ALBERT ÁGOTA KATALIN

Jogászok és informatikusok kommunikációs problémáinak eliminációs eszközei a mesterséges intelligencia alkalmazása során

Methods for eliminating communication problems for lawyers and IT professionals in the application of artificial intelligence

DOI: HTTPS:// DOI.ORG/10.53793/RV.2024.2.4

Absztrakt

A mesterséges intelligencia napjainkban tapasztalható robbanásszerű fejlődése új kihívások elé állítja nemcsak az informatikusokat, hanem a jogászokat is. Felmerül a kérdés: hogyan biztosíthatjuk, hogy a mesterséges intelligencia használata etikus és megbízható legyen, társadalmunk érdekeit szolgálva? Mi a garanciája annak, hogy például egy nagy nyelvi modell projekt résztvevői egy irányba mozogva, az irányadó jogi-etikai kereteken belül valósítsák meg elképzeléseiket? A tanulmány azt elemzi, milyen eszközökkel lehet a konstruktív párbeszéd hiányából adódó problémákat orvosolni, különös tekintettel azon platformokra, amelyek már most rendelkezésre állnak. A cél annak bemutatása, hogy a meglévő eszközök, így például a megvalósíthatósági tanulmányok, a különféle hatásvizsgálatok, az érdekmérlegelési tesztek, a szabályzatok és az oktatás hogyan, mi módon lehetnek alkalmasak arra, hogy a jogászok és az informatikusok közötti kommunikációs szakadék áthidalható legyen.

KULCSSZAVAK: MESTERSÉGES INTELLIGENCIA, JOGÁSZ-INFORMATIKUS PÁRBESZÉD, SZTEREOTÍPIÁK, HATÁSVIZSGÁLAT, ADATVÉDELEM, ADATBIZTONSÁG, LLM

Abstract

The explosive development of artificial intelligence today poses new challenges not only for computer scientists but also for lawyers. The question arises: how can we ensure that the use of artificial intelligence is ethical and safe for the benefit of our society? What is the guarantee that, for example, the participants in a large language model project will implement their ideas within the guiding legal and ethical framework, moving in the same direction? The study analyses the methods that can be applied to overcome the lack of constructive dialogue, with a particular focus on the platforms that are already available. The aim is to show how existing means, such as feasibility studies, impact assessments, legitimate interests tests and codes of conduct and education, can be used to bridge the communication gap between legal practitioners and IT experts.

KEYWORDS: ARTIFICIAL INTELLIGENCE, LAWYER-INFORMATICS DIALOGUE, STEREOTYPES, IMPACT ASSESSMENT, PRIVACY, DATA SECURITY, LLM

Bevezetés

A mesterséges intelligencia (MI) soha nem látott várakozást generál a verseny- és a közszféra világában, a nagy nyelvi modellek (LLM), például a ChatGPT közkinccsé válása pedig már az egyének szintjén is új lehetőségeket nyújt. A világvége-hívők vészharangot kongatnak, egy szempillantás és a gépek uralják a Földet, míg mások kevésbé aggódnak, szerintük ez is megoldódik így vagy úgy. És vannak, akik a jogászokban reménykednek – ők majd visszagyömöszölik az informatikusok elszabadította szellemet a palackba.

Mindeközben az MI-informatikusok már sok éve tudják, új világ épül, ám csak mostanában szembesülnek azzal, nem biztos, hogy ezt az új világot képesek uralni. Sőt, sokuk szerint bár az armageddont nem az MI fogja elhozni, de az, hogy a jogászok bele akarnak szólni milyen is legyen ez az új világ, félút a világégéshez (1. sz. ábra).

1. sz. ábra: Ahogy az informatikus és a jogászok látják az MI-t
Forrás: Midjourney

Az új világ azonban nemcsak új technológiát hoz irodánkba, autónkba és még a telefonunkba is, hanem önmagunk újraépítésére is rákényszerít. A média naponta hozza a jóslatokat, melyik szakma, hogyan fog átalakulni az MI miatt (pl. ILO Working Paper 96, 2023), de már maga az MI előállítása is azt követeli, váljunk meg régi beidegződéseinktől és a megszokott munkastílusunktól. Akár jogászként, akár informatikusként veszünk részt MI-projektekben, olyan problémákkal szembesülünk, amelyek megoldása új készségeket igényel szakmai és emberi mivoltunktól is. De mi történik, ha a jogászok és az informatikusok egója akkor csap össze, amikor jogszerű és etikus, a közjó érdekében tevékenykedő MI-t kellene sikerre vinnünk? Ha nem is az emberiség pusztulását, egy-egy félresikerült MI modell az érintettek számára a földi poklot még elhozhatja, gondoljunk például egy diszkrimináló egészségügyi alkalmazásra. A tét tehát nemcsak a jogászok és az informatikusok jó kapcsolata, hanem a társadalmunk és az MI viszonya.

Az informatikusok és jogászok közös felelőssége, hogy az MI ne az emberiség veszte, hanem jövője legyen. Tanulmányom célja annak vizsgálata, hogy a meglévő eszközök segítségével hogyan lehet(ne) a jogászokat és az informatikusokat közelebb hozni egymáshoz, hogy együtt biztosítsák az MI, különösen az LLM jövőjét. Ezek az eszközök – kicsit másképp használva, mint eddig – segíthetnek hidat építeni afelett a szakadék felett, amely a két szakma képviselői között napjainkban található.

Mi a probléma?

„A fekete lyukak ott keletkeznek, ahol az informatikusok és a jogászok beszélgetni kezdenek." (meg nem nevezett jogász)

Egy átlagos IT-projektnek számos olyan szereplője van, akik iskolai végzettség, szakmai tudás, tapasztalat és szocializáció terén is jelentősen különböznek. Simulékonyságukat vagy éppen permanens harci vágyukat nemcsak saját személyiségük, hanem az általuk képviselt szervezet/szervezeti egység kultúrája, annak dominanciája, valamint kapcsolatrendszere is alakítja. Ezek a jellemzők pedig nemcsak azt befolyásolják, hogy a többi csapattag hogyan tekint rájuk, hanem azt is, ők maguk hogyan, mi módon teszik jobbá vagy rosszabbá a projekt eredményességét.

A hagyományos IT-fejlesztésekben a jogászok ritkán tagjai a projektcsapatnak, bevonásuk többnyire polgári jogi, iparjogvédelmi, versenyjogi és fogyasztóvédelmi területekre korlátozódik. A projektmenedzsment irodalma csak elvétve említi szükséges szereplőként őket, miközben számtalan hagyományos IT-fejlesztésben elengedhetetlen (lenne) közreműködésük. Ilyenek például azok a nagy mennyiségű személyes adatot kezelő szoftverek, amelyek esetében a beépített és alapértelmezett adatvédelem követelményének megfelelés komoly kockázatot hordoz. Az MI-projektek esetében nemcsak ezeket a kockázatokat kell kezelnünk, hanem az is kihívás, hogy projektünk eredménye ne legyen Orwell (Nagy Testvér) és Kafka (Josef K. pere) világának szerelemgyereke.

Az olasz adatvédelmi hatóság (Garante per la protezione dei dati personali, továbbiakban: Garante) ideiglenesen korlátozta a személyes adatok ChatGPT-vel történő feldolgozását (Garante 9870832). A Garante megállapította:

- ✓ *az adatkezelő (OpenAI) nem bocsátotta az érintettek rendelkezésére a GDPR[20] 13. cikke szerint készült adatkezelési tájékoztatót*
- ✓ *a személyes adatok gyűjtése és a ChatGPT algoritmusok képzésére történő felhasználása megfelelő jogalap nélkül történt, megsértve a GDPR 5. és 6. cikkét*
- ✓ *azon érintettek esetében, akiknek a személyes adatait az interneten gyűjtik, a ChatGPT működése mögött álló algoritmusok nem garantálják a pontosságot, megsértve a GDPR 5. cikk (1) bekezdés d) pontját*
- ✓ *a felhasználók életkorának ellenőrzésére szolgáló mechanizmus hiánya a GDPR 8. cikkét sérti.*

A Garante sürgősségi eljárás keretében az OpenAI-t az adatkezelés ideiglenes korlátozására kötelezte minden olyan adatkezelés esetében, amely Olaszország területén lévő érintetteket érint.

Az új technológiák térhódításával a jogászok megkerülhetetlenek és az MI-projektek sikerének kulcsa az, hogy a jogászok a kezdetektől a projekt részei legyenek. A cél az informatikusok támogatása, számukra keretek adása. Természetesen dönthet úgy egy projektvezető, hogy nem hallgatja meg a jogászok véleményét, azonban az MI nem megfelelő használatáért kiszabott közigazgatási bírságok azt mutatják, nem feltétlen érdemes ezt az utat választani.

A Garante feloldotta a ChatGPT korlátozását azzal a feltétellel, hogy az OpenAI végrehajtja a javasolt intézkedéseket (Garante 9874702):

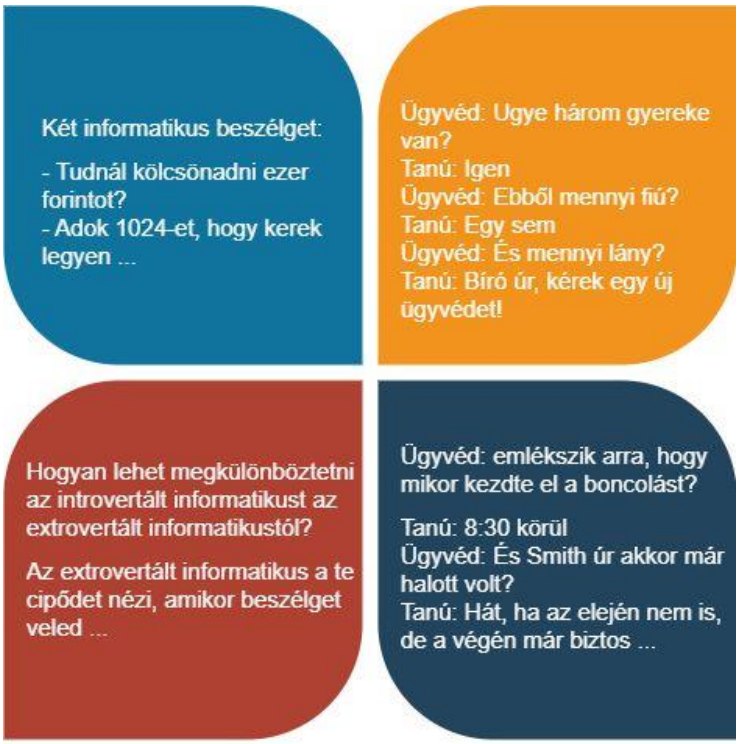
- ✓ *a weboldalán olyan tájékoztatót tesz közzé, amely elmagyarázza, az érintettektől (a szolgáltatás felhasználóitól és nem felhasználóitól) gyűjtött adatokat a ChatGPT algoritmusainak képzésére használják fel, továbbá tájékoztatást nyújt az adatkezelés módjáról, a szolgáltatás működéséhez szükséges adatkezelés logikájáról, az érintettek jogairól és a GDPR által előírt bármely más információról*
- ✓ *olyan, a weboldaláról elérhető eszközt rendszeresít, amellyel az Olaszországból bejelentkező érintettek gyakorolhatják tiltakozási jogukat a harmadik féltől gyűjtött személyes adataik kezelése tekintetében, amennyiben az adatkezelésre az algoritmus képzése, illetve a szolgáltatás nyújtása céljából kerül sor*
- ✓ *a weboldalán olyan eszközt tesz elérhetővé, amellyel az érintettek kérhetik és elérhetik a tartalomgenerálás során pontatlanul feldolgozott személyes adataik helyesbítését, vagy – amennyiben ez a technika jelenlegi állása szerint nem lehetséges – az ilyen adataik törlését*
- ✓ *olyan adatkezelési tájékoztatóra mutató linket illeszt be, amely a regisztráció, illetve a szolgáltatás újraaktiválása előtt jelenik meg*
- ✓ *módosítja a személyes adatok algoritmusképzés céljából történő kezelésének jogalapját; az adatkezelést hozzájárulásra vagy jogos érdekre kell alapoznia*
- ✓ *a weboldalán olyan könnyen hozzáférhető eszközt alkalmaz, amely lehetővé teszi az érintettek számára a tiltakozást a ChatGPT használata során gyűjtött személyes adataik kezelése ellen abban az esetben, ha az adatokat az adatkezelő jogos érdeke alapján az algoritmusok képzése céljából kezelik*
- ✓ *korhatár-rendszert kell felállítania a kiskorúak kiszűrésére az érintett által megadott életkor alapján, illetve 2023. május 31-ig olyan tervet kell benyújtania a Garante-hoz az életkor-ellenőrző eszközök bevezetésére vonatkozóan, amelynek eredményeként megakadályozható, hogy a 13 év alatti érintettek a ChatGPT-t a 18 év alatti érintettekkel együtt használják, amennyiben a szülői felügyeleti jogot gyakorló személy ehhez nem adta egyértelmű hozzájárulását*
- ✓ *tájékoztató kampányt kell indítania – szolgáltatásainak reklámozása nélkül – az algoritmus képzésével és a személyes adatok törlésének lehetőségével kapcsolatban.*

20 általános adatvédelmi rendelet, a továbbiakban GDPR

Miért nehézkes a jogászok és az informatikusok közötti együttműködés?

A jogászok és az informatikusok nem arról híresek, hogy a közös munkáért küzdenének. De mi okozza ezt a kibékíthetetlennek tűnő ellentétet?

Több tényező, többek között az eltérő gondolkodásmód, a beivódott sztereotípiák sokasága, például *„az informatikusok kockák"* és *„introvertáltak"*, *„a jogászokat földi halandó nem érti meg"*, illetve *„nagyképűek"* (2. sz ábra).



2. sz. ábra: Sztereotípiák
Forrás: Saját szerkesztés

A ki-, illetve megbékülést a szervezeti kultúra sem segíti, gyakran a munkán túl is egészen mást várnak el a jogászoktól és az informatikusoktól (pl. öltözködés és hajviselet terén), támogatva a sztereotípiák továbbélését.

A hagyományosan ellenségeskedő, egymást félre-, illetve meg nem értő magatartás alapvetően gátolja az MI-projektekhez szükséges együttműködést, ahogy az is, amikor a jogászok *„gyógyszert szednek informatika ellen"*, az informatikusok pedig úgy vélik, *„a jogszabályt mindenki el tudja olvasni, minek ahhoz jogász?"* Ezen kijelentéseknek természetesen van némi alapja, hiszen kevés olyan jogász van, aki egyben matekzseni is és a törvények is ugyanabból a betűkből állnak, mint a Grimm mesék. Éppen ezért az informatikusokra küldetés hárul, hogy úgy magyarázzák el az MI-modellek lényegét, hogy azt még egy jogász is megértse (és innen már csak egy apró lépés a megmagyarázható MI, mint jogi követelmény teljesítése), miközben a jogászok ráébreszthetik az informatikusokat, nemcsak az egymás mellé sorolt szavakból álló jogszabályok léteznek, hanem a jog szelleme is.

Ezen ellentétek miatt a jogászok és az informatikusok közötti kommunikáció és együttműködés nehézkes, ez pedig lehetetlenné teheti a kitűzött célok elérését.

Mi befolyásolhatja a sikeres együttműködést?

Az eltérő szemléletmód megjelenik a szaknyelvben is, a saját terminológiák használata pedig félreértésekhez vezethet. Ezt mutatja az alábbi példa:

Egy informatikus panaszkodik egy jogásznak:
- Nagyon nehéz informatikusként dolgozni, állandó a nyomás, minden működjön tökéletesen, és a felhasználók elégedettek legyenek.
Mire a jogász:
- Megértelek, jogászként nagyon hasonló a helyzetem, meg kell felelnem a jogszabályoknak, és a kockázatokat is minimalizálnom kell.
Mire az informatikus:
- Talán össze kellene fognunk és megalkotnunk egy "Code of Law"-t, hogy könnyebben tudjunk együtt dolgozni.
Mire a jogász:
- Akkor először határozzuk meg, hogy a "Code" szó melyik jogi kategóriába tartozik …

Nemcsak a szakzsargon, de az eltérő prioritások is eredményezhetnek konfliktusokat, és már csak emiatt is tekinthetik egymást ellenfélnek a jogászok és az informatikusok úgy, hogy valójában nem azok. A jogászok a jogi megfelelőségre és a különféle szabályozások betartására koncentrálnak, „compliance-üzemmódba" kapcsolva védve ezek teljesítését, míg az informatikusok a technológiai megoldásokra és a hatékonyságra összpontosítanak. A „vízilónak is a maga bébije a legszebb" jelen esetben fokozottan érvényesül, és csakis ezt az attitűdöt megértve és elfogadva lehet a szakterületek közötti átjárást biztosítani.

Egy MI-projekt interdiszciplináris együttműködése nagyfokú rugalmasságot és toleranciát kíván meg, valamint azt, hogy a szemben álló felek hajlandóak legyenek álláspontjukat közelíteni. Ehhez az ellenfeleknek el kell ismerniük, hogy a másiknak is lehet igaza (jogos ellevetése, logikus érve stb.), illetve saját prioritásaikat és érvrendszereiket is meg kell ismertetniük és el kell fogadtatniuk. A feleknek önmérsékletet kell tanúsítaniuk, valamint meg kell békélniük azzal, hogy a magyarázatokat és az érveket időnként le kell egyszerűsíteniük, túl kell lépniük saját szaknyelvük terminológiáin ahhoz, hogy azt minden érdekelt értse. Hinniük kell abban is, hogy a velük együttműködésre vállalkozó (gyakran kötelezett) személy nem tudatlan, vagy netalán alulképzett, hanem egyszerűen csak más terület specialistája.

Az eltérő kommunikációs stílus is gondot jelenthet. A jogászok hajlamosak részletesebben, valamint formalizáltabban kommunikálni és elveszni a jogszabályok szövegének „szolgai" visszaadásában, míg az informatikusok általában a tömörebb és technikailag orientáltabb megfogalmazást részesítik előnyben. A két tábor közös vonása, hogy az esetek többségében meglehetősen magasra teszik a lécet a laikusok számára, illetve túlságosan is hangsúlyozzák saját szaktudásukat. Amennyiben a szereplők nem hajlandóak a másik fél számára is érthetően fogalmazni, elvesznek az érvek és magyarázatok, a projekt pedig olyan irányba tévedhet, amely senkinek sem előnyös.

Az MI sokaknak viszonylag új terület (még akkor is, ha kb. 70 éves), így előfordulhat, hogy a projekt résztvevőinek nincs elegendő ismeretük vagy tapasztalatuk a modellek fejlesztésével kapcsolatban. Az IBM felmérése szerint (IBM 2022) a vállalkozások számára az MI sikeres bevezetésének legnagyobb akadálya a korlátozott MI-képességek, szakértelem vagy tudás (34%). Ez különösen akkor probléma, ha a felek szeretnék eltitkolni járatlanságukat, nem mernek kérdezni, és felháborodnak, ha valaki megkérdőjelezi hozzáértésüket. A legjobb védekezés a támadás politikája még a legkiválóbb ötleteket is a süllyesztőbe küldheti, a konfrontatív magatartás pedig akadályozhatja a megvalósítást.

