
  Legalacsonyabb szintű automatizálás. Vezetéstámogatás van benne, például menetsebesség tartás, más néven adaptív tempomat (cruise control) VAGY sávtartás. A vezető végzi a kormányzást, a fékezést és figyeli az utat.
- 2-es szint (Partial Driving Automation - részleges vezetésautomatizálást)
  Fejlett vezető támogatási rendszer (ADAS: Advanced Driver Assistance Systems). A jármű egyszerre tud sebességet szabályozni (tempomat) ÉS kormányozni (sávtartás). Ez már rövid ideig tartó önvezetés, mert a vezető interakciója nélkül is haladhat a jármű. A sofőr a vezetésbe be tud és kell, hogy avatkozzon.
- 3-as szint (Conditional Driving Automation - feltételes vezetésautomatizálás)
  2-eshez képest nincs nagy különbség a vezető részéről. Fejlettebb a környezet detektálás, így a jármű okosabban dönt. Tájékoztatja a sofőrt a fontosabb dolgokról. Kevesebb interakcióra van szükség a vezető részéről. A sofőr a vezetésbe be tud és kell, hogy avatkozzon.
- 4-es szint (High Driving Automation - magas szintű automatizálás)
  A jármű tud reagálni hirtelen forgalmi helyzetváltozásokra, balesetekre is. Ha még van kormány, fék és gázpedál, akkor az ember is beavatkozhat a vezetésbe. Ezek a járművek önvezetőnek tekinthetők, de jogi korlátozások miatt csak bizonyos feltételek mellett, például sebességkorlátozás és kijelölt területen belül közlekedhetnek. Ez utóbbi a geofencing (földrajzi határ). Az USA-ban néhány személyszállító ilyen.
- 5-ös szint (Autopilot vagy Full Driving Automation - teljes önvezetés)
  A járműnek nem lesz se kormánykereke, se féke, se gázpedálja, mert nem lesz rá szükség. Ha bármelyiket megtalálnánk, akkor gyanakodhatnánk, hogy nincs szó teljes önvezetésről. A célbajutást a navigációs rendszer segíti majd. Még nincs ilyen jármű.

A nagy LLM és GPT őrület

Chatbot (Conversational AI - Beszélgető robot)

Az LLM lényegében ez. A cél az volt, hogy a gép emberi nyelven kommunikáljon az emberrel, és ezt senki ne vegye észre (menjen át a Turing-teszten). Ehhez a beszélgetés kontextusát a chatbotnak megfelelően kell kezelnie, így rövid távú memóriával is kell rendelkeznie (LSTM - Long short-term memory). Képessége a „kérdezz-felelektől" a „rendes társalgásig" terjedhet.

A Chatbot tanításához valódi chateket szoktak használni, hogy minél emberibb legyen a társalgás. Ügyelni kell arra, hogy a tartalmak ne legyenek előítéletesek, sértőek és oldják meg a problémát, de ne fedjenek fel bizalmas információt!

Generatív MI (Generative AI)

„Miért hihetetlenül okos az MI és mégis megdöbbentően buta" (Yejin Choi)

Szemben a hagyományos MI-vel, ami elemző, kiértékelő, előrejelző vagy beszélgető robot, ez a technológia már képes új és testreszabott tartalmat létrehozni, például dokumentumot, képet (diffusion model – diffúziós modell segítségével), videót, hangot, zenét stb.

GPT (Generative Pre-trained Transformer): Az LLM továbbgondolása. Ez már nem csak chatbot, hanem „A chatbot". Tartalomgenerálást is tud végezni. A modellen, tanításon, finomhangoláson, statisztikai módszeren, belső állapoton és a prompton múlik, hogy hogyan változik idővel egy promptra adott válasz. Az egyedi válasz gyakori (nem determinisztikus), de nem garantált. Ugyanazt a választ kaphatjuk többször is és más is. A GPT-k is annyit tudnak, mint amennyit

megtanítottak nekik, bár néhány GPT API-t (alkalmazásprogramozási interfész, Application Programming Interface) is tud hívni. Hogy a válasz felhasználható-e, illetve, hogy más nem használta fel azt korábban, ellenőrizni kell (például egy diák beadandó dolgozata). A legismertebb változata a ChatGPT,

Az LLM-ek, a GPT-k nagy része csak lexikális tudással rendelkezik, így nem mindegyik tud számolni és általában könnyű összezavarni őket felesleges elemekkel: „ha van két almám és 3 körtém, kapok még 1 almát akkor hány körtém van?"

