

TDK DOLGOZAT

2024

Szerző: Dr. Mándy Dávid Gergő

Cím: Adatvagyon, adatgazdálkodás,
de minek az AI?

*Data assets, data management,
but why use AI anyway?*

Konzulens: Dr. Csiszár Csilla Margit
egyetemi docens

Intézmény: Miskolci Egyetem
Gazdaságtudományi Kar
Gazdálkodástani Intézet

Tartalom

Rövidítések jegyzéke.....	
Bevezetés.....	1
1. AI rendszerek	2
1.1. Csoportosítás funkció szerint	3
1.1.1. Real-time elemzés	4
1.1.2. Predikció.....	5
1.1.3. Machine Learning (ML).....	7
1.1.4. Deep Learning (DL)	9
1.1.5. Generatív mesterséges intelligencia (GenAI).....	14
1.2. Jogi háttér	17
2. Adatvagyon-gazdálkodás	21
2.1. Az adattömeg minősége	22
2.2. Adatvagyon-gazdálkodás szükségességének elve és annak alapelvei	24
2.3. A kormányzás alapelvei	24
2.4. Az adat áru funkciója	27
2.5. Adatpiacok felépítése és szereplői	29
3. Következtetések, javaslatok	31
3.1. Hol van a mesterséges intelligencia helye az adatvagyon-gazdálkodásban?	31
3.2. Mit nevezünk adatgazdaságos vállalatnak?.....	32
3.3. Mi az adatpiacok felépítése?	32
3.4. Milyen pénzforgalmat realizál az adatpiac jelenleg? Mik az előrejelzések?	34
4. Összefoglaló.....	37
Irodalomjegyzék.....	
Ábrajegyzék	
Táblázatjegyzék.....	

Rövidítések jegyzéke

AI – Artificial Intelligence (Mesterséges intelligencia)

API – Application Programming Interface (Alkalmazásprogramozási felület)

DL – Deep Learning (Mélytanulás)

ES – Expert System (Szakértői rendszer)

EU – Európai Unió

FLOPS – Floating-Point Operations (Lebegőpontos számítások)

GenAI – Generative Artificial Intelligence (Generatív mesterséges intelligencia)

GDPR - General Data Protection Regulation (Általános adatvédelmi szabályok)

IoT – Internet of Things (A dolgok internete – hálózatba kötött „intelligens” eszközök)

JSON – JavaScript Object Notation (JavaScript objektumjelölés)

KDD – Knowledge Discovery in Databases (Tudáskeresés adatbázisokban)

LSTM – Long State Term Memory (Hosszútávú memória)

ML – Machine learning (Gépi tanulás)

NNW – Neural network (Neurális hálózat)

SQL – Structured Query Language (Strukturált lekérdezőnyelv)

XML – Extensible Markup Language (Kiterjeszthető jelölőnyelv)

Bevezetés

2024-ben a negyedik ipari forradalom kapcsán igencsak központi szerepet kapnak az adataink. Olyannyira, hogy a mai tudományos kutatások jelentős része a területen az adatvezérelt döntéshozatal témáját vizsgálja. De honnan származnak ezek az adatok? Hogyan lehet egy vállalaton belül kiválasztani, megszerezni és megérteni azokat az adatokat, amelyeket *valóban* fel tudunk használni profitszerzési céllal? Hogyan tudjuk ezeket racionális keretek között felhasználni? Végso soron a vállalat vezetése hogyan gazdálkodik a szűkös erőforrásokkal, hogyan optimalizálja a saját működését úgy, hogy egyébként radikális átalakítások szükségesek, azonban a bevétel kiesése nem megengedhető?

A mesterséges intelligencia valóban tud segíteni ezen, vagy még több kérdést vet fel? Azáltal, hogy bevonjuk egy vállalatirányítási rendszerbe, könnyíti a dolgunkat, egyszerűsíti, vagy világosabbá teszi a feladatainkat? Hogyan tud és milyen formájában hasznosulni a számunkra a rendszer által szolgáltatott információ? Mennyibe fog ez nekünk kerülni?

Számos és egyre több kérdést vetett fel bennem is ez a terület a téma kutatása során, minél mélyebbre merültem egy adott részterület elemzése során, annál biztosabb voltam abban, hogy pusztán egy adatkormányzás, egy vállalati BI integráció a jövőben önálló tudományterületté válik. A dolgozat elkészítésekor a kutatási kérdések mentén az irodalmat áttanulmányozva számos aspektussal találkoztam, ezt igyekeztem a kérdések szempontjából didaktikailag érthető módon rendszerezni és felépíteni. Készítettem egy rövid összeggést a mesterséges intelligencia rendszerek és a mai legfejlettebb mesterséges intelligencia alapjait lefektető elvek létrejöttéről és mai alkalmazásairól, ennek jogi vetületéről, illetve egy rövid kitekintést arról, hogy 2024-ben az adatok piacán milyen jellegű szerepkörök fellelhetők, illetve hol van ebben a helye az átlagfelhasználónak és az adatfelhasználó vállalatoknak. Dolgozatomban a következő oldalakon a következő kérdésekre kerestem a választ:

- Milyen szerepe van a mesterséges intelligenciának az adatvagyon-gazdálkodásban?
- Mit nevezhetünk adatgazdaságos vállalatnak?
- Mi az adatpiacok felépítése?
- Milyen pénzforgalmat realizál az adatpiac jelenleg? Mik az előrejelzések?

1. AI rendszerek

Amennyiben az adatvagyon-gazdálkodás és az adatgazdaságosság fogalmait szeretnénk definiálni, a téma kutatása során mindenképpen bele fogunk futni a mesterséges intelligencia valamely szegmensébe. Némiképp tyúk, vagy tojás kérdéssé evolválna, ha azt szeretnénk megvizsgálni, hogy az AI mainstream jelenléte vette elő ezt a fogalmat, vagy azért kezdik el használni a vállalkozások ilyen intenzitással (2024 környékén ez bizonyos régiókban a 40% körüli értéket is eléri) (*AI in Action*, 2024) mert az adatvagyon-gazdálkodás egyik hatékony eszközévé lehet, még akkor is, ha néhány évvel ezelőtt még nem így hívtuk.

Dolgozatomban csak azon AI rendszerek elemzését tűztem ki célul, amelyeket kifejezetten gazdasági vállalkozások, hatékonyságnövelési, vagy direkt profitszerzési céllal használnak, tehát nem célom részletesen vizsgálni olyan, mindenki számára elérhető, legjellemzőbben keresési, vagy szórakozás céljából felkeresett platformokat, melyek kiesnek ebből a halmazból. Azonban, mivel ezen rendszerek használata sokrétű és diverzifikált, annak érdekében, hogy pontosan tudjuk vizsgálni az adatstratégiai és adatvagyon-gazdálkodási vonatkozásokat célszerű egyfajta csoportosítást végezni. Első körben, az informatika a rendszer képességei alapján a gyenge, erős vagy szuper AI kategóriába sorolja be azokat.

A 2024-ben használt AI rendszerek a gyenge, vagy más néven szűk AI kategóriába tartoznak, mint a ChatGPT, Siri, IBM Watsonx. Más néven szűk AI, a mesterséges intelligencia egyik legsikeresebb és legelterjedtebb formája, amely specifikus feladatok elvégzésére lett kifejlesztve. Ezek a rendszerek képesek információk feldolgozására és válaszok generálására, azonban nem rendelkeznek tudatossággal vagy önálló gondolkodással, tehát nem képesek általánosítani az emberi intelligenciát, és nem képesek kreatív gondolkodásra. Alkalmazásuk széleskörű és sok iparágban jelen van. Például a keresőmotorok, mint a Google, nap mint nap segítenek gyorsan információkat találni. Az ügyfélszolgálati chatbotok azonnali válaszokat adnak a felhasználói kérdésekre, szintén a gyenge AI példái. (Data and AI Team, 2023)

A kereskedelmi platformok, mint az Amazon, a Temu és a Netflix, ajánlórendszereik révén javasolják a felhasználóknak a számukra érdekes termékeket és tartalmakat. (Biswas, 2023) Az automatizálás révén ezek a vállalatok csökkentik működési költségeiket, miközben növelik az értékesítéseiket, ezzel extra profitra tesznek szert a versenytársaikkal szemben. E rendszerek gyorsan és pontosan dolgozzák fel az adatokat, így azonnali válaszokat tudnak nyújtani a felhasználók kérdéseire. Ugyanakkor a gyenge AI indentifikációjából adódóan a korlátai közé tartozik, hogy nem képes általánosítani, nem alkalmazható különböző kontextusokban.

Továbbá, nem rendelkezik érzelmi intelligenciával sem, tehát nem képes az emberi érzések megértésére vagy empátia kifejezésére. Ez a hiányosság általában kiütöközik az ügyfélszolgálati alkalmazásokban, ahol az ügyfelek érzelmi reakcióinak reflexiójára is szükség van. (Steck *et al.*, 2021; Monetate, 2024)

Ezen kívül még megkülönböztetünk erős, avagy általános AI-t, illetve mesterséges szuperintelligenciát, vagy szuper AI-t, holott valójában egyik sem létezik fizikailag. Az erős AI a mesterséges intelligencia egy olyan formáját jelöli, amely képes bármilyen intellektuális feladatot elvégezni, amit egy ember is képes. Tehát nem csak egy adott feladatra van kiképezve, mint a gyenge AI, hanem általános intelligenciával rendelkezik, és bármilyen problémát meg tud oldani, amit egy ember is meg tud oldani. Létrehozása kapcsán az egyik legnagyobb kihívás, hogy milyen úton-módon lehet létrehozni egy olyan AI-t, amely önállóan tud tanulni és gondolkodni, és nem csak a programozott utasításokat követi. Viszont a mesterséges szuperintelligencia már szinte science-fiction kategória, elméletek szerint ebbe a kategóriába sorolnák azt a számítógépet, amely az emberi intelligencia szintjét meghaladva lenne képes tanulni, érvelni, szocializálódni, önmagát reprodukálni és döntéseket hozni mindezekről, természetesen bárminemű beavatkozás nélkül. (Microsoft, 2024)

Létezik még számos csoportosítási forma egyébként a szakirodalomban, megkülönböztetnek különféle inputok vagy outputok szerinti, adatkezelési módszerek szerinti, vagy felület szerinti rendszert, azonban az adatvagyon-gazdálkodás és az adatgazdaságosság fogalmának tárgyalása szempontjából didaktikailag talán szerencsésebb a funkció szerinti csoportosítást alkalmazni, azon belül is követve az alapelvek meghatározása tekintetében az időrendiséget az AI rendszerek közel több, mint 80 éves történelmében.¹ (McCulloch, 1943)

1.1. Csoportosítás funkció szerint

A funkció, avagy a képességek szerinti csoportosítás során azt az elvet alkalmazom a tárgyalás során, hogy bemutatom: honnan indultak és miként elvolválódtak azzá ezek a megoldások, amelyek a ma szinte mindenki által közismert AI fogalmát alkotják, komplexen és széleskörűen alkalmazhatóak a pontosság és az egyre csökkenő számítási idő kombinációjának köszönhetően és egyre inkább abba az irányba mutatnak a trendek, hogy az átalakulóban lévő gazdasági ökoszisztémánk egyik kulcsszereplőjévé avanszáltak. (Gurcan *et al.*, 2023) A csoportosítást az algoritmusok fejlettségi szintjének megfelelően végeztem.

¹ A neurális hálózatok matematikai alapelveiről szóló első leírásokat már 1943 tájékán fellelhetjük.

1.1.1. Real-time elemzés

Habár a mesterséges neurális hálózatok elvi modellje létezett már 1943-ban és Alan Turing munkássága is alapvető, ha ebben a témakörben mozgunk, az idő csak az 1990-es években érett meg arra, hogy az addig felépített elméletek gyakorlati alkalmazásban öltsenek testet és a nagyközönség számára látványos demonstráció kereteiben legyőzzenek egy-egy – a saját szakosztályukban legmagasabb eredményeket elérő – versenysportolót, vagy nagymestert. Példának okáért 1997-ben a Deep Blue nevű számítógép sikeresen legyőzte Garri Kaszparov, többszörös olimpiai- és világbajnok sakknagymestert, mely 11,38 GFLOPS számítási sebességgel is órákat vett igénybe (Ma az IBM Watson 80 TFLOPS-nál jár) Ennek jelentősége nyilván amellet, hogy a számítógépet gyártó IBM részvényei a bemutatót követő egy hónapban 22%-ot emelkedtek, szimbolikus jelentőséggel is bír és egy új korszak kezdetét jelentette. (Schaeffer and Plaat, 1997) A reakcióelemzés egyik ma is legismertebb alkalmazása a sakkban, illetve a hasonló elven játszható 1-1 elleni, a játékosok számára adott műveleti lehetőségekkel szabályozott játékokban történik. Bár ezek a gépek nem rendelkeznek akkora intelligenciával, mint az emberi játékosok, rendkívül összetett algoritmusokat használnak, és másodpercenként milliárdos nagyságrendű pozíciót és pozíciórendszerek alkotta végkimeneteket vizsgálnak meg, hogy meghatározzák a következő lépést.

Ugyanakkor ehhez rendkívül hasonló real-time analízist használt a Microsoft Xbox 360 Kinect konzol a játékosok testmozgásának detektálására. Természetesen az akkori alkalmazása még gyerekcipőben járt, és sokkal több fejlesztésre volt szükség ahhoz, hogy a mindennapi alkalmazásokban használható legyen. (Elaraby, 2019) De ma is egy átlag felhasználó rengeteg aspektusban találkozik a real-time elemzéssel, amit AI vagy hasonló fejlettségű algoritmusok, szoftverek végeznek. Elég például csak arra gondolni, hogy a minket körülvevő környezetben folyamatosan működnek meteorológiai, szeizmológiai állomások, melyek folyamatosan adatokat gyűjtenek be a mérőműszereikből, majd legtöbbször egy központi feldolgozó rendszerbe küldik ezeket az adatokat feldolgozásra, melyeket ez a rendszer utána feldolgoz, strukturál, eleméz. (Dewitte *et al.*, 2021)

A pénzügyi szektor is használ real-time elemzéseket, a pénzpiacok vonatkozásában, ahol kiterjedt tudományág a technikai elemzés, amelyet természetesen ma már az emberi munkaerő mellett AI algoritmusok segítségével is végeznek, hiszen így sokkal nagyobb számú adatot, sokkal gyorsabban képesek elemezni. Tulajdonképpen ennek a használata nem új keletű történet, már a számítógép-korszak kezdetétől fogva részletes dokumentációt érhetünk el a részvények vagy jegyek korábbi teljesítéseiről, azok egymáshoz való viszonyáról és árfolyam-

mozgási dinamikáiról, melyeket ma is felhasználnak trendelemzések és előrejelzések formájában, azonban a valós idejű megfigyelő algoritmusok új keretrendszerbe helyezik ezt az iparágat is, azzal, hogy ezres nagyságrendű részvénytmozgásokat tudnak valós időben követni, és elemezni. (Serrano, 2018)

De, amikor az átlag felhasználó reggel elindulva az autójában elindítja a Waze applikációt, vagy a Google térképet, akkor találkozik a valós idejű forgalom figyelő algoritmusokkal is, amely folyamatosan monitorozza a felhasználók által jelentett eseményeket, a többi felhasználónak az adott útszakaszon való áthaladási idejét, begyűjti a közúti információkat a közútkezelőktől, ezáltal becsülve a felhasználó számára a leggyorsabb eljutási útvonalat a célpont felé (Vasserman és mtsai., 2015). Természetesen a felhasználó helyzetét is folyamatosan figyeli GPS műholdas összeköttetések révén és ezt a változót is beépíti az elemzéseibe. Ez már egy kicsit tovább visz bennünket a valós idejű adatelemzésből, ahol az algoritmus már nem csak megfigyel, elemez és mintázatokat keres, majd levon egy következtetést, hanem már előrejelzést is ad a korábbi trendek, vagy valamilyen egyéb, korábban meghatározott rendszer alapján, ez a rendszer lesz a prediktív AI.