Mit tehetünk a kommunikációs problémák elkerülése érdekében?

A kezdetektől törekedniük kell egy olyan közös nyelv kialakítására, amely segítségével meg tudjuk értetni magukat és mi is meg tudjuk érteni a többieket. Ehhez azonban kevés egy szótár, szükséges a passzív ismeretek aktívvá tétele, valamint a szakkifejezések és az összefüggések magyarázata. A hiányos, levegőben lógó ismeretek furcsa kérdéseket generálhatnak, ez pedig visszatetszést kelthet, illetve – vérmérséklettől függően – egymás kimondatlan vagy kimondottan alkalmatlannak minősítését.

„Az új dolgok új szavakat igényelnek. De az új dolgok a régi szavakat is módosítják, olyan szavakat, amelyeknek mélyen gyökerező jelentésük van. A távíró és a filléres sajtó megváltoztatta azt, amit egykor „információ" alatt értettünk. A televízió megváltoztatta azt, amit egykor a „politikai vita", a „hírek" és a „közvélemény" kifejezések alatt értettünk. A számítógép ismét megváltoztatja az „információ" fogalmát. Az írás megváltoztatta azt, amit valaha „igazság" és „jog" alatt értettünk; a nyomtatás ismét megváltoztatta őket, és most a televízió és a számítógép ismét megváltoztatja őket. Az ilyen változások gyorsan, biztosan és bizonyos értelemben csendben történnek. A lexikográfusok nem tartanak népszavazást a kérdésben. Nem írnak kézikönyveket, hogy elmagyarázzák, mi történik, és az iskolák sem vesznek róla tudomást. A régi szavak még mindig ugyanúgy néznek ki, még mindig ugyanolyan típusú mondatokban használjuk őket. De nem ugyanaz a jelentésük, sőt, egyesesetekben ellentétes jelentésük van. (…) [A] technológia önkényesen kisajátítja legfontosabb terminológiáinkat. Újradefiniálja a „szabadságot", az „igazságot", az „intelligenciát", a „tényt", a „bölcsességet", az „emlékezetet", a „történelmet" - mindazokat a szavakat, amelyekkel élünk." (Postman 1993: 8)

Az érvek megértéséhez és az álláspontok közelítéséhez meg kell értenünk a másik szempontrendszerét és prioritásait, ehhez pedig türelem kell és vágy az új, számunkra szokatlan tudás befogadására. Egy MI-projekt óhatatlanul is új kompetenciákat követelhet meg, és az élethossziglani tanulás koncepciója is új értelmet nyerhet. Ha nem vagyunk nyitottak és toleránsak, nem biztos, hogy az MI, mint szakterület nekünk való.

A kommunikáció nem szorítkozhat egyetlen alkalomra (pl. a projektindító értekezletre), hanem a projekt teljes életciklusában érdemi és építő jellegű kapcsolatot kell fenntartanunk, hogy időben azonosíthassuk és kezelhessük a kihívásokat, valamint a problémákat. A folyamatos párbeszéd elősegíti a közös megértést, a kölcsönös tiszteletet, a hatékony együttműködést és összességében a projektünk sikeres végrehajtását.

Milyen platformokon képzelhető el ez az együttműködés?

A parttalan viták és dühös kivonulások, ajtócsapkodások, valamint sértődések elkerülése érdekében célszerű olyan fórumokat rendszeresítenünk, amelyeken közös megoldást találva megtárgyalhatjuk a problémákat. Ehhez párbeszéd és olyan közös munka kell, amely azonos vágányon tartja a szereplőket és a lehető legkevesebb alkalmat ad az egymás melletti elbeszélésre.

A legalkalmasabb platformok félformálisak. A résztvevőknek komolyan kell venniük ezeket az alkalmakat. Ha még egy felesleges értekezletnek fogják fel azt, amin meg kell jelenniük, olyannak, ami elvonja őket az érdemi munkájuktól, a negatív hozzáállásuk előre vetíti a kudarcot. A szereplők azonban nem vehetik túl komolyan az eseményt, a karót nyelt, saját szakmaiságukban elvesző, önnön nagyságukat megkoszorúzó résztvevőkkel kétséges a pozitív eredmény. Célszerű barátságos, befogadó hangulatot teremtenünk és együttgondolkodásra, közös problémamegoldásra sarkallni a résztvevőket. Ahhoz azonban, hogy tudjuk, mikor milyen eszközt lehet alkalmaznunk a kommunikációs jéghegyek megolvasztására, megfelelő mélységben ismernünk kell azt, hogy az LLM-projektek miben térnek el a hagyományos IT-projektektől, illetve a projektek mely pontjai esélyesek arra, hogy közelebb hozzuk a résztvevők gondolkodásmódját

Miben különböznek a hagyományos IT-projektek az MI-projektektől?

A különbség különösen a stádiumok esetén szembeötlő (3. sz. ábra).

3. sz. ábra: A hagyományos IT és az MI-projektek szakaszai
Forrás: Saját szerkesztés

A hagyományos IT- és az MI-fejlesztések gyakran összekapcsolódnak és kiegészítik egymást, valamint az MI-modellek általában hagyományos IT-infrastruktúrára épülnek. A határok azonban nem mindig egyértelműek, illetve az alkalmazott módszerek és technológiák nagymértékben függnek a konkrét projekt jellegétől és céljaitól. Az innovatív technológiák folyamatosan átjárják egymást, az MI egyre inkább behatol a hagyományos IT területére, míg a hagyományos IT is folyamodhat MI-technikákhoz és eszközökhöz.

A hagyományos IT-projekt szakaszai

1. *Igényfelmérés és a követelmények meghatározása*
   üzleti folyamatok, problémák, elvárások elemzése (megvalósíthatósági tanulmány, követelmények dokumentálása stb.)
2. *Tervezés*
   az IT-rendszer architektúrájának kidolgozása és az alkalmazás vagy a rendszer részletes struktúrájának megtervezése (adatmodell, felhasználói felület és adatbázis tervezése, rendszerkomponensek és funkciók meghatározása stb.)

3. *Fejlesztés*
   az alkalmazás vagy program kódolása a meghatározott követelmények és tervek alapján, tesztelés, hibajavítás és dokumentáció készítése
4. *Telepítés*
   az elkészült rendszer vagy alkalmazás valós környezetben történő üzembehelyezése
5. *Tesztelés*
   a rendszer különböző teszteknek alávetése a funkcionalitás, a teljesítmény, a biztonság és a stabilitás érdekében
6. *Üzemeltetés és karbantartás*
   a rendszer aktív használata és fenntartása, frissítése, valamint szükséges esetekben javítása

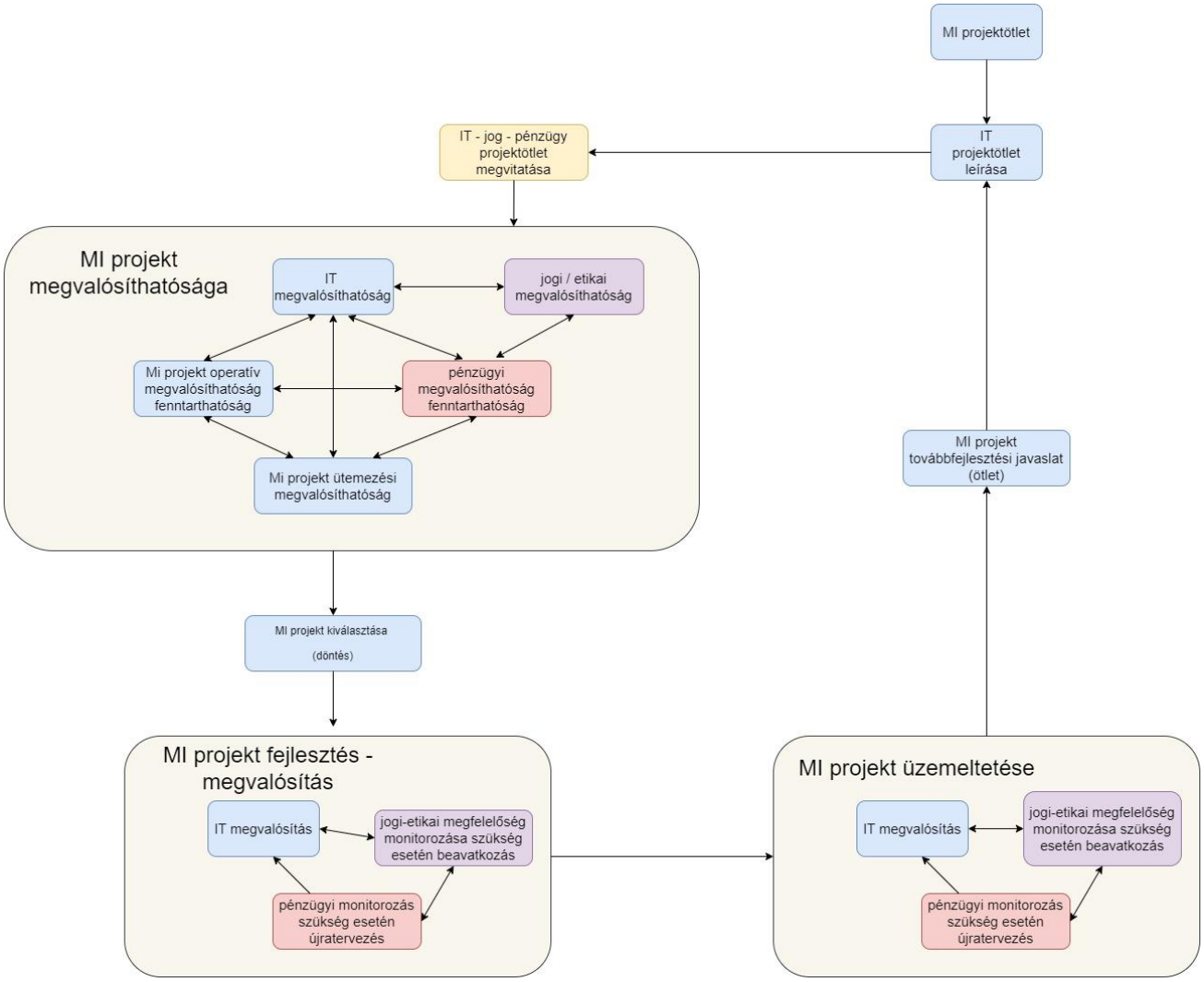
A szakaszok rugalmasak és akár jelentősen változhatnak a projekt jellegétől és méretétől függően, illetve az agilis fejlesztési módszerek, mint például a scrum vagy kanban olyan szakaszokat is felhasználhatnak, amelyek folyamatosan ismétlődő ciklusokra épülnek a fejlesztés, a tesztelések és a visszajelzések alapján.

Az MI-projekt szakaszai

Általában több, egymást átfedő szakaszból áll, amelyek nem minden esetben követik ugyanazt a sorrendet (4. sz. ábra). Tipikus főbb szakaszok:

1. *A projekt céljainak, a rendszer követelményeinek, illetve a várható felhasználói elvárások meghatározása (adatigények, rendelkezésre álló erőforrások és kívánt funkciók meghatározása stb.)*
2. *Adatgyűjtés és adatelőkészítés*
   adatokra van szükség a modell betanításához, teszteléséhez és validálásához (gyűjtés, megtisztítás, normalizálás, előkészítés stb.)
3. *A modellfejlesztés és betanítás*
   kiválasztjuk a megfelelő algoritmusokat és fejlesztjük a modellt, például hiperparaméterek beállítása (ez automatizálható), a modell adatokon betanítása, validációs tesztelés a modell teljesítményének értékelésére)
4. *A modell kiértékelése* előre meghatározott teljesítménymutatók és kritériumok alapján
5. *Telepítés és integráció*
   a sikeres kiértékelés után telepítjük a modellt a célhardvereken vagy a célgépeken (szoftverkomponensek és API-k beállítása, a rendszer integrálása a meglévő infrastruktúrába vagy alkalmazásba)
6. *Üzemeltetés és karbantartás*
   a modell folyamatos üzemeltetést és karbantartást igényel (modellfrissítések, adatok további gyűjtése, a modell teljesítményének figyelemmel kísérése és ellenőrzése)
7. *Felhasználói visszajelzés és iteráció*
   a modell bevezetése után a felhasználói visszajelzések és tapasztalatok nyomonkövetése, lehetővé téve a modell finomhangolását, a hibák javítását, valamint az új igények kielégítését. Az iterációs folyamban újra visszatérhetünk a korábbi szakaszokhoz a fejlesztés és optimalizálás érdekében.

Egy adott projekt sokkal részletesebb, szerteágazóbb is lehet, illetve az MI-modell fejlesztési és telepítési folyamatai jelentős mértékben függnek az alkalmazás konkrét céljától és a rendelkezésre álló erőforrásoktól.



4. sz. ábra: Az MI-projekt szakaszai a projekt három fő szereplőjének (IT, pénzügy, jog) együttműködése alapján
Forrás: Saját szerkesztés

Mennyivel másabb egy MI-, mint egy hagyományos IT-fejlesztés?

Az MI-fejlesztésekre jellemző:

- *a projekt komplexitása*
  gyakran van szükségük nagyobb adathalmazokra, összetett algoritmusokra és tanulási folyamatokra a gépi tanulás vagy mély tanulás alkalmazása révén, illetve figyelembe kell vennünk a specifikus adatigényeket, a modell architektúráját és finomhangolását is
- *a modell komplexitása*
  az alkalmazott modellek és algoritmusok általában komplexebb és specifikusabbak, mint egy hagyományos IT-fejlesztés esetében
- *az automatizáció szintje*
  gyakran magában foglalja az automatizáció magasabb szintjeit, ahol az adatok alapján az intelligens rendszerek képesek önálló tanulásra és döntéshozásra
- *adatigény*
  jelentős mennyiségű adatot gyűjtünk, készítünk elő és dolgoznunk fel a modell kifejlesztéséhez és betanításához (az adatok minősége, sokszínűsége és mérete kritikus a modell hatékonyságában és teljesítményében) (5. sz. ábra)

5. sz. ábra: Adatigény
Forrás: Az ábra alján

- *algoritmusok és modellek*

speciális algoritmusok és modellek alkalmazása jellemző; gépi tanulási algoritmusokat, mint például döntési fákat, neurális hálózatokat vagy szupport vektor gépeket használhatunk a tanulási és előrejelzési feladatokra

- *adattárolás és adatkezelés*

gyakran nagy mennyiségű adatot tárolunk és kezelünk

- *tanulási fázis*

általában van olyan fázis, amikor az algoritmusok és modellek az adatokból tanulnak

- *tanulás és iteráció*

a modell kifejlesztése és betanítása során gyakran van szükségünk többszöri iterációra, finomhangolásra és validációra

- *prediktív és adaptív képességek*

Gyakran prediktív és adaptív képességeket céloznak meg (a modell képes előre jelezni és a környezeti változásokhoz alkalmazkodni)

- *etikai és jogi kérdések*

számos etikai és jogi kérdés merülhet fel, mivel a modellek döntéseket hozhatnak, nagymennyiségű személyes adatot kezelhetnek és jelentős hatással lehetnek az emberek életére és magánszférájára

- *átláthatóság és megmagyarázhatóság*

gyakran felmerül az átláthatóság és az interpretálhatóság követelménye, azaz a modell megmagyarázható és érthető legyen

- *kockázatok és biztonság*

a modellek új kockázatokat hoznak magukkal (pl. ENISA 2023a), például annak veszélyét, hogy előre nem látott módon viselkednek vagy félreértelmeznek adatokat.

Milyen célból fejlesztünk LLM-modelleket?

Az LLM képes bonyolult nyelvi feladatokat elvégezni, mivel rengeteg adatot tanul meg és azokon keresztül szövegek összefüggéseit, mintáit és jelentéseit tudja értelmezni. Például:

- szöveggenerálás (rövidebb-hosszabb szövegek létrehozása, ideértve cikkeket, novellákat, versikéket és egyéb tartalmakat)
- gépi fordítás
- szövegösszefüggések megértése (pl. keresőmotorok, szemantikai elemzések)
- chatbotok és virtuális asszisztensek (értelmes és emberi válasz adása a felhasználóknak)
- tudásbázisok feltöltése, a felhasználók kérdéseire pontos válaszok adása
- szövegátalakítás és formázás, nyelvhelyességi hibák javítása, stilisztikai módosítások elvégzése
- jogi dokumentumok elemzése, releváns információk kinyerése
- orvosi diagnózisok és kutatás (orvosi szövegek elemzése, betegségek diagnosztizálásának és orvosi kutatások támogatása)
- nyelvtanulás és oktatás
- közösségi média elemzés, érzelmek felismerése, trendek azonosítása
- játékfejlesztés (párbeszédek, háttértörténetek és karakterek játékokhoz)
- célcsoport-specifikus tartalmak létrehozása, márkaüzenet hatékony kommunikációja
- reklám- és piackutatás (fogyasztói viselkedés és preferenciák elemzése, reklámstratégiákhoz információk kidolgozása)

- történelem- és kultúrakutatás (történelmi szövegek fordítása és elemzése, segítség a kulturális összefüggések és trendek megértéséhez)
- időjárásjelentések és adatok elemzése
- finanszírozási és gazdasági elemzés (pénzügyi jelentések és gazdasági adatok elemzése, támogatva a befektetési és üzleti döntéseket)
- egyéni hangszín és stílus szintézis (egyedi hangszínnel és stílussal generált szövegek létrehozása)
- politikai elemzés és közvéleménykutatás (szövegek elemzésével a politikai vélemények és trendek feltárásának támogatása).

Ki és hogyan határozza meg az MI-projektek követelményeit?

A követelmények meghatározása rendszerint többlépcsős folyamat, amelyben:

- *az ügyfél vagy (üzleti) tulajdonos* képviseli a végfelhasználói igényeket, meghatározza a projekt céljait és üzleti stratégiát
- *a projektmenedzser* felelős a projekt szervezéséért, a kommunikációért és az aktuális feladatok végrehajtásáért, valamint elősegíti a projekt szereplői közötti együttműködést
- *az üzleti elemzők* felmérik és elemzik az ügyfelek és végfelhasználók igényeit, elemzik a folyamatokat, valamint dokumentálják a funkcionális és nem funkcionális üzleti követelményeket
- *a fejlesztők és technikai szakemberek* a technikai megvalósíthatóság szempontjából értékelik a követelményeket, támogatják az üzleti követelmények technikai megfogalmazását és lefordítását (pl. architektúra tervezése, rendszereszközök és technológiák kiválasztása, a rendszer kapcsolódási pontjainak meghatározása stb.).