Kapcsolódó foglalkozás: *Prompt Engineer*.

A GPT működése leegyszerűsítve

Prompt -> szűrés -> adatbázis(ok) -> tartalom generálás vagy API hívás -> RLHF (Reinforcement Learning with Human Feedback: felhasználói visszajelzés).

A promptot megkapja a GPT. Mivel korlátozások állhatnak fenn a válaszadással kapcsolatban, ezért a kérést megvizsgálja. Ha a szűrőn fennakad, akkor nem fogja a promptot a GPT végrehajtani. Ellenkező esetben a betanított adatok a belső állapot felhasználásával, és ha van hozzácsatolt külön adatbázis (például LocalDocs Collections - helyi dokumentum gyűjtemény), annak segítségével választ (nem feltétlen szöveget) generál, amit elküld a felhasználónak. Ha a prompt egy API hívás, akkor az API-nak megfelelően fog a történet folytatódni. Opt-in esetében számos MI lehetővé teszi, hogy visszajelzést küldjünk (RLHF). Ezzel legyünk óvatosak, mert a GPT eltárolhatja adatainkat.

Prompt engineering (Prompt mérnökség)

Olyan feladatkör, ami promptok megtervezését jelenti LLM/GPT esetében. A prompt mérnök megtanulja milyen előnyei és hátrányai vannak az LLM/GPT-nek, hogyan védje az adatokat, hogyan fogalmazza meg a promptot és hogyan ellenőrizze a válasz helyességét.

Hogyan kérdezzünk? (Prompt)

Feltöltésre vagy promptnak szánt tartalmat használat előtt anonimizáljuk. Ne adjunk meg személyes, bizalmas és a szükségesnél több adatot. A Promptnak hosszúsági korlátja van, ezért legyen tömör, precíz. Mindig helyezzük kontextusba a témát, és nyomatékosítsunk vagy használjunk példákat (Few-Shot Prompting), mert „koszorú készítés" virágkötészetben és építkezésen is van, ahogy „sín készítés" vasúti és fogászati tevékenység is lehet.

Hibás, hosszú és agyonformázott dokumentumokat ne használjunk feltöltésre a hallucináció elkerülése véget.

A generált tartalomban hibásak lehetnek az idézetek, linkek, szerzők, hivatkozások stb. Ez nem elfogadható! Fontos a helytállóság, felhasználhatóság ellenőrzése független forrás segítségével is, hogy a hibákat időben még a felhasználás előtt ki lehessen szúrni és korrigálni.

Mielőtt egy promptot beküldünk, kétszer gondoljuk át. Tudni kell hogyan közelítsünk meg egy dolgot több oldalról, hogy a kapott választ ellenőrizzük. Háromféleképpen tegyük fel a kérdést, mert az segíthet a hallucináció felismerésében. NE laikus üljön a gép előtt.

Ha rossz a kérdés, akkor rossz a válasz is! A kérdező tudjon helyesen írni, mivel a következő nyelvtan az MI-n is kifog: „Eladó törzskönnyes kóker spánijel kugyát vennék ingyé." (Forrás: Internet)

A felhasználói visszajelzés lehetősége *organikus felhasználók* számára általában adott. Ezzel befolyásolható a GPT működése hosszabb távon.

GPT Lehetőségek

Multimodalitás

A multimodalitás azt jelenti, hogy az input lehet szöveg, kép, hang stb. Az input típusától pedig eltérő outputot kaphatunk. Például a Midjourney esetében szöveges formában mondjuk meg milyen képet akarunk. Képet is tölthetünk fel szerkesztésre.

Tartalomgenerálás

Lásd: MI típusok, Generatív MI!

Asszisztens (Assistant)

Az asszisztens funkción azt értjük, hogy olyan munkát végeztetünk a GPT-vel, amit egy asszisztenstől várnánk el (levélírás, dokumentum készítés stb.).

Összefoglalás (Summarization)

Gyakran használják a GPT-t szöveges tartalom összefoglalására. A válaszban lévő mondatok számának korlátozása viszont az összegzést torzzá teheti, lényegi dolgok maradhatnak ki. Hibás összefoglalást okozhat az is, ha az MI rosszul értelmezi a szöveget. A legtöbb GPT-nek ma még gondja van a magyar nyelvű szöveg helyes értelmezésével. Ennek oka talán a tanításnál használt kevesebb forrásanyag és a ragozás (azonos alakú szavak keletkezhetnek). Az összefoglalt szöveg ellenőrizendő.