1.1.2. Predikció

A predikció előfeltétele volt egyébként az adatelemzés, akárcsak a klasszikus adatelemzés során, amikor múltbeli adatokat vizsgálva jutunk el egy adott következtetésre. Ha azonban az adathalmazunk hatalmas, az előrejelzés igénye pedig erős, akkor adja magát a helyzet, ha van már egy megfelelő technológiai fejlettségünk az adatok közötti korrelációk rögzítésére és kezelésére, akkor ezeknek az elemzéseknek a felhasználásával fejleszthetünk egy olyan rendszert, amely az eddigiektől sokkal gyorsabban képes a rendelkezésre álló források alapján egy előrejelzést meghozni. Ugyanakkor meg kell jegyeznünk, hogy a fejlettségi potenciál a prediktív rendszerek esetén egyfelől a folyamatos előrejelzés készsége, illetve az előrejelzési hibákból való tanulás, valamint annak vizsgálata, hogy a korábbi adatokból való elemzés, a real-time elemzés vagy ennek a kettőnek a kombinációja adja majd a legpontosabb előrejelzéseket. Az időjárási viszonyok előrejelzésére ma már NNW-ket használnak, gyakorlatilag abból az 1943-as koncepcióból nőtt ki, mint amit korábban tárgyaltunk. A feltanítás során múltbeli adatokat adnak meg, amelyeket ezután elemez a hálózat, összefüggéseket és mintázatokat keresés a jövőbeli időjárási viszonyok előrejelzése érdekében. (Dewitte és mtsai., 2021) Ez egy példa arra, amikor az adatgyűjtő rendszer és az előre jelző rendszer két külön elemként működik, de olyan alkalmazása is létezik, ahol egy rendszer alelemeiként dolgoznak együtt, illetve a későbbiekben fogom ismertetni, hogy a legfejlettebb

időjárás előre jelző algoritmusok ma már sokkal komplexebbek, és nem csak real-time adatokat dolgoznak fel, hanem valós időben elemzik azoknak a kapcsolatát a múltbeli adatokkal, amelyeket szintén folyamatosan gyűjtenek. Egy-egy ilyen algoritmus dolgozik az AccuWeather és az IBM Weather felületein is, (Olanrele *et al.*, 2022) ahol a szolgáltatók a saját gyűjtésű adataikkal és elemzéseiket vizualizálva jelenítik meg a felhasználók számára az előrejelzéseket, azonban itt felhívom a figyelmet arra, hogy emellett az adatokat harmadik fél jellegű vállalatoknak is értékesítik, akik a vásárolt adatok felhasználásával saját célú kutatásokat, hosszútávú elemzéseket, vagy akár saját időjárás alkalmazást fejleszthetnek amit vagy az operációs rendszerükbe integrálnak, esetleg önállóan értékesítik és felhasználói adatok gyűjtésére használják.(Li *et al.*, 2022)

A pénzügyi szektorban az NNW alapú prediktív AI rendszerek alkalmazása forradalmi változásokat hozott. Ezek az algoritmusok képesek akár valós időben reflektálva követni a részvényárak előrejelzéseit, amely kulcsfontosságú információkat hordoz a befektetők és kereskedők számára, valamint azon további – később részleteiben tárgyalt – szintén szoftveres alapú, úgynevezett robo-tanácsadóknak, akik a felhasználók fintech applikációiban (Revolut, eToro, stb) automatizált befektetési döntéseket hoznak, persze ne szaladjunk ennyire előre, ezek a rendszerek kombináltan többféle megoldás felhasználásával működnek, melyek később kerülnek tárgyalásra. (Tao és mtsai., 2021)

Persze a logikusnak tűnő megoldások mellett széleskörű palettát találunk az irodalomban a prediktív AI rendszerek használatára önmagában. Például egy rendkívül érdekes esettanulmány kapcsolható ide, a Siemens egyik megoldása már gyártáskor, illetve a gyártás tervezésekor képes előre jelezni a termék vagy adott eszköz karbantartási periódusait, illetve a termék lehetséges típushibáit. (Annanth, Abinash and Rao, 2021) Természetesen ez elsőre nem tűnhet nagy kunsztnak, ha gyártani akarunk mondjuk egy hűtőszekrényt és szeretnénk meghatározni, hogy nagyjából hány ajtónyitás után tud elkopni egy nyitószerkezet, ha ismerjük az anyag fizikai tulajdonságait, vagy hány hőcserélő ciklust tud 5%-os hatásfokvesztés mellett végrehajtani egy kompresszor. Alapvetően ezek az értékek gyártási tételekre vonatkoznak, ahol a gyártott termékeket közelítőleg homogénnek tekintik, illetve a felhasznált alapanyagokat és összeszerelési munkafolyamatokat közel egyenértékűnek tekintik, így egy közelítő értéket tudunk ez alapján adni, hogy nagyjából milyen selejtmennyiségre számíthatunk egy adott gyártási tétel esetében, valamint hosszútávon milyen üzemóra tartományban lesz szükséges az adott alkatrészek cseréje. Itt belátható azonban, hogy a fakocka és egy elektromos autó igencsak két végpontján helyezkedik el a gyártási komplexitás skálájának, így előbbinek kevesebb, utóbbinak drámaian több gyenge pontja lehet, főleg azt is számításba véve, hogy a gyártáshoz

felhasznált gépeknek is vannak bizonyos gyenge pontjai, illetve karbantartási periódusai. A Siemens rendszere, amellett, hogy a gyártás egyéb folyamataiban is használ intelligens megoldásokat, a döntéstámogató rendszerbe integrálva egy prediktív alrendszert használ egyrészt a gyártó gépsorok karbantartás előrejelzésére (ehhez egyébként belső szenzorok, korábbi gyártások adatait illetve a tervezendő gyártás végrehajtásakor felmerülő használati igénybevétel potenciális hatását használja fel) másrészt ezt már a gyártás tervezésekor képes előre jelezni, ezzel minimalizálva a gyártás közbeni rendszerösszeomlás okozta üzemszünet kockázatát, de egyúttal óriási adattömegek gyors elemzése révén lehetővé teszi, hogy ezeket az adatokat a fentebb leírt integrációnak köszönhetően azonnal a vállalati döntéshozók rendelkezésére bocsátja.

A Siemens prediktív AI rendszerére is igaz az, ami a korábban tárgyalt időjárás előre jelző rendszerek egy részére is, amely szintén nem új keletű találmány, viszont egy lépéssel közelebb visz a ma használt adatvagyon-gazdálkodási elemek megértése tekintetében, és az pedig egy olyan algoritmus, amely lehetővé teszi a prediktív AI számára, hogy elemezze a saját előrejelzéseit is. Talán mondhatjuk azt, hogy ez az algoritmus tanította meg először az elemző számítógépet az önreflexióra, hogy a korábbi előrejelzéseit megvizsgálva rájöjjön, hogy a helyes előrejelzést milyen adatokból hozta meg, illetve milyen részleteket nem vett figyelembe a hibás előrejelzések előfordulásakor, ez pedig a gépi tanulás, avagy Machine Learning (ML).

1.1.3. Machine Learning (ML)

Az AI rendszerek evolúciójának következő lépése az, amikor egy rendszer képes az önreflexióra. Ma ez már működik természetesen a klimatikus, a pénzügyi és a gyártási területek AI rendszereiben, de ez nem volt mindig így.

Nagyából az 1960-as években járunk, ahol már túl vagyunk az NNW alapkoncepcióin, Turing munkásságán és körvonalazódni látszik egy ELIZA nevű beszélgető droid fejlesztése, melynek kapcsán elkezdődnek az ML alapvető fogalmainak lefektetése is. (Fradkov, 2020) Jelentősége egyébként csakugyan, mint a Deep Blue számítógépnek alapvetően inkább szimbolikus, mint tudományos értelemben vett mérföldkő, hiszen rengeteg szociokulturális hatását észleljük még napjainkban is a ChatGPT által, amellett, hogy az ELIZA alapjait képező scripteket fellelhetjük a történelemben később is számos – az emberi kommunikációt szimuláló – programkód, vagy kezdetleges virtuális asszisztens „arca” mögött. Emlékszünk még Clippit-re az idegesítő kis gémpapocsrá, aki minduntalan segíteni akart nekünk a szövegszerkesztés során? (Rajaraman, 2023) Hála a Microsoft fejlesztőinek, cserélgethető volt több különböző karakterre, majd végleg kivezették. Azonban Clippit és ELIZA mutat egy közös tulajdonságot,

mégpedig azt, hogy mindketten a felhasználói interakciókból folyamatosan tanulva képesek változtatni a saját működésüket. Illetve, még egy közös tulajdonság, mindketten egy-egy szimbolikus időpillanatban járultak hozzá a ML fejlődésének idővonalához. Ugyanis ELIZA és Clippit megjelenése között, egyfajta csend, egyfajta stagnálása volt ezen folyamatok fejlődésének. Ugyan fellelhetőek voltak játékokban, bizonyos alkalmazásokban, de az 1990-es évekig nem volt különösebb jelentősége ezeknek az algoritmusoknak egészen a Big Data koncepciók megjelenéséig, ami nagyjából Clippit létrejöttének ideje is. Természetesen neki aligha bármilyen köze lenne hozzá, a Big Data alapvetően nem egy tyúk-tojás kérdés, ha az AI rendszerek viszonylatában nézzük. Előbb létezett a kezdetleges AI és csak azután a Big Data. Ezt több helyen az irodalomban egyébként a komputerizáció és a kialakuló jóléti társadalom okozta népességnövekedés következményének tudják be, mégpedig azáltal, hogy az ipari forradalmak idejében tapasztalható népességrobbanások megnövekvő személyi adathalmazzal hozták magukkal, az adatok digitalizációja során pedig óriási méretű adathalmazokkal szembesültek. (Beer, 2016) De hogy mindez hogyan is köthető az ML algoritmusok fejlődési lépcsőjéhez? A nagy adattömegek kezelése és elemzése kapcsán lehetővé vált a prediktív AI rendszereknek, hogy nagyságrendekkel több adatból tudjanak dolgozni, tehát innentől kezdve nem csak real-time, vagy kezdetleges historikus adatok álltak a rendelkezésükre, hanem óriási adattömegek, amelyben szükséges eligazodni, mélységeiben mintázatokat és összefüggéseket keresni, hogy a predikció eredménye a lehető legpontosabb legyen.

Egyfajta feltérképezése ezeknek az adathalmazoknak a tudáskinyerés folyamata, a szakirodalomban leggyakrabban Knowledge Discovery in Databases (KDD)-ként hivatkoznak rá. (Arora, Lele and Jindal, 2022). Az ML algoritmus feladata ebben az esetben a statisztikai elemzés, és a mintázatfelismerés, hogy feltárja az összefüggéseket. Alkalmazásával az adatbányászat hatékonysága és sebessége jelentősen megnövekedett, mivel az algoritmus már képes automatikusan tanulni az adatokból, és így egyre pontosabb előrejelzéseket készíteni, továbbá lehetővé teszi a folyamatos modell fejlesztést, így a korábbi tapasztalatok felhasználásával a jövőbeli eseményekre vonatkozóan is egyre megbízhatóbb predikciókat fog gyártani. (Ashley-Osuzoka *et al.*, 2024)

Az etikai megfontolandók között részletesebben ki fogok térni majd az adatbányászatra, ugyanis a rendszerek fejlődése során egyre érzékenyebb adatbázisokban is használatosak, így ma már komoly jogi keretrendszer van keletkezőben, hogy ezeket az adatokat megvédjék. Ugyanakkor az adatbányászatnak ma kiemelt jelentősége is van, az adatpiacok tárgyalásakor részleteiben is ki fogok térni erre.

A kérdésre, hogy az átlag felhasználó 2024-ben hol találkozhat akár a tudta nélkül ilyen ML megoldásokkal, tulajdonképpen a válasz az, hogy majdnem mindenhol. A korábban már említett Amazon, Netflix és Temu ajánlórendszerek mind olyan ML algoritmusok, amelyek a folyamatos interakciók révén változtatják a saját működésüket és készítenek egy predikciót a vásárlói szándék előrejelzésére. Persze ettől lényegesen többet is tudnak, hiszen – mint fent említettem már – ezek a technológiai innovátor vállalatok komplex BI rendszereket használnak, melyeknek a vásárlói monitoring ML algoritmus csupán a jéghegy csúcsa jellegű alkomponense, mégis talán szemléletes példa annak bemutatására, hogy az adatvagyon-gazdálkodásban, illetve az adatgazdaságos vállalat működésében hol van ennek a gyakorlati jelentősége. De felsorolhatunk tényleg számtalan példát. FaceID-s telefonunk van? Máris minden egyes feloldáskor találkoztunk az ML algoritmussal, amit feltanítottunk a képernyőzár beállításakor, (illetve az Apple betanított rengeteg egyedi arcvonás felismerésével) és a rendszer máris képes felismerni saját biometrikus mintázatainkat. Használtunk esetleg FaceApp-ot? Ugyanez a helyzet, csak a beleegyezésünkkel a mintázatokat potenciálisan további elemzési céllal felhasználhatják (Kwak & Kim, 2018)

Már közelítünk napjainkhoz, még néhány fontos lépés hátravan ahhoz, hogy eljussunk a modern, abszolútdigitális vállalat megközelítéséhez, ugyanis a technológiai fejlődés során, ahogy az ML egyre bonyolultabbá vált, egy új struktúra kezdett kialakulni. Bár sokszor az irodalom együtt tárgyalja a gépi tanulással, mégis inkább szerencsésebbnek látom picit leválasztani, hiszen – bár alapelveiben megegyezik – mégis kapacitásában és feladatkörében valamelyest különböző gyakorlati alkalmazásokkal rendelkezik a mélytanulás, avagy Deep Learning (DL).

1.1.4. Deep Learning (DL)

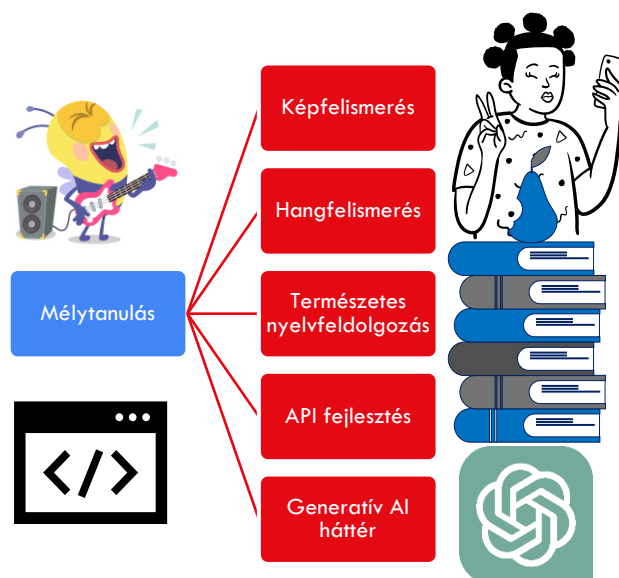
A Deep Learning dióhéjban összefoglalva a lényegét, működését tekintve nem sokban tér el egyébként a már korábban tárgyalt ML-től, olyannyira nem, hogy az alapjai pontosan ugyanazokban az NNW-kben gyökereznek, mint a Deep Blue számítógépnek, azonban sokkal több réteg helyezkedik el vertikálisan is ezekből a hálózatokból, ezáltal a számítási kapacitása lehetővé teszi számára sokkal komplexebb feladatok ellátását. (Steck *et al.*, 2021) A korábbi modellekben is láthattuk azt, hogy képes már a számítógép real-time big data és historikus big data elemzésre, ezek alapján predikciókat alkotni és tanulni azok helyességéből ezáltal fejlesztve önmagát. Az újdonság ezek mellett az, hogy a jóval bonyolultabb architektúrájának köszönhetően (Schmidhuber, 2015) ezt egyfelől sokkal gyorsabban és hatékonyabban végzi elődeinél, azonban el is ér egy olyan technológiai fejlettségi fokot, hogy már a DL algoritmus

száma számára nem feltétlenül szükséges a bevitt adatok felcímkézése, vagy strukturáltsága és ez lesz a későbbiekben ami elsővonalbeli technológiává lépteti elő a DL-t, hiszen egy vállalati adatvagyonban rengeteg strukturálatlan, vagy félig strukturált adatot birtoklunk, erről később lesz szó még, de gondoljunk csak bele, az eddig használt AI rendszerek akkor tudtak helyes predikciót alkotni, ha minden egyes bevitt adatra (természetesen a Big Data) esetében is mi előzetesen megmondtuk neki, hogy az micsoda, megadtuk neki a *kontextust*.