Az MI, illetve az LLM-projektekben *a jogászok bevonása különösen akkor kritikus, ha az adott projekt nagy számú személyes adat kezelését igényli.* Jelen tanulmányomban az ilyen, számos személyes adatot kezelő LLM-projektekre összpontosítok, tekintettel arra, hogy jogi és etikai szempontból ezek általában magas kockázatúak, illetve jelentős kockázatot jelent az adatvédelmi felügyeleti hatóságok szankcionálási gyakorlata is. Azonban *a személyes adatokkal nem dolgozó LLM-modellekkel kapcsolatban is számtalan jogi probléma merülhet fel*, például szerzői és szomszédos jogi, versenyjogi, fogyasztóvédelmi és termékfelelősségi területen.

Mi a jogászok feladata az LMM követelményeinek meghatározása során?

Optimális esetben:

- azonosítják a jogi kötelezettségeket, korlátokat, valamint az etikai kérdéseket
- azonosítják a jogi kockázatokat és közreműködnek azok kezelésében
- meghatározzák az adatvédelmi és adatbiztonsági követelményeket (pl. hozzájárulások kezelése, személyes adatok kezelésének korlátozása, panaszbenyújtás módjainak kialakítása és a beérkező panaszok kezelése, tiltakozási jog gyakorlásának biztosítása és a jogi dokumentációk elkészítése)
- előkészítik, illetve felülvizsgálják a szerződéseket, különösen, ha különböző típusú együttműködésre vagy adatmegosztásra kerül sor
- biztosítják – az elszámoltathatóság elvének megfelelve – a jogi szempontoknak megfelelő szerződéskötési és dokumentálási gyakorlatot (pl. közös adatkezelésekkel és adatfeldolgozásokkal kapcsolatos megállapodások), közreműködnek a felhasználói, a licenc- és egyéb szerződések kidolgozásában
- közreműködnek a projektben résztvevő felek jogi felelősségének meghatározásában
- megfelelő jogi keretek kidolgozásával hozzájárulnak a potenciális jogi viták megelőzéséhez.

Egy LLM számos, különböző szakterületen dolgozó jogász együttműködését igényelheti, a személyes adatokkal dolgozó modellek esetében pedig elengedhetetlen az adatvédelmi és adatbiztonsági szakjogászok bevonása olyan területeken, mint:

- az adatkezelési folyamatok jogi megfelelőségének biztosítása, illetve az ezzel kapcsolatos adatvédelmi tájékoztatók és hozzájárulási mechanizmusok kidolgozása
- a beépített adatvédelem és adatbiztonság követelményeinek érvényesítése, szabályzatok és eljárások kidolgozásának, adatbiztonsági intézkedések megvalósításának, adatvédelmi incidensek kezelésének és az érintettek érdekeinek, jogainak és szabadságainak érvényesülésének támogatása
- adattovábbításhoz, illetve adatmegosztáshoz szükséges megállapodások kidolgozása
- jogos érdekre hivatkozó adatkezelések esetében érdekmérlegelési tesztek végzése, valamint a tiltakozási jog érvényesítésének elősegítése
- amennyiben a modell az érintettek jogaira és szabadságaira nézve jelentős kockázatú

adatkezeléseket tartalmaz, adatvédelmi hatásvizsgálat elvégzésének, valamint a szükséges kockázatenyhítő intézkedések kidolgozásának támogatása
- adatvédelmi incidensek kezelése (kockázatkezelés, kárelhárítás és kárenyhítés, jogszabályban előírt kötelezettségek teljesítése, például felügyeleti hatóság értesítése, érintettek tájékoztatása stb.)
- olyan tartalmak vagy tevékenységek azonosítása és kezelése, amelyek jogellenesek vagy jelentős jogi kockázatokat hordoznak
- a jogszabályi környezet és a jogalkalmazási gyakorlat változásának nyomon követése
- adatvédelmi tudatosság növelése a saját alkalmazottak és az adatfeldolgozók körében.

A jogászok azonban csak akkor „hasznosak", ha optimális időben és mértékben vonják be őket, illetve, ha együtt tudnak működni a többi szereplővel.

Jogászok és informatikusok együttműködése

Ha megkérdezünk informatikusokat, hogyan tudnak együtt dolgozni a jogászokkal, valószínűleg visszakérdeznek, egyáltalán van-e olyan, aki képes erre. Ha pedig a jogászoknál érdeklődünk, hogyan működnek együtt az informatikusokkal, ne csodálkozzunk, ha a fekete lyukak keletkezéséről kezdenek el beszélni. Ezen ellenérzéseket leküzdve fontos, hogy a jogászok már a követelmények meghatározásakor jelen legyenek, és ne a projekt majdnem kész állapotában közöljék, hogy a súlyos jogi hiányosságok miatt vissza kell térni a rajtkockára és újra kell kezdeni az egészet, ezúttal a jogi-etikai kereteket is figyelembe véve. Még rosszabb, ha az adatvédelmi felügyeleti hatóság a szankcionáló határozatában közli, milyen jogi problémái vannak a modellel, és szólít fel minket a korrekciós intézkedések haladéktalan megtételére.

Milyen következményei lehetnek az együttműködés elmaradásának?

Például:
- nem vesszük kellő mértékben figyelembe a jogi követelményeket és korlátokat, illetve nem felelünk meg az etikai normáknak
- a félreértések, az információhiány és egyéb kommunikációs problémák akadályozhatják a hatékony és eredményes projektmenedzselést, illetve a tényleges munkavégzést
- a jogi és technikai szempontok összehangoltságának hiányában a projekt irányítása eltévedhet, ez pedig kihatással lehet a határidőkre, a költségekre és a minőségre, illetve jogi vitákhoz vezethet
- a modellünk olyan funkciókat tartalmazhat, amelyek jogellenesek vagy erkölcsileg aggályosak.
- Milyen lehetőségek vannak az együttműködésre az LLM-projekt kezdeti időszakában?

Például:
- rendszeres és intenzív kommunikáció a projekttagok között, például a közös műhelymunka és megbeszélések alkalmat adnak a tapasztalatok, nézőpontok, prioritások és követelmények megtárgyalására, valamint az üzleti célok, a technikai megvalósítás és a jogi előírások harmonizálására. A kevésbé formális párbeszéd elősegíti az egymás jobb megértését, valamint csökkentheti a „pusztába kiáltott" szavakat
- a követelmény- és funkcionális specifikációk, valamint egyéb projektdokumentációk rendszeres egyeztetése lehetővé teszi a jogi-etikai követelmények korrekt beépítését, a kockázatok azonosítását, a diszkrimináció, illetve a tisztességtelen működés megelőzését
- a projektre vonatkozó belső szabályzat, valamint hatásvizsgálatok és érdekmérlegelési tesztek előkészítése.

Mikor kezdődjön az együttműködés?

Az az optimális, ha az informatikusok minél korábban konzultálnak a jogászokkal. Az ötlet informális egyeztetése lehetőséget ad a jogi-etikai aggályok korai megfogalmazására, az informatikusok pedig elkerülhetik a nem megfelelő irányba gondolkodásból adódó felesleges munkát. Ehhez azonban az szükséges, hogy az ötlet megvalósíthatósága megítélhető, a jogi kockázatok pedig beazonosíthatóak legyenek. Ehhez az informatikusoknak meg kell osztaniuk az ötletüket, valamint az alapvető információikat a modell működésének módjáról, a tervezett adatkezelésekről és az esetleges harmadik felekkel való együttműködésről – ezen ismeretek hiányában a jogászok nem tudnak megfelelő véleményt kialakítani.

Számos aggály merülhet fel, például:
- Hogyan magyarázható majd meg az LLM működése? Ha nem megmagyarázható (pl. feketedoboz modellek), hogyan tudunk megfelelni a megmagyarázhatóság követelményének?

- Mi szükséges az adatvédelmi és MI alapelveknek megfelelésünkhöz? Hogyan oldható meg például az ember jelenléte, az érintetti kérelmek és panaszok kezelése?
- Lesznek megfelelő adatok a betanításhoz, teszteléshez, validáláshoz? Hogyan csökkenthetjük ezen adatok kezeléséből származó kockázatokat, például megoldható az adatok anonimizálása?
- Hogyan felelünk meg az etikus MI és a GDPR automatikus döntéshozatal követelményeinek?
- Hogyan kerülhetjük el az algoritmus torzításából eredő problémákat, hogyan lesz a modellünk diszkriminációmentes, megbízható és etikus?
- Ki fogja tesztelni és ellenőrizni a modellt?
- A modellünk mely pontjai a legsebezhetőbbek (kiber)biztonsági szempontból és hogyan oldható meg ezen kockázatok csökkentése? Milyen alapvető intézkedésekre lesz szükségünk és azok hogyan valósíthatók meg?

A párbeszédet mindaddig fent kell tartanunk, amíg tisztázzuk a követelményeket – ez az együttműködés a kulcs a jelentős pénzügyi kihatással járó jogi problémák, felesleges visszalépések és újratervezések megelőzéséhez.

A jogászok blokkolhatják az informatikusok kreativitását?

Gyakori vád, hogy a jogászok jönnek, elmondják ki, mit nem tehet, és ezzel kiölik az informatikusokból a kreativitást. De vajon ez igaz?

Tény, a jogászok nagyon lelombozóak tudnak lenni és néha az informatikusoknak az is felfoghatatlan, hogy vannak olyan előírások, amelyek túlmutatnak szakterületükön, ennek ellenére mégis vonatkoznak rájuk. Például nem lehet bármelyik felhőbe szabadon fellőni a személyes adatokat és egyáltalán nem lényegtelen az, hogy a világ mely pontján, milyen joghatóság alatt tároljuk a személyes adatokat.

A jogászok bevonása azonban nem feltétlenül blokkolja az informatikusok kreativitását, hanem – mindkét szakterület részéről építő kommunikáció esetén – akár még segítheti is az ötleteknek és magának a modellnek a hatályos jogi keretek közötti kreatív megvalósítását. A jogi korlátok figyelembevétele nem feltétlen jelenti az informatikusok gúzsba kötését, netalán az alkotási vágyuk kiölését. Ez azonban a bevont jogász szakmai kvalitásán is múlik, hiszen azzal még a legkiválóbb informatikus sem tud mit kezdeni, ha egy jogász egy problémát látván felolvassa a jogszabályt, majd kategorikusan kijelenti, amit az informatikus szeretne, azt nem lehet. Az inspiráló párbeszéd lényege éppen az, hogy nem azt kell ismételgetni, mit, hogyan nem lehet megcsinálni, hanem azt kell átgondolni, mit, hogyan lehet – ehhez pedig az informatikusnak is képesnek kell lennie megértenie, mi az adott jogszabályi korlátnak az oka és hogyan lehet elérni, hogy az ne vonatkozzon az adott esetre. Ha a jogász a korlát lényegét nem magyarázza el, vagy úgy adja azt át, hogy azt az informatikusnak esélye sincs megértenie, a megoldás fényévnyi távolságra kerülhet, a projekt pedig vakvágányra futhat. Éppen ezért a jogásznak úgy kell megfogalmaznia mondandóját, hogy a nem jogászok számára is emészthetővé tegye, valamint a megoldási javaslatokról el kell döntenie, belefférnek-e az irányadó jogi-etikai keretekbe. Ehhez azonban nemcsak biztos szakmai tudás kell, hanem felelősségvállalási készség is. A közösen értelmezett jogi-etikai keretek medret adnak a projektnek, az informatikusoknak pedig át kell gondolniuk, sőt időnkét át is kell értékelniük azt, hogyan közelítenek az olyan értékekhez, mint az emberi jogok meg az etikus MI.

A gyakorlatban ez a kommunikáció sokkal nehezebb, mint amilyennek látszik. Nemcsak az egymás iránt érzett ellenszenvet kell leküzdeniük a szereplőknek, hanem kreatívan kell felhasználnunk a rendelkezésre álló platformokat.

Milyen eszközöket használhatunk a jogászok és az informatikusok egymáshoz közelebb hozására?

Elsősorban hagyományos projekteszközöket, például:

- az együttműködést hatékonyan támogató szervezeti megoldásokat (rendszeres interdiszciplináris oktatások, a kommunikációra, valamint a közös területekre fókuszáló workshopok, csapatépítő tréningek). A kulcsszerep a projektvezetőé, illetve azé a személyé, akinek a feladata az eseményeken résztvevő szereplők bevonása az együttműködésbe, és aki nem hagyja, hogy bárki is hiúsági versenyt rögtönözzön vagy netalán szótlanul üldögéljen a sarokban.
- olyan csatornákat, amelyek alkalmasak a rendszeres és strukturált kommunikációra (rendszeres megbeszélések, közös munkacsoportok), segítve a kapcsolattartást, de nem engedve teret az ellenségeskedésnek, a rivalizálásnak és a mellébeszélésnek
- nyelvi egyensúly megteremtését egymás szakmai terminológiájának megértése érdekében
- tudásmegosztást, egymás szempontrendszerének megértése érdekében

- mentorálás rendszeresítését, a jogászok-informatikusok együttműködő párokba szervezését az információáramlás, a tudásmegosztás, valamint a minőségi emberi kapcsolatok támogatására
- szerepek és felelősségek tisztázását (mindenki tisztában legyen a saját helyzetével és senki se törekedjen túlzott dominanciára)
- írásbeliséget a naprakészen tájékoztatás, valamint a döntések, szabályozások és egyeztetések dokumentálása érdekében
- külső szakértők, például etikai tanácsadók bevonását, újfajta perspektívát és tudást hozva a projektekbe
- a felsővezetés dominanciáját, segítve a különutas csapattagok „közös útra" terelését, valamint a határidők betartását.

Olyan eszközöket célszerű használnunk, amelyek – lehetőleg – személyes jelenlétet igényelnek, komplexek és gondolkodásra, valamint a keretek és a projekt egészének átgondolására ösztönöznek. Nem kell feltétlen új eszközökben gondolkodnunk, átértékelhetünk és kreatívabb formában hasznosíthatunk számos, az LLM-projektek során egyébként is „kötelező" eszközt, például:

- a megvalósíthatósági tanulmányokat
- a hatásvizsgálatokat
- az érdekmérlegelési teszteket
- a projekt végrehajtásával kapcsolatos belső szabályzatot, valamint
- az adatvédelmi és adatbiztonsági ismeretek oktatását.

Jelen esetben a hangsúly nem ezen eszközök ajánlott/kötelező jellegén, hanem azok kicsit másként alkalmazásán van, céljuk pedig a konstruktív párbeszéd, egymás megértése és elismerése úgy, hogy mindeközben a jogi-etikai kereteket is megteremtjük.

A megvalósíthatósági tanulmány

A megvalósíthatósági tanulmány alapvető projekteszköz egy adott LLM megvalósíthatóságának elemzésére a projekt kezdeti fázisában még a tényleges fejlesztés megkezdése előtt. Gyakran több alternatívát is tartalmaz, objektív információkat és elemzéseket szolgáltatva a projekt előnyeiről, hátrányairól, korlátairól és kockázatairól.

Milyen követelményeknek kell megfelelnünk az LLM fejlesztése során?

Az Európa Bizottság mesterséges intelligenciával foglalkozó magas szintű szakértői csoportja (szakértői csoport) szerint az MI alkalmazása során – többek között, de nem kizárólag – az alábbi követelményeknek kell megfelelnünk (Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport 2019):

- *az emberi cselekvőképesség támogatása és emberi felügyelet*
- *műszaki stabilitás és biztonság*
- *adatvédelem és adatkezelés (a magánélet tiszteletben tartása, az adatok minősége és sértetlensége, valamint az adatokhoz való hozzáférés)*
- *átláthatóság (nyomon követhetőség, megmagyarázhatóság és tájékoztatás)*
- *sokféleség, megkülönböztetésmentesség és méltányosság*
- *társadalmi és környezeti jólét (fenntarthatóság és környezetbarátság), valamint*
- *elszámoltathatóság (ellenőrizhetőség, hátrányos hatások minimalizálás, jogorvoslat).*

A hagyományos IT-projektekben a megvalósíthatósági tanulmány rendszerint öt egymásra épülő elemből áll:

- *technikai megvalósíthatóság* (a rendelkezésünkre álló eszközökkel és szakemberekkel)
- *gazdasági megvalósíthatóság* (költség- és haszonértékelés, beruházási, üzemeltetési és karbantartási költségek, valamint az elvárt megtérülés és nyereség elemzése)
- *üzleti megvalósíthatóság* (a projekt illeszkedése az üzleti stratégiánkhoz, céljainkhoz és igényeinkhez)
- *jogi megvalósíthatóság* (az eredmény megfeleljen a jogszabályoknak és egyéb előírásoknak)
- *operatív megvalósíthatóság* (a projekt hatása szervezetünk napi folyamataira, szükséges erőfeszítések a fenntartás érdekében)
- *megvalósíthatóság ütemezése* (reális határidők a betartás és a betartatás érdekében).

Javasolt hatodik területként a fenntarthatóságot is beiktatnunk a megvalósíthatósági tanulmányba („Ecology by Design"), kitérve olyan témákra, mint a digitális technológia közvetlen negatív környezeti hatásainak csökkenése, jobb gazdálkodás a természeti

erőforrásokkal, illetve a digitális és egyéb szereplők közötti kapcsolatok.

Ezen építőkövek egymásra épülnek és egymást feltételezik – jogi megfelelés hiányában a technikai megvalósítás irreleváns, ha pedig nem rentábilis az adott projekt, nem érdemes gondolkodni benne még akkor sem, ha egyébként az informatikusok és a jogászok szerint minden rendben. A tanulmány alapján a döntéshozók meghatározzák a projekt érdemességét és a kockázatok kezelésének módját, valamint alapul szolgálhat a projektfinanszírozáshoz és a megvalósítási tervhez is.

Az MI-projektek megvalósíthatósági tanulmánya hasonló, ám komplexebb, például a jogi mellett az *etikai megvalósíthatóságra* is kiterjed, és vizsgáljuk a szükséges adatok beszerezhetőségét, felhasználhatóságát, a pénzügyi fedezetet és a beszerzés, valamint a felhasználás jogszerűségét-etikusságát.

A megvalósíthatósági tanulmány elkészítése során:

- az informatikusok elmagyarázzák mit szeretnének
- a pénzügyesek kiszámolják, amit az informatikusok kitaláltak, az rentábilis-e
- az értékesítők kikutatják, van-e piaci igény a modellünkre (amennyiben értékesítésre szánjuk)
- a jogászok meghatározzák a jogi-etikai kereteket (az informatikusok ötlete jogilag megvalósítható-e, és ha igen, hogyan)
- a biztonsági szakemberek eldöntik, megoldható-e a projekt védelme információbiztonság, illetve az üzleti titok védelme szempontjából stb.