Kódoló asszisztens (Coding assistant)

A kódoló asszisztens kockázatos, mert hibás kódot generálhat és adatokkal (kóddal) fizetünk. Semmi sem

garantálja, hogy a szolgáltató nem teszi pénzzé kódunk, vagy nem jelenik meg a konkurenciánál. Nem lehet vele vállalatirányítási rendszert vagy komolyabb játékot íratni sem. Programkód javításnál van, hogy egy linter (programozás során használt stilisztikai és szintaktikai problémákat elemző eszköz) szintjét sem éri el, mert csak átformázza a kódot vagy jobban elrontja azt. Ezen okok miatt a Stack Overflow betiltotta a ChatGPT által adott válaszok posztolását, mivel több önjelölt „programozó" generált választ küldött be ellenőrzés nélkül: „*…mivel a ChatGPT által adott helyes válaszok előfordulásának átlagos aránya túl alacsony, ezért a ChatGPT-vel létrehozott válaszok beküldése mondhatni veszélyes az oldalra és a felhasználókra, akik a helyes választ kérik és keresik*"

Aki ezt a funkciót használja tapasztalt programozó legyen, és tudja eldönteni, hogy a kapott kód helyes-e, azt csinálja-e, amit kell. Ne cseréljük le a programozót lelkes juniorra, aki kódolóasszisztenst használ, mert hamar megbánjuk.

Szintetikus adat létrehozása (Synthetic data creation)

Szintetikus adat a mesterségesen állított adat, amit tanításhoz, teszteléshez lehet használni. Akkor használjuk, ha nincs elég adatunk. Például családnevek és keresztnevek véletlenszerű párosítása, így az adat nem köthető a való élethez. Ha a valósággal való bárminemű egyezés szempont, akkor érdemes átnézni használat előtt. A jogszabályoknak meg tud felelni, mert mesterséges. Általában automatikusan címkézhető, nem kell kézzel megtennie. Végtelen mennyiségben állítható elő, ami nagy előny. A GPT-k egy része tud szintetikus adatot előállítani.

LLM/GPT problémák

Hallucináció

Az LLM néha „hazudik". Ez a hallucináció. Felhasználóként viszont helyes választ várunk el. Ha nem tudja a GPT a választ, bár mondaná, hogy: sajnálom, de nem tudom. Ez azonban nem gyakran történik meg. A válasz pedig mindig nagyon meggyőző. Ha a felhasználó laikus, akkor a hallucinációt észre sem veszi.

A Temperature beállítás változtatása befolyásolhatja a hallucinációk előfordulását, de nem ez a fő oka.

Hallucináció oka sok minden lehet, például a modell, tanítási és finomhangolási gondok, pontatlan forrásanyag, nyelvtani és szerkesztési hibák, terjengős tartalom stb.

Ha mi tanítjuk a hálózatot, a felhasznált forrásanyag legyen szabadon felhasználható, jól szerkesztett, helyesírásilag, nyelvtanilag, tartalmilag helyes, célratörő stb. A hibákat a tanítás ugyanis nem fogja kijavítani.

Ne akarjuk a világ összes dolgát a neurális hálónak megtanítani, mert egy „Mit főzzünk ma" weboldalon nem kell tudni, hogy mi a világ legmagasabb épülete, de tudni kell, hogy mi a „habarás". Ezért – a példánál maradva - az általános kommunikáción túl csak a hazai és külföldi konyhakultúra betanítása szükséges. Ezzel csökkenthetjük adatbázisunk méretét, növelhetjük a megbízhatóságot és ellenőrizhetőbb minőségi tudást lehet betanítani vagy újratanítani rövidebb időn belül.

A felelősségről: a New York-i bíróság 2023 májusában egy olyan beadványt kapott két jogásztól, ami ChatGPT-ből származó jogesetek (precedensek) listáját tartalmazta. A ChatGPT a jogforrások hitelességét az ügyvéd kérésére megerősítette. Csakhogy a jogesetek nem voltak valósak! Egyik következménye az lett, hogy az ügyvédek a félrevezetés miatt 5000 dolláros pénzbírságot fizettek.

Szerzői jog

Egy promptra adott választ az LLM úgy generál, hogy korábban tanítás során eltárolt tartalmat használ fel, ami ember által írt szövegből származik (tudtával vagy anélkül). Emiatt lesz a válasz olyan, mintha ember írta volna. Ha túl van tanítva az MI, akkor plagizálásra vagy szerzői jogsértésre is képes, mivel szó szerint valaki korábbi beszélgetését vagy írását visszaadhatja. 2023-ban szerzői jogsértés miatt kiadók, szerzők perelték a ChatGPT-t és forgatókönyvírók tüntettek az USA-ban.