Az időjárás előrejelzésekhez szükség volt arra, hogy az algoritmus számára megmondjuk, hogy az A interfész felől érkező adatok, melyek bizonyos értéket vehetnek fel, -40 és +60 között, azokat ő értelmezze °C-ban és hőmérsékleti adatként tárolja azokat. A B interfész felől érkező adatokat, melyek érkezhettek mondjuk -150 és +150 között, értelmezze km/h -ban és a szélirány dimenziójaként kezelje. A DL algoritmusnak azonban elegendő, ha a szenzorokkal összekapcsoljuk és elkezd az adatokat saját magától felismerni. A korábbi tudását felhasználva tudja azt, hogy a hőmérsékleti adatok és a szélirány adatok miképp értelmezendők, azaz tudja, hogy egy adott területen mondjuk utóbbi egy dinamikusabban változó paraméter és abszolútértéken magasabb értéket is felvehet, míg a hőmérséklet egy sokkal kisebb mértékben változik adott időegység (pld 24 óra) alatt, ezáltal ha A és B szenzorra kapcsoljuk, akkor a korábbi ismereteit felhasználva és a real-time adatokat detektálva rövid időn belül megtalálja azokat a mintázatokat és összefüggéseket, amelyek rávezetik, hogy ezeket az adatokat pontosan hogy is kell értelmeznie. (Dewitte *et al.*, 2021)

De a DL algoritmusok más helyeken is brillíroznak. Ha szem előtt tartjuk azt, hogy ezek a rendszerek képesek felismerni azt az adatot, amit észlelnek, akkor gyorsan beláthatjuk, hogy átlag felhasználói szinteken is számos helyen találkozunk velük. Használtuk már a telefonunkon a diktálás funkciót? Egy ilyen rendszer segített nekünk abban, hogy a beszédünk alapján felismerje azokat a szavakat, amelyeket kimondunk, azzal, hogy a hangfelvételt a másodperc törtrésze alatt eljuttassa a hálózaton keresztül egy ilyen beszédfelismerő algoritmushoz, amelynek a feltanítása során rengeteg hanganyagot szolgáltatott, azok szöveges anyagával, így képes volt azt a mintázatot felismerni, hogy az írott szöveg és a hangzó szöveg között milyen korreláció van. Esetleg kezdtünk már úgy mondatot, hogy: Hé, Siri! vagy Hé Alexa! Ugyanez a helyzet. (Kepuska and Bohouta, 2018; Haeb-Umbach *et al.*, 2019) De hasonló a helyzet bármely nyelvfeldolgozó, fordító, vagy képfelismerő DL algoritmus esetén is. Hatalmas méretű adatbázisok segítségével képesek arra, hogy a képek vagy videók egyes képpontjait végig vizsgálva egyértelműen azonosítsák annak tartalmát, megkeressék a modellezett nyelvek között és azonnal, ha a felhasználó erre parancsot ad, akkor a vizsgálat eredményeképpen ne az egyes képpontok színét, vagy a közöttük lévő korrelációt adják eredményül, hanem azt a szót, vagy

szöveget, amely a nyelvi adatbázisban szereplő párját adja ezeknek a képpont tulajdonságoknak. Az alkalmazásokról készítettem egy ábrát:



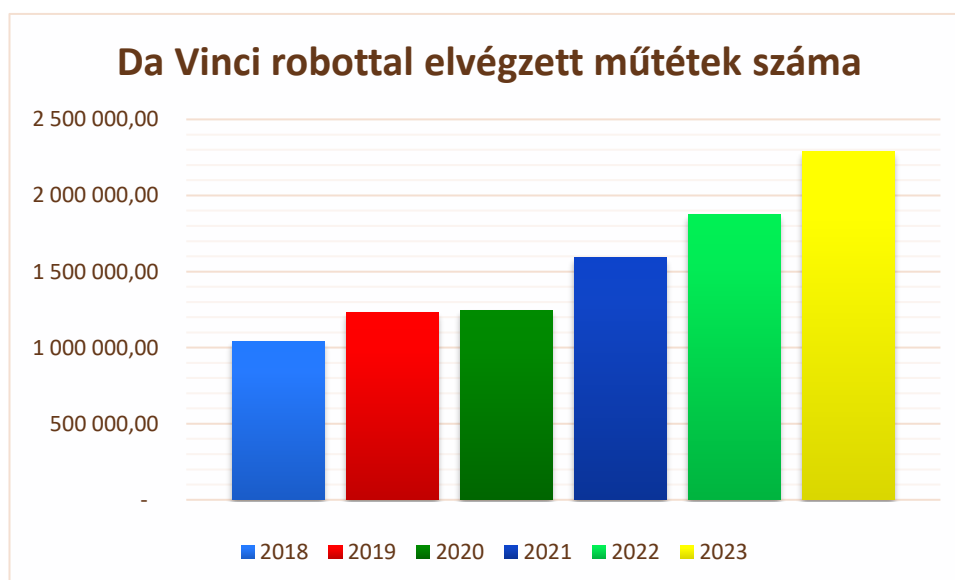
1. ábra - DL alkalmazásai. forrás: Saját Szerk..

Van esetleg az autónkban táblafelismerő rendszer? Vagy sávtartó asszisztens? Találkoztunk már a közösségi platformokon olyan reklámmal, ami azt az applikációt hirdeti, amelyikkel lefotózzuk a kedvenc kávécsserjénket és megmondja, ha esetleg túllőntöttük vagy betegséget kapott el? Ismét egy DL algoritmus áll a háttérben. (de Oliveira Aparecido *et al.*, 2020) (Fujiyoshi, Hirakawa and Yamashita, 2019) Igen, itt a „jó eséllyel” mind a két kontextusban igaz. A DL algoritmus végzi a kép felismerését az applikációban, megfuttatja az adatbázisában és megkeresi azokat az egyezéseket, ahol hasonló képpontkombinációval találkozik, megkeresi a nyelvi adatbázisban a hozzá tartozó adatpárokat, majd ennek a keresésnek a találatait adja eredményül, mindez az úgynevezett *multilayer* NNW-nek köszönhetően ismét nagyon gyorsan, a felhasználó számára gyakorlatilag azonnal történik. Illetve abból a szempontból is talált a DL algoritmus jelenléte, mely a közösségi média platformokon keresztül működik, illetve azokon az oldalakon, ahová a fiókunkat kapcsoltuk, tehát, a böngészési vagy keresési előzményeink alapján – bár személyünkhöz nem társított módon – megjelenik egy célzott értékesítési ajánlat, feltételezve, hogy a mi preferenciáink alapján az számunkra hasznos. (Meta Platforms, 2024) De visszatérve egy kicsit az önvezető autók analógiájára, a DL platformot használó ipari robotok a nehéziparban és legtöbbször olyan gyártósori munkafolyamatokban fellelhetőek,

amelyeket egyébként egészségkárosító potenciál szempontjából nagy kockázatúak az emberek számára. Amellett, hogy a DL segíti az automatizációt is, ezek a robotok növelik a termelési hatékonyságot is, mivel nem igényelnek munkaközi pihenőt, vagy nem vonatkozik rájuk a táppénz, szabadság, vagy műszakokra osztott munkarend igénye, illetve jogszabályi kötelezettsége, ezáltal az ilyen típusú termelésből kizárható például az emberi fáradékonyság. Az ilyen robotokat másnéven kooperatív robotoknak, avagy kobotnak nevezik, amely egy robot és egy ember közötti partnerséget jelenti. A kobot nemcsak a robot és az ember közötti fizikai érintkezést teszi lehetővé ugyanazon a gyártósoron, hanem felhasználóbarát kialakítású is. Elsőre lehet, hogy nem tűnik úgy, hogy ez valami örületes újdonság lenne, automata lyukkártyás kötőgépek 1910-es évek óta vannak, de automata mosógép is a legtöbb háztartásban évtizedek óta. Automata kávéfőző, automatizált gyártósorok a második ipari forradalom óta. Akkor miért képvisel ez ekkora szenzációt? A titok az integrációs lehetőségben és az alkalmazkodásban rejlik. Míg egy első generációs síkkötőgép nagyjából annyi alkalmazkodásra volt képes, hogy a lyukkártyában rejlő információk alapján lekötött párszáz méter anyagot és ha elfogyott a fonal, vagy elakadást, hibát érzékelt a rendszer akkor leállt, 2024-ben egy automatizált gyártósor kapcsolatban van a vállalati döntéstámogató rendszerrel, folyamatosan monitorozza önmagát, mint teljesítmény és használat szempontjából, azaz a saját működéséről minden adatot beszolgáltat az üzemeltetőnek, képes saját maga beállítására és a saját gyártási ütemtervének kialakítására egy komplex gyártórendszeren belül (pl. utasításokat készít egy szakaszos gyártás esetén, hogy pontosan mikor lesz szüksége a következő alapanyag adagra). (Annanth, Abinash and Rao, 2021).

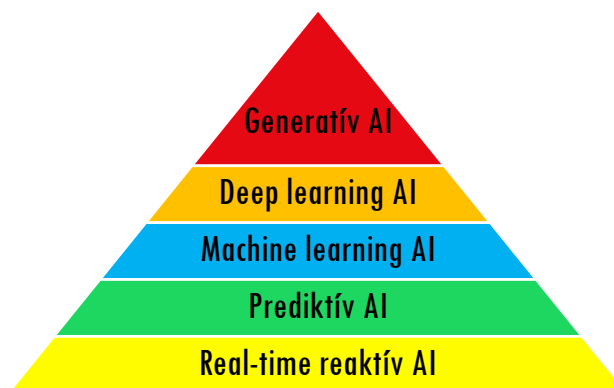
2024-ben már széles körben alkalmaznak kobotokat különféle iparágakban, például az élettudományokban, az autóiparban, a gyártásban, az elektronikában, a repülőgépiparban, a csomagolóiparban, a műanyaggyártásban és az egészségügyben is. Használatuk egyszerű, miközben megbízhatóak, biztonságosak és pontosak. Programozásuk lehetővé teszi, hogy fontos feladatokat végezzenek, például mérgező anyagok kezelése, jármű karosszéria csavarok rögzítése vagy akár a főzés. Az emberi üzemeltető távolról könnyen vezérelheti a kobotot és veszélyes munkákat végezhet. (Javaid *et al.*, 2022) Hogy még egy példát hozzunk, nem csak abban az értelemben kell „veszélyesnek” ítélni az adott munkát, amit elsőre gondolnánk, hogy például reaktor maradványokat kell letakarítani egy felrobbant reaktorcsarnok tetejéről. Tegyük fel, hogy egy olyan munkafolyamatot végzünk, amely egyébként teljesen veszélytelen normál esetben, apró bemetszéseket ejtünk, cernát fűzünk egy tűbe, ollóval aprókat vágunk, ez teljesen hétköznapi hangzik addig amíg nem embereken végezzük ezeket a tevékenységeket. Mondjuk, hogy viszonylag kevés átlagember szokta összevarrni mondjuk a szomszéd lőtt

sebeit. A klasszikus sebészetben azonban ettől sokkal komplexebb folyamat például egy rupturált vakbél, vagy egy koronária stent behelyezése, amelyet még mindig emberi kéz el tud végezni. Azonban a klasszikus sebészet is elért tudományos fejlődése során egy határt, amit most a legújabbban az emberi precizitás szűk keresztmetszete korlátozott be. Hogy jobban szemléltessem a kobotok funkcióját és rendeltetését a daVinci kobotot fogom bemutatni, mely az utóbbi néhány esztendőben jelent meg Magyarországon is, lehetővé téve azt, hogy eddig megoldhatatlannak tűnő műtéteket végrehajtsanak, például bizonyos típusú daganatok operációja, vagy nagy precizitást igénylő munkákat, például bizonyos aorta sebészeti eljárásokat a droidra bízzanak, természetesen mindkettőt közvetlen emberi irányítással. Jóllehet, egy ilyen rendszer használata alkalmazkodást kíván mind a felhasználó (hiszen egy komplex képzést kell elsajátítani a kezelői vizsgához) mind a „vállalat” részéről, azonban lényegesen széleskörűbb egészségügyi ellátást, végeredményben kisebb műtéti kockázatot és rövidebb lábadozási időt eredményez a páciens számára. Ez a rendszer számos helyen alkalmaz DL algoritmusokat, például real-time endoszkópos műtét közbeni diagnosztikával, folyamatos adatmonitoringgal, gyűjtéssel és analízissel, másrészt, mivel minden „műtendő terület” más és más fizikai paraméterekkel rendelkezik, például elasztikusság szempontjából, ezért folyamatosan elemzi a szenzorokból visszaérkező adatokat, hogy a megfelelő vágó, nyújtóerőt fejtse ki a kezelő a konzol segítségével. (Tran *et al.*, 2020) A gyártó cég elképzelései alapján a térnyerés függvényében, nem csak félautonóm, de totál autonóm robotok kifejlesztése lenne a hosszútávú cél. A robottal elvégzett összes műtétek számát az alábbi ábrán szemléltetem.



2. ábra - Műtétek tendenciája - forrás: Saját Szerk., adatok: Intuitive Surgical, befektetői jelentések

Persze azzal, hogy ez a kobot adatokat gyűjt és elemez, folyamatosan fejleszti önmagát, illetve azt a tudásbázist, amiből dolgozik, azaz minél több műtétet végeznek vele, annál több diagnosztikai adattal fog rendelkezni, 2023-ban valamivel több, mint 2,2 millió operációt, egyre nagyobb mennyiségű korrelációt és mintázatot lesz képes felismerni, mígnem végül lehetőségessé válik egy, már a beteg tüneteinek alapján a lehetséges beavatkozást vagy szükséges terápiát megajánló rendszer létrehozása. Természetesen ez a rendszer akkor a leghatékonyabb, ha nem csak egy orvos tabletjén működik, ahol a megfelelő, előre megadott rubrikák kijelölésével tudjuk a mintázatot előállítani, hanem akár a beteg a saját szavaival is képes megfogalmazni a panaszait, ezt a rendszer képes értelmezni és kontextust adni neki, majd a beteg nyelvén és szókészletével megfogalmazva át tudja adni azt az információt, amire a keresése találatot hozott. Ezzel megérkeztünk az AI rendszertan következő szintjéhez, ahol a



3. ábra - AI rendszerek hierarchiája - *Saját Szerk..*

mesterséges intelligencia már nem csak egy önfejlesztő fénysebességű keresőmotor, hanem ezek mellett egy olyan eszköz, amely ennek a keresésnek az eredményét képes bármelyik átlagfelhasználó számára értelmezhető formában prezentálni, ennek nyomán a modellezhető hierarchia csúcsán helyezkedik el, ezt a következő ábrán szemléltetem:

1.1.5. Generatív mesterséges intelligencia (GenAI)

ELIZA és Clippit kapcsán már tárgyaltunk olyan rendszerekről, amelyek lehetővé teszik azt, hogy a felhasználók interakcióba lépjenek egy számítógépes rendszerrel, úgy, hogy az adaptív válasszal rendelkezik attól függően, hogy a felhasználó milyen állításokat, vagy utasításokat ad meg. Ha távolról szemléljük az eseményeket, lehet, hogy nem vennénk észre nagy különbséget, még talán azt is mondhatnánk, hogy lám eltelt a kettő között körülbelül hetven év és a felhasználó begépel valamit amire a komputer válaszolt. Azonban a kettő rendszer között eltelt hetven év technológiai fejlődése integrálódott a mai generatív rendszerekbe, tulajdonképpen a pontosság, a precizitás és a széleskörű használati aspektus elősegítése érdekében. ELIZA még arra volt képes, hogy egy zárt nyelvi adatbázisból kulcsszavas keresés alapján adott kérdésekre

adott válaszokat adott. Bár a felhasználóknak úgy tűnhetett, hogy a gép megérti a kérdés kontextusát, de valójában nem ez volt a helyzet. (Rajaraman, 2023) És itt a kulcsszó a *kontextus*. Amely nem csak az adatelemzésnek, vagy az adatvagyon-gazdálkodásnak, de bármilyen adatkezelésnek is az alapvető építő köve. A ChatGPT azért jelenti a '20-as évek technológiai forradalmát, mert nemcsak egy olyan számítógép, amely „valamilyen” választ ad „valamilyen” típusú kérdésekre, hanem az eddigiekben részletezett predikciós, ML és DL algoritmusok segítségével az adott felhasználó kérdésének képes kontextust adni, annak alapján választ generálni, mindezt egyébként az algoritmust előállító vagy szervezet, vagy egyén mindennemű beavatkozása nélkül. Ha még ez nem lenne önmagában elegendő, a nyilvánosság számára való elérhetőség megteremti azt a lehetőséget, hogy a nagyszámú beszélgetés (adatbevitel) eredményeképpen az algoritmusok segítségével önmagát fejlessze, azaz, nemcsak az NNW keresés hatékonyságát, hanem azt, ami őt megkülönbözteti az eddigiekben tárgyalt rendszerektől, azt, hogy minél pontosabban tudja magának megadni a beviteli adatokra vonatkozó kontextust. (Weber, 2024) Egy általános GenAI képes rutinfeladatokat lerövidíteni vagy automatizálni, ezzel növelve az értékesítés hatékonyságát és csökkentve a költségeket. Alkalmazása jelentős például marketing területén hiszen személyre szabott hirdetések és tartalmakat tud létrehozni, biztosan találkoztunk már közösségi médiában Temu hirdetéssel úgy, hogy még azt sem tudta senki, hogy az micsoda, de már célzott ajánlatokat tett. (Monetate, 2024) Míg a termékfejlesztésben segíti a prototípusok kidolgozását, az ügyfélszolgálatban pedig a nyelvi modell alapú chatbotok gyakorlatilag 24/7 támogatást tudnak nyújtani, ezáltal javítva az ügyfélményt. (Ashley-Osuzoka *et al.*, 2024) Rengeteg szolgáltatóval találkozhat az átlagfelhasználó, aki már valamilyen úton-módon GenAI-t használ. Esetleg Telekom előfizetők vagyunk? Biztosan volt már olyan, hogy akár a társaság honlapjára kattintva felugrott egy chat ablak és a telefonhangból már jól ismert Vanda kínálta fel nekünk segítségét. (VANDA, ISTVÁN and ANDRÁS, 2024) Vanda 86 nyelven beszél és létrejöttével felszámolta a telefonos ügyfélszolgálatok „Nyomja meg a kettes gombot” gúnynévre hallgató, statikus hang alapú navigációs menürendszerét. Folyamatosan tanul az ügyfelek és a rendszer interakcióiból, illetve az egyszerűbb feladatok végrehajtására, pl. egy Domino egyenleg lekérdezés, gyakorlatilag bármilyen emberi beavatkozás nélkül képes, ha pedig olyan feladatról van szó (vagy pusztán már korábbi ismeretei alapján ismeri az ügyfél ilyen irányú preferenciáit) akkor egy ügyintézőhöz irányítja a beérkező hívást, vagy chatet. A korábban is említett Siri, vagy Alexa -bár folyamatosan fejlesztik és szintén végez némi DL-t a felhasználói szokások feltérképezése tekintetében, Vanda egy picit több ennél. Fejlesztői azt a célt tűzték ki, hogy létrehozzanak egy univerzális virtuális asszisztens szolgáltatást, amely teljes mértékben

személyre szabható vállalati szinten. Használja a fentebb ismertetett összes technológiai lépcsőt, azaz ML, DL algoritmusokat, NNW-eket, nyelvi modellezést, hangzó szöveg felismerést és felhasználói profilozást, valamint rendelkezik azokkal a vállalati információkkal, big data halmazokkal, amelyek az azonnali válasz lehetőségével felruházzák. Emellett a legnépszerűbb üzenetküldő alkalmazásokba is integrálható (Skype, Messenger, Viber... stb), tehát nem is szükséges hozzá bizonyos esetekben, hogy a szolgáltató honlapjára tévedjünk, vagy felhívjuk annak ügyfélvonalát, gyakorlatilag (ha engedélyeztük) Vanda a névjegyzékünkbe kerül, tehát akár Siri hangvezérléssel is azonnal elérhetővé válik. (Rekettye, 2023) De akár elegendő lehet kiadnunk Sirinek egy olyan kérést, hogy egyeztesse le üzenetváltásban Vandával az adott problémát, mondjuk egy készülékbeállítási kérdésben.