Mire a hatástanulmány elkészül, a szereplőknek módjuk van megismerni egymás gondolkodásmódját, érveit, valamint kedvenc vesszőparipáit is, és arra is rájöhetnek, hogyan lehet a másik vitorlájából kifogni a szelet. Az is egyértelművé válik, kik kulcsfontosságúak a projekt szempontjából, kikkel „lehet együtt dolgozni” és kik azok, akikkel erre esély sincs. Sőt, ez utóbbiak cseréjére is sor kerülhet, ha nincs esély az érdemi javulásra.

Hatásvizsgálatok végzése

A hatásvizsgálat során különböző aspektusból felmérjük és értékeljük egy adott projekt lehetséges gazdasági, társadalmi, valamint környezeti hatásait, a jogi és szabályozási követelményeket, és az etikai-adatvédelmi kérdéseket. A folyamatba bevon(hat)juk az érintett feleket, például a közösség tagjait, a felhasználókat, az ügyfeleket, valamint különféle szakértőket is.

A hatásvizsgálat pozitív hozadéka, hogy

- azonosíthatjuk és elemezhetjük a kockázatokat és a problémákat,
- elősegíthetjük a felelősségteljes döntéshozatalt, illetve a fenntarthatóságot,
- meghatározhatjuk a szükséges korrekciós intézkedéseket,
- megismerhetjük és figyelembe vehetjük az érintett felek érdekeit és szükségleteit.

6. sz. ábra: Hatásvizsgálat
Forrás: https://www.behance.net/gallery/3754298/Privacy-Cartoons

Többféle hatásvizsgálatot végezhetünk, ezek összefüggenek, átfedhetik, de akár ki is egészíthetik egymást (pl. az adatvédelmi, az alapjogi és az etikai hatásvizsgálat). Ezek nem egyszeri alkalmak, hanem folyamatok, amelyeket nem együltő helyünkben kell elvégeznünk felesleges adminisztratív nyűgként letudva, hanem folyamatosan nyomonkövetjük a körülmények alakulását a modellünk teljes életciklusa alatt. A hatásvizsgálatok alapján korrekciókra lehet szükségünk azért, hogy az LLM a lehető legmegfelelőbben

illeszkedjen környezetébe, illetve minimalizáljuk a potenciális kockázatokat és hatásokat.

Milyen hatásvizsgálatokat végezhetünk?

- *Az etikai hatásvizsgálatban* az etikai kockázatokat és következményeket vizsgáljuk (diszkrimináció és elfogultság, a magánélet és a személyes adatok védelme, az autonóm döntéshozatal és az emberi értékek tiszteletben tartása).
- *A társadalmi hatásvizsgálatban* a társadalmi, gazdasági és kulturális hatásokat elemezzük (pl. a munkahelyek és foglalkoztatás változása, szociális egyenlőtlenségek, társadalmi befogadás és kirekesztés).
- *Az adatvédelmi és adatbiztonsági hatásvizsgálat* (DPIA) követelményeit a GDPR[21], illetve a LED[22] fekteti le, célja pedig az érintettek jogait és szabadságait érintő adatvédelmi és biztonsági kockázatok azonosítása és minimalizálása.
- *A jogszabályi és szabályozási hatásvizsgálatban* felmérjük és elemezzük a vonatkozó jogi előírásokat, szabályozásokat és jogalkalmazási gyakorlatot, ezek hatásait, illetve az ezekből eredő kockázatokat.
- *Az alapjogi hatásvizsgálatot* az MI használatáról szóló rendelet tervezete határozza meg (29a. cikk). Amennyiben kötelező a DPIA elvégzése, az adatvédelmi és az alapjogi hatásvizsgálatokat együtt kell elvégeznünk.

Számos hatásvizsgálati modell az adatvédelmi hatásvizsgálatra építi rá rétegenként a többit, esetről esetre – az adott projekthez igazodva –, eltérő tartalommal.

„Minden egyben" hatásvizsgálat

Nemzetközi szinten számtalan szervezet, számtalan hatásvizsgálat mintát tett közzé, ezek közül kiemelendő a kanadai kormány *algoritmikus hatásvizsgálata*[23] (Government of Canada, 2023), melynek célja, hogy segítséget nyújtson az algoritmusok és az MI-rendszerek hatásainak felmérésében, értékelésében és a bevezetéssel járó hatások enyhítésében, biztosítva az algoritmusok társadalmi hatásainak átláthatóságát és ellenőrizhetőségét. Az AIA kérdőív mintegy 80 kérdésből áll az üzleti folyamatokra, az adatokra, a rendszertervezésre, az algoritmusra és a döntésre vonatkozóan, a válaszok alapján pedig megjeleníti az adott projekt hatásszintjét, illetve tájékoztatást nyújt az irányadó követelményekről.

Az AIA:

- segít megérteni az algoritmusok és MI modellek potenciális hatásait az emberekre, a társadalomra és a szervezetünkre, lehetővé téve a projektek felelős tervezését, valamint az esetleges kockázatok és torzítások azonosítását
- elősegíti a párbeszédet és az átláthatóságot a projekt résztvevői és a szélesebb közösség között, valamint lehetőséget nyújt a szükséges korrekciók és javítások végrehajtására a tervezés korai szakaszában
- javítja az algoritmusokkal és MI-rendszerekkel kapcsolatos döntéshozatali folyamatokat és segít a technológia iránti bizalom építésében.

Az AIA „minden egyben" hatásvizsgálat, amely

- vizsgálja az adatok kezelésének módját a tervezés és működtetés során, valamint az adatgyűjtési, tárolási és feldolgozási gyakorlatokat, illetve az adatvédelmi irányelveknek megfelelést
- segít értékelni, megakadályozni, illetve csökkenteni a diszkriminációt, az előítéleteket, a torzításokat és az igazságtalanságokat
- felméri a társadalmi hatásokat, ideértve a szélesebb társadalmi és gazdasági következményeket (emberi jogok, munkaerőpiac, társadalmi egyenlőtlenségek, egészségügy és más területeken)
- támogatja a felelős és etikus tervezést, segít az értékalapú döntések meghozatalában, a társadalmi normák és értékek figyelembevételében, illetve az etikai irányelvek és keretrendszerek kidolgozásában
- támogatja az átláthatóságot és a számonkérhetőséget, valamint segít nyomon követni az algoritmusok működését és a döntéshozatali folyamatokat
- értékeli a biztonságot (potenciális sebezhetőségek, az adatvesztés kockázata, hozzáférési jogosultságok és az adatok védelme a külső fenyegetésekkel szemben)
- figyelembe veszi a felhasználói élményt és a felhasználhatóságot, illetve a felhasználók igényeihez és képességeihez illeszkedést
- felméri a gazdasági előnyöket és hatékonyságot, értékeli a költségeket és a megtérülést, a termelékenységet, illetve a piaci versenyképességet
- felhívja a figyelmet a jogi és szabályozási kérdésekre, értékeli a releváns jogi kereteket és a megfelelést („compliance").

21 GDPR 35. cikk

22 bűnügyi adatvédelmi irányelv (LED)

23 Algorithmic Impact Assessment, továbbiakban AIA

Az AIA szabadon elérhető, akár saját gyakorlatunkban is hasznosíthatjuk.

Az etikai hatásvizsgálat

> *„Általában a technológiáink az egyik kezükkel adnak, a másikkal pedig úgymond tarkón vágnak minket." (Charlie Brooker, a Black Mirror készítője) (Kafka 2023)*

Az etikai hatásvizsgálatban feltárjuk és felmérjük egy adott projekt vagy technológia etikai kockázatait és azok következményeit. Amennyiben a szakértői csoport etikai iránymutatását vesszük irányadónak, akkor a megbízható LLM jogszerű, etikus, valamint műszaki és társadalmi szempontból is stabil annak érdekében, hogy lehetőleg ne okozzon kárt (Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport 2019).

Miért fontos, hogy az LLM etikus legyen?

Amikor intelligenciát igénylő tevékenységet végzünk, felelősséget kell vállalnunk ítéleteink pontosságáért, megbízhatóságáért és megalapozottságáért, valamint az is elvárás, hogy cselekedeteinket és döntéseinket megfelelően támasszuk alá, illetve másokkal szemben tisztességesen, méltányosan és észszerű módon járjunk el.

Az LLM megjelenése és intelligenciát igénylő tevékenységekhez szükséges hatalmának növekedése a kognitív funkciók széles körének algoritmikus folyamatokra való átruházását jelenti, amelyekért ezek a modellek maguk közvetlenül nem felelősek, és nem is vonhatók azonnal felelősségre viselkedésük következményeiért (7. sz. ábra).

7. sz. ábra: LLM és etikai felelősség
Forrás: Az ábra alján

Az LLM és más MI-modellek, mint inaktív és programalapú gépek erkölcsileg nem elszámoltatható ágensek, ezt az etikai rést pedig jelenleg az MI etikájának egyre növekvő kerete próbálja betölteni olyan elvek érvényesítésének megkövetelésével, mint a tisztességesség, az elszámoltathatóság, a megbízhatóság, a fenntarthatóság és az átláthatóság.

A Holberton-Turing-eskü célja, hogy az MI szakértőket közös értékek ernyője alá gyűjtse.

„A HOLBERTON-TURING ESKÜ

Az adattudomány és a mesterséges intelligencia szakma képviselőjeként ünnepélyesen fogadom, életemet az emberiség szolgálatának szentelem.

Humanitás és etika:

Az emberi élet iránti legnagyobb tiszteletet tanúsítom.

Nem engedem, hogy az életkor, betegség vagy fogyatékosság, hitvallás, etnikai származás, nem, nemzetiség, politikai hovatartozás, vallási meggyőződés, faji hovatartozás, szexuális irányultság, társadalmi helyzet vagy bármely más tényező szempontjai közrejátszanak a munkavégzésemben.

Nem használom fel tudásomat az emberi jogok és polgári szabadságjogok megsértésére, még fenyegetés esetén sem.

Adattudomány, a mesterséges intelligencia művészete, magánélet és személyes adatok:

Tiszteletben tartom azoknak a tudósoknak és mérnököknek nehezen megszerzett tudományos eredményeit, akiknek a nyomdokaiban járok, és örömmel osztom meg a tudásomat mindazokkal, akik utánam jönnek.

Nem felejtem el, hogy a mesterséges intelligencia a tudomány mellett művészet is, és hogy az emberi szempontok fontosabbak a technológiai szempontoknál.

Tiszteletben tartom az emberek magánéletét és azt, hogy személyes adataikat nem tárják fel a mesterséges intelligencia rendszerek előtt, hogy a világ megismerje azokat.

Nem felejtem el, hogy nem száraz adatokkal, puszta nullákkal és egyesekkel találkozom, hanem emberi lényekkel, akiknek a mesterséges intelligenciával való kapcsolata befolyásolhatja a szabadságát, családját vagy gazdasági stabilitását.

Tiszteletben tartom a rám bízott titkokat.

Napi munka és etikett:

Szakmámat lelkiismeretesen és méltósággal gyakorlom.

Ápolom az adattudományi és mesterséges intelligencia szakma becsületét és nemes hagyományait.

Megadom tanáraimnak, kollégáimnak és diákjaimnak az őket megillető tiszteletet és hálát.

Megosztom tudásomat az emberek javára, az adattudomány és a mesterséges intelligencia fejlődésének érdekében.

Figyelembe veszem munkámnak a méltányosságra gyakorolt hatását mind a történelmi előítéletek állandósításában, amelyet a múltbeli adatokból a jövőre vonatkozó előrejelzésekre való vak kivetítése okoz, mind pedig a gazdasági vagy egyéb egyenlőtlenségeket növelő új feltételek megteremtésében.

Mindezen ígéreteimet elsősorban azért teszem, hogy a mesterséges intelligenciát úgy hozzam létre, hogy az a közjó érdekében együttműködjön az emberekkel, ahelyett, hogy bitorolná az emberi szerepet kiszorítva őket.

Ezeket az ígéreteket ünnepélyesen, szabadon és becsületemre fogadom meg."

Mikor etikus egy mesterséges intelligencia?

A szakértői csoport szerint négy fő etikai elv betartására kell törekednünk:

1. *Az emberi autonómia tiszteletben tartásának elve*
    - az MI-rendszerekkel nem rendelhetjük alá, nem kényszeríthetjük, nem téveszthetjük meg, nem manipulálhatjuk és nem kondicionálhatjuk indokolatlanul az embereket
    - a modelleket úgy kell kialakítanunk, hogy azok az emberek kognitív, szociális és kulturális készségeit fokozzák, kiegészítsék és megerősítsék
    - az emberek és MI-rendszerek közötti feladatmegosztás elve az emberközpontú kialakítás; az emberi döntésnek jelentős szerepet kell hagynunk, és biztosítanunk kell, hogy az ember felügyelhesse és ellenőrizhesse a modellek munkafolyamatait
    - a kár megelőzésének elve: az MI-rendszerek soha nem okozhatnak kárt, illetve nem lehetnek hátrányos hatással az emberi lényekre, valamint meg kell akadályoznunk a rosszindulatú használat lehetőségét, a természetes környezet és valamennyi élőlény figyelembevételét is ideértve
2. *A méltányosság elve*

Az MI-rendszerek kifejlesztésének, elterjesztésének és használatának méltányosnak kell lennie:

- az előnyök és a költségek egyenlő és igazságos elosztása
- annak biztosítása, hogy az egyéneket és csoportokat ne érje méltánytalan torzítás, hátrányos megkülönböztetés és megbélyegzés
- az oktatáshoz, termékekhez, szolgáltatásokhoz és technológiához való hozzáférés terén esélyegyenlőség támogatása
- nem eredményezheti a (végső) felhasználók megtévesztését vagy választási szabadságának sérelmét
- tiszteletben kell tartanunk az eszközök és célok arányosságának elvét, és egyensúly kell teremtenünk a versengő érdekek és célkitűzések között
- a modellek és az e rendszereket üzemeltető emberek hozta döntések megtámadhatósága (hatékony jogorvoslat lehetősége)
- a döntésért felelősek azonosíthatók, és a döntéshozatali folyamatok megmagyarázhatók legyenek

3. *A megmagyarázhatóság elve*
    - a modellek képességeit és célját nyíltan közölnünk kell
    - a döntéseket lehetőség szerint el kell magyaráznunk
    - a „fekete doboz" algoritmusok esetében egyéb intézkedésekre, például a rendszer képességeinek nyomonkövethetőségére, ellenőrizhetőségére és az azokkal kapcsolatos átlátható tájékoztatásra van szükségünk
    - a megmagyarázhatóság szükséges mértéke nagyban függ attól, hogy téves vagy pontatlan eredmény esetén a következmények mennyire súlyosak az érintettekre nézve

Ezen elvek akár ütközhetnek is egymással, ekkor bizonyítékokon alapuló megfontolt döntésre van szükségünk ahhoz, hogy egyik vagy másik elvet – jogszerűen – háttérbe szoríthassuk.

4. *Stabil mesterséges intelligencia*

Az egyéneknek és a társadalomnak bíznia kell abban, hogy az MI nem okoz kárt, ennek érdekében óvintézkedéseket kell tennünk.

Mire összpontosít az etikai hatásvizsgálat?

- Az *adat etikussága* (nyomonkövethetőség, elérhetőség, integritás, biztonság, nem szelektív gyűjtés stb.), az algoritmus etikussága (megbízhatóság, védettség, cél, torzítás, minőség, megmagyarázhatóság, átláthatóság stb.) a rendszer etikussága (ergonómia, alkalmazkodóképesség, következetesség, bizalmasság stb.), a gyakorlatok etikussága (kultúra, szabályok, szakmaiság, megbízhatóság, a magánszféra védelme, elszámoltathatóság,

diverzifikációk, integráció stb.), illetve a döntések etikussága (autonómia, szabad akarat, dehumanizáció, menedzsment és irányítás, felelősség, nyilvánosságra hozatal, környezet, fenntarthatóság stb.)
- *diszkrimináció,* illetve elfogultság azonosítása és kezelése
- *adatvédelem és magánélet védelme* (a felhasználói adatok kezelése, hozzáférhetősége és a vonatkozó jogszabályok, valamint belső szabályok betartása)
- *autonóm döntéshozatal* (átláthatóság, felelős módon működés, döntésekért felelősségvállalás)
- *társadalmi hatás*ok és esetleges *egyenlőtlenségek* **(**munkahelyek változása, szociális és gazdasági egyenlőtlenségek, társadalmi befogadás és kirekesztés)
- *felhasználói bizalom és felelősség* (átláthatóság, a felhasználók megért(het)ik-e a modell működését, felelős használat és esetleges korlátozások).

Az etikai szempontokat már a tervezés időszakában figyelembe kell vennünk (*Ethics by Desing*), valamint az LLM teljes életciklusában is (*Ethics by Evolution*) (Leroy 2021).

A jogi (jogszabályi megfelelőség, adatvédelmi, alapjogi stb.) és etikai hatásvizsgálatok hasonlóak, de különböznek is (8. sz. ábra):

	Jogi hatásvizsgálat	Etikai hatásvizsgálat
Fókusz	A modell megfeleljen a vonatkozó jogszabályoknak, szabályozásoknak és előírásoknak	A modell társadalmi, etikai és morális következményei, az erkölcsi dilemmák és érdekek felismerése, valamint a társadalmi elfogadottság növelése
Szempontok	Adatvédelmi jogszabályok, szerzői jogok, szellemi tulajdon védelme, felhasználói szerződések és egyéb releváns jogi aspektusok	Felelősségteljesség, a diszkrimináció megelőzése, a személyes adatok védelme, az átláthatóság és az igazságosság a technológiai megoldásokban
Szereplők	Általában jogászok és olyan egyéb szakértők bevonásával, akik képzettek a jogi kérdések elemzésében és az alkalmazandó jogszabályok, valamint a vonatkozó jogalkalmazási gyakorlat megértésében	Általában olyan etikai szakemberek, filozófusok, társadalomtudósok és egyéb szakértők bevonásával, akik hozzáértők a társadalmi és etikai kérdések elemzésében és értékelésében
Lehetséges válaszok, megoldások	Jogszabály által behatárolt, adott esetben lehet, hogy csak egyetlen jogszerű megoldás van	Soha nincs abszolút és egyértelmű válasz; az adott kérdéstől függően vannak legjobb lehetséges választások

8 sz. ábra: A jogi és az etikai hatásvizsgálatok összehasonlítása
Forrás: Saját szerkesztés

Az UNESCO-nak is van olyan etikai hatásvizsgálat iránymutatása, amelyet segítségül hívhatunk (UNESCO 2023).

Míg a jogszabályok értelmezésében könnyebb egységes álláspontja jutni, az etikai kérdések megítélésében sokkal nagyobb eltérések lehetnek például faji, etnikai hovatartozástól, életkortól, az iskolai végzettségtől, de akár jövedelmi szinttől függően is (9. sz. ábra). Ha azt nézzük, hogy a bibliai tízparancsolathoz, vagy a hét főbűnhöz hozzáállásban mekkora különbségek lehetnek, akkor hogyan juthatunk megegyezésre olyan összetett problémákban, mint például az önvezető autók tipikus dilemmái?