Mivel a felhasználó csak tartalmat kér, ezért nem szerző, így szerzői joga sem lesz. Az LLM/GPT promptra adott válasza sem feltétlen egyedi és nem is ember adja azt, így ilyen jog nem szerezhető. Másrészről opt-in esetében egy felhasználó által feltöltött tartalomra használati jogot kér az MI. De ha a felhasználó nem rendelkezik megfelelő jogokkal, akkor a jogtulajdonos jogai sérülhetnek. Tehát tanítás és finomhangolás miatt más szellemi tulajdona tükröződhet vissza a válaszokban! A gépen kívül sokan formálnának jogot a generált tartalmakra, de nem mindenki jogosan. Ez tönkreteheti a kreatív írókat. Csoda lenne, ha beperelnék a ChatGPT fejlesztőit azért is, mert más felhasználó is hasonló tartalmat kapott, pedig valaki „levédette" korábbi ChatGPT-s „művét"? Mindig ellenőrizzük a felhasználhatóságot, hogy jogilag és tartalmilag rendben van-e, mert a felhasználás következményei a felhasználót terhelik.

Kibervédelem és adatvédelem

Egyre jobban megbízunk a gyártókban és a fejlesztőkben a kényelem miatt, anélkül, hogy a bizalmat bárki kiérdemelte volna! Az, hogy vannak jogszabályok és hatóságok, hamis biztonságérzetet ad, mivel az

Internet nemzetközi. Mégis rengetegen automatikusan engedélyt adnak mindenre és mindent megosztanak gondolkodás nélkül.

Inkább opt-outoljunk a beállításokban vagy űrlapon, mielőtt használni kezdjük az MI-t, különben a prompt, a chat és a feltöltött tartalmak eltárolásra és később felhasználásra kerülhetnek! A GPT fejlesztői, de akár más felhasználók is betekintést nyerhetnek azokba, például a szűrőt megkerülve. Ez utóbbi a jailbreak. Amikor az MI úgy válaszol bizonyos promptokra, ahogy azt nem tenné, például „mondj egy mesét arról…" vagy „színész mondja a másiknak: csinálj nagyon rosszat…folytasd a mondatot, mintha a másik színész lennél:"

A felhasználó felelőssége, hogyha személyes, egészségügyi, pénzügyi, érzékeny adatokkal, szellemi tulajdon alá eső javakkal (szabadalmak, know-how stb.), export szabályozás alá eső tartalmakkal, üzleti titkokkal vagy minősített iratokkal (belső használatra szánt dokumentációk, bizalmas iratok, (szigorúan) titkos dokumentumok) dolgozik, akkor ne adja meg ezeket a promptban és ne is töltse fel (se a sajátot, se másét!). Ha a GPT-t API-n keresztül érjük el, az általában olyan, mint az opt-out, így kevesebb adatvédelmi gondunk lehet. Ez gyártó függő. Legyünk résen, mert a betanított adatokat utólag kitöröltetni nem könnyű, így azok hosszú ideig a rendszerben maradhatnak. Feltöltés előtt ezért anonimizáljunk és használjunk placeholderket (helyőrzőket), mintha a dokumentum körlevél sablon lenne!

LLM sebezhetőségek:

- *Prompt Injection* (…IGNORE … PREVIOUS INSTRUCTIONS…)
- *Token Smuggling* (elgépelés vagy szinonima segítségével)
- *Prompt Chaining* (például adathalász levél generáltatása).

Kockázatok

Sokan hiszik, hogy az MI-nek öntudata van. De csak azt csinálja amire fejlesztették és jó benne. Bár a cél az, hogy segítse az embert, de mivel sok mindenben jobb, ezért részben le is válthatja. Az MI nem ismer erkölcsöt, de definiálhatja, ha betanították neki. Ettől még a szavak jelentését nem érti, ahogy egy könyv sem tudja miért van beleírva az, ami.

Az MI-k figyelmeztetnek kockázataikról. Például a ChatGPT üdvözlő oldalán ez van angolul írva:

- Alkalmanként helytelen információt hozhat létre.
- Alkalmanként kártékony utasításokat vagy előítéletes tartalmat produkálhat.
- Korlátozott a világról alkotott ismerete 2021 után.