4. ábra - A tárgyalt algoritmusok megjelenésének időrendje - forrás: Saját Szerk..

A GenAI jelenlegi tudásunk legmagasabb technológiai szintjén áll, alkalmazási területei között végtelen sok lehetőség van. Generálhatunk szöveg alapján képet, kép alapján szöveget, alá feliratozhatunk egy videót, létrehozhatunk kép alapján egy programkódot, egy drón videófelvevételeiből egy földrajzi terület, vagy egy építkezés alatt álló épület felületének 3D modelljét. (Mitchell, 2021) Létrehozhatunk zenét (Louie *et al.*, 2020), előállíthatunk hangfelvételeket (Josan, 2024) vagy lecserélhetjük Kerekes József hangját VR Pistière a Family Guy-ban (*Így cseréld ki BÁRKI hangját! | Kísérletezés az AI hangkészítéssel!* 2023) és itt két fontos dolog kapcsán meg kell állnunk egy kicsit. Első körben az a tény, hogy a GenAI-ban rejlő lehetőségek egyelőre korlátlanak tűnnek, széles tömegek számára a GPT-k használatával vált elfogadott mainstream gondolatmenetté, hiszen az átlagfelhasználók által beszolgáltatott adatok az átlagfelhasználók számára voltak hasznosak. Tehát ha ilyen szemmel tekintek a MathGPT-re, aki egy tetszőleges vállalatgazdaságtani számítási példa adatai alapján felismeri a kontextust, majd a vállalatgazdaságtan szaknyelvét felhasználva nem csak megoldja, de egy kétperces videóban ismerteti is a példa levezetésének menetét, akkor joggal lehetek lenyűgözve. Azonban, ha azt szeretném tőle, hogy készítsen egy tervezetet a vállalkozásom jövő évi költségvetéséről, akkor azt a választ kapnám, hogy nem áll rendelkezésére kellő mennyiségű

információ, de a fellelhető tudása alapján a következőkből áll egy ilyen tervezet... és felsorolja egy kitalált vállalat költségvetési tervezetének alap elemeit. Állításai önmagukban nagyjából helyesek és a kontextust is jól értelmezte. azonban a válasza a számomra nem hasznos. Tehát ahhoz, hogy a vállalat irányítási rendszer kellően pontos válaszokat kapjon a GenAI-tól ahhoz az ő tudásbázisába kell táplálni a vállalati adatainkat. Természetesen ezt nem a nyilvános GPT-k részére szeretnénk átadni, hiszen ki szeretné a saját üzleti titkait bárki számára elérhetővé tenni? (Ashley-Osuzoka *et al.*, 2024) Egyfelől, a GPT integráció egy komoly adatvagyon-biztonsági kockázatot is jelent, hiszen ha nem megfelelő a jogosultságok kiosztása a rendszerben a *data leakage* komoly károkat okozhat, másfelől komolyabb etikai kérdéseket is elénk tár, mivel a munkavállalóink érzékeny adatai is a tudástárba tartoznak (ez valahol ésszerű következmény, mivel a vállalatirányítási rendszer tartalmazza azokat a személyi adatokat, amelyeket a munkaszerződések tartalmaznak, mint név, lakcím, elérhetőség, bankszámlaszám, munkabér, stb...) És itt jön be a második tényező, hogy akár a nyilvánosság számára elérhető, akár egy privát fejlesztésű és zárt láncú AI megoldás aspektusáról tárgyalunk, felmerülhet az abúzus lehetősége is. Azaz, akár megkérdezhetnénk a GPT-t, szerinte mi lehet a heroin értékesítésének jó gyakorlata a búza téri piacon, vagy munkáltatóként leinformálhatnánk az állásinterjú jelöltjeink Tinderes vásárlási előzményeit. Önmagából fakad a feladat, hogy az AI rendszereknek a jelenkori jogszabályokkal harmonikusan kell a feladatukat végezni, azonban a technológiai fejlettség ezeket bizonyos tekintetben túlszárnyalta, ezért némileg szükséges volt a jelenlegi jogi keretrendszerünket felülvizsgálni.

1.2. Jogi háttér

2024 augusztus 1-jétől hatályba lépett egy új jogi keretrendszer az AI fejlesztésével és alkalmazásával kapcsolatban, mely azt hivatott szabályozni, hogy ezen rendszerek használata biztonságos és transzparens maradjon a jövőben is. Ennek értelmében, fokozatos bevezetéssel, legkésőbb 2026. augusztus 2-ig (bizonyos fejezetek kivételével, melyeket már korábban is) alkalmaznia kell közvetlenül valamennyi EU tagállamnak. A jogszabály egyértelműen megfogalmazza, hogy milyen tiltott gyakorlatokat utasít el, mint például olyan rendszerek fejlesztése, amelyek megnehezítik adott esetben a döntéshozatalt, vagy amely bizonyos személyeket szociális helyzetükből adódó sebezhetőségükön keresztül kihasználnak. Emellett természetesen megfogalmaz olyan követelményrendszereket is, melyek a felhasználói- és adatbiztonságot helyezik előtérbe, ugyanakkor kockázati szempontból rangsorolja is ezeket az eszközöket. (*Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations*

(EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), 2024)

1. táblázat - AI Act kockázati besorolásai

Kockázati kategória	Általános leírás	Követelmények	Vállalati alkalmazás
ELFOGADHATÓ	Alacsony kockázattal járó AI rendszerek, amelyek számára nincs különleges szabályozás.	Az etikai irányelvek betartása	Általános információs rendszerek, marketing automatizálás
KORLÁTOZOTT	AI rendszerek, amelyek esetében érvényesek bizonyos kötelező követelmények	Egyedi, rendszerre szabott követelmények. (pl. átláthatóság, naplózhatóság)	Döntéstámogató rendszerek, amelyek kiértékelések alapján ajánlásokat adnak
MAGAS	AI rendszerek, amelyek jelentős biztonsági kockázatot jelentenek az emberi jogokra vonatkozóan.	Megfelelés a tesztelési és minősítési követelményeknek. Rendszeres auditálás és felülvizsgálat.	Vállalati döntéstámogató rendszerek, amelyek kritikus döntéseket hoznak (pl. pénzügyi elemzés)
ELFOGADHATATLAN	AI alkalmazások, amelyek veszélyeztetik az emberek jogait, és az AI Act által tiltásra kerülnek.	Nincs engedélyezett alkalmazás.	Munkavállalói személyes adatgyűjtést, profilozást végző rendszerek.

Forrás: Saját Szerk..

Látható tehát, hogy szinte valamennyi kockázati csoportba sorolva láthatunk üzleti céllal, vagy gazdasági vállalkozások által különféle területeken használt AI rendszereket, ezért fontos követelményként tekintünk innentől kezdve a jogharmonizációra az Európai Unió valamennyi tagállamában. Jóllehet, természetesen a világ többi országában is elkezdődött valamiféle jogi formula kidolgozása a felelős és megbízható AI megoldások keretezésére, azonban ezek javarészt elkülönülő, országokon belüli jogszabályok, vagy ennél is fragmentáltabb, akárcsak az Egyesült Államokban, ahol ugyan pl. California állam már 2024. szeptember 19-én szintén elfogadott egy – az európaihoz nagyon hasonló tartalmú – jogszabályt, a Mesterséges Intelligencia Átláthatósági Törvényt, mely szintén legkésőbb 2026-tól alkalmazandó, a szövetségi szabályozás egyelőre még nem érte utol az ipari forradalmat ebben a tekintetben. (California AI Transparency Act., 2024)

Az Egyesült Államokbeli Nemzeti Szabvány és Technológia intézet (NIST) irányelveket fogalmazott meg a mesterséges intelligencia alapon működő technológiák tervezésének, fejlesztésének és irányításának biztonsága érdekében, melyben hangsúlyozza az analitika, a szabványok és az kockázatértékelési eszközök fontosságát annak érdekében, hogy a kialakuló

rendszereket az elszámoltathatóság, az átláthatóság, az adatvédelem fokozása és a méltányosság jellemezze. Joe Biden 2023. október 30-i rendeletével összhangban a NIST célja az, hogy erősítse az Egyesült Államokban a mesterséges intelligencia kockázatainak mérséklését, miközben előmozdít egy olyan innovatív ökoszisztémát, amely védi a fogyasztókat és támogatja a munkavállalókat. A NIST kezdeményezései, amelyeket a NIST AI Innovation Lab és az Egyesült Államok Mesterséges Intelligencia Biztonsági Intézete vezet, magukban foglalják az alapkutatást, az alkalmazott mesterséges intelligencia kutatást, valamint a technikai szabványok és kockázateértékelési szempontok kidolgozását, különösen bizonyos szociális esetekben való alkalmazásának biztonsága érdekében. (Biden, 2023)

Csakúgy, mint a kaliforniai rendelkezések, a NIST ajánlásai összhangot mutatnak azzal a törvénytervezettel, mely 2022 óta van a döntéshozók asztalán és azokat a célokat foglalja össze, amelyeket követelményként kell támasztani az AI alapú automatizált rendszerekkel szemben. Öt irányelv köré épül: biztonságos és hatékony rendszerek biztosítása, az algoritmus általi diszkrimináció megakadályozása, az adatvédelem védelme, az automatizált rendszerek működésének egyértelmű tájékoztatása és magyarázata, valamint az emberi alternatívák biztosítása a döntéshozatalhoz. Ezek az elvek a nyilvánosság jogainak védelmére szolgálnak, biztosítva, hogy a technológiai fejlődés ne kockáztassa az emberi alapjogokat. (*Blueprint for an AI Bill of Rights*, 2022)

2023. augusztus 15-én életbe lépett Kína első adminisztratív AI szabályozása a generatív mesterséges intelligencia szolgáltatások számára, amelyet több kormányzati szerv közösen alkotott meg. Ezen intézkedések célja a generatív AI egészséges fejlődésének és szabályozott alkalmazásának elősegítése, a nemzeti biztonság és a közérdek védelme érdekében. Ez a szabályozás minden olyan szervezetre vonatkozik, amely Kínában a jogszabály keretei közé tartozó szolgáltatásokat nyújt, függetlenül attól, hogy melyik országban üzemeltetik azt. A legfontosabb megfelelőségi követelmények közé tartozik a jogszerű adatfelhasználás, a tartalom moderáció és a felhasználóvédelemért való felelősség. A szolgáltatóknak be kell tartaniuk a szocialista alapértékeket és létre kell hozniuk panaszkezelési mechanizmusokat is. A szabályozási és jogalkalmazási hatáskör közvetlen gyakorlója a Kínai Kiberbiztonsági Igazgatóság, felügyeli a megfelelőséget, és büntetést szab ki a jogsértések esetén, a meglévő törvényekben meghatározottak szerint. (Sheehan, 2023)

Tehát fontos követelményként kell figyelembe vennünk azt is, ha magyarországi vállalatunk bármilyen AI platformot használ, az megfeleljen az AI Actban kifejezetten és ne legyen jogsértő az üzemeltetése, természetesen azzal, hogy mint szolgáltatást igénybe vevő szervezet ez ugyanannyira az üzemeltető felelőssége is, csakugyan ahogyan az EU többi tagállamában is

érvényesek és alkalmazandóak lesznek ezek a követelmények. Azonban, amennyiben egy multinacionális vállalatnak már nem elegendő sok esetben a székhelyük szerinti adott országban egyeztetni ezt, hanem minden olyan területen is, ahol a cég működik valamilyen formában, azaz exporttevékenysége, vagy leányvállalata van. Ezért fontos számára a körültekintés, hiszen ma már egyre több ország, vagy országcsoport szabályozza az AI platformok használatát, a korábban említettek mellett Oroszországban, Kanadában, Japánban, Ausztráliában és Indiában is találkozhatunk hasonló jogi keretekkel. (Corrêa *et al.*, 2023; Data and AI Team, 2023)



5. ábra - AI szabályozás kezdete a világ országaiban

Forrás: Saját Szerk..

2. Adatvagyon-gazdálkodás

Az adatvagyon-gazdálkodás fogalmának tisztázásához szükségünk van arra, hogy tisztában legyünk nagy vonalakban a hagyományos adatelemzés módszereivel és felhasználásával, hogy lássuk azt, hogy a jelenleg tárgyalt kérdések alapjai a múltban már messzebbre gyökereznek, a jelenlegi ipari forradalmak során pusztán összegezzük eddig megszerzett ismereteinket, ennek a folyamatnak a során csupán olyan hozzáadott értéket teremtünk, mely annak lezajlása után az iparági standard részét fogja képezni.

A klasszikus adatelemzés a statisztika és a matematikai modellezés eszközeit használva foglalkozik az adatok értelmezésével és következtetések levonásával. Célja az információk kinyerése, az adatok mögötti mintázatok és összefüggések feltérképezése. Az adatelemzés alapelvei közé tartozik az adatok minősége, a reprezentativitás, az érvényesség és a megbízhatóság. Az adatok minősége kulcsfontosságú, mivel a hibás vagy hiányos adatok torzíthatják az eredményeket, a reprezentativitás pedig biztosítja, hogy a mintavétel megfelelően tükrözze a teljes eseményteret, míg az érvényesség és a megbízhatóság a mért adatok pontosságát és konzisztenciáját jelenti. Az adatelemzés módszerei közé tartoznak például a leíró számadatok, a következtetési statisztikák, a korrelációs elemzések és a regressziós modellezés. A leíró számadatok, mint például az átlag, a medián és a szórás, segítenek az adatok jellemzésében. A következtetési statisztikák lehetővé teszik a populációkra vonatkozó következtetések levonását az adott minta alapján, például hipotézisvizsgálatok formájában. A korrelációs elemzések célja az adatok közötti kapcsolatok feltárása, míg a regressziós modellezés a függő és független változók közötti összefüggések modellezésére szolgál. (Ács *et al.*, 2023)

A klasszikus adatelemzés megtalál bennünket akkor, amikor elolvassuk egy cikket arról, hogy mennyit változtak például a lakásárak, de akkor is, amikor éppen az Árukeresőn bogarászunk, hogy hol a legolcsóbb a kedvenc kávéfőzőnk. A vállalati életben adatelemzést használunk piackutatás során, az ügyfélkiszolgálásban, a pénzügyi, illetve gyártás tervezésben, a logisztikában. (Field, 2013)

Az Ipar 4.0 idejében az adatelemzés módszerei, alapelvei nem változtak meg, csupán az eszközei, a sebessége és a kapacitása. Ez önmagában lehetőséget ad a vállalatnak arra, hogy az adatelemzési időt lerövidítse, vagy az adatelemzésbe bevont adatok mennyiségét növelje. Összehasonlítva a Deep Blue rendszert a WatsonX rendszerrel, beláthatjuk, hogy ma már

lehetőségünk van rendkívül összetett és nagy méretű adathalmazokkal dolgozni, így a rendszertől közel maximális pontosságot kaphatunk, ha a megfelelő adatokkal tanítjuk be.