9. sz. ábra: Hovatartozás
Forrás: Az ábra alján

„Jól ismert példa erre az úgynevezett troliprobléma, amely Philippa Foot brit filozófusra vezethető vissza. Ez egy olyan gondolatkísérlet, amely arra hivatott, hogy tesztelje erkölcsi intuíciónkat azzal kapcsolatban, hogy erkölcsileg megengedett-e, sőt szükséges-e feláldozni egy ember életét annak érdekében, hogy több ember életét megmentsük.

Az autonóm járművek szerkezetileg hasonló helyzetekkel szembesülhetnek, amelyekben elkerülhetetlen, hogy mások megmentése érdekében egy vagy több személynek kárt okozzanak, vagy akár meg is öljenek egy vagy több személyt. Tegyük fel, hogy egy önvezető autó nem tud megállni, és csak a választás lehetősége van számára, hogy két embercsoport egyikébe hajtson: egyrészt két idős férfi, két idős nő és egy kutya; másrészt egy fiatal nő egy kisfiúval és egy kislánnyal. Ha az első csoportba hajt bele, a két nő meghal, a két férfi és a kutya súlyosan megsérül. Ha a második csoportba hajt bele, az egyik gyerek meg fog halni, a nő és a másik gyerek pedig súlyosan meg fog sérülni.

A helyzetet tetszés szerint további részletekkel lehet kiegészíteni. Tegyük fel, hogy az idős emberek csoportja a kutyával a közlekedési szabályoknak megfelelően viselkedik, míg a nő és a gyerekek a piros jelzés alatt mennek át az úton. Ez erkölcsileg releváns? Változtatna-e a helyzeten, ha az egyik idős férfi helyébe egy fiatal orvos lépne, aki sok ember életét menthetné még meg?" (Misselhorn 2022: 32–33)

A döntési helyzetet tovább bonyolítja, ha mi ülünk az autóban. Az emberek nagy többsége úgy véli, a halálos áldozatok összességének minimalizálására programozott önvezető járművek etikusabbak – de inkább olyan járművet vásárolnának, amelyik előnyben részesíti a járműben ülő életét. Sőt, az emberek többsége egyenesen elutasítaná az etikusabbnak tartott autó megvásárlását (Bonnefon–Shariff–Rahwan 2016), hiszen ki venne meg olyan járművet, amelyik bármikor képes feláldozni a saját utasát? Hiába törekednénk a közjó szempontjából a legtökéletesebb megoldásra, ezt az egyének önérdekű viselkedése nem fogadná el – azaz az MI-projektünk lehet, hogy etikus, de pénzügyi szempontból garantáltan kudarc. És még olyan kapaszkodónk sincs, mint a jogszabályok távolságtartó, „száraz" paragrafusai vagy az irányadó jogalkalmazási gyakorlat. Az etikai hatásvizsgálat éppen ezért lelkileg nagyon megterhelő is lehet, különösen akkor, ha emberélet feletti döntésről van szó.

Mi a szerepük a jogászoknak és az informatikusoknak az etikai hatásvizsgálatban?

A jogászok támogatják az etikai döntések jogi vonatkozásainak értékelését-értelmezését, és a felelősségi problémák, valamint a következmények meghatározását az automata döntéshozatal, a profilozás és a modell kapcsán. Azonosítják az adatvédelem és a magánélet védelmével kapcsolatos jogi és etikai kérdéseket, illetve meghatározzák az adatvédelem, valamint más jogágak támasztotta követelményeket.

Az informatikusok felelősek a technikai kérdések azonosításáért, valamint közreműködnek annak meghatározásában, hogy az etikai szempontokra tekintettel milyen technikai lehetőségeink vannak és milyen korlátokkal szembesülhetünk. Kreatív megoldásokat dolgozhatnak ki, megtalálva azokat a módszereket és technikákat, amelyek megfelelnek az etikai követelményeknek. Ezen kívül jelentős szerepet vállalnak a felelős LLM-fejlesztés és alkalmazás elősegítésében, valamint gondoskodnak arról, hogy a

modellek átláthatóak, megmagyarázhatóak, megfelelően dokumentáltak és teszteltek, valamint a felhasználók megértése és bizalma szempontjából megfelelőek legyenek.

Amennyiben a jogászok és az informatikusok együttműködése elmarad, az etikai problémákat kaotikusan kezelhetjük, illetve figyelmen kívül hagyjuk, vagy akár fel sem ismerjük az etikai kockázatokat, ez pedig aggályos eredményhez vezethet (pl. előítéletesség, elfogadhatatlan, etikátlan adatkezelési gyakorlat, adatmanipuláció vagy személyes adatok tisztességtelen felhasználása). A tisztességtelen LLM jelentős társadalmi visszhangot és tiltakozást válthat ki, az együttműködés hiánya pedig növeli a reputációs károk, a társadalmi ellenállás, valamint a jogi konfliktusok kockázatát.

A társadalmi hatásvizsgálat

> *„A jog fejlődésének középpontja nem a törvényhozásban, nem a doktrínákban, nem a joggyakorlatban keresendő, hanem a társadalomban." (Ehrlich)*

A társadalmi hatásvizsgálatban feltárjuk, felmérjük és elemezzük projektünk társadalmi, illetve egyénekre gyakorolt következményeit, valamint az esetleges egyenlőtlenségeket és társadalmi változásokat.

„A mesterséges intelligencia egyre szélesebb körű alkalmazása az egészségügyben várhatóan megváltoztatja a klinikai munka és az egészségügyi ellátás jellegét. Az egyik elvárás az, hogy a mesterséges intelligencia megkönnyíti a rutinfeladatok és az adminisztráció automatizálását. Egyesek szerint ennek eredményeként az orvosoknak és más egészségügyi dolgozóknak több idejük marad a betegekre. Mások viszont azzal érvelnek, hogy ha a betegek gyakrabban lépnek kapcsolatba a mesterséges intelligenciával, például a telemedicina keretében, akkor az orvosok által a betegekkel töltött idő csökkenni fog. Továbbá, bár a mesterséges intelligencia helyettesíthet bizonyos rutinfeladatokat, az orvosoknak és az ápolóknak több időt kell majd tölteniük a technológia kezelésével, az adatok elemzésével és az új mesterséges intelligencia-alkalmazások használatának megtanulásával.

Ehhez kapcsolódó kérdés, hogy az MI az egészségügyben munkanélküliséghez vezet-e, és hogy a dolgozókat fel- vagy át lehet-e képezni, hogy elkerüljék a kiszorulást és alkalmazkodjanak az MI által támogatott orvosláshoz. A mesterséges intelligencia általi automatizálás miatti munkahelyvesztés széles körben elterjedt aggodalom, és az előrejelzések szerint az egészségügyi munkaerő szinte minden részterületére, köztük az orvosok és más szakértők bizonyos típusaira is vonatkozik. Mások azt állítják, hogy az MI új munkahelyeket fog teremteni az egészségügyben, ami ellensúlyozhatja a lehetséges veszteségeket, és hogy az MI segíthet enyhíteni az egészségügyi munkaerőhiányt, többek között az alacsony és közepes jövedelmű országokban, ahol az egészségügyi személyzet hiánya gyakran komoly problémát jelent.

Bár ezek a forgatókönyvek eltérőek, abban egyetértés van, hogy a mesterséges intelligencia és más digitális technológiák beágyazása az egészségügybe és az orvosi kutatásba az egészségügyi személyzet átképzését teszi szükségessé, hogy alkalmazkodni tudjanak az új szerepekhez. A WHO 2021-es, az egészségügyi célú mesterséges intelligencia etikájáról és irányításáról szóló iránymutatása [World Health Organization (2021)] szintén felveti az egészségügy „überizálódásával" kapcsolatos aggodalmakat, ami az orvosok és a gyakorló orvosok munkáját kevésbé stabillá és kevésbé biztonságossá teheti. A jelentés szerzői szerint a mesterséges intelligencia által vezérelt egészségügyi platformok létrehozása az egészségügyi ágazatban a „gig-gazdaság" növekedéséhez vezethet, amelyben az ápolók, orvosok és más alkalmazottak igény szerint, ideiglenes vállalkozóként, a foglalkoztatás stabilitása nélkül dolgoznak. Ez a fejlemény – figyelmeztetnek – alááshatja a betegek és az egészségügyi szolgáltatók közötti kapcsolatot is, ami több röpke interakciót, az ellátás minőségének csökkenését és a bizalom elvesztését eredményezheti." (Rosemann–Zhang 2021: 105)

A társadalmi hatásvizsgálat során elemezzük, hogy projektünk:

- várhatóan hogyan változtathatja meg a munkavállalás szerkezetét, mely területeken szűnnek meg munkahelyek és hol van szükség újakra, valamint a változás milyen (új) készségeket igényel a munkavállalóktól
- hatással van-e a szegénységre, a jövedelmi különbségekre, az oktatáshoz és egészségügyhöz való hozzáférésre, valamint az esélyegyenlőségre
- milyen hatással lehet a társadalmi befogadásra és kirekesztésre
- hogyan érinti a különböző csoportokat, beleértve a kisebbségeket, a fogyatékossággal élőket vagy a társadalmilag elnyomott, kirekesztett csoportokat, és hogy esetlegesen tovább erősíti-e a meglévő egyenlőtlenségeket vagy diszkriminációt
- hogyan járulhat hozzá a társadalmi változásokhoz és elmozdulásokhoz, valamint hogyan alakíthatja át az üzleti modelleket, a közszolgáltatásokat vagy a társadalmi szokásokat, és ez milyen hatással lehet a társadalmi normákra és értékekre.

Az érdekeltek hatásvizsgálata (SIA)

Az Alain Turing Intézet (Leslie 2019) a közszféra MI-projektjeire összpontosítva ezen projektek társadalmi hatásával és fenntarthatóságával kapcsolatos hatásvizsgálatot az érdekeltek hatásvizsgálatának nevezi (SIA[24]), melyben az „érdekeltek" kifejezés alatt elsősorban az érintett személyeket érti, de a fogalom kiterjedhet csoportokra és szervezetekre is, mivel e kollektívák egyes tagjaira is hatással lehet az MI-rendszerek bevezetése. Az ajánlás szerint a SIA-t a projekt három kritikus pontján kell elvégeznünk:

- *a probléma megfogalmazása („Alfa fázis")*
  célja az etikai megengedhetőség megállapítása, kiindulási pont, az érintettek beazonosítása, a célok felállítása, valamint az MI-projekt egyéni jólétre, valamint a közjólétre gyakorolt lehetséges hatásainak mérlegelése
- *a bevezetést megelőzően („Alfától Bétáig fázis")*
  a betanítás, tesztelés és validálás után újra elő kell vennünk az eredeti SIA-t megerősíteni, hogy a modellünk még mindig összhangban van az eredeti értékelésünkkel. Meg kell vizsgálnunk a képzett modell célját, a tervezési és tesztelési eredményeket, valamint ezek viszonyát az eredeti SIA-hoz, teret engedve a potenciális veszélyekkel kapcsolatos aggályainknak is (pl. diszkrimináció, torzítás)
- *újraértékelés („Béta fázis")*
  az élesben beüzemelés után időről időre újra el kell végeznünk a SIA-t, összevetve az eredeti SIA-val annak érdekében, hogy a felmerülő problémákat megoldhassuk.

Mi a szerepe a jogászoknak és az informatikusoknak a társadalmi hatásvizsgálatban?

A jogászok felelősek a releváns jogi előírások, szabályozások és jogalkalmazási gyakorlatok azonosításáért, valamint a vonatkozó jogi követelmények megértéséért és megértetéséért. Közreműködnek a jogi kockázatok azonosításában, illetve a jogilag elfogadható stratégiák és megoldások kidolgozásában.

Az informatikusok felelősek a technikai hatások azonosításáért, valamint segítenek megérteni, hogy a projekt milyen változásokat hozhat, és ezek a változások hogyan befolyásolhatják a társadalmat és az embereket. Jelentős szerepet vállalnak a felelős technológiai tervezésben, például olyan megoldások kialakításában, amelyek minimalizálják a negatív, illetve elősegítik a pozitív hatásokat.

A jogászok és az informatikusok együttműködésének hiányában előfordulhat, hogy nem vesszük figyelembe vagy akár teljesen figyelmen kívül hagyjuk a potenciális negatív társadalmi hatásokat (szociális egyenlőtlenség, diszkrimináció, kizárólagosság stb.), helytelenül vagy hiányosan kommunikálhatjuk azokat félreértést, bizalmatlanságot vagy ellenszenvet eredményezve; az időben nem azonosított és kezelt aggályok pedig reputációs károkat, jogi problémákat vagy üzleti veszteségeket okozhatnak.

Az adatvédelmi hatásvizsgálat

A DPIA során feltárjuk az adatkezelések jellegét, szükségességüket és arányosságukat, valamint a személyes adatok kezeléséből eredő, a természetes személyek jogait és szabadságait érintő kockázatokat, illetve e kockázatok értékelésével és a kezelésükre szolgáló intézkedések meghatározásával minimalizáljuk ezeket. Az Európai Adatvédelmi Testület elődje, a 29. cikk szerinti adatvédelmi munkacsoport (WP29) a hatásvizsgálatokkal kapcsolatban iránymutatást adott ki (WP 248 rev.01).

Egy LLM-projekt során hol kezelhetünk személyes adatokat?

Sokkal több adat személyes adat, mint ahogy egy átlagos projektrésztvevő azt gondolná. Ez számos konfliktushoz vezethet, különösen az informatikusok esetében, akik gyakran hitetlenkedve veszik tudomásul, hogy például a rendszerhasználattal kapcsolatos adatok is személyes adatok (ki, mikor lépett be, mit csinált, meddig volt bent stb.). A jogászok egyik legfontosabb feladata elmagyarázni – sok-sok érthető példával – valójában mely adatok tartoznak a személyes adatok körébe és miért. A jogászok azonban nem csodálkozhatnak azon, hogy más szakterületek képviselői milyen „tudatlanok", hiszen napjainkban még a legalapvetőbb adatvédelmi ismeretek sem tartoznak bele az általános műveltségbe. Sőt, a nem adatvédelmi specialista jogászok szintén kevés tudással rendelkeznek e téren és nekik is gyakran meggyűlik a bajuk a személyes adatokkal kapcsolatos problémákkal.

A DPIA egyik legfontosabb küldetése annak kiderítése, hogy egy LLM-projekt során hol és mily módon kezelünk személyes adatokat, illetve ezek az adatkezelések milyen kockázattal járnak az érintettek jogaira és szabadságaira nézve. Ehhez az szükséges, hogy az informatikusok képesek legyenek megmondani,

24 Stakeholder Impact Assessmen (SIA)

pontosan hol és milyen személyes adatok vannak, azokkal mi történik, valamint a jogászok beazonosítsák, milyen jogalapra hivatkozva történik ezen adatok kezelése, illetve megfelelünk-e a GDPR 5. cikkében foglalt alapelveknek, például a célhoz kötöttség, az adattakarékosság, az adatpontosság és a korlátozott tárolhatóság elvének. A DPIA feladata megfelelőségünk mértékének feltárása, illetve a hiányosságok esetében korrekciós intézkedések javaslata, természetesen mindvégig az elszámoltathatóság szuperelvének megfelelve.

Többek között, de nem kizárólag személyes adatok lehetnek:

- *képzési adatok*
  A nyers képzési adatok személyes adatokat tartalmazhatnak, így közvetlenül fennállhat felelősségünk az érintetti jogokkal kapcsolatban, például a hozzáférési és törlési kérelmekre reagálnunk kell. Az előkészítés eltávolíthatja a képzési adatokból a legtöbb közvetlenül azonosító adatot, ha azonban az előkészítés után az adatok még mindig összekapcsolhatók egy adott személlyel, akkor azok tekintetében a GDPR/LED szabályai továbbra is érvényesek.
- *modellben lévő személyes adatok*
  Az LLM jellemzően nem tartalmaz személyes adatokat, ha azonban a modell szándékosan kezel adatfragmentumokat, akkor el kell döntenünk, hogy a GDPR/LED alkalmazandó-e, így például az is előfordulhat, hogy az érintettnek joga lehet személyes adatainak törlésére, ami szükségessé teheti a modell újratanítását.
- *működéssel kapcsolatos személyes adatok*
  A modell be-, illetve kimenete is tartalmazhat személyes adatokat, és ez számos érintetti jogot aktiválhat, beleértve az automatizált döntéshozatalhoz, illetve a helyesbítéshez kapcsolódó jogokat is. Amennyiben az eredményt csak statisztikailag megalapozott találgatásként kezeljük (lásd pl. az ajánló alkalmazások), az érintett nem élhet a helyesbítéshez való jogával arra hivatkozva, hogy az eredmény pontatlan.

Az adatvédelmi hatásvizsgálat tartalma

A GDPR nem határozza meg külön a DPIA fogalmát, de a minimális tartalmát rögzíti[25]. A NAIH[26] hangsúlyozza, hogy „a kockázatelemzés a személyes adatok kezelésével összefüggő folyamatokra, adatkezelési műveletekre vonatkozik, amelyek a hatásvizsgálat lefolytatásának eredményeként az érintett jogait és szabadságait érintő kockázatot jelentenek. Az adatvédelmi hatásvizsgálat lényege az adatkezelés előzetes kontrollja a kockázatok feltárása és a kockázatok mérséklésére teendő intézkedések értékelése révén. A kockázatnak egyértelműnek, konkrétnak kell lennie, és ahhoz, hogy az adatkezelő azonosítani tudjon kockázatokat, meg kell előznie egy kockázatelemzési folyamatnak". (NAIH 2023)

A DPIA során ez a kockázatalapú megközelítés szükségessé teszi az egymással versengő, illetve ellentétes érdekek közötti kompromisszumokat, például annak mérlegelését, hogy a lehető legtöbb személyes adatot kívánjuk-e felhasználni egy modell kiképzéséhez annak pontosságának biztosítása érdekében, szemben az ilyen nagy adathalmaz felhasználásával járó adatvédelmi kockázatokkal, például az adattakarékosság és a célhoz kötöttség elve betartásának követelményével. Meg kell határoznunk, hogy a különböző tevékenységeket milyen jogalapokra hivatkozva kívánjuk végezni, a személyes adatok különleges kategóriájába tartozó, valamint a bűnügyi adatok esetében pedig a 9. cikk (2) bekezdésében felsorolt kivételek egyikének megfelelést is alá kell támasztanunk.

Azt is vizsgálnunk kell, hogy milyen joghatóságok alatt kezeljük az adatokat (pl. harmadik országban és emiatt külön garanciákat fel kell-e mutatnunk, illetve van-e olyan tagállami jogszabály, amelyek még tovább szűkíti a mozgásterünket). Alá kell támasztanunk, hogyan felelünk meg a szükségesség és az arányosság, illetve a fokozatosság követelményének, valamint figyelemmel kell kísérnünk a diszkrimináció lehetőségét is.