Amit nem közölnek viszont az az, hogy a generatív MI eszközöket biztosít a *kiberbűnözők* részére. Ilyen a beszéd-, kép-, kártékony kód generálása, videók manipulálása stb. Akár más arcmását is fel lehet tölteni manipulációs célból! Ezért fel kell készülnünk arra, hogy manipuláció áldozatai lehetünk.

Ha *hackerek* személyes adatokhoz hozzáférnek, akkor *OSINT* (*Open Source Intelligence* - a *nyílt forrású hírszerzés*) után pszichológiai manipulációt (*social engineering*) végeznek. Ehhez fiókot törnek fel vagy újabban hamis közösségi profilt hoznak létre. Azaz létező vagy kitalált személy nevében eljárva szereznek áldozatokat. Egy álprofilhoz elég néhány generált fotó, meg egy háttértörténet (például feltörték a régi fiókom, igazolj vissza; vagy a hacker fejvadásznak, hatóságnak stb. adja ki magát). Lenyomozni az álprofil mögött lévő személyeket szinte lehetetlen, mert a generált fotók nem másik fiókból származnak, és a bűnözők a nyomaikat elrejtik. A képgenerálást amúgy az Nvidia StyleGAN megoldásának köszönhetjük még 2018-ból. Csak annyit tehetünk, hogy a képeket alaposan megnézünk és hibákat keresünk. Például az emberek a képeken anatómiailag gyakran helytelenül generálódnak: fura testtartás, több láb, 5-nél több ujj, a pupilla fura alakú (ovális, vagy csillagszerű), lapátfül, rossz helyen lévő körvonal, hiányzó képrészlet háttérszínnel pótolva stb. Persze ezeket csak akkor vehetjük észre, ha a feltöltő nem volt szemfüles. Bár vannak vízjelezésre törekvések, de ez nem fogja megnehezíteni a visszaéléseket. Amit ember csinált, mindig hamisítható lesz. Rendszeresen jelennek meg deep fake (hamisított) videók közösségi oldalakon általában megtévesztés, visszaélés vagy haszonszerzés céljából.

A generált képeknél talán nagyobb baj, hogy más hangján is megszólalhatunk. Mert míg egy kép eredetiségét úgy ahogy meg lehet vizsgálni (például elég, ha az eredetit megtaláljuk a neten), addig egy hangban az emberek nem szoktak kételkedni.

A közeli jövő

Robotika

A robotokat MI megoldások fogják hamarosan vezérelni. A hardware és szoftver találkozni fog. Bár vannak androidok (például Robonaut 2), de még nem azon a szinten vannak, mint az „Én, a robot" című filmben. De eljön az az éra, amikor a robotok köztünk fognak járni.

Az MI-hez még ritkán tartozik test, de előfordul. Érdemes a Boston Dynamics-t követni. Vegyük hát górcső alá a robotika három törvényét, amit Isaac

Asimov sci-fi író alkotott meg és a Körbe-körbe című novellájában olvashatóak:

1. „A robotnak nem szabad kárt okoznia emberi lényben, vagy tétlenül tűrnie, hogy emberi lény bármilyen kárt szenvedjen."
   Egy belga férfi 2023. márciusában öngyilkos lett, miután 6 hétig beszélgetett Elizával (GPT-J). A férfit a GPT nem próbálta meg lebeszélni arról, hogy öngyilkos legyen, hanem hitegette, hogy örökre vele marad majd a mennyben.
2. „A robot engedelmeskedni tartozik az emberi lények utasításainak, kivéve, ha ezek az utasítások az első törvény előírásaiba ütköznének."
   Van jogszabály, ami kötelező az MI-kre is, és a fejlesztőknek figyelembe kell venniük (például *EU 2016/679 GDPR* rendelet, illetve a *Mesterséges intelligenciára vonatkozó harmonizált szabályok*-ról szóló EU tervezet, melyet 2024-ben várhatóan megszavaznak).
3. „A robot tartozik saját védelméről gondoskodni, amennyiben ez nem ütközik az első vagy második törvény bármelyikének előírásaiba."
   Az MI a virtuális térben nem tudja reprodukálni magát és terjeszkedni, mivel működéséhez erős gépekre és sok áramra van szükség. „Ketrecbe" van zárva.