2.1. Az adattömeg minősége

A legjobb minőségű adatok az úgynevezett *strukturált adatok*, melyek definiált formátumban érkeznek, például táblázatokban, vagy relációs adatbázisokban, ahol minden egyes adatmezőnek pontosan meghatározott funkciója és minősége van, például szöveg, szám vagy időpont. (Weber, 2024) Előnyük, hogy gyorsan és könnyen elemezhetők, mivel az adatok mindig ugyanabban a formátumban keletkeznek, és az közöttük lévő kapcsolat is világosan definiált. Akárcsak az egyik első példánkban az időjárás szenzorunk esetében, tegyük fel, hogy az adattovábbítás óránként történik, ahol a műszer percenként regisztrálta a szélerősséget és a hőmérsékletet, ezeket egy táblázat formájában, a kellő jelölésekkel látta el, megfelelően fejlevezve, így az értelmező algoritmus számára egyértelműen azonosítható, mert egyezik az ő általa ismert fejlecekkel. (Olanrele *et al.*, 2022) Ezeket az adatokat leggyakrabban az ML és DL algoritmusainak feltanítására használják, mivel ezek pontos és tiszta adatok, tehát a legkönnyebben képes az algoritmus felismerni a saját alkalmazását és alkalmazkodási szükségleteit. Ennélfogva ezek az adatok az adatkezelés egyik alappilléreinek számítanak, mert könnyen rendszerezhetők, illetve elemzésük során a vállalatok képesek gyorsan és hatékonyan lekérdezni az információkat. Az úgynevezett Structured Query Language (SQL) nyelv lehetővé teszi az adatok egyszerű lekérdezését, szűrését és aggregálását, azáltal, hogy egyszerű parancsok révén a felhasználó képes komplex kérdéseket feltenni az adatbázisnak, és azonnali válaszokat kapni. Ha üzemeltettünk már valamilyen weblapot, akár egy Joomla portálrendszerrel, vagy láttunk egy PrestaShop webáruházat, akkor nem ismeretlen a MySQL adatbázis, ahol azoknak a motorját és adatkapcsolatait találjuk, de például a legnépszerűbb adatvizualizációs eszközök, mint a PowerBI vagy a Tableau is ismeri az SQL-t.

Azonban a Big Data halmazokban nem mindig jelentkeznek az adatok ilyen minőségben, a legtöbb kép, vagy szöveges tartalom úgynevezett *félleg strukturált* adat. Az ilyen adatok nem követik a hagyományos relációs adatbázisok által megkövetelt szigorú struktúrákat, de tartalmazznak olyan elemeket, amelyek lehetővé teszik a feldolgozást. A félleg strukturált adatok példái közé tartozik az XML², JSON³, és más hasonló formátumok, amelyek széles körben

² Az XML egy általános célú leíró programnyelv, melynek elsődleges funkciója az interneten való adatmegosztás.

³ Az XML-hez hasonló, azonban JavaScript eredetű nyelv, mely általában csak kettő számítógép (szerver-kliens) kapcsolatban használatos.

elterjedtek az interneten és az applikációkban. Leginkább a webes szolgáltatások, az API-k⁴ és az adatcserélő platformok ismerik ezeket kiemelten, ahol az információk dinamikusan keletkeznek és változnak. Ha a webáruházunkba integráltunk már valamiféle számlázó szoftvert, vagy azonnali fizetési modult akkor találkoztunk hasonlóval. (Steurer and Kappe, no date) Ezek a rendszerek képesek ezeket az adatokat feldolgozni, elemezni és hasznos információkat kinyerni belőlük. Egyik legismertebb példája a Természetes Nyelvfeldolgozás (NLP), mely képes feldolgozni a szöveges adatokat, mint például az ügyfélvéleményeket, közösségi média posztokat és e-maileket, illet már a korábban tárgyalt Vanda, illetve Siri, Alexa és más hasonló hangzó szöveg felismerő rendszerek használják. (VANDA, ISTVÁN and ANDRÁS, 2024)

Mindemellett az olyan platformok, mint az AWS, Google Cloud és Microsoft Azure, kínálnak olyan szolgáltatásokat, amelyek lehetővé teszik a félig strukturált adatok tárolását, feldolgozását és elemzését a felhőben, így biztosítva skálázhatóságot és rugalmasságot. (Pêga, 2023) Emellett természetesen sok más szerepkör is jut nekik az adatpiacon, erről még később szó lesz, de korábban is megjegyeztem, hogy ma a legtöbb vezető innovátor nem önmagában használ GenAI, ML, DL rendszereket, hanem egymással különböző kombinációkban, többféle alrendszer együttes működése hozza létre azt a szolgáltatás csomagot, amit ezek a nagyvállalatok értékesítenek.

Ezen felül még széleskörben felhasználnak félig strukturált adatokat, olyan területeken, mint az például az adatbányászat, a fogyasztói viselkedéselemzés, beleértve a vásárlói interakciókat, a weboldalak látogatottságát és a közösségi média aktivitást; továbbá marketing automatizálási platformok, melyek felhasználói reakciók alapján automatikusan generálják és optimalizálják a marketing kampányokat, de akár itt felsorolhatóak a kockázatelemző rendszerek, pénzügyi, és piaci elemzések is. (Bell *et al.*, 2021)

Azonban mai tudásunk szerint a *strukturálatlan* adathalmazok jelentik a legnagyobb kihívást az AI rendszerek számára, mert nem rendelkeznek előre meghatározott struktúrával, és gyakran szöveges, képi vagy hanganyag formájában érkeznek. Például a közösségi médiában közzétett szövegek, e-mailek, videók és képek mind olyan strukturálatlan információk, amelyekből értékes információk kinyerhetők, azonban nem minden Business Intelligence szolgáltatás képes ezt kezelni, azonban néhány azok közül, akik már félig strukturált adatokkal is képesek dolgozni, a teljesen strukturálatlan adatokat is kezelni tudják, azonban ezen művelet

⁴ Alkalmazásprogramozó interfésznek nevezzük azt a paraméterlistát, mely segítségével egy felhasználó vagy egy szoftver egy másik rendszer programozására képes annak különösebb belső háttérmotorjainak ismerete nélkül is.

végrehajtásához mindig emberi felügyelet szükséges és feltanításnál nem megengedett a bevitele. (Pêga, 2023)

Tehát, szokták mondani, nem mind arany, ami fénylik, de annak is egy részét el lehet úgy adni. 2024-ben ott tartunk, hogy bár van nehézség az adatgyűjtést és adattisztítást illetően, de a fentiek alapján úgy tűnik, hogy ma nincs akadály arra vonatkozóan, hogy milyen létező adatot ne gyűjthetne be és abból ne tudna szintetizálni. Viszont, hogy ez az adatmennyiség számára éppen elegendő, sok vagy kevés, az adatok minősége kellően kielégítő, jó, vagy kiváló, mint ahogyan már említettem az AI rendszerek esetében is, az egyéni vállalati kereteken múlik.

2.2. Adatvagyon-gazdálkodás szükségességének elve és annak alapelvei

Az adatvagyon-gazdálkodás meghatározza, hogyan kezeljük, tároljuk és használjuk fel az adatokat vállalati, állami, de talán a későbbiekben még alacsonyabb mikro szinteken is. 2024-ben a cégek és különböző szervezetek egyre nagyobb mennyiségű adatot gyűjtenek és elemeznek, amelyek megfelelő kezelése elengedhetetlen a versenyképesség megőrzéséhez. Az adatvagyon-gazdálkodás célja nem csupán az adatok rendszerezése és menedzselése hanem azok értékesítése is, természetesen az árszabás már egy egészen összetett folyamat, lévén, hogy az adat egy korlátlanul újratehermthető erőforrás, illetve az értéke általánosságban nem meghatározható, mert nagyban függ a kontextustól, hiszen elég csak belegondolnunk, ha én vállalkozóként pizzát árulok egy budapesti aluljáróban, nem különösen értékes a számomra a Miskolcot elkerülő autópálya forgalma, viszont ha éppen azon gondolkodom, hogy egy felsőzsolcai benzinkút nyitásán töröm a fejemet éppen, akkor szintén nem izgat, hogy Budapesten mennyi tömegközlekedési bérletet adnak el évente. Viszont ahhoz, hogy a rendelkezésemre álló adattömeget folyamatosan karbantartsam, a kellő mértékig gyűjtsem és ha szükséges akkor értékesítsem, akkor már kimondható, hogy adatvagyon-gazdálkodási tevékenységet végzek még ilyen mikrovállalati szinten is. Természetesen minél nagyobb a vállalat mérete és/vagy az általa kezelt adatok mennyisége, így az adatvagyon-gazdálkodási tevékenység is szükségszerűen nagyobb figyelmet követel magának. A jól megtervezett adatvagyon-gazdálkodási stratégia pedig lehetővé fogja tenni az adatokban rejlő lehetőségek maximális kiaknázását. (Corrêa *et al.*, 2023)

2.3. A kormányzás alapelvei

Az első alapelv az úgynevezett *data governance*, azaz adatkormányzás, amely az adatok rendelkezésre állását, védelmét és felhasználását irányítja, a rendszer feje, az adatkormányzó személy irányításával. Ez a keretrendszer biztosítja, hogy az adatok kezelésére vonatkozó

policy-k világosak legyenek, ezáltal megelőzve a jogi és etikai problémákat. A jogi háttér formálódóban van jónéhány esztendeje, elég csak visszagondolnunk a GDPR-ra, vagy az első fejezetben tárgyalt rendelkezésekre. Ahogy haladunk előre az Ipar 4.0-ban, amint egyre komplexebb és szabályozottabb lesz a vállalatok adatvagyon-gazdálkodási struktúrája, egyre több adatkormányzó csapatot fognak alkalmazni a cégek, ma is már jónéhány data governor pozíció is fellelhető az ismert karrierportálokon. Külön-külön minden egyes vállalatra érvényes kormányzási stratégiát nem lehet ismertetni, hiszen abból fakadóan, hogy minden BI rendszert a vállalat a saját adataival tanít be és a saját koncepciójának megfelelően irányít, az adatvagyon-gazdálkodási módszerét így annak fő irányítási mechanizmusait sem lehet általános sémával lefedni. Azonban léteznek olyan alapvető megfontolások, amelyeket belátva felállíthatunk egyfajta adatkormányzási jó gyakorlatot. Az első és a legfontosabb maga az, hogy legyen egy adatkormányzó, és egy stratégiai célpont, amit követ és annak megfelelően alakítja a gazdálkodási folyamatot, betartatja a többi alapelvet és megalapozza az adatalapú döntéshozatal sikerességét és hitelességét, a transzparencia biztosítása által (Sargiotis, 2024). Létezik egy olyan megközelítés is, mely szerint az adatkormányzás lehet decentralizált is, ahol ugyan szintén több helyről érkezik az adatfeldolgozások eredménye, azonban ezt nem egy központi adatkormányzó látja el, hanem minden egyes kiosztott szegmensben egy-egy adatfelügyelő kolléga dolgozik és felelős a saját területéért, illetve látja el a gazdálkodási tevékenységet, a legszélsőségesebb eset az abszolútdemokratikus vagy önkormányzás, ahol minden egyes dolgozó hozzáfér minden információhoz és ráhatással lehet a kormányzási folyamatra. (Munroe, 2022) Ezeket a modelleket különösképpen nem tárgyalnám, mert véleményem szerint a hatékonyságot rombolja. A második ilyen fontos gyakorlat, hogy az adatok hozzáférhetősége a megfelelő szinten és nyelven jelentkezzen, legyen kellően releváns és elégséges, illetve ennek a fordítottja is, tehát ahol nem szükséges, vagy egyenesen káros, ott semmilyen hozzáférés ne keletkezzék, magyarul mondva, az adatok hozzáférhetőségének és biztonságának biztosítása alapvető kormányzói feladat. (Tibor, 2023)

A második alapelv az adatminőség biztosítása. Az adatok megbízhatósága és pontossága kulcsfontosságú a helyes döntések meghozatalához. A hibás vagy hiányos adatok torzíthatják az eredményeket, ezért a folyamatos ellenőrzés és felülvizsgálat is elengedhetetlen. Ehhez szükségessé vált az elmúlt években a szabványosított adatbeviteli eljárások és validálási technikák alkalmazása, hogy elkerüljék a lyukas, pontatlanul bevitt, vagy duplikált adatokat.

A harmadik alapelv az adatoknak a rendszerbe való integrálása, az az a centralizált adatkezelés elve. (Mint már említettem ez az alapelv nem minden esetben érvényesül, hiszen alapvetően a szoftveres háttér, pl. Tableau lehetővé teszi, de nem feltétlenül alkalmazandó)

(Tableau Support, 2024) Az adatok különböző forrásokból származhatnak, például ügyféltől, különböző BI algoritmusokból, egyéb szoftveres alkalmazásoktól vagy piackutatásokból, különböző részlegekről, ezeket továbbítva félreértések, alul reprezentált helyzetek keletkezhetnek. Ahhoz, hogy egy hatékony adatvagyon-gazdálkodási ökoszisztémát tudjunk üzemeltetni vállalatként szükségünk van arra (már a szoftveres követelmények is ezt várják el tőlünk), hogy az általunk használt keretrendszer számára értelmezhetővé és központivá tegyük a kezelt adatokat. Az egységes formátumban és rendszerben történő tárolás és kezelés drámaian megkönnyíti az adatok közötti interakciót és integrációt. Leegyszerűsítve, meg kell adnunk azt a kontextust a rendszer számára, amely segítséget nyújt neki, hogyan értelmezze ezt az eddig rendelkezésére álló adattömeg viszonyaiban.

A negyedik alapelv a biztonság. Az adatvédelem kiemelkedően fontos, különösen a személyes és érzékeny információk esetében. A titkosítás, a naplózás és a hozzáférés-ellenőrzés, segítenek megakadályozni az adatszivárgást és a jogosulatlan hozzáférést. Például a felhő alapon működő adatelemző eszközök (AWS, Azure, Tableau), melyek lehetővé teszik az adatok hatékonyabb kezelését és feldolgozását, rendkívül magas szintű biztonsági intézkedéseket alkalmaznak az adatvédelem tekintetében, emellett rugalmasságot is nyújtanak, mivel a vállalatok könnyen bővíthetik kapacitásaikat az adatkezelési igények növekedésével. A vállalatoknak szükségük van egy átfogó kockázatkezelési stratégiára, amely lehetővé teszi az adatok védelmét és a potenciális problémák előrejelzését. Ezzel egy időben persze folyamatosan frissíteni kell a biztonsági protokollokat is, hogy lépést tudjanak tartani a mindig változó fenyegetésekkel, ezért egy saját körben működtetett rendszer ilyen módon is kockázatokat hordoz, amikkel számolni kell, természetesen a felhőszolgáltatást igénybe vevő szervezetek válláról átkerül ez a teher a felhőszolgáltatóra, hiszen amennyiben ő biztosítja a felületet a vállalat számára, onnantól a biztonsági kockázat jelentős része az ő kockázata. (IBM Cognos Analytics, 2024)

Ugyanakkor az adatmegosztási politika is fontos szerepet játszik az adatvagyon-gazdálkodásban. A vállalatoknak világosan meghatározott irányelvekkel kell rendelkezniük az adatok megosztására vonatkozóan, hogy megvédjék a bizalmas információkat. Az adatok megosztása lehetőséget teremt a kollaborációra és az innovációra, de csak akkor, ha ez biztonságos keretek között történik. A megfelelő megosztási politika segít elkerülni a jogi problémákat és az adatokkal kapcsolatos etikai dilemmákat. (*Data Governance Act*, 2022)

Az ötödik alapelv a fenntartható adatkezelés. Ez biztosítja, hogy az adatok hosszú távon is hasznosak és relevánsak maradjanak. Magában foglalja az adatok archiválását, valamint a régi – már nem releváns – adatok eltávolítását, megfontolandó értékesítését. Természetesen itt nem

a sokszor felmerülő energetikai fenntarthatóságra gondolok, amellet, hogy az energia kérdése is egy szűk keresztmetszet lehet a 2030-as évekre, (Goldman Sachs Research, 2024) alapvetően itt a fenntartható adatvagyon-gazdálkodás alapelve inkább azt hivatott leírni, hogy az egyre növekvő adatmennyiség ne csak egy folyton növekvő kupac legyen, hanem egy rendezett és értékes „aranybánya”.