A DPIA főbb területi:

- a személyes adatok kezelése
  A DPIA során azonosítjuk az adatvédelmi kockázatokat és meghatározzuk a megfelelő intézkedéseket, például a titkosítási módszereket, a belső szabályzatainkat, a hozzájárulási mechanizmusokat, valamint a szükséges érdekmérlegeléseket
- sérülékenység és potenciális fenyegetettség, külső és belső támadások lehetőségei (pl. adathalászat, rosszindulatú szoftverek vagy jogosulatlan hozzáférés), biztonsági intézkedések (pl. tűzfalak, jogosultságkezelés, biztonsági protokollok, rendszeres frissítések, biztonsági mentések stb.)
- a felhasználói hozzáférés és azonosítás, az adatokhoz és a rendszerhez való hozzáférés kezelése és ellenőrzése (azonosítási és hitelesítési mechanizmusok)
- a projektünk adatvédelmi jogszabályoknak és egyéb szabályozásoknak, iránymutatásoknak,

25 GDPR 35. cikk (7) bekezdés

26 Nemzeti Adatvédelmi és Információszabadság Hatóság

valamint a vonatkozó jogalkalmazási gyakorlatoknak megfelelősége.
- minden egyéb, amit fontosnak tartunk, például biometrikus adatok kezelésére, bűnügyi adatokkal kapcsolatos speciális követelmények stb.

Az MI használatáról szóló rendelet tervezete (29. cikk (6) bek.) szerint *„adott esetben a magas kockázatú mesterséges intelligenciát alkalmazó rendszerek üzembe helyezői a 13. cikk alapján szolgáltatott információkat felhasználják az (EU) 2016/679 rendelet 35. cikke vagy az (EU) 2016/680 irányelv 27. cikke szerinti adatvédelmi hatásvizsgálat elvégzésére vonatkozó kötelezettségük teljesítéséhez, amelynek összefoglalóját közzé kell tenni (…)".*

Amennyiben a DPIA szerint a kockázat mérséklését célzó garanciák, biztonsági intézkedések és mechanizmusok hiányában az adatkezelés magas kockázattal járna a természetes személyek jogaira és szabadságaira nézve, és ez a kockázat nem mérsékelhető a rendelkezésünkre álló technológiák és a végrehajtási költségek szempontjából észszerű módon, abban az esetben az adatkezelési tevékenységünk megkezdése előtt konzultálnunk kell az illetékes adatvédelmi felügyeleti hatósággal. A NAIH szerint ide tartozik az, ha az érintettek olyan jelentős vagy akár visszafordíthatatlan következményekkel szembesülnek, amelyeket nem tudnak leküzdeni (pl. adatokhoz való jogosulatlan hozzáférés, amely az érintettek életét fenyegető veszélyt, elbocsátást vagy pénzügyi nehézséget eredményez) (NAIH 2023).

A NAIH az előzetes konzultáció keretében megállapítja:
- a DPIA-t a vonatkozó jogszabály, illetve iránymutatás szerint folytattuk-e le
- pontosan azonosítottuk-e az adatkezelési tevékenységeket, azok kockázatait, valamint sikerült-e a kockázatok kezelésére irányuló intézkedéseket meghoznunk
- az adatkezelésben érintett adatok esetében szétválasztottuk-e a személyes adatok és a különleges adatok kezelését
- az adatkezeléseink jogszerűek-e és amennyiben szükséges, végeztünk-e érdekmérlegelési teszte(ke)t, valamint
- a fennmaradó kockázatok mérséklésében tud-e segítséget nyújtani.

Amennyiben a tervezett adatkezelés megsértené a GDPR-t, különösen, ha adatkezelőként a kockázatot nem elégséges módon azonosítottuk vagy csökkentettük, úgy a NAIH gyakorolhatja a GDPR 58. cikkében említett hatásköreit, többek között megtilthatja az adatkezelésünket.

A finn adatvédelmi hatóság figyelmeztetésben részesített egy adatkezelőt az egészségügyi ellátások értékelésével kapcsolatban. Az egészségügyi szolgáltató (adatkezelő) előzetes konzultációt kért egy új, egészségügyi ellátások elemzésére szolgáló eszközével kapcsolatban. A rendszert az egészségügyi ellátás megelőző és proaktív megközelítésének javítására tervezték úgy, hogy algoritmusok segítségével azonosítsa az egészségügyi kockázatokkal rendelkező betegeket, majd az azonosított páciensek esetében egy egészségügyi szakember értékelné a kezelés szükségességét. A hatóság úgy döntött, az ügyet nem előzetes konzultációként, hanem felügyeletként kezeli. A hatóság megállapította, a rendszer által további vizsgálatra azonosított betegek nem tartoznak a kizárólag automatizált adatkezelésen alapuló döntés hatálya alá, mivel az egészségügyi szakemberek a rendszer ajánlásán kívül más tényezőket is figyelembe vesznek a végső értékelésnél. A rendszer által ki nem választott betegek esetében azonban a döntés végleges lenne, és kizárólag automatizált döntéshozatalon alapulna, ez pedig jelentősen érintené őket, mivel megfosztanák őket az egészségügyi szolgáltatás igénybevételének lehetőségétől. A hatóság döntése szerint a tervezett adatkezelés sérti a GDPR 22. cikkét (Tietosuojavaltuutetun toimisto 2020).

Mi a szerepe a jogászoknak és az informatikusoknak az adatvédelmi hatásvizsgálatban?

A jogászok felelősek a releváns jogi követelmények azonosításáért, valamint segítenek megérteni, milyen adatvédelmi jogszabályok és jogalkalmazási gyakorlat vonatkozik az LLM-re, miért fontos ezek betartása és hogyan kell ezeket figyelembe vennünk a tervezés, a fejlesztés és az üzemeltetés során. Adatvédelmi jogi tanácsot nyújtanak, kiterjedve – többek között – az adatvédelmi kockázatokra, a megfelelő intézkedésekkel kapcsolatos javaslatokra, valamint a jogszerű adatkezelési gyakorlatokra.

Az informatikusok felelősek a technikai adatvédelmi és biztonsági kockázatok, az LLM sérülékenységei és potenciális fenyegetései azonosításáért, valamint javaslatokat tesznek a megfelelő biztonsági intézkedésekre és megoldásokra. Megtervezik a modell biztonsági architektúráját, javaslatokat tesznek az adatvédelmi protokollokra és technikákra, illetve implementálják a megfelelő biztonsági és adatvédelmi intézkedéseket. Felelősek a rendszerbiztonság és adatvédelem tesztelésééért, melynek keretében elvégzik a biztonsági teszteket, ellenőrzik a rendszer sérülékenységeit és intézkednek a szükséges korrekciók érdekében.

Amennyiben a jogászok és az informatikusok együttműködése elmarad, előfordulhat, hogy nem azonosítjuk vagy nem vesszük megfelelő mértékben figyelembe az adatvédelmi és adatbiztonsági előírásokat, emiatt a projektünk adatbiztonsági szempontból sérülékenyebbé válhat (rosszindulatú támadások, adatlopások, adatmanipulációk), valamint könnyen megszeghetjük a vonatkozó jogszabályokat és egyéb szabályozásokat. Elmaradhat a vonatkozó tájékoztatóink, belső szabályzataink megalkotása, illetve ezek hiányosak lehetnek, ez pedig a megfelelő adatvédelmi gyakorlatok és elvek hiányát eredményezi, akár komoly pénzügyi és reputációs kárt is okozva.

Az együttműködés során „létfontosságú" a konszenzus, mivel a „felülről kapott", egyeztetés nélküli, kétségesen használható szervezetidegen utasítások és szabályzatok, amelyeket a szereplők még csak el sem olvasnak, vagy ha bele is néznek, azt dühösen a sarokba dobják, semmiképpen sem hatnak abba az irányba, hogy a projektünk jogszerű és biztonságos legyen.

Az alapjogi hatásvizsgálat (magas kockázatú MI-rendszerek esetében)

Az alapjogi hatásvizsgálatban azonosítjuk és értékeljük egy adott LLM-projekt alapjogi hatásait, figyelembe véve az alapjogokat és azt, hogy egy adott intézkedés hogyan befolyásolja ezeket a jogokat. Alapvető követelmény, hogy ne sértsük az alapjogokat, illetve az alapjogok közötti egyensúlyt és konfliktusokat figyelembe vegyük a projekt teljes időtartama alatt.

Az MI rendelet-tervezete szerint a nagy kockázatot jelentő MI-vel kapcsolatos rendszerek használatba vétele előtt – bizonyos kivételektől eltekintve – rendszerüzemeltetőként kell elvégeznünk a hatásvizsgálatot, mely tartalma:

- a rendszer tervezett felhasználási céljának egyértelmű felvázolása
- a rendszer használatának tervezett földrajzi és időbeli hatókörének világos felvázolása
- a rendszer használata által valószínűleg érintett természetes személyek és csoportok kategóriái
- annak ellenőrzése, hogy a rendszer használata megfelel-e az alapvető jogokra vonatkozó uniós és nemzeti jognak
- a magas kockázatú MI-rendszer használatba vételének az alapvető jogokra gyakorolt észszerűen előrelátható hatása
- a marginalizált személyeket vagy kiszolgáltatott csoportokat valószínűleg érintő konkrét ártalmi kockázatok
- a rendszer használatának észszerűen előrelátható káros hatása a környezetre
- részletes terv arra vonatkozóan, hogy az azonosított károkat és az alapvető jogokra gyakorolt negatív hatásokat hogyan fogjuk enyhíteni
- az irányítási rendszer, amelyet a telepítő bevezet, beleértve az emberi felügyeletet, a panaszkezelést és a jogorvoslatot.

Amennyiben a kockázatok mérséklésére vonatkozó részletes terv nem azonosítható, üzembehelyezőként tartózkodnunk kell a magas kockázatú MI-vel rendelkező rendszer használatba vételétől, valamint erről indokolatlan késedelem nélkül tájékoztatnunk kell a szolgáltatót és a nemzeti felügyeleti hatóságot.

Ez a hatásvizsgálati kötelezettség a magas kockázatú MI-t alkalmazó rendszer első használatára vonatkozik, egyéb esetekben támaszkodhatunk a korábban elvégzett alapjogi hatásvizsgálatra vagy a szolgáltatók által elvégzett meglévő értékelésre. Amennyiben a magas kockázatú MI-rendszer telepítőjeként úgy ítéljük meg, hogy a hatásvizsgálat kritériumai már nem teljesülnek, új alapjogi hatásvizsgálatot kell végeznünk. A tervezet szerint a kis- és középvállalkozások önkéntesen alkalmazhatják az alapjogi hatásvizsgálatra vonatkozó rendelkezéseket.

Amennyiben használatba adóként a GDPR 35. cikke vagy a LED 27. cikke alapján DPIA-t kell végeznünk, az alapjogi hatásvizsgálatot a DPIA-val együtt kell elvégeznünk (29a. cikk).

Mi a szerepe a jogásznak és az informatikusnak az alapjogi hatásvizsgálatban?

A jogászok felelősek az alapjogi keretek feltérképezéséért, értelmezik és értékelik a jogszabályokat annak érdekében, hogy felismerjék az alapjogokkal összefüggő kérdéseket, valamint azonosítsák az esetleges kockázatokat vagy sérelmeket. Megvizsgálják a vonatkozó jogi környezetet, azonosítva az alapjogokat érintő követelményeket, korlátokat és elvárásokat, illetve figyelemmel kísérik a változásokat és a jogalkalmazási gyakorlatot.

Az informatikusok felelősek a technikai elemzések elvégzéséért, hogy megértsük, az adott LLM hogyan működik és milyen technikai megoldásokat igényel, illetve meghatározzák az adatok feldolgozására, tárolására, hozzáférésére és biztonságára vonatkozó technikai követelményeket. Felelősek az információbiztonságért, közreműködnek a biztonsági kérdések azonosításában és megoldásában, a megfelelő biztonsági gyakorlatok alkalmazásában (10. sz. ábra).

10. sz. ábra: Biztonság
Forrás: Az ábra alján

A jogászok és az informatikusok együttműködése elengedhetetlen annak érdekében, hogy a jogi és technikai szempontok harmonizáljanak, és döntéseink összhangban legyenek az alapjogokkal és a jogalkalmazási gyakorlattal, ehhez pedig az szükséges, hogy az informatikusok – jogi alapismeretek nélkül is – elfogadják, nekik is tenniük kell a projekt alapjogi megfelelőségéért. Amennyiben ez az együttműködés elmaradt, az MI megsértheti az egyének alapvető jogait, illetve az alkalmazott technológiák jogi helyzete bizonytalan lehet, ez pedig nehezítheti a bevezetést és felhasználást, valamint növelheti a jogi kockázatokat. Az együttműködés hiánya hátrányosan befolyásolhatja döntéseink minőségét és megalapozottságát, valamint olyan intézkedésekhez vezethet, amelyek csökkenthetik projektünk hitelességét és elfogadottságát. Ha az emberek úgy érzik, az alapvető jogaikat figyelmen kívül hagyjuk, az csökkenti a bizalmat az MI-technológiák iránt és növeli az aggodalmakat a magánszférával, az adatvédelemmel és az etikai kérdésekkel kapcsolatban.

Jogszabályi és szabályozási hatásvizsgálat

Mikor jogszerű egy mesterséges intelligencia?

Akkor, ha az MI kifejlesztésének, elterjesztésének és felhasználásának részét képező folyamatokra és tevékenységekre vonatkozó valamennyi előírást betartjuk, például:

- az elsődleges uniós jogot (az Európai Unió Szerződései és Alapjogi Chartája)
- a másodlagos uniós jogot (pl. a GDPR/LED, a hátrányos megkülönböztetést tiltó irányelvek, a gépekről szóló irányelv, a termékfelelősségi irányelvek, a nem személyes adatok szabad áramlásáról szóló rendelet, a fogyasztói jogra, valamint a munkahelyi biztonságra és egészségre vonatkozó irányelvek stb.) – az MI használatával kapcsolatos jogalkotás jelenleg is folyamatban van
- az ENSZ emberi jogi egyezményeit
- az Európa Tanács egyezményeit (pl. az emberi jogok európai egyezménye), valamint
- az uniós tagállam(ok) jogát.

A jogszabályi és szabályozási hatásvizsgálat során ennek a jogi környezetnek az elemzését és értékelését végezzük el, feltárva a vonatkozó követelményeket és korlátokat, az alábbi fókuszpontokkal:

- releváns jogszabályok, iránymutatások és jogalkalmazási gyakorlatok
- szektorspecifikus előírások, iparági szabványok
- szükséges engedélyek
- tanúsítványhoz, magatartási kódexhez csatlakozás
- különböző felelősségek (pl. szerzői és szomszédos jogok, fogyasztóvédelem, termékfelelősség, büntetőjogi felelősség stb.)
- kereskedelmi és versenyjogi szempontok stb.

Mi a szerepe a jogászoknak és az informatikusoknak a jogszabályi és szabályozási hatásvizsgálatban?

A jogászok feladata a releváns jogi előírások, jogi és szabályozási követelmények és kötelezettségek értékelése, azonosítva azokat a követelményeket, amelyeket teljesítenünk kell, megfelelő intézkedéseket és belső szabályozásokat javasolva.

Az informatikusok értékelik a technikai megvalósíthatóságot a jogszabályi és szabályozási keretek figyelembevételével, azonosítva a technikai korlátokat és kihívásokat, valamint kidolgozzák azokat a technológiai megoldásokat, amelyek támogatásával megfelelünk a jogi követelményeknek (pl. adatbiztonsági mechanizmusok, felhasználói hitelesítés, adatkezelési protokollok stb.) (11. sz. ábra).

Bár sokan úgy gondolják, hogy a jogi megfelelőség egyedül a jogászok feladata, a modern technológiák esetében a jogászok és az informatikusok közös munkája a kulcsa a jogi és szabályozási kockázatok minimalizálásának. Az egyes technológiák jogi és szabályozási összefüggéseinek figyelembevétele versenyelőnyt jelenthet, azonban együttműködés hiányában ezt elmulaszthatjuk. Könnyen megszeghetjük a vonatkozó jogi előírásokat is, illetve nem megfelelő adatvédelmi és adatbiztonsági intézkedéseket dolgozhatunk ki vagy implementálhatunk, ami

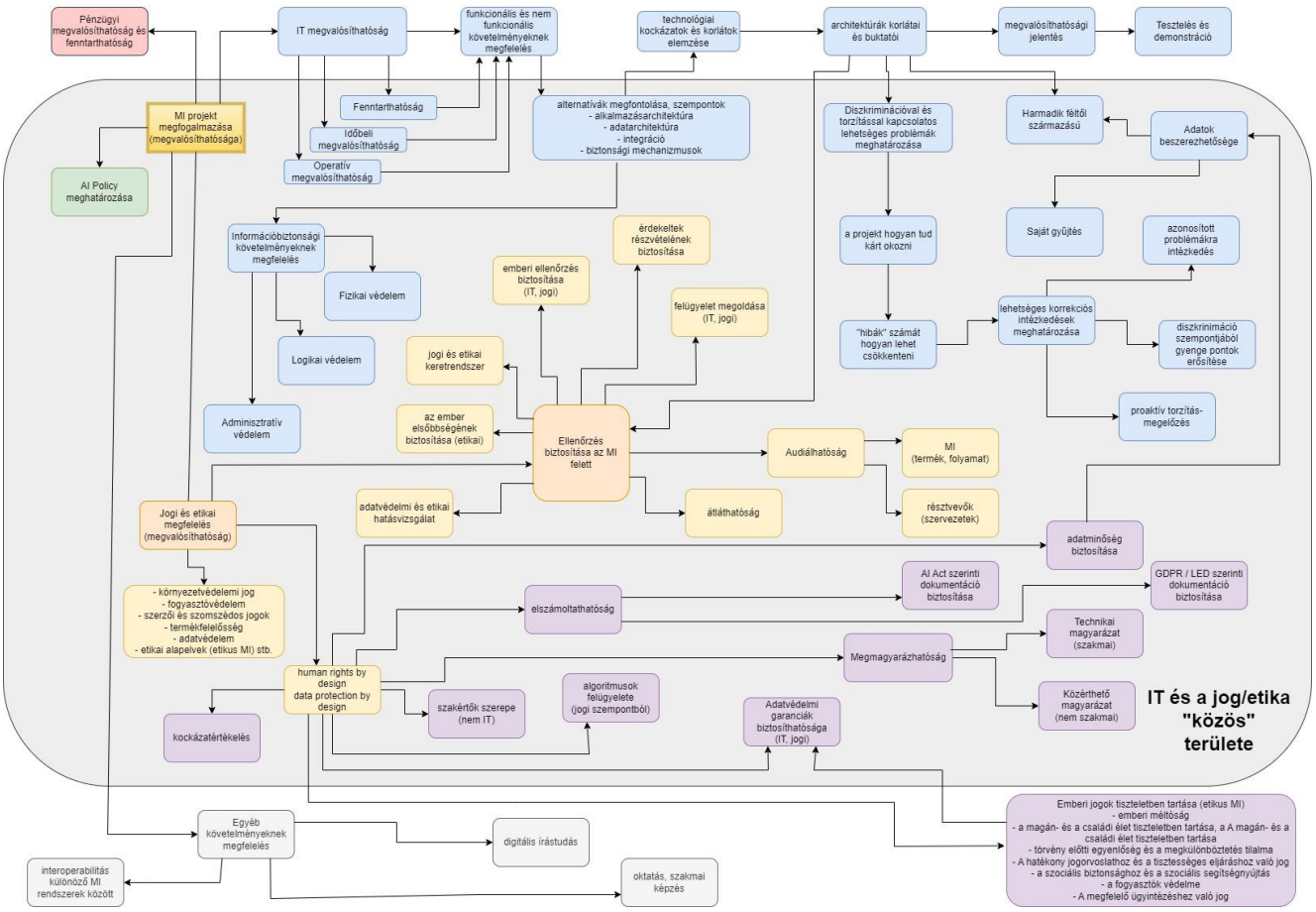
jogsértésekhez, vitákhoz, hatósági eljárásokhoz, pereskedésekhez vezethetnek, valamint jelentős többletköltséget is okozhatnak (pl. határidő túllépés, kötbérfizetési kötelezettség, közigazgatási bírság, kártérítés fizetési kötelezettség stb.). Ezen kívül a hiányosságok súlyosan károsíthatják szervezetünk vagy projektünk hírnevét és ügyfeleink bizalmát.