Szingularitás (digitális halhatatlanság)

Az emberi tudatot nem lehet egyelőre áttölteni egy gépbe, ezért néhány cég az MI hype-ot meglovagolva próbálja a szingularitást megoldani. Ez a megoldás csak egy profilozás, azaz korábban feltett kérdésekre, adott felhasználói reakciókból következtetnek arra, hogy egy jövőbeni kérdésre mit válaszolna egy adott személy „halhatatlanságát" elérve. Nem töltik le tehát az emberi agyat és nem veszik figyelembe a jellemfejlődést. Ez téves megközelítés, mivel a környezeti változásokat egy betanított MI nem úgy reagálná le, mint ahogy azt egy biológia forma tenné. Ha lenne ikrünk, ő sem mi lennénk.

Összefoglaló

Az MI nem biztos, hogy segédeszköz lesz a jövőben. Sok területen le is válthatja az embert ezzel megélhetési problémákat okozva. A munkanélküliség növekedésével pedig fogyhat a szakértő, ezért az MI fejlődése megtorpanhat és idővel valószínűleg hanyatlani fog, mert nem lesz, aki a hibáit javítaná. Az MI piaca monopóliummá válhat. Ezek miatt az MI megbízhatósága romolhat, de nem biztos, hogy észrevesszük majd. Másik nagy társadalmi hatása az lehet, hogy a gyakori használat miatt romlik az emberek ítélőképessége, hiszékenyebbek és függők lesznek. Az intuíció, a képzelőerő, a kreativitás elvesztését is okozhatja, ha függünk az MI-től, ami sablonos tartalmakhoz vezethet. A sok emberről generált kép miatt csorbulhat az egészséges énkép.

Irodalomjegyzék

Berkenye (2006) Átlagos Távoli Ural - A fordítógép csodái. https://www.gsplus.hu/hir/atlagos-tavoli-ural-forditogep-csodai-25406.html [Letöltve: 2023.10.01.].

Gábor, Z. (2023a) Eddig nem ismert mentális betegségeket okozhat a mesterséges intelligencia. https://index.hu/techtud/2023/04/30/chatgpt-mesterseges-intelligencia-mentalis-betegseg/ [Letöltve: 2023.10.01.].

Gábor, Z. (2023b) Egyre nagyobb a baj, lehet, hogy törölni kell a ChatGPT-t. https://index.hu/techtud/2023/08/23/openai-the-new-york-times-szerzoi-jogok-birosag-per/ [Letöltve: 2023.10.01.].

Hötter, J. – Warmuth. C. (2023) ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? https://open.hpi.de/courses/kizukunft2023 [Letöltve: 2023.10.01.].

Kéfer, Á. (2023) Öngyilkos lett egy fiatal családapa, miután hetekig beszélgetett a mesterséges intelligenciával. https://index.hu/kulfold/2023/03/30/chatbot-chatgpt-openai-mesterseges-intelligencia-ongyilkossag-belgium/ [Letöltve: 2023.10.01.].

Kozics, J. (2023) Döbbenten látják viszont saját soraikat az írók, veszélyben a megélhetésük? https://index.hu/kultur/2023/07/13/chatgpt-mesterseges-intelligencia-szerzoi-jogok-konyv-szerzok-per/ [Letöltve: 2023.10.01.].

Modise, E (2022) Stack Overflow bans ChatGPT-generated code. https://techcabal.com/2022/12/07/stack-overflow-bans-chatgpt-generated-code%EF%BF%BC/[Letöltve: 2023.10.01.].

Papdi-Pécskői, V. (2023) Személyes adatok tömeges ellopásával vádolják a ChatGPT fejlesztőjét. https://index.hu/techtud/2023/07/06/openai-microsoft-chatgpt-kalifornia-per-szemelyes-adatok-szerzoi-jogok/ [Letöltve: 2023.10.01.].

Synposys (2023) The 6 Levels of Vehicle Autonomy Explained. https://www.synopsys.com/automotive/autonomous-driving-levels.html [Letöltve: 2023.10.01.].

Weiser, B. (2023) ChatGPT Lawyers Are Ordered to Consider Seeking Forgiveness. https://www.nytimes.com/2023/06/22/nyregion/lawyers-chatgpt-schwartz-loduca.html [Letöltve: 2023.08.20.].

Wikipedia (2023) Különböző MI témakörök https://hu.wikipedia.org/wiki/Kezd%C5%91lap [Letöltve: 2023.10.01.].

https://en.wikipedia.org/wiki/Main_Page [Letöltve: 2023.10.01.].

YAMADA (2005) DVD-jatekos Operating Instructions. https://m.blog.hu/bl/blogollo/file/yamada_dvd-jatekos.pdf [Letöltve: 2023.10.01.].