

Brüsszel, 2019.4.8.
COM(2019) 168 final

**A BIZOTTSÁG KÖZLEMÉNYE AZ EURÓPAI PARLAMENTNEK, A
TANÁCSNAK, AZ EURÓPAI GAZDASÁGI ÉS SZOCIÁLIS BIZOTTSÁGNAK ÉS A
RÉGIÓK BIZOTTSÁGÁNAK**

Az emberközpontú mesterséges intelligencia iránti bizalom növelése

1. BEVEZETÉS – A MESTERSÉGES INTELLIGENCIÁRA VONATKOZÓ EURÓPAI STRATÉGIA

A mesterséges intelligencia (AI) lehetőséget teremt arra, hogy jobbra tegyük világunkat: javíthatja az egészségügyi ellátást, csökkentheti az energiafogyasztást, biztonságosabbá teheti a gépjárműveket, és lehetővé teheti a mezőgazdasági termelők számára a víz és a természeti erőforrások hatékonyabb felhasználását. A mesterséges intelligencia felhasználható továbbá a környezet- és az éghajlatváltozás előrejelzésére és a pénzügyi kockázatkezelés javítására, valamint eszközöket kínál az igényekhez igazított termékek kevesebb hulladékkal történő előállításához. Emellett segíthet a csalás és a kiberbiztonsági fenyegetések észlelésében, és a bűnüldöző hatóságok számára lehetőséget teremt a bűnözés elleni hatékonyabb fellépésre.

A mesterséges intelligencia az egész társadalom és a gazdaság számára szintén előnyöket tartogat. Olyan stratégiai technológiáról van szó, amelyet jelenleg világszerte gyors ütemben fejlesztenek és egyre szélesebb körben alkalmaznak. Ugyanakkor a mesterséges intelligencia új kihívásokat is jelent a munka jövője szempontjából, valamint jogi és etikai kérdéseket vet fel.

E kihívások kezelése, illetve a mesterséges intelligencia által kínált lehetőségek lehető legjobb kihasználása céljából a Bizottság 2018 áprilisában európai stratégiát¹ tett közzé a témában. Ez a stratégia a mesterséges intelligencia fejlesztésének középpontjába az embert állítja (**emberközpontú mesterséges intelligencia**). Ez egy három részből álló megközelítés, amelynek célja az EU technológiai és ipari kapacitásainak erősítése, a mesterséges intelligencia hasznosításának terjesztése az Unió gazdaságának egészén belül, a társadalmi-gazdasági változásokra való felkészülés, valamint a megfelelő etikai és jogi keret biztosítása.

Az AI-stratégia megvalósítása érdekében a **Bizottság a tagállamokkal közösen kidolgozott egy, a mesterséges intelligenciára vonatkozó összehangolt tervet²**, melyet 2018 decemberében azzal a céllal mutatott be, hogy szinergiákat hozzon létre, adatokat gyűjtsön (melyeket számos AI-alkalmazás nyersanyagként kezelhet), valamint növelje a közös beruházásokat. A terv további célja, hogy előmozdítsa a határokon átnyúló együttműködést, és a szereplők mobilizálásával a következő évtizedben évente **legalább 20 milliárd euróra** növelje a köz- és magánberuházásokat³. A Bizottság a „Horizont 2020” program keretében megduplázta a mesterséges intelligenciával kapcsolatos beruházásait, és az Európai horizont kutatási és innovációs keretprogramon, valamint a Digitális Európa programon keresztül évente 1 milliárd eurót tervez fordítani különösen az egészségügy, a közlekedés és a gyártás közös adattereinek, a nagy kísérleti létesítményeknek – például az intelligens kórházaknak és az automatizált járművekkel kapcsolatos infrastruktúráknak –, valamint a stratégiai kutatási tervnek a támogatására.

Egy ilyen közös stratégiai kutatási, innovációs és bevezetési terv végrehajtása érdekében a Bizottság intenzívebb párbeszédbe kezdett az ipart, a kutatóintézeteket és a hatóságokat képviselő **valamennyi érintett féllel**. Az új Digitális Európa programnak nagy lesz a jelentősége annak elősegítése szempontjából is, hogy a mesterséges intelligencia digitális innovációs központok, megerősített tesztelési és kísérleti létesítmények, adatterek és képzési programok révén elérhetővé váljon valamennyi tagállam kis- és középvállalkozásai számára.

¹

² COM(2018) 795.

³ E cél elérése érdekében a Bizottság a következő, 2021 és 2027 közötti programozási időszakra vonatkozóan azt javasolta, hogy az Unió évente legalább 1 milliárd eurót különítsen el az Európai horizont és a Digitális Európa programok költségvetéséből a mesterséges intelligencia területén történő beruházásokra.

Európa mesterséges intelligenciával kapcsolatos etikai megközelítése, amely az európai termékek biztonságosságának és magas minőségi színvonalának jó hírnevére alapoz, erősíti a polgárok digitális fejlődésbe vetett bizalmát, és versenyelőnyt kíván teremteni a mesterséges intelligenciával foglalkozó európai vállalatok számára. E közlemény célja egy, az érdeket felek legszélesebb körét bevonó, átfogó kísérleti időszak elindítása a mesterséges intelligencia fejlesztésére és használatára vonatkozó etikai iránymutatások gyakorlati végrehajthatóságának vizsgálata érdekében.