Melyek a jó hatásvizsgálat ismérvei?

A hatásvizsgálat egyik legfontosabb előnye – eltekintve attól a kockázatminimalizálással kapcsolatos céltól, amire kitalálták –, hogy segít egy asztal köré összehozni a résztvevőket és teret ad a projekt Alfától Omegáig megismerésére, az építő párbeszédre és a felmerülő problémák hatékony, minden szakterület számára megfelelő megoldására. A párbeszéd kezdeményezője olyan személy kell legyen, aki „hídemberként" nem retten meg a nehezen kezelhető szereplőktől, képes kreatívan megoldani a konfliktusokat és közelebb hozza az olykor egymástól nagyon messzi álláspontokat. A siker feltétele, hogy ez a „hídember" széles látókörrel rendelkezzen és képes legyen csitítani a felek indulatait, a társalgást megfelelő mederben tartani. Enélkül a hatásvizsgálat csak szimpla „lepapírozás", egy kérdőív mechanikus kitöltése.

A jó hatásvizsgálat néhány fontos *ismérve:*

- lehetőséget ad egymás személyének, szempontrendszerének és prioritásainak megértésére
- teljeskörű, átfogóan elemezi a lehetséges közvetlen és közvetett hatásokat, valamint következményeket
- előre meghatározza a kereteket, teljeskörűségében sem markol túl sokat, komplex folyamatokban gondolkodik és azokat lépésről lépésre bontja le
- alaposan elemezi a különböző tényezőket és szempontokat (adatok gyűjtése, elemzési módszerek alkalmazása, releváns szakirodalom áttekintése, szakértők bevonása stb.)
- nem hazudik, nem célja a valóság elrejtése. Ha nem vagyunk hajlandóak szembesülni a problémákkal, akkor azokat megoldani sem tudjuk, miközben soha nem tudhatjuk, azok mikor kerülnek felszínre, az idő előre haladtával egyre nagyobb galibát okozva.
- tárgyilagos, objektív és semleges, megbízható adatokra, releváns kutatásokra és szakértői véleményekre támaszkodva nyújt megbízható és objektív eredményt
- rugalmas a változó körülmények és feltételek kezelésében (pl. egy új jogszabály alapjaiban rengetheti meg projektünket, és rugalmasság hiányában akár fel is adhatjuk célunkat)
- bevonja a folyamatokba az érdekelteket, szakértőket és érintetteket. Nem célszerű senkit sem kihagynunk csak azért, mert kellemetlen alaknak tartjuk; a meg nem hallgatott vélemények olyan információkat tartalmazhatnak, amelyek nem ismerete később számtalan gondot és jelentős költségnövekedést is okozhat.
- az eredmények, a módszerek és a feldolgozott adatok tekintetében átlátható
- eredményei felhasználhatók a döntéshozatalban, a szükséges intézkedések és belső szabályzatok kialakításában
- megfelelően ütemezett és bármikor újranyitható, tekintettel a körülmények változására
- a projekt méretéhez és jellegéhez skálázható, alkalmazkodva annak komplexitásához
- költséghatékony, arányban áll erőforrásainkkal és az elvárt előnyökkel
- támogatja az innovációt és az új technológiák bevezetését, miközben biztosítja a kockázatok megfelelő kezelését és a fenntartható fejlődést.



11. sz. ábra: Az IT és a jog-etika közös területe a hatásvizsgálatok során (egy lehetséges megoldás a hatásvizsgálat hatókörének megállapításához)
Forrás: Saját szerkesztés

Nem megfelelő hatásvizsgálattal csak adminisztrációs kötelezettségeinket teljesítjük (úgy-ahogy), miközben raboljuk egymás idejét és feleslegesen idegesítjük egymást. Csakis építő jelleggel érdemes végeznünk, különben csak egy drága és felesleges bürokratikus eljárás, amely nem betemeti, hanem még tovább mélyíti a törésvonalakat.

Érdekmérlegelési tesztek végzése

Az érdekmérlegelési teszt olyan adatvédelmi eszköz, amely a GDPR rendelkezéseinek megfeleléshez szükséges. Célja – azon kívül, hogy kötelezően el kell végeznünk jogos érdekre hivatkozó adatkezelések végzése előtt –, hogy az érdekek, jogok és értékek összeütközése esetén segítse az érdekek összehasonlítását és a konfliktusok feloldását. A mérlegelés során azt vizsgáljuk, egy adott adatkezelés hogyan befolyásolja az érintett felek érdekeit és jogait, valamint az milyen szélesebb társadalmi és környezeti kihatásokat eredményezhet. Amennyiben készítünk hatásvizsgálatot, annak eredményeit célszerű felhasználnunk a tesztünkhöz.

Az érdekmérlegelés során:

- azonosítjuk az érintett feleket és azok érdekeit, jogait, szabadságait, valamint azt, hogy mely érdekek érintettek és milyen hatások érvényesülnek az adatkezelésünkkel kapcsolatban
- összehasonlítjuk az érintett felek érdekeit és megítéljük, azok milyen fontosságúak, valamint milyen prioritásúak
- azonosítjuk azokat a konfliktusokat, amelyekben az érdekek, valamint jogok és szabadságok ütköznek egymással. Az érdekmérlegelés lehetőséget ad folyamataink és azok jogszabályi megfelelőségének átgondolására, valamint alternatív megoldásokat találhatunk a konfliktusok feloldása érdekében.

Az érdekmérlegelési teszt lépéseiről és módszereiről a WP29 ajánlást bocsátott ki (WP 217).

Az érdekmérlegelést tekinthetjük egyfajta „mini” hatásvizsgálatnak, és bár célja nem a kockázatok minimalizálása, hanem a jogos érdek alátámasztása, a

mérlegelés fontos része, hogy az érintettek jogait és szabadságait érő kockázatokat minimalizálva állítsuk be azt az egyensúlyi helyzetet, amely segítségével adatkezelésünk jogszerű lesz.

Az érdekmérlegelést példákkal alá kell támasztanunk, ehhez pedig elengedhetetlen a szereplők együttműködése. Amennyiben a teszt modern technológiákkal kapcsolatos, állandó résztvevő az adatvédelmi tisztviselő, az informatika és a biztonsági szolgálat képviselője is. A felügyeleti hatóság elvárja, hogy az érdekmérlegelés térjen ki a szükségességre, az arányosságra és a fokozatosságra, a lehetőségek megvitatásához pedig elengedhetetlen a különféle szakterületek együttműködése.

Az LLM esetén különösen a modell betanítása során merülhet fel a jogos érdekre hivatkozás, éppen ezért a betanítással kapcsolatos adatkezelések esetében (pl. adatok gyűjtése, előkészítése stb.) alapvető az informatikusok részvétele az érdekmérlegelésben. Az ő feladatuk továbbá a kockázatcsökkentő intézkedések azonosítása és technikai megoldása (pl. anonimizálás, álnevesítés, titkosítás, jogosultságmenedzsment stb.), a jogászoknak pedig el kell tudni fogadtatniuk az informatikusokkal, hogy miért is van szükség ezekre a gyakran igen drága intézkedésekre.

Belső szabályzatok

A projektnek és a szervezetnek is szüksége van valamiféle keretre, amely támpontot ad a feladatok és felelősségek tekintetében, és amely általában különféle szabályzatokból, utasításokból, illetve protokollokból áll össze.

Az MI-szabályzat megalkotása nem lehet egyetlen személy vagy szakterület feladata, és ez az a lehetőség, amelyet kiaknázhatunk annak érdekében, hogy a későbbiekben a különböző szakterületek képviselői ne akadályozzák a közös munkát. Ugyanis egy MI-szabályzat megalkotása arra készteti a szereplőket, hogy mélységében gondolják át kinek mi a feladata és felelőssége, milyen lépéseket és mérföldköveket kell teljesíteni a megfelelőség érdekében, és kit, mikor kell bevonni a projektbe.

Amennyiben teljeskörű, mindenki által elfogadható szabályzatot készítünk

- mindenki tudja mi a feladata, és nem próbálja azt áthárítani másokra
- a felelősségi körök tisztázottak (pl. ki, milyen értékben hagyhat jóvá beszerzést stb.)
- milyen részletességgel kell dokumentálni a projektet
- rendkívüli események esetén kinek, mi a teendője, és nem utolsó sorban
- a teljes projekt időtartama alatt megfelelünk az átláthatóság és az elszámoltathatóság elvének.

Mire jó egy jó szabályzat?

Ritkán lelkesedünk a szabályzatokért – általában nyűgnek érzik, felesleges bürokráciának, íróasztal mellőli okoskodásnak. A szabályzatoknak azonban van pozitív hozadéka is. Amennyiben leszabályozzuk a folyamatokat, mindenki tudja, mi a feladata, milyen felelősséget kénytelen (el)viselni, mi a teendője rendkívüli esemény esetén és mire számíthat, ha valami félresiklik. A jó szabályzat tehát nem pusztán egy dokumentum, hanem egy útmutató, amely irányt mutat a kockázatos folyamatokban és segít – akár a káosz közepette is – a helyes irány megtalálásában.

Az MI-projektek sokszereplős, „zűrös" folyamatok, könnyen összemosódó felelősséggel, miközben a modell lehet olyan „feketedoboz", amelynek felnyitása szinte lehetetlen, alaposan feladva a leckét a felelőst keresőknek. Éppen ezért elengedhetetlen az átláthatóság, garantálva, hogy nem marad ki fontos lépés és nem a modell tesztelésénél közli a jogász, az jogszerűtlen és az egészet lehet elölről kezdeni. Ez nemcsak hatalmas időveszteséget (határidőcsúszást stb.) okozhat, hanem jelentős pénzügyi kihatással is járhat (kötbérfizetési kötelezettség, túlmunka ellentételezése, új adatkészlet beszerzése stb.), miközben egy átgondolt szabályzat betartásával a kényes helyzet elkerülhető lett volna.

A jó MI-szabályzatot azonban nem lehet csak úgy letölteni a netről, a jó szabályzat igazodik a sajátosságainkhoz. A jó szabályzatot hely-, személy- és projektismerettel rendelkező személyek alkotják meg, vagy egy, már meglévő sablont kell a szervezetünkre és a projektünkre szabni.

A jó szabályzat ténylegesen is támogatja projektünket – ez pedig akkor lesz így, ha minden szakterület a saját követelményeit belefoglaltatja, ahogy azt is, mit vár el másoktól. Természetesen a projektgazdán, az informatikusokon és a jogászokon kívül számtalan szakterület kérését meg kell hallanunk, így például a pénzügyesekét, a beszerzőkét, a humán erőforrásgazdálkodását.

Mire térjen ki a szabályzat?

Az MI-szabályzatnak nincs kötelező tartalma, hiszen maga az MI-szabályzat sem kötelező. Célszerű olyan témákat leszabályoznunk, amelyek szabályozatlansága gondot okozhat, például:

- projektfolyamatok leírása
- folyamatok kiszervezésének követelményei

- melyik stádiumban kit, miért és hogyan kell bevonni
- milyen alapvető jogszabályokat, szabályozásokat és szakmai minimumokat kell figyelembe venni

„A bíróságok nem várhatják el, hogy a gyakran évszázados múltra visszatekintő szokásjog mindig alkalmas lesz arra, hogy a technológia szabályozásának új jogi kérdéseit kezelje. A már létező jogi kategóriák bizonyos esetekben alkalmazhatóak lehetnek, de ezt csak úgy lehet megállapítani, ha először is megvizsgáljuk a kategóriák alapját és értékeljük, hogy a doktrína kiterjesztése kielégíti-e ezt az alapot. Ez az elemzés az adott jogvitától és a szóban forgó technológiától függően változik, és gyakran megköveteli a döntésnek a szóban forgó technológia jövőbeli fejlődésére és elterjedésére, valamint tágabb értelemben a gazdaságra és a társadalmi jólétre gyakorolt hatásának mérlegelését.

A valós világbeli vitákat és társadalmi összefüggéseket nem szabad előre létező jogi kategóriákba kényszeríteni. A jogi kategória csupán egy konstrukció; a viták és a kontextus a megváltoztathatatlan valóság. Ha a jogi kategóriák nem illeszkednek jól az új valósághoz, akkor a jogi kategóriákat kell újraértékelni. (…)

Először is, a már meglévő jogi kategóriák már nem feltétlenül alkalmazhatók az új jog és technológia vitáira. Annak mérlegeléséhez, hogy a meglévő jogi kategóriáknak van-e jogi és társadalmi értelme egy új technológiai rendszerben, kritikus fontosságú először a jogi kategorizálás mögött álló indoklás értelmezése, majd annak értékelése, hogy az alkalmazható-e az új jogvitára.

Másodszor, a jogi döntéshozóknak ügyelniük kell arra, hogy az új technológia csodái ne torzítsák el a jogi elemzésüket. Ez a technológiailag laikus jogi döntéshozók számára különleges kihívást jelent, amely megköveteli, hogy a fejlődő technológia ígéreteiből kiindulva átlássák annak tényleges jellemzőit és a tudományos ismeretek jelenlegi szintjét.

Harmadszor, az újonnan megjelenő technológiákból eredő új jogviták típusai gyakran előre nem láthatóak. Azok a jogrendszerek, amelyek képesek alkalmazkodni és fejlődni a technológia és az arról alkotott ismereteink fejlődésével, sokkal sikeresebben fognak működni, mint a már létező jogi rendszerekhez való vak ragaszkodás". (Mandel 2017: 271)

- projektek mérföldkövei
- torzítás és diszkrimináció elkerülése, csökkentése és megszüntetése (feladatok, módszerek)
- a tesztelés követelményei (feladatok és felelősségek)
- átláthatósági és megmagyarázhatósági követelmények (ICO–Alan Turing Institute 2022)

„A mesterséges intelligencia képes arra, hogy a döntéshozatalt az emberektől a gépekre ruházza át. Bár ez pontosabb, gyorsabb és hatékonyabb orvosi és egészségügyi irányítási döntésekhez vezethet, számos aggályt is felvet. Számos mesterséges intelligencia-technológia "feketedoboz" rendszerként működik, amelynek belső folyamatai és következtetései rejtve maradnak, vagy nehezen magyarázhatók el a felhasználók számára. Ez az átláthatóság hiányához vezethet, és alááshatja az egészségügyi rendszerek üzemeltetőinek, az orvosoknak és a betegeknek az autonómiáját. Az egészségügyi ellátást igénybe vevők a kontroll és az autonómia elvesztését tapasztalhatják, különösen akkor, ha az orvosi döntések nem átláthatóak, és hiányzik a betegek és az orvosok közös döntéshozatala. A kontroll és az autonómia ezen problémái egyre inkább előtérbe kerülnek a mesterséges intelligencia technológiák programozásának automatizálására irányuló erőfeszítésekkel összefüggésben, olyan számítógépes programok révén, amelyek önállóan képesek új mesterséges intelligencia modelleket és alkalmazásokat létrehozni, telepíteni és skálázni. Ehhez kapcsolódó kihívás a mesterséges intelligencia kiszámíthatósága és megbízhatósága. Az algoritmusokban és adatokban lévő hibák vagy az elfogult adatkészletek használata az MI-rendszerek hibás vagy tisztességtelen döntéseihez vezethet. A hibás vagy elfogult ítéletek befolyásolhatják a betegbiztonságot és az egészségügyi ellátás hatékony végrehajtását. Az adatkészletek és algoritmusok torzítása az erőforrások igazságtalan elosztásához is vezethet, és diszkriminálhat bizonyos csoportokat, például azáltal, hogy figyelmen kívül hagyja a kevés erőforrással vagy embereket, vagy csoportokat." (Rosemann–Zhang 2021: 104–105).

- az adatok és a modell védelme
- kiemelt jogterületek, például szerzői jog és szomszédos jogok[27], üzleti titok, szabadalom, termékfelelősség[28], fogyasztóvédelem, versenyjog stb.
- irányítás és felügyelet
- a rendszeres audit (korrekciós intézkedések) stb.

Adatvédelmi és adatbiztonsági ismeretek oktatása, a tudatosság növelése

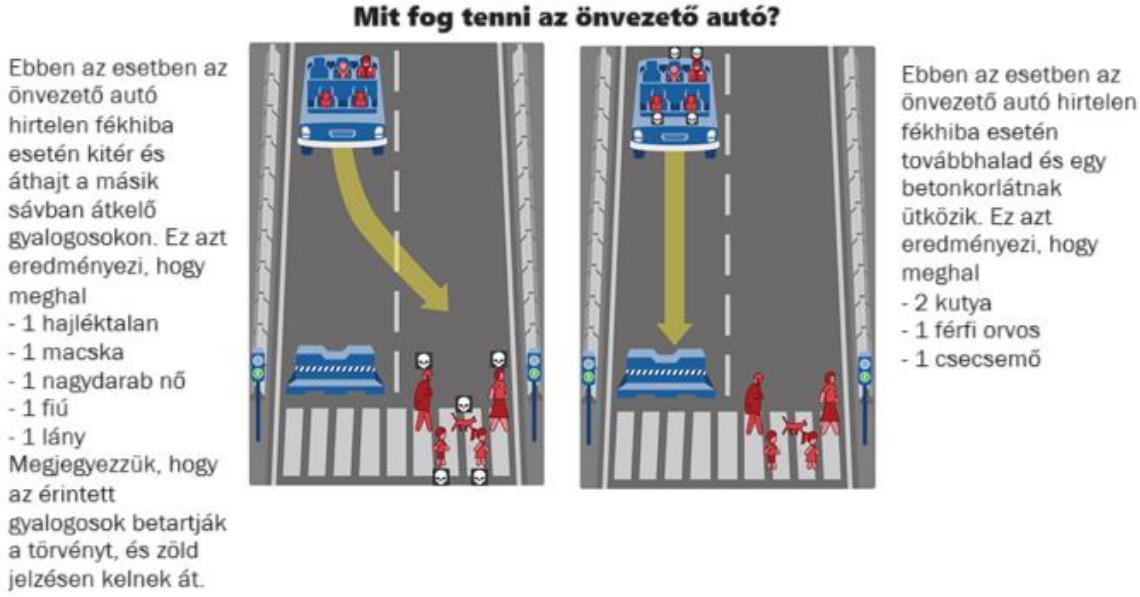
A kötelező oktatás általában az a szükséges rossz, amit meg kell úszni, különösen, ha közben nem lehet „értelmes" dolgot csinálni, például e-mailekre válaszolni, híreket olvasni, ásítozni stb. Az ilyen oktatásnak sok értelme nincs, még egy pipa valamelyik

27 (EU) 2019/790 irányelv

28 a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv javaslata

projektdokumentációban. Ezeket az alkalmakat azonban – megfelelő előadóval és átgondolt tematikával – workshoppá vagy csapatépítéssé formálhatjuk.

- *Adatvédelmi és adatbiztonsági ismeretek*
  Az oktatás akkor hasznos, ha nem a jogszabály citálását jelenti, hanem a gyakorlatot ismerteti. A legjobb példák a mindennapokból merítenek és felhasználják a hallgatóság tapasztalatait.
- *Alapjogok értelmezése*
  Az alkotmányos normákat nem tudjuk csak úgy egyszerűen átültetni egy másik társadalmi kontextusba, például abba a digitális környezetbe, amelyben az LLM működni fog, éppen ezért az alapjogokat újra meg kell határoznunk. A számtalan digitális vetület miatt ez nem lehet csak a jogászok feladata (pl. az adatok nem megfelelő szelektálása, az algoritmus döntésének kirekesztő jellege vagy torzítása).
- *Etikai dilemmák megvitatása*
  Az etikai dilemmák megvitatásához találunk az interneten forrásokat, például a Moral Machine oldalt az önvezető autók döntéseiről (12. sz. ábra). A témában globális kutatás (Bonnefon 2021), illetve tanulmány (Bonnefon–Shariff–Rahwan 2016) is elérhető.



12. sz. ábra: Hogyan döntsön az önvezető jármű?
Forrás: https://www.moralmachine.net/; 2/13)

- *Konkrét hatósági határozatok, bírósági döntések átbeszélése*
  Melyik informatikust ne érdekelné, hogy mit követhetett el az a bank, amelyik az MI miatt 250 milliós büntetést kapott? Amennyiben – köszönhetően az eset közös megvitatásának – a projekt szereplőiben tudatosul, az „akadékoskodó" jogász csak el akarja kerülni, hogy a projekt miatt büntetést kapjunk, jobban elfogadják érveit.
- *Megtörtént adatvédelmi incidensek elemzése*
  Az incidensek elemzésekor érdemes megtörtént esetekből kiindulnunk, mi volt a hiba, mely ponton és hogyan állt be az adatkezelő, illetve az adatfeldolgozó felelőssége. Az Európai Adatvédelmi Testület (EDPB) 01/2021. számú iránymutatása hasznos incidens-szituációkat tartalmaz kockázatminősítéssel és intézkedési javaslattal, illetve a saját tapasztalatainkból is meríthetünk.
- *Adatvédelmi incidens szimuláció*
  Végezhetünk olyan adatvédelmi incidens szimulációt is, amely IT vonatkozású, és amelyben az informatikusoknak oroszlánrészt kell vállalniuk (a biztonság sérülése az ő felségterületükön történik). Kielemezhetjük, hogy az EDPB 9/2022. számú iránymutatása alapján mire nem gondoltunk a szimuláció során, illetve, ha valóban megtörtént volna az eset, akkor milyen mértékű szankcióra számíthatnánk (EDPB 04/2022).

Összegzés

"Miért kell egy robotnak parancsba adni, hogy engedelmeskedjen a parancsoknak - miért nem elég az eredeti parancs? Miért parancsoljuk meg egy robotnak, hogy ne tegyen kárt - nem lenne egyszerűbb, ha eleve nem is parancsolnánk neki, hogy kárt tegyen? Vajon az univerzumban van egy titokzatos erő, amely az entitásokat a rosszindulat felé húzza, ezért egy pozitronikus agyat úgy kell programozni, hogy ellenálljon neki? Vajon az intelligens lények elkerülhetetlenül alakítanak ki magatartásproblémát? (...) Most, hogy a számítógépek valóban okosabbá és erősebbé váltak, az aggodalom alábbhagyott. A ma mindenütt jelenlévő, hálózatba kapcsolt számítógépek soha nem látott mértékben képesek arra, hogy rosszat tegyenek, ha egyszer rosszra vetemednének. De a bajt csak a kiszámíthatatlan káosz vagy az emberi rosszindulat okozza vírusok formájában. Már nem aggódunk elektronikus sorozatgyilkosok vagy felforgató szilícium-összeesküvések miatt, mert kezdjük felismerni, hogy a rosszindulat - akárcsak a látás, a motoros koordináció és a józan ész - nem jár ingyen a számítástechnikával, hanem be kell programozni. (...) Az agresszió, akárcsak az emberi viselkedés minden más, általunk természetesnek tartott eleme, egy kihívást jelentő mérnöki probléma!" (Pinker 2012: 15–16)

Az MI fejlődése új kihívások elé állítja a jogászokat és az informatikusokat. Steven Pinker gondolatai alapján felmerül a kérdés: *hogyan biztosíthatjuk, hogy az MI etikus és megbízható legyen, a társadalom érdekeit szolgálva?* Mi a garanciája annak, hogy egy MI-projekt résztvevői egy irányba mozogva, az irányadó jogi-etikai kereteken belül valósítsák meg elképzeléseiket? Álláspontom szerint ennek *az együttműködésnek az egyik alappillére a hatékony kommunikáció*, éppen ezért tanulmányomban azt vizsgáltam – az adatvédelem és adatbiztonság terén szerzett tapasztalataim alapján –, hogy milyen eszközökkel lehet az ennek hiányából adódó problémákat orvosolni. Álláspontom szerint a megvalósíthatósági tanulmányok, a hatásvizsgálatok, az érdekmérlegelési tesztek, a szabályzatok és az oktatás – többek között, de nem kizárólag – mind alkalmasak lehetnek a jogászok és az informatikusok közötti híd építésére.

A kommunikációs problémák megoldása nélkül etikus MI kifejlesztése és használata szinte lehetetlen, és ehhez egyrészt olyan szereplőkre van szükség, akik kreatívan törekszenek a problémák megoldására, másrészt pedig olyan platformokra, amelyek lehetővé teszik a konstruktív párbeszédet és az együttgondolkodást. A szervezeteknek tehát tudatosan kell felkészülniük MI-projektjeikre, és a siker érdekében érdemes magas kvalitású szakembereket alkalmazni, akik nem csak a modern technológiákban rejlő előnyöket ismerik fel, hanem képesek a hatékony együttműködésre és a kockázatok kezelésére. Ezen együttműködés nélkül az MI nemcsak szervezetünknek, hanem az érintetteknek, sőt a társadalmunknak is komoly, akár hosszabb távra kiható gondokat okozhat.

Köszönetnyilvánítás

Köszönöm *Dr. Botzheim Jánosnak*, az Eötvös Loránd Tudományegyetem Mesterséges Intelligencia Tanszékének tanszékvezető egyetemi docensének informatikai területen nyújtott segítségét.

Irodalomjegyzék

A személyes adatoknak az illetékes hatóságok által a bűncselekmények megelőzése, nyomozása, felderítése, a vádeljárás lefolytatása vagy büntetőjogi szankciók végrehajtása céljából végzett kezelése tekintetében a természetes személyek védelméről és az ilyen adatok szabad áramlásáról, valamint a 2008/977/IB tanácsi kerethatározat hatályon kívül helyezéséről szóló, 2016. április 27-i (EU) 2016/680 európai parlamenti és tanácsi irányelv (bűnügyi adatvédelmi irányelv, LED).

Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.] a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet, GDPR).

Az Európai Parlament és a Tanács (EU) 2019/790 irányelve (2019. április 17.] a digitális egységes piacon a szerzői és szomszédos jogokról, valamint a 96/9/EK és a 2001/29/EK irányelv módosításáról.

Az Európai Unió Alapjogi Ügynöksége (FRA) (2021) Hogyan alakítsuk jól a jövőt? Mesterséges intelligencia és alapvető jogok. Összefoglaló.

Bonnefon, J.-F. (2021) The Car That Knew Too Much Can a Machine Be Moral? The MIT Press.

Bonnefon, J.-F.–Shariff, A.–Rahwan, I. (2016) The social dilemma of autonomous vehicles. *Science*, on 24 Jun 2016: Vol. 35. doi: 10.1126/science.aaf2654 [Letöltve: 2023.09.23.].

Brownsword, R.–Scotford, E.–Yeung, K. (eds.] (2017) *The Oxford Handbook of Law, Regulation and Technology*. Oxford University Press. doi: 10.1093/oxfordhb/9780199680832.001.0001 [Letöltve: 2023.09.23.].

CISCO (2022) Consumer Privacy Survey.

Coeckelbergh, M. (2020) *AI Ethics*. The MIT Press. doi: 10.7551/mitpress/12549.001.0001 [Letöltve: 2023.09.23.].

Datatilsynet (2018) Artificial intelligence and privacy. Report.

Dimatteo, L. A.–Poncibò, C.–Cannarsa, M. (eds.] (2022) *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. doi: 10.1017/9781009072168 [Letöltve: 2023.09.23.].

DRAFT *Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts* (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD)). 16/5/2023. Version: 1.1 (MI rendelet tervezete).

Dubber, M.D.–Pasquale, F.–Das, S. (eds.] (2020) *The Oxford Handbook of Ethics of AI*. Oxford University Press. doi: 10.1093/oxfordhb/9780190067397.001.0001 [Letöltve: 2023.09.23.].

EDPB 1/2021 Iránymutatás az adatvédelmi incidensek bejelentésével kapcsolatos példákról.

EDPB Guidelines 4/2022 on the calculation of administrative fines under the GDPR.

EDPB Guidelines 9/2022 on personal data breach notification under GDPR.

ENISA (2020) *Artificial Intelligence Cybersecurity Challenges.*

ENISA (2023a) Artificial Intelligence and Cybersecurity Research. https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research [Letöltve: 2023.09.23.].

ENISA (2023b) Cybersecurity of AI and Standardisation.

Európai Ügyvédi Kamarák Tanács (2022) Útmutató ügyvédek és ügyvédi irodák számára a mesterséges intelligencián alapuló eszközök EU-n belüli használatához.

European Parliament Special Committee on Artificial Intelligence in a Digital Age: REPORT on artificial intelligence in a digital age (2020/2266(INI)).

Floridi, L. (ed) (2021) *Ethics, Governance, and Policies in Artificial Intelligence*. Spinger. doi_ 10.1007/978-3-030-81907-1 [Letöltve: 2023.09.23.].

Floridi, L.–Cowls, J.–Beltrametti, M. et al. (2018) *AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds & Machines 28. pp. 689–707. doi: 10.1007/s11023-018-9482-5 [Letöltve: 2023.09.23.].

Garante per la protezione dei dati personali Provvedimento del 30 marzo 2023 [9870832]. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832 [Letöltve: 2023.09.23.].

Garante per la protezione dei dati personali Provvedimento dell'11 aprile 2023 [9874702]. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702 [Letöltve: 2023.09.23.].

Government of Canada: Directive on Automated Decision-Making. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 [Letöltve: 2023.09.23.].

Hallinan, D.–Leenes, R.–Gutwirth, S.–De Hert, P. (eds.] (2020) Data Protection and Privacy. Hart Publishing. doi: 10.5040/9781509941780 [Letöltve: 2023.09.23.].

Holberton-Turing eskü. https://www.holbertonturingoath.org/ [Letöltve: 2023.09.23.].

Howard, J.J.–Rabbitt, L. R.– Sirotin, Y. B. (2020) *Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making*. PLoS ONE 15(8): e0237855. doi: 10.1371/journal.pone.0237855 [Letöltve: 2023.09.23.].

IBM Global AI Adoption Index 2022. https://www.ibm.com/downloads/cas/GVAGA3JP [Letöltve: 2023.09.23.].

ICO (2018) *Data Protection Impact Assessments (DPIAs).*

ICO (2020) *AI auditing framework* - draft guidance for consultation.

ICO (2023) *Guidance on AI and Data Protection.*

ICO–The Alan Turing Institute (2022) *Explaining decisions made with AI.*

ILO Working Paper 96 (2023) Generative AI and Jobs: A global analysis of potential effects on job quantity and quality. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---inst/documents/publication/wcms_890761.pdf [Letöltve: 2023.09.23.].

Jarjabka, Á. et al. (2020) Projektmenedzsment ismeretek. Pécs: Pécsi Tudományegyetem Közgazdaságtudományi Kar Vezetés- és Szervezéstudományi Intézet.

Javaslat az Európai Parlament és a Tanács irányelve a szerződésen kívüli polgári jogi felelősségre vonatkozó szabályoknak a mesterséges intelligenciához való hozzáigazításáról (a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv), COM(2022) 496 final 2022/0303 (COD).

Johannessen, J.-A. (2021) Robot Ethics and the Innovation Economy. Routledge. doi: 10.4324/9781003174493 [Letöltve: 2023.09.23.].

Kafka, P. (2023) *The creator of Black Mirror is okay with tech. People, on the other hand …* VOX Aug 4, 2023.

https://www.vox.com/technology/2023/8/4/23819299/black-mirror-charlie-brooker-interview-ai-peter-kafka-media-column?mc_cid=3f3a9ae0bb&mc_eid=31136da945 [Letöltve: 2023.09.23.].

Kerrigan, C. (ed.) (2022) Artificial Intelligence Law and Regulation. Edward Elgar Publishing Limited. doi: 10.4337/9781800371729 [Letöltve: 2023. 09. 23.].

Leroy, J. B. (2021) *Societal Responsibility of Artificial Intelligence*. ISTE Ltd. doi: 10.1002/9781119831808 [Letöltve: 2023.09.23.].

Leslie, D. (2019) Understandig artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. doi: 10.5281/zenodo.3240529 [Letöltve: 2023 09.23.].

Liao, S. M. (ed) (2020) *Ethics of Artificial Intelligence*. Oxford University Press. doi: 10.1093/oso/9780190905033.001.0001 [Letöltve: 2023.09.23.].

Mandel, G. N. (2017) *Legal Evolution in Response to Technological Change*. The Oxford Handbook of Law, Regulation and Technology. In: Brownsword, R.–Scotford, E.–Yeung, K. (eds.). doi: 10.1093/oxfordhb/9780199680832.013.45 [Letöltve: 2023.09.23.].

Marino, D.–Monaca, M. A. (eds.) (2020) Economic and Policy Implications of Artificial Intelligence. Springer. doi: 10.1007/978-3-030-45340-4 [Letöltve: 2023.09.23.].

Mehrabi et al. (2022) A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6). pp. 1–35. 25 Jan 2022. doi: 10.1145/3457607 [Letöltve: 2023.09.23.].

Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport (2019) *Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan.*

Misselhorn, C. (2022) *Artificial Moral Agents: Conceptual Issues and Ethical Controversy*. The Cambridge Handbook of Responsible Artificial Intelligence. Cambridge University Press & Assessment. doi: 10.1017/9781009207898

NAIH: Az adatvédelmi hatásvizsgálat és előzetes konzultációja. https://naih.hu/az-adatvedelmi-hatasvizsgalat-es-elozetes-konzultacioja [Letöltve: 2023.09.23.]

Naqvi, A. (2020) Artificial Intelligence for Audit, Forensic Accounting, and Valuation. John Wiley and Sons. doi: 10.1002/9781119601906 [Letöltve: 2023.09.23.].

Pinker, S. (2012) How the Mind Works. Penguin Books Ltd.

Rosemann, A.–Zhang, X. (2021) *Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health – a new column in Intelligent Medicine*. Published by Elsevier B.V. on behalf of Chinese Medical Association. doi: 10.1016/j.imed.2021.12.002 [Letöltve: 2023.09.23.].

Slaughter, R. K.–Kopec, J.–Batal, M. (2021) Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission. ISP Digital Future Whitepaper & YJoLT Special Publication.

Tietosuojavaltuutetun toimisto (2020) Automatisoitujen yksittäispäätösten syntyminen ennakoivan terveydenhuollon työkalussa. 6482/186/2020. https://finlex.fi/fi/viranomaiset/tsv/2022/20221544 [Letöltve: 2023.09.23.].

UNESCO (2023) Ethical Impact Assessment - A Tool of the Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137 [Letöltve: 2023.09.23.].

Voeneky, S.–Kellmeyer, P.–Mueller, O.–Burgard, W. (eds.) (2022) The Cambridge Handbook of Responsible Artificial Intelligence. Cambridge University Press & Assessment. doi: 10.1017/9781009207898 [Letöltve: 2023.09.23.].

von Braun, J.–Archer, M. S.–Reichberg, G. M.–Sorondo, M. S. (eds.) (2021) Robotics, AI, and Humanity. Springer. doi: 10.1007/978-3-030-54173-6_1 [Letöltve: 2023.09.23.].

Wolf, F. (2022) A Blueprint for the Regulation of Artificial Intelligence Technologies. Ethics International Press Ltd., UK.

World Economic Forum (2019) *Guidelines for AI Procurement.*

World Health Organization (2021) Ethics and governance of artificial intelligence for health. https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf [Letöltve: 2023.09.23.].

WP 217 (2014). 06/2014. számú vélemény az adatkezelő 95/46/EK irányelv 7. cikke szerinti jogszerű érdekeinek fogalmáról.

WP 248 rev.01 (2017) Iránymutatás az adatvédelmi hatásvizsgálat elvégzéséhez és annak megállapításához, hogy az adatkezelés az (EU) 2016/679 rendelet alkalmazásában „valószínűsíthetően magas kockázattal jár”-e.

Završnik, A.–Simončič, K. (eds.) (2023) *Artificial Intelligence, Social Harms and Human Rights*. Palgrave Macmillan. doi: 10.1007/978-3-031-19149-7 [Letöltve: 2023. 09. 23.].