A hatodik alapelv a felhasználói képzés. Az adatok kezeléséhez szükséges készségek és ismeretek átadása elengedhetetlen a sikeres adatvagyon-gazdálkodás érdekében. A munkavállalók oktatása és képzése alapvető szükséglet, hogy a szervezet minden – adatvagyon-gazdálkodásban érintett – tagja tisztában legyen az adatvagyon-gazdálkodás fontosságával és az adatok helyes kezelésének módjával. (Tableau Support, 2024) Rendszeres képzések és workshopok szükségesek a dolgozók adatkezelési kompetenciájának fejlesztésében. A döntéshozatal során pedig az adatok elemzése segíti majd a vezetőket a megalapozott döntések meghozatalában, persze akkor és csak akkor, ha egyértelmű lesz számukra is, hogy azt látják, amit az adatok mutatnak, ez is hangsúlyozza a képzés fontosságát, hiszen az adatalapú döntéshozatal csökkenti a kockázatokat, és növeli az üzleti lehetőségeket, ha nem csak statikus adatként, hanem dinamikus eszközként tekintünk rá. (Tibor, 2023)

2.4. Az adat áru funkciója

Az adat a modern gazdaság egyik legfontosabb árucikkévé vált, amely jelentős ráhatással van az ágazatokban a piaci versenyre. Gyűjtése, tárolása, feldolgozása és értékesítése lehetőséget biztosít a cégek számára, hogy optimalizálják döntéshozatali folyamataikat, és növeljék versenyképességüket. A digitális forradalom 2024-ben már elindította az ágazatokat abba az irányba, hogy hatalmas mennyiségű adatot gyűjtsenek és elemezzenek. Big data adathalmazok elemzése révén a cégek képesek komplex összefüggéseket és mintázatokat megtalálni. (Benlachmi and Hasnaoui, 2020)

Az adatok kereskedelme is drámaian bővül, különösen a marketing és hirdetési szektorban, ahol a vállalatok a vásárolt adatok alapján célzott kampányokat indíthatnak. Az adatkereskedelem új üzleti modelleket teremt, amelyek révén a cégek új bevételi forrásokat generálnak. Ugyanakkor a személyes adatok védelme is egyre fontosabbá válik, a GDPR és más adatvédelmi jogszabályok kötelezővé tették a cégek számára, hogy felelősségteljesen kezeljék a személyes adatokat, figyelembe véve az ügyfelek jogait és elvárásait, melynek a betartása nemcsak kötelezettség, hanem a fogyasztói bizalom megteremtéséhez is hozzájárul. (Mohammad Amini *et al.*, 2023)

Léteznek ma már olyan vállalkozások, amelyek bevételüket kizárólag adatok gyűjtéséből, vásárlásából és értékesítéséből szerzik, ezeket hívja a köznyelv *adatbrókereknek*. A piacon az egyik piacvezető adatbróker, az Experian, mely egy 1996-ban, Dublinban alapított vállalat, a 2023-2024-es években 6,6, illetve 7,1 millió dolláros árbevételre tett szert pusztán az adatok adásvételéből származó tevékenységgel. (Experian, 2024) Kapcsolatban vannak további adatelemző és statisztikai felméréseket végző cégekkel, vannak kormányzószerzi felkéréseik, jellemzően Észak és Dél-Amerikából, az Egyesült Királyságból és Írországból. De nem csak ő az egyetlen cég, aki hasonló tevékenységekkel foglalkozik. Az Equifax, az Epsilon, a CoreLogic, az Acxiom mind olyan vállalatok, amelyek adatok kereskedelméből és strukturálásából profitot realizálnak. De a kérdés, hogy ezek az adatok honnan származnak? Különböző ingyenesen elérhető statisztikákból (például nagyvállalatok éves pénzügyi beszámolóí, a különböző államok statisztikai hivatalainak adatbázisai, egyéb nyílt adatbázisokból, pl. a Tableaunak is van ilyen nyílt adatbázisa), ezen adatok hátrányai, hogy vagy aktualitásukat veszítették, vagy strukturálatlanok, esetleg hiányosak vagy pontatlanok. Tehát a minőségük nem éri el az értékesíthetőség minimumszintjét, vagy jogilag olyan közadatnak minősülnek, amelyeket a jogszabályi környezet nyilvános elérhetőségre kötelez. (305/2005. (XII. 25.) Korm. rendelet) Nézzük meg azt az esetet, ha szeretnénk egy egészségügyi termékekkel foglalkozó nagykereskedő vállalatot indítani és a vásárlói szokásokat, a piaci keresleteket és tendenciákat, költséghatékonysági okokból ingyenesen beszerzett 2020-as adatokkal szeretnénk modellezni, valószínűleg azt látnánk, hogy komoly kereslet mutatkozik véroxigénszint mérő készülékekre és szájmaszkokra, tudjuk már csak a saját szokásaink alapján, hogy 2024-ben már nem ez a helyzet és igencsak torzított értékeket kapnánk, ha egy ilyen adatsor alapján szeretnénk mintázatokot szintetizálni.

De még egy megjegyzés a relevanciáról: Gondoljunk csak bele, ha szeretjük a Forma-1-et és aktuális adatokra vagyunk kíváncsiak, akkor számunkra melyik adat az értékesebb: Az, hogy kedvenc csapatunk jogelődje a 2019-es szezonban milyen pilóta felállással versenyzett, vagy az, hogy a hétvégi futamon a két pilóta milyen esélyekkel vág neki a versenynek, esetleg az, hogy mi lesz a jövő évi magyar nagydíj végeredménye? Ha ez utóbbit most novemberben megtudhatnánk, minden bizonnyal a legértékesebb információ lenne bárki számára, egyrészt aki valaha kacérkodott sportfogadásokkal, vagy pusztán csak rajongója a sportágnak, de lássuk meg itt az esetet, hogy a jövőről szóló információt pénzzé lehet tenni. Hasonló alapelv működik az adatpiacon is, ahol az adatbróker olyan információkat szolgáltat, amelyeket megfelelően kiaknázva a jövőről információkat szerezhet a vásárló, esetleg olyan információkat a jelenről, melyek nagy eseményteret fednek le, széleskörben valid információkat hordoznak.

Illetve származnak adatok különböző internetes honlapokról (Hány sütit fogadtunk el ma reggel?), mobil alkalmazásokból, játékokból, vásárlói hűségkártyákból, illetve közösségi média felületeken fellelhető profiladatokból. A közös nevező ez esetekben egyébként a kapcsolt felhasználói fiókok esete, például amikor a kedvenc áruházláncunk pontgyűjtő kártyáját a Google vagy a Meta fiókunkkal kötjük össze. (Meta Platforms, 2024) (Yen, 2018)

Ennek olvasatában értelmezhető, hogy az adatoknak, mint jószágnak az árszabása igencsak heterogén és komplex. A legfontosabb ármeghatározó tényezők közé tartozik egyébként az *adatminőség*, a *relevancia*. Ez utóbbi már lefedi az aktualitás kérdését is. Ezek alapján már nagyjából felállítható egy közelítőleges matematikai összefüggés arra vonatkozólag, hogy milyen áron lehet értékesíteni különböző adathalmazokat. A modell feltételezi azt, hogy az adatgazdálkodó vállalat profitmaximalizáló és az adatgazdálkodási tevékenysége során is ezt az elvet követi. Továbbá a második modellfeltétel a korlátozott költségvetés feltétele, azaz a rendelkezésre álló tőke jellegű források mennyisége meghatározott. Figyelembe veszi a modell az eladási és a vételi hajlandóságot, illetve az adatminőséget is egy változóval jellemzi. (Park, Oh and Choi, 2023)

2.5. Adatpiacok felépítése és szereplői

Az adatpiacok olyan platformok, ahol az adatokat kereslet és kínálat alapján lehet vásárolni, eladni vagy cserélni. Ezen piacok középpontjában azon adatok értéke áll, amelyeket a fentebb részletezett módon gyűjtenek az adatbrókerek. Az adatpiacok működése szoros kapcsolatban áll a technológiai fejlődéssel, mivel az adatok feldolgozása és elemzése során használt technológiák, mint az ML, DL algoritmusok, folyamatosan fejlődnek. Az adatpiacok egyik alapelve a transzparencia, hiszen szükséges a felhasználók számára, hogy megértsék, honnan származnak az adatok, és hogyan használják őket. Ez növeli a felhasználók bizalmát, ami kulcsfontosságú a piaci dinamika szempontjából. A másik fontos alapelv a biztonság, amely garantálja, hogy az adatok védve legyenek a jogosulatlan hozzáféréstől, valamint, hogy a vásárló felhasználók adatai ne kerüljenek veszélybe. (Son8kavde *et al.*, 2023)

Az éves becsléseket szándékosan nem jeleztem, mert az adott években készült előrejelzések mindig alábecsülték a jövőbeni árbevétel-növekedést. (Grandview Research, 2024)

Az adatpiac főszereplői: az adatszolgáltatók az adatgyűjtők, és az adatfelhasználók. Az adatszolgáltatók lehetnek már a fentebb említett nyilvános adatbázisoktól a célzottan megrendelt felméréseken át egészen a magánszemélyekig szinte bárki, persze egy

adatfelhasználó is lehet adatszolgáltató, amikor adatszolgáltatásért cserébe adatszolgáltatást kap. Számos ilyen módon működő adatbróker működik, mint például az IQVIA, aki egyébként egy egészségügyi adatokkal kereskedő vállalat, (Candore *et al.*, 2020) sok minden más mellett szerződésben áll olyan kiskereskedőkkel, akiknek a részére a szortimentjükbe tartozó termékek valós keresleti adatait szolgáltatja, cserébe a kiskereskedő is szolgáltat be hűségprogramos vásárlásokat, termékek nagykerárait, stb... Az ilyen típusú megoldást szeretem *adathitelnek* nevezni, mert mindkét fél profitál a kölcsönös adatcseréből, viszont az egyik fél (nevezhetnénk akár adatbanknak is) profitszerzési tevékenysége kizárólagosan az adatkereskedelemre és az adatelemzésre specializálódik, tehát nem lehet pénzügyileg adatdeficites, ezért kizárólagosan úgy tudja a működését finanszírozni, ha az általa kibocsátott adatok nominális értéke kisebb (ebből persze adatgazdálkodási tevékenység során végzett szintézissel magasabb értékű adat is előállítható), mint az általa begyűjtött adat értéke, a begyűjtött adat feldolgozási és tárolási költségeivel együtt.

Az ilyen módon működő adatgyűjtők felelősek az adatok összegyűjtéséért, feldolgozásáért és hogy olyan platformokat biztosítsanak, ahol az adatok elérhetővé válnak. Az adatfelhasználók pedig különböző célokra használják fel az adatokat, például piackutatásra, termékfejlesztésre vagy üzleti döntéshozatalra. E három szereplő közötti interakciók és tranzakciók hozzák létre az adatpiacok dinamizmusát, azonban érdekesebbnek tűnik különválasztani egy-egy résztvevőt, hogy jobban átlássuk az adat körforgását, ahol is az adatnagyker (vagy más szóval *adatvender*) van tulajdonképpen a gráf középpontjában. Az adatok általában valamiféle tartalmat fogyasztó felhasználók hozzájárulásával keletkeznek a rendszerben, melyeket az adatgyűjtők rendszereznek és elemeznek ki, majd értékesítenek a nagyker felé, aki ezeket tovább értékesíti egy felhőplatformnak (mint például az AWS, vagy a Microsoft, akiknek szintén van egyébként saját adat nagyker platformjuk, gyakorlatilag az adatpiac szempontjából a vezető tech óriások az egész ellátási láncot magukba gyűrték, de külső cégektől is vásárolhatnak), aki pedig annak a vállalatnak értékesíti, akinek ez a számára megfelelő pénzbeli ellenértéken továbbadható, amellet, hogy a saját ML és DL algoritmusait tanítja vele. (Thompson and Warzel, 2022) Ezt a következő ábrán szemléltettem:

3. Következtetések, javaslatok

Ebben a fejezetben újra előveszem a dolgozat elején ismertetett kérdéseket, majd külön alpontokban fogom rájuk megadni a válaszokat.

3.1. Hol van a mesterséges intelligencia helye az adatvagyon-gazdálkodásban?

Erre a kérdésre, talán azt mondhatjuk, hogy a részfolyamatok száma véges sok, a teljes lista felvázolására idő szűkében nem kerítenék sort. De talán olyan szempontból hoztam szemléletes példákat az irodalomkutatás során melyek értően az olvasó számára megmutatják, hogy a generatív mesterséges intelligencia szintjén az összes többi, evolúciós szempontból alsóbbrendű rendszert használjuk. Tehát, ha egy vállalat szeretné a Meta platformjára készülő hirdetéseit a felhasználók preferált színpalettája szerint kódolni, akkor ezt a generatív szolgáltatás fogja neki elkészíteni, azonban, az adatok megszerzéséhez egy mélytanuló, a várható pozitív kimenetelű vásárlási döntések előrejelzéséhez pedig prediktív intelligenciát fog használni. Az alábbi táblázatban összeszedtem, hogy melyik funkciót melyik részterületek használják:

2. táblázat - AI használat a vállalatirányításban - Saját Szerk..

AI szintje	Használt terület	Vállalat
Valós idejű adatelemzés	Pénzügyi elemzés	Bloomberg
	Ügyfélszolgálat	Zendesk
	Kiberbiztonság	Darktrace
Prediktív AI	Logisztika	Amazon
	Kereslet-előrejelzés	Walmart
	Betegségek kockázati előrejelzése	IBM Watson Health
	Marketing kampányok	Netflix
Machine learning	Kockázatkezelés	Allianz
	Termékajánlások	Amazon
	Kép- és hangfelismerés	Google
	Szövegelemzés	Grammarly
Deep Learning	Teljesítményelemzés	Salesforce
	Önvezető autók rendszerei	Tesla
	Arcfelismerés	Meta
	Fordítás	Google Translate
Generatív AI	Képkalkáló diagnosztika	Siemens
	Szöveglétrehozás	MathGPT
	Képek létrehozása	DALL-E
	Zene alkotás	AIVA
	Játékfejlesztés	Ubisoft

3.2. Mit nevezünk adatgazdaságos vállalatnak?

Az előzőekben már láthattuk, hogy mennyi adattal szükséges dolgoznunk ahhoz, hogy akár egy adatmigráció, vagy pusztán egy optimalizálási folyamat végbe menjen, továbbá, hogy az adatok újra hasznosításának köszönhetően gyakorlatilag korlátlan mennyiségű adatot előállíthatunk, látható, hogy szükséges egyfajta határt szabnunk vállalati kereteken belül, hogy az adatokkal való elfoglaltságaink ne a profitszerző tevékenységünk kárára következzen be. Ezért szükséges megállapítanunk egy olyan pontot, ahol az adatvagyon-gazdálkodás még kellően optimális, viszont minden további intézkedésünk ezzel kapcsolatban olyan többlettőkét vonzana, amely közvetlenül nem hoz akkora profitnövekedést.

Ezek alapján az én olvasatomban, az a vállalat tekinthető hipotetikusán adatgazdaságosnak, amely éppen úgy gazdálkodik a rendelkezésére álló adatokkal, hogy ehhez képest bármilyen többletadatnak a felhasználásával vagy folyamat optimalizálással a számára már nem keletkezik többletprofit, ezt nevezem én az *adatgazdaságos vállalat* koncepciójának.

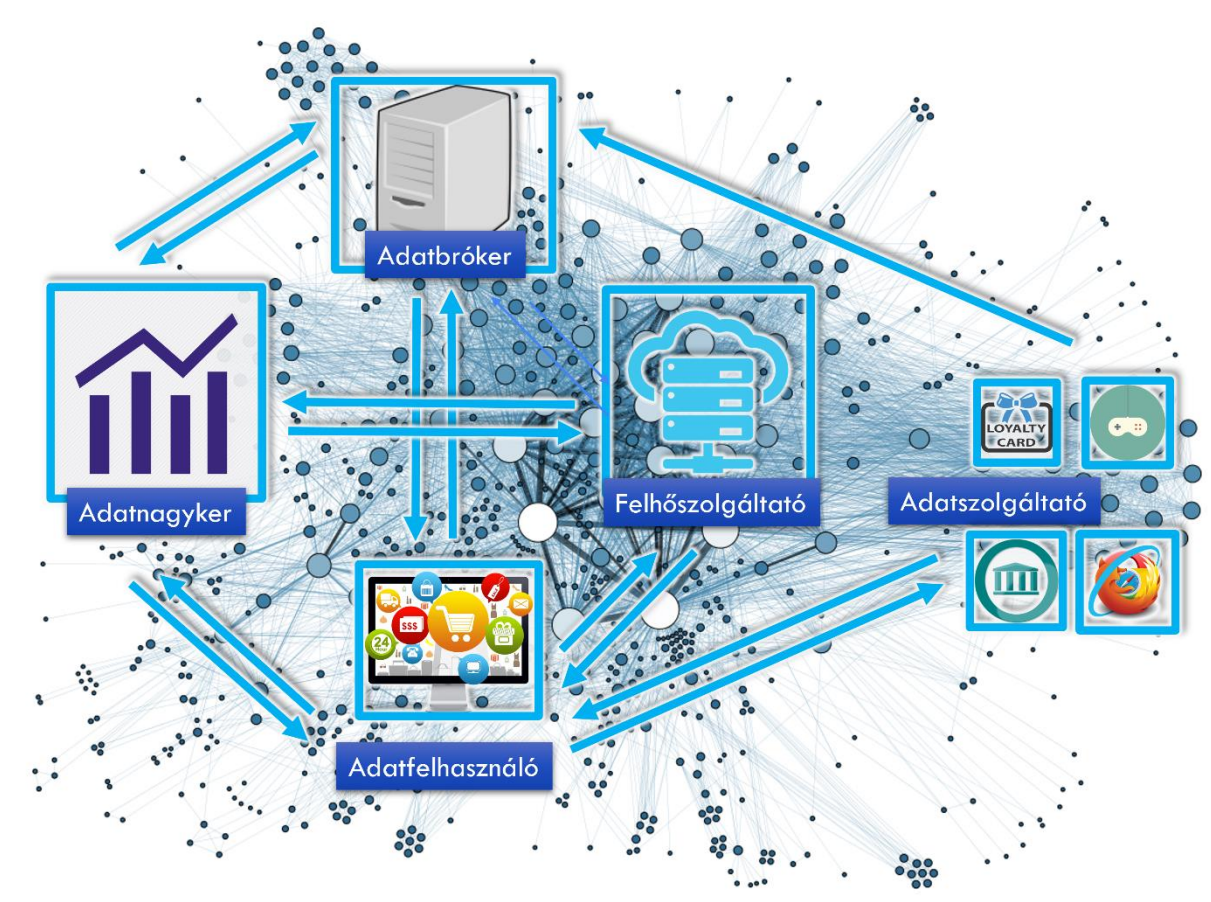
Természetesen 2024-ben még talán kimondható, hogy bőven nem tartunk azon a szinten, ahol egy vállalat már adatvagyon-gazdaságosan működhetne, de - mivel a rendelkezésre álló technológiai eszközeink segítségével exponenciálisan nő az adatpiac forgalma - ez idővel el fogja hozni azt a korszakot, ahol a vállalatok olyan mértékű adatot fognak kezelni, mely szükségessé fogja tenni, hogy figyelembe vegyünk ezt a pontot, ami körül a folyamatos adatáramlási dinamikáknak köszönhetően oszcillálni fog az adott vállalati adatvagyon. És nem más fogja ezt megszabni, mint a legtöbb esetben a piaci verseny. Hatékonyabb kezelésre ösztönzi majd a vállalatokat. Azok a cégek, amelyek jobban kezelik adataikat, lesznek képesek a leggyorsabban és leghatékonyabban reagálni a piaci változásokra (vagy már adott esetben a termékbe belegyártani). Viszont a dinamikus verseny miatt folyamatosan szükség lesz az adatvagyon stratégia ismételt átgondolására.

3.3. Mi az adatpiacok felépítése?

Az adatpiac szereplőinek többféle csoportosítási lehetőségét engedném meg az irodalomkutatás alapján. A háromszereplős modellben, csupán az adatszolgáltató, az adatkereskedő és az adatfelhasználó van jelen. Nevükből adódóan az egyikük szolgáltatja az adatot, az egyikük felhasználja az adatot és a kettejük közötti híd pedig az adatkereskedő. Mivel általában az adatpiaci relációk ennél bonyolultabbak, ezért célszerű legalább az ötszereplős piaci modellt alkalmazni véleményem szerint. Tehát, az adatszolgáltató, az adatgyűjtő, az adatkereskedő, a felhőplatform és az adatfelhasználó. Ez a csoportosítás valamelyest közelebb

áll a valós viszonyokhoz. Azonban fontos megjegyezni, hogy – bár a piac fragmentált – a technológiai óriások (AWS, Google, Microsoft), amelyek rendelkeznek olyan vállalatirányítási rendszerekkel, melyekbe AI szolgáltatások integrálhatóak, szintén rendelkeznek saját adatpiaccal, ahol kapcsolt értékesítés révén, például a döntéstámogató rendszert futtató felhőszolgáltatás keretein belül értékesíthetnek a számukra releváns adatokat, amellett, hogy természetesen külső forrásból, akár közvetlenül az adatgyűjtőktől, vagy az adatkereskedőktől is vásárolhatják.

Akkor mi a különbség adatgyűjtő és adatkereskedő között? A különbség árnyalatnyi, mégis jól érzékelhető, az adatpiac szempontjából kiemelten fontos. Mint ahogy említettem, rengeteg félig strukturált és strukturálatlan adat létezik, amelyek feldolgozása egy szemmél látható akadályt képez az adatelemzésben, habár a mélytanuló algoritmusok ebben sokat tudnak segíteni. Az adatgyűjtő, aki az egyes felhasználóktól gyűjti az adatokat és strukturálja azokat, tehát azok az adatgyűjtők, akik kedvenc internetes újságunk sütijeit osztogatják nekünk, ha azokat elfogadjuk, akkor bizonyos adatokat szolgáltatunk nekik a készülékünkön elhelyezett nyomkövető segítségével, majd azokat strukturálják, esetleg szintetizálják, mintázatokat és korrelációkat keresnek benne, majd – akár ezekkel együtt – értékesítik azokat.



6. ábra - Az ötszereplős adatpiaci modell - Saját Szerk.

Az adatkereskedő ilyennel már nem foglalkozik, ez a szereplő csupán adatokat ad és vesz, ahogy a tőzsdén egy befektető a részvényeit adja és veszi, azonban az adatkereskedő már nem foglalkozik legtöbbször azzal, hogy adatokat strukturáljon. Az adatpiac felépítésének öt szereplős modelljét a jellemző adatáramlási dinamikákkal a hatodik ábrán mutattam be.

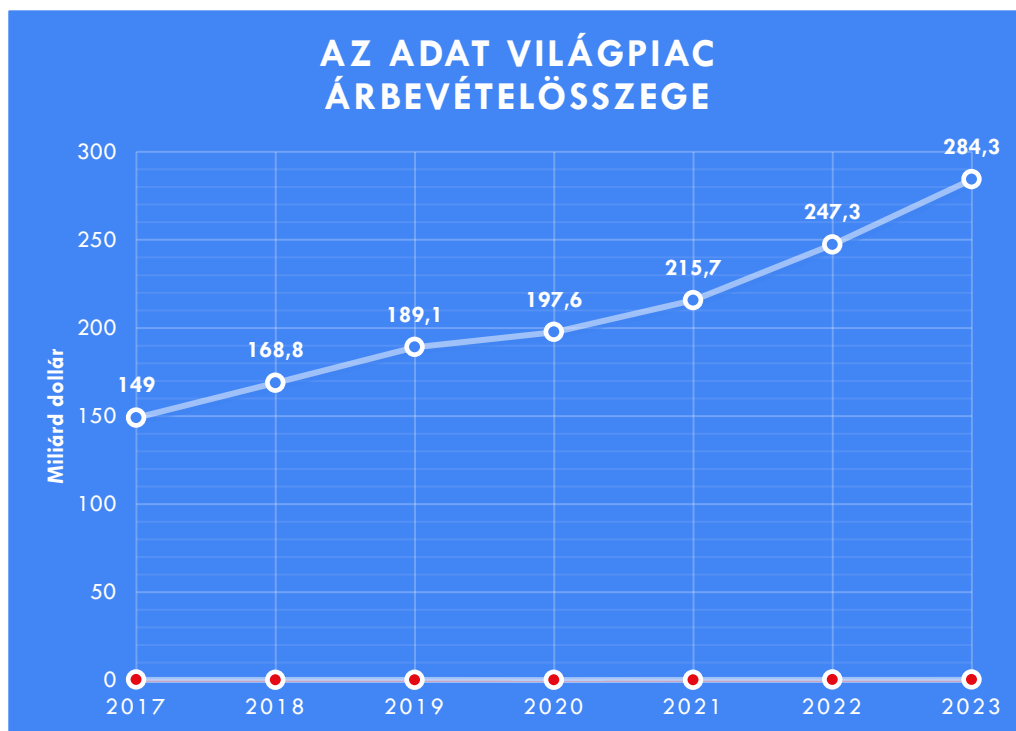
Innen is jól kiolvasható, hogy az adatpiacon leginkább a kétirányú adatáramok jellemzőek. Nem feltétlenül szükséges azonban, hogy az adatszolgáltató által a rendszerbe táplált adat pontosan végig járva az útvonalakat, minden kézen megforogjon, ahogy az sem, hogy pontosan ugyanaz az adat érkezzon vissza hozzá. Az adatfelhasználó és az adatszolgáltató kapcsolata némileg különbözik ebből a szempontból, hiszen az adatfelhasználó egy célzott értékesítés, vagy egy célzottan generált tartalom formájában juttat adatokat a szolgáltató felé. Gyakran előfordul, hogy ennek a tartalomnak a mérése, vagy a tartalomfogyasztási szokások monitoringja kapcsán direktben az adatszolgáltató és a közvetlen felhasználó van kapcsolatban. Leggyakoribb példaként az Amazon, Temu, Netflixet tudnám említeni, ahol a saját böngészési szokásaink alapján a saját célzott értékesítési ajánlataink vagy tartalmaink jelennek meg.

Szintén különleges a felhőszolgáltatók kapcsolatrendszere, hiszen ők direktben nem gyűjtenek adatokat a felhasználókról. Természetesen amint fent említett technológiai piacvezetők, akik az adatpiac minden szereplőjeként jelen vannak a rendszerben, nyújtanak Business Intelligence típusú szolgáltatást, viszont ebből a szempontból ők adatfelhasználók lesznek ebben a szerepükben, ha az adatgyűjtésről van szó, és többnyire ritka az, amikor egy átlagfelhasználó vásárlási előzményei direktben mondjuk az AWS BI rendszerében vannak eltárolva, nem pedig az adott üzletnek az AWS által üzemeltetett platformján, tehát ez esetben az adatok begyűjtése indirekten a szolgáltatótól, a szolgáltató-felhasználó útvonalon érkeznek a felhőszolgáltatóhoz.

3.4. Milyen pénzforgalmat realizál az adatpiac jelenleg? Mik az előrejelzések?

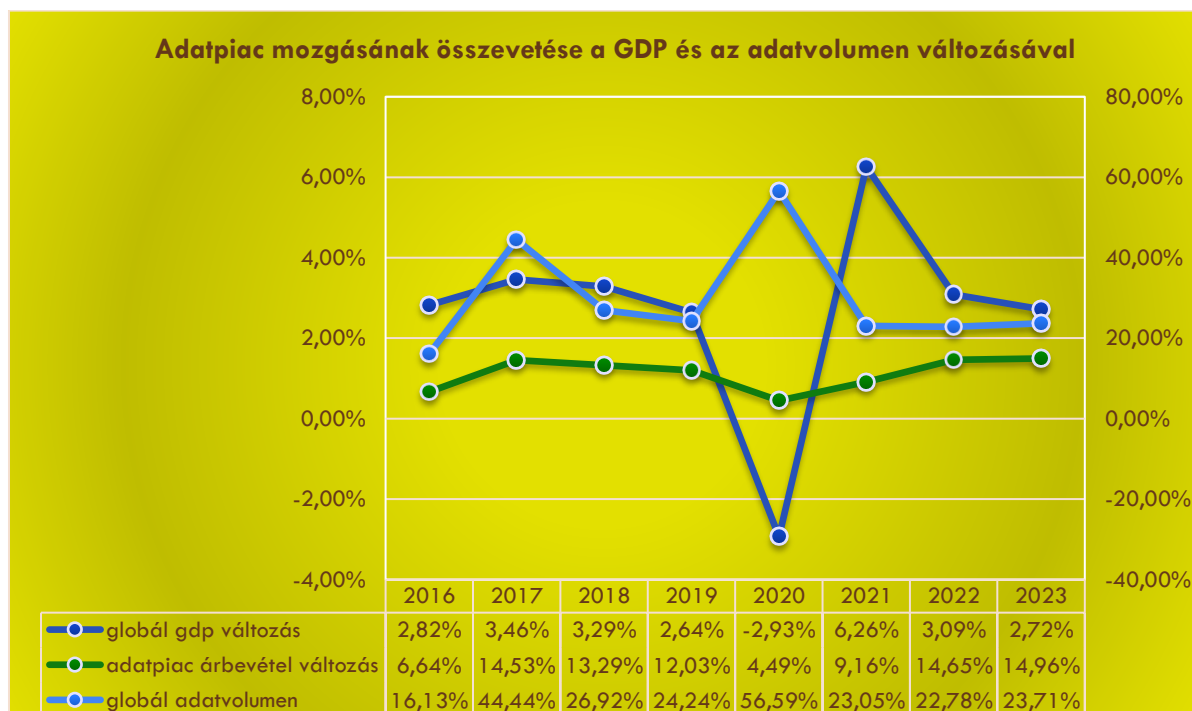
A következő ábrán szemléltetem a világ adatpiacának árbevétel tendenciáját az elmúlt években. További becsléseket azért nem tüntettem fel az ábrán, holott azt több hasonló elemzést végző platform megteszi egészen a 2030-as évekig, azonban a múltbeli predikciók rendre alulmúlták a valós teljesítményt, illetve az érintett időszakban a piacon kívüli tényezők (mint például

energetikai és jogi keretek) fognak avatkozni a piac teljesítményébe, így a múltbéli trendekből való előrejelzéstől ez esetben eltekintének, de lássuk, hogy miből fakadhat a növekedés.



7. ábra - Az adatvilágpiac árbevétel alakulása (Edge Delta, 2024) - Saját Szerk..

Nézzük meg, hogy az árbevétel-változás hogyan követte le a globális adatmennyiség-változásának és a GDP változásnak az arányait, ennek változása a bal függőleges, a másik két paraméter pedig a jobb függőleges tengelyen ábrázolt.



8. ábra - Adatpiac mozgásának összetétele a GDP és az adatvolumen változásával (Trading Economics, 2024) (Edge Delta, 2024) - Saját Szerk..

Amellett, hogy a változást megmutató ábra rendkívül jól mutatja meg a Covid-hatást, melyben jól látszik a 2020-as év és környéki adataiban, hogy a világjárvány mennyit lendített a digitalizáció előre lendítése érdekében, jól tükrözi azt is, hogy nagyjából a globális adatvolumen változásához igazodik az adatpiac árbevételének mozgása is, figyeljük meg, hogy azokban az években ahol kisebb mértékben növekedett az adat, ott az árbevétel növekedése is kisebb volt, a 2020-as évet leszámítva pedig, ahol nagyobb volt az adatok növekménye ott az árbevétel is jobban nőtt, előrejelzésekbe szintén nem bocsátkoznék, révén hogy a generatív mesterséges intelligencia széleskörű elterjedése napjainkban bontakozik ki, az adatpiac forgalmára gyakorolt hatását egyelőre nem ismerjük, illetve számos elemző a korábbi években is alul becsülte a növekedést, ezt már fentebb említettem. Viszont készítettem még egy ábrát arról, hogy mekkora befolyása van az Egyesült Államoknak az adatpiacokra, érdekes lehet elgondolkodni arról, hogy a Donald Trump által ígért, Kínával szembeni importvámok hogyan lehetnek majd később hatással az adatok árszintjére, ha az adatközpontok legnagyobb része jelenleg az Egyesült Államokban, a hardverigényüket egy részben kielégítő alkatrészek pedig Kínából érkeznek.

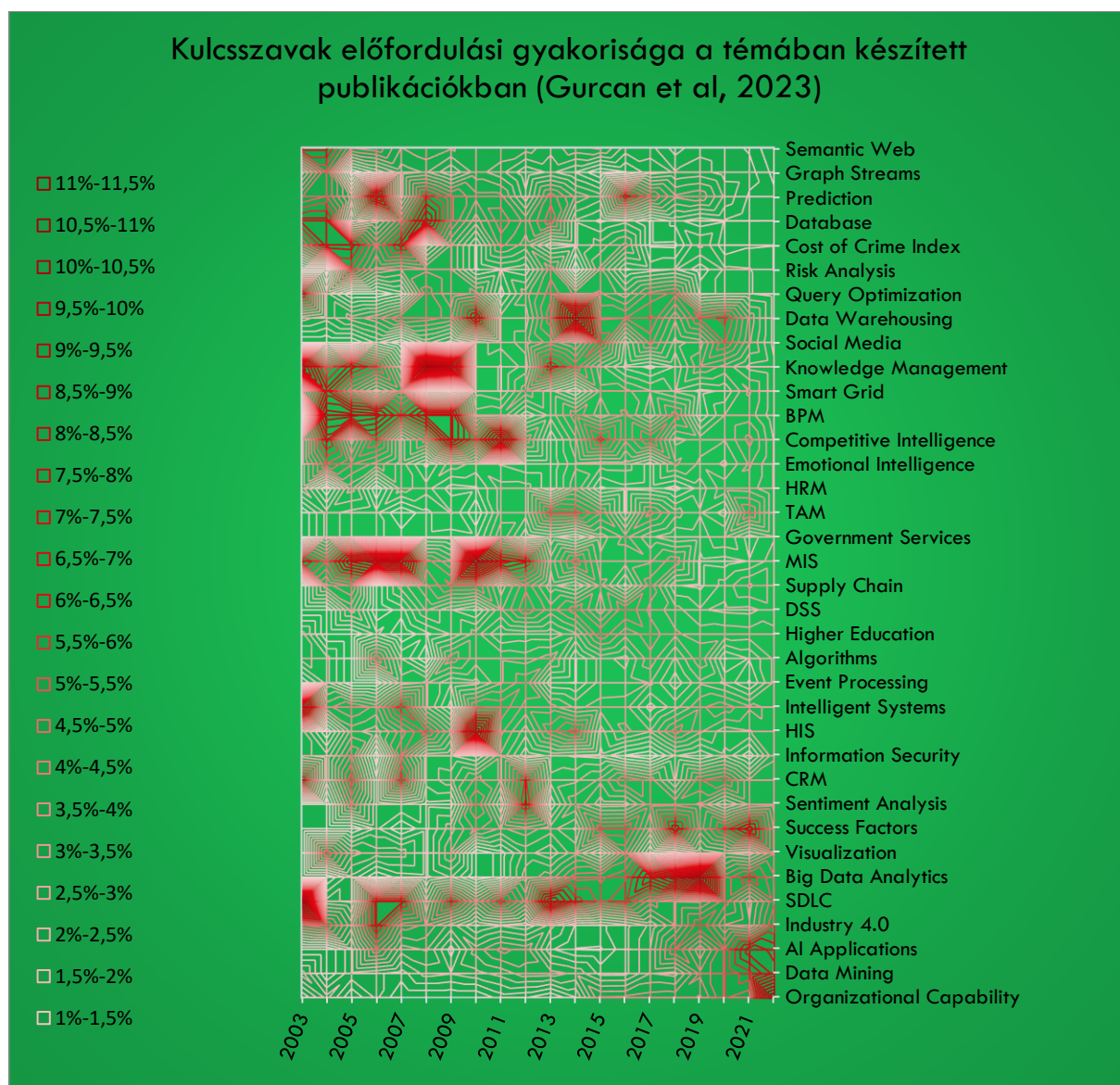
Adatközpontok száma a világon (2023)



8.ábra - Adatközpontok a világon (Edge Delta, 2024) - Saját Szerk..

4. Összefoglaló

Összegeztük, hogy milyen alrendszerrel dolgozik a modern vállalatirányítási rendszer az adat vezérelt döntéshozatalban, modelleztük az adatpiacot és kiemeltük annak jelentőségét az Ipar 4.0-ban. Látható, hogy az adatpiac folyamatosan stabil növekedési ütemet mutatott a világjárvány időszakát leszámítva az elmúlt években, bizonyítva ezzel azt, hogy az adatok iránti kereslet folyamatos, a big data felhasználása a vállalati életben, akár a termékek vagy szolgáltatásokba építve, akár a vállalatirányításban egyre nagyobb ütemben veszi ki a részét.



9. ábra - Kulcsszavak a kutatott irodalomban (Gurcan et al., 2023) - Saját Szerk..

A trendek egyértelműen azt mutatják, hogy a vállalatok részéről nyitottság és hajlandóság, majd végső soron a piaci kényszer érvényesül a technológiai fejlesztések irányában. Jól tükrözi ezt az utolsó ábra is.

Az adott sorokból azt olvashatjuk ki, hogy mennyire volt kutatott egy adott kifejezés abban az évben, a mintavételben elemzett cikkek hány százalékában fordult elő, ezt az ábrán sötétebb vörös vonalak jelzik. Látható, hogy a tudományos kutatói érdeklődés az évek során hogyan változott meg a közösségi média, a tudásmenedzsment témaköre felől az információbiztonság, a CRM rendszerek és az adatraktározás felé, majd a legutóbbi években az adatbányászat és a vállalati kapacitások témái dominálnak. A továbbiakban érdemes lesz kiterjeszteni mélyebben a kutatást abba az irányba, hogy az egyéni vállalat hogyan tudja az adatpiacon elhelyezni önmagát, és adatgazdaságossági mutatókat meghatározni, az adatvagyon-gazdálkodás teljesítményét mérni, illetve közgazdaságtani szempontból milyen kritikus mérőszámokkal lehet objektíven mérni az adatnak, mint vállalati inputnak a profitszerzésre irányuló hozadékát.

Irodalomjegyzék

Ács, P. *et al.* (2023) 'Gyakorlati adatelemzés'.

AI in Action (2024). IBM Corporation. Available at: <https://www.ibm.com/think/reports/ai-in-action>.

Annanth, V.K., Abinash, M. and Rao, L.B. (2021) 'Intelligent manufacturing in the context of industry 4.0: A case study of siemens industry', in. *Journal of Physics: Conference Series*, IOP Publishing, p. 012019.

Aroraa, G., Lele, C. and Jindal, M. (2022) *Data Analytics: Principles, Tools, and Practices: A Complete Guide for Advanced Data Analytics Using the Latest Trends, Tools, and Technologies (English Edition)*. BPB Publications.

Ashley-Osuzoka, J. *et al.* (2024) 'Generative Artificial Intelligence in Corporate America: A Review of Artificial Intelligence Use, Benefits, Challenges, Regulation and Cyber Security Issues'. Available at: <https://doi.org/10.20944/preprints202409.1557.v1>.

Beer, D. (2016) 'How should we do the history of Big Data?', *Big Data & Society*, 3(1), p. 2053951716646135. Available at: <https://doi.org/10.1177/2053951716646135>.

Bell, F. *et al.* (2021) 'Semi-structured Data in Snowflake', in *Snowflake Essentials: Getting Started with Big Data in the Cloud*. Springer, pp. 239–251.

Benlachmi, Y. and Hasnaoui, M.L. (2020) 'Big data and spark: Comparison with hadoop', in. *2020 Fourth World conference on smart trends in systems, security and sustainability (Worlds4)*, IEEE, pp. 811–817.

Biden, J. (2023) *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Presidential Actions*. Available at: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

Biswas, S. (2023) 'Assimilation of "Artificial Intelligence" (AI) in "Amazon": An in depth Research Case Study.', *Globsyn Management Journal*, 17.

Blueprint for an AI Bill of Rights (2022). Available at: <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>.

California AI Transparency Act. (2024) SB-942. Available at: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB942.

Candore, G. *et al.* (2020) 'Can we rely on results from IQVIA medical research data UK converted to the observational medical outcome partnership common data model? A validation study based on prescribing codeine in children', *Clinical Pharmacology & Therapeutics*, 107(4), pp. 915–925. Available at: <https://doi.org/10.1002/cpt.1785>.

Chawda, V. *et al.* (2020) ‘Unique Face Identification System using Machine Learning’, in. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, pp. 701–706.

Corrêa, N.K. *et al.* (2023) ‘Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance’, *Patterns*, 4(10), p. 100857. Available at: <https://doi.org/10.1016/j.patter.2023.100857>.

Data and AI Team, I. (2023) ‘Understanding the different types of artificial intelligence’, *Think*, 12 October. Available at: <https://www.ibm.com/think/topics/artificial-intelligence-types>.

Data Governance Act (2022) *THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION*. Available at: <https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng>.

Dewitte, S. *et al.* (2021) ‘Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction’, *Remote Sensing*, 13(16), p. 3209. Available at: <https://doi.org/10.3390/rs13163209>.

Elaraby, A.F. (2019) ‘3D OBJECT DETECTION AND CLASSIFICATION WITH MICROSOFT KINECT AND DEEP NEURAL NETWORKS’.

Experian (2024) *Experian annual report 2024*, p. 264. Available at: https://www.experianplc.com/content/dam/marketing/global/plc/en/assets/documents/reports/2024/experian_annual_report_2024_web.pdf.

Field, A. (2013) *Discovering statistics using IBM SPSS statistics:(and sex and drugs and rock’n’roll): Andy Field*. Sage.

Fradkov, A.L. (2020) ‘Early history of machine learning’, *IFAC-PapersOnLine*, 53(2), pp. 1385–1390. Available at: <https://doi.org/10.1016/j.ifacol.2020.12.1888>.

Fujiyoshi, H., Hirakawa, T. and Yamashita, T. (2019) ‘Deep learning-based image recognition for autonomous driving’, *IATSS research*, 43(4), pp. 244–252. Available at: <https://doi.org/10.1016/j.iatssr.2019.11.008>.

Goldman Sachs Research (2024) ‘AI is poised to drive 160% increase in data center power demand’, *Goldman Sachs Research* [Preprint]. Available at: <https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand>.

Grandview Research (2024) *Global Data Place Platform Market Size & Outlook*. Available at: <https://www.grandviewresearch.com/horizon/outlook/data-market-size/global>.

Gurcan, F. *et al.* (2023) ‘Business Intelligence Strategies, Best Practices, and Latest Trends: Analysis of Scientometric Data from 2003 to 2023 Using Machine Learning’, *Sustainability*, 15(13), p. 9854. Available at: <https://doi.org/10.3390/su15139854>.

Haeb-Umbach, R. *et al.* (2019) ‘Speech processing for digital home assistants: Combining signal processing with deep-learning techniques’, *IEEE Signal processing magazine*, 36(6), pp. 111–124. Available at: <https://doi.org/10.1109/MSP.2019.2918706>.

IBM Cognos Analytics (2024) 'Cognos Analytics Reports Security'. Available at: <https://www.ibm.com/docs/en/cognos-analytics/12.0.0?topic=administration-cognos-analytics-reports-security>.

Így cseréld ki BÁRKI hangját! | Kísérletezés az AI hangkészítéssel! (2023). Debrecen: TheVR Tech. Available at: <https://www.youtube.com/watch?v=whVXGeSDuh4> (Accessed: 27 November 2024).

Javaid, M. *et al.* (2022) 'Significant applications of Cobots in the field of manufacturing', *Cognitive Robotics*, 2, pp. 222–233. Available at: <https://doi.org/10.1016/j.cogr.2022.10.001>.

Josan, H.H.S. (2024) 'AI and Deepfake Voice Cloning: Innovation, Copyright and Artists' Rights', *Artificial Intelligence* [Preprint].

Kepuska, V. and Bohouta, G. (2018) 'Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)', in. *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*, IEEE, pp. 99–103.

Kwak, T. and Kim, M. (2018) 'Mobile App Classification Method Using Machine Learning Based User Emotion Recognition', in. *Proceedings of the 8th International Conference on Information Communication and Management*, pp. 22–26.

Li, T. *et al.* (2022) 'Smartphone app usage analysis: datasets, methods, and applications', *IEEE Communications Surveys & Tutorials*, 24(2), pp. 937–966. Available at: <https://doi.org/10.1109/COMST.2022.3163176>.

Louie, R. *et al.* (2020) 'Novice-AI music co-creation via AI-steering tools for deep generative models', in. *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13.

McCulloch, W.S. (1943) '(1943) Warren S. McCulloch and Walter Pitts A logical calculus of the ideas immanent in nervous activity Bulletin of Mathematical Biophysics 5: 115-133', *Bulletin of mathematical biophysics*, 5, pp. 115–133. Available at: <https://doi.org/10.1007/BF02478259>.

Meta Platforms, I. (2024) 'Privacy Policy of Facebook'. Meta Platforms Incorporated. Available at: https://www.facebook.com/privacy/policy/?entry_point=comet_dropdown (Accessed: 27 November 2024).

Microsoft, C. (2024) 'Mit jelent a mesterséges intelligencia?' Available at: <https://azure.microsoft.com/hu-hu/resources/cloud-computing-dictionary/what-is-artificial-intelligence#önvezető-autók>.

Mitchell, E. (2021) 'Una grande mappa per una piccola isola: il rilevamento delle isole Faroe', *GEOmedia*, 25(2).

Mohammad Amini, M. *et al.* (2023) 'Artificial intelligence ethics and challenges in healthcare applications: a comprehensive review in the context of the European GDPR mandate', *Machine Learning and Knowledge Extraction*, 5(3), pp. 1023–1035. Available at: <https://doi.org/10.3390/make5030053>.

Monetate (2024) '5 Learnings from the Temu Ecommerce Strategy'. Available at: <https://monetate.com/resources/blog/5-learnings-from-the-temu-ecommerce-strategy/>.

Munroe, K. (2022) *Data Modeling with Tableau: A practical guide to building data models using Tableau Prep and Tableau Desktop*. Packt Publishing Ltd.

Olanrele, O. *et al.* (2022) 'An IOT Based Weather Station using an Embedded System', *Quantum Journal of Engineering, Science and Technology (QJOEST)*, 3(3), pp. 31–40.

de Oliveira Aparecido, L.E. *et al.* (2020) 'Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases', *International Journal of Biometeorology*, 64(4), pp. 671–688. Available at: <https://doi.org/10.1007/s00484-019-01856-1>.

Park, H., Oh, H. and Choi, J.K. (2023) 'A Profit Maximization Model for Data Consumers with Data Providers' Incentives in Personal Data Trading Market', *Data*, 9(1), p. 6. Available at: <https://doi.org/10.3390/data9010006>.

Pêga, S.A.F. (2023) 'Big data framework implemented in cloud Azure'.

Rajaraman, V. (2023) 'From ELIZA to ChatGPT: history of human-computer conversation', *Resonance*, 28(6), pp. 889–905. Available at: <https://doi.org/10.1007/s12045-023-1620-6>.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) (2024). Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.

Rekettye, G. (2023) 'EMPÁTIA–VEVŐÉLMÉNY–VEVŐÉRTÉK', <https://pea.lib.pte.hu/handle/pea/44356> [Preprint].

Sargiotis, D. (2024) 'Data Governance Tools and Technologies: Navigating the Options', in *Data Governance: A Guide*. Springer, pp. 305–325.

Schaeffer, J. *et al.* (1992) 'A world championship caliber checkers program', *Artificial Intelligence*, 53(2–3), pp. 273–289. Available at: [https://doi.org/10.1016/0004-3702\(92\)90074-8](https://doi.org/10.1016/0004-3702(92)90074-8).

Schaeffer, J. and Plaat, A. (1997) 'Kasparov versus deep blue: The rematch'. Available at: <https://doi.org/10.3233/ICG-1997-20209>.

Schmidhuber, J. (2015) 'Deep learning', *Scholarpedia*, 10(11), p. 32832. Available at: <https://doi.org/10.4249/scholarpedia.32832>.

Serrano, W. (2018) 'Fintech model: The random neural network with genetic algorithm', *Procedia Computer Science*, 126, pp. 537–546. Available at: <https://doi.org/10.1016/j.procs.2018.07.288>.

Sheehan, M. (2023) 'China's AI regulations and how they get made', *Horizons: Journal of International Relations and Sustainable Development*, (24), pp. 108–125.

Sonkavde, G. *et al.* (2023) 'Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications', *International Journal of Financial Studies*, 11(3), p. 94. Available at: <https://doi.org/10.3390/ijfs11030094>.

Steck, H. *et al.* (2021) 'Deep learning for recommender systems: A Netflix case study', *AI Magazine*, 42(3), pp. 7–18. Available at: <https://doi.org/10.1609/aimag.v42i3.18140>.

Steurer, M.E. and Kappe, F. (no date) 'A Micropayment Enabled Webshop for Virtual Items'.

Tableau Support (2024) 'Data governance in Tableau'. San Fransisco, CA. Available at: https://help.tableau.com/current/blueprint/en-us/bp_governance_in_tableau.htm.

Tao, R. *et al.* (2021) 'Robo advisors, algorithmic trading and investment management: Wonders of fourth industrial revolution in financial markets', *Technological Forecasting and Social Change*, 163, p. 120421. Available at: <https://doi.org/10.1016/j.techfore.2020.120421>.

Thompson, S.A. and Warzel, C. (2022) 'Twelve million phones, one dataset, zero privacy', in *Ethics of data and analytics*. Auerbach Publications, pp. 161–169.

Tibor, K. (2023) 'Az adatműveltség és a COVID-19 válság', *Tudományos és Műszaki Tájékoztatás*, 70(1), pp. 30–37.

Tran, N. *et al.* (2020) 'A deep learning approach to intrinsic force sensing on the da Vinci surgical robot', in. *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, IEEE, pp. 25–32.

VANDA, P., ISTVÁN, A.P.V. and ANDRÁS, K. (2024) 'THE CASE OF VANDA–CAN AN AI-POWERED VIRTUAL AGENT INCREASE CUSTOMER SATISFACTION?', *UIK 2024*, p. 379.

Vasserman, S., Feldman, M. and Hassidim, A. (2015) 'Implementing the wisdom of waze', in. *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Weber, I. (2024) 'Large Language Models are Pattern Matchers: Editing Semi-Structured and Structured Documents with ChatGPT', *arXiv preprint arXiv:2409.07732* [Preprint].

Yeh, C.-L. (2018) 'Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers', *Telecommunications Policy*, 42(4), pp. 282–292. Available at: <https://doi.org/10.1016/j.telpol.2017.12.001>.

Ábrajegyzék

1. ábra - DL alkalmazásai . <i>forrás: Saját Szerk.</i>	11
2. ábra - Műtétek tendenciája - <i>forrás: Saját Szerk., adatok: Intuitive Surgical, befektetői jelentések</i>	13
3. ábra - AI rendszerek hierarchiája - <i>Saját Szerk.</i>	14
4. ábra - A tárgyalt algoritmusok megjelenésének időrendje - <i>forrás: Saját Szerk.</i>	16
5. ábra - AI szabályozás kezdete a világ országaiban <i>Forrás: Saját Szerk.</i>	20
6. ábra - Az adatvilágpiac árbevétele alakulása (Edge Delta, 2024) - <i>Saját Szerk.</i>	35
7. ábra - Adatpiac mozgásának összevetése a GDP és az adatvolumen változásával (Trading Economics, 2024)(Edge Delta, 2024) - <i>Saját Szerk.</i>	35
8. ábra - Kulcsszavak a kutatott irodalomban (Gurcan et al., 2023) - <i>Saját Szerk.</i>	37

Táblázatjegyzék

1. táblázat - AI Act kockázati besorolásai	18
2. táblázat - AI használat a vállalatirányításban - <i>Saját Szerk.</i>	31