2. AZ EMBERKÖZPONTÚ MESTERSÉGES INTELLIGENCIA IRÁNTI BIZALOM NÖVELÉSE

A mesterséges intelligenciára vonatkozó európai stratégia és az összehangolt terv egyértelművé teszi, **hogy a bizalom előfeltétele a mesterséges intelligencia emberközpontú megközelítésének**: A mesterséges intelligencia nem maga a cél, hanem egy, az embereket szolgáló eszköz, melynek végső célja az emberi jólét növelése. Ennek megvalósítása érdekében **gondoskodni kell arról, hogy a mesterséges intelligencia megbízható legyen**. A társadalmaink alapjául szolgáló értékeket teljes mértékben be kell építeni a mesterséges intelligencia fejlesztésébe.

Az Unió **az emberi méltóság tiszteletben tartása, a szabadság, a demokrácia, az egyenlőség, a jogállamiság, valamint az emberi jogok** – ezen belül a kisebbségekhez tartozó személyek jogai – tiszteletben tartásának értékein alapul⁴. Ezek az értékek az összes olyan tagállam osztozik, amelyet a pluralizmus, a megkülönböztetésmentesség, a tolerancia, az igazságosság, a szolidaritás és az egyenlőség jellemez. Emellett az **Európai Unió Alapjogi Chartája** egyetlen szövegben határozza meg az EU-ban élő polgárokat megillető személyiségi, polgári, politikai, gazdasági és szociális jogokat.

Az EU **szigorú keretszabályokkal** rendelkezik, amelyek meghatározók lesznek az emberközpontú mesterséges intelligenciára vonatkozó globális normák kialakításában. Az általános adatvédelmi rendelet (GDPR) biztosítja a személyes adatok magas szintű védelmét, és olyan intézkedések végrehajtását írja elő, amelyek biztosítják a beépített és alapértelmezett adatvédelmet⁵. A nem személyes adatok szabad áramlásáról szóló rendelet felszámolja a nem személyes adatok szabad mozgásának akadályait, és Európa-szerte biztosítja valamennyi adatkategória feldolgozását. A közelmúltban elfogadott kiberbiztonsági jogszabály pedig hozzá fog járulni az online világba vetett bizalom megerősítéséhez. Ugyanezt a célt szolgálja az elektronikus hírközlési adatvédelmi rendeletre irányuló javaslat⁶ is.

Ugyanakkor a mesterséges intelligencia új kihívásokkal is jár, mivel lehetővé teszi a gépek számára a „tanulást”, valamint döntések emberi beavatkozás nélküli meghozatalát és végrehajtását. Az ilyen típusú funkciók hamarosan szabványos részévé válnak az áruk és szolgáltatások számos típusának, az okostelefonoktól kezdve az automatizált autók és a robotokon át az online alkalmazásokig. Az algoritmusok azonban nem teljes – és ezért nem megbízható – adatok alapján is hozhatnak döntéseket, kibertámadás áldozatává válhatnak,

⁴ Az EU a fogyatékkal élő személyek jogairól szóló ENSZ-egyezményhez is csatlakozott.

⁵ (EU) 2016/679 rendelet. Az általános adatvédelmi rendelet (GDPR) biztosítja a személyes adatok szabad áramlását az Unión belül. Rendelkezéseket határoz meg a kizárólag automatizált feldolgozáson alapuló döntéshozatalra vonatkozóan, ideértve a profilalkotást is. Az érintett személyeknek joguk van ahhoz, hogy tájékoztatást kapjanak az automatizált döntéshozatal létezéséről, illetőleg ahhoz, hogy érdemi tájékoztatást kapjanak az automatizált döntéshozatal során alkalmazott logikáról, valamint adataik feldolgozásának jelentőségéről és várható következményeiről a saját szempontjukból. Ezekben az esetekben joguk van emberi beavatkozást kérni, álláspontjukat kifejteni és a döntést vitatni.

⁶ COM(2017) 10.

döntéseik pedig torzulhatnak vagy egyszerűen tévesek lehetnek. Az AI-technológiai fejlesztések megfontolatlan, azonnali alkalmazása tehát problémákhoz vezetne, a polgárok pedig vonakodnának elfogadni és használni a technológiát.

Ehelyett az AI-technológiát emberközpontúan kell fejleszteni, hogy az méltóvá váljon a emberek bizalmára. Ez azt jelenti, hogy a mesterséges intelligencián alapuló alkalmazásoknak nemcsak a jogszabályokkal kell összhangban lenniük, hanem az etikai elveknek is meg kell felelniük, és biztosítaniuk kell, hogy az alkalmazásuk ne okozzon nem kívánt károkat. A mesterséges intelligencia fejlesztésének minden szakaszában figyelemmel kell lenni a nemek, a faji vagy etnikai származás, a vallás vagy meggyőződés, a fogyatékoság és az életkor szerinti sokféleségre. Az AI-alkalmazásoknak a polgárok érdekeit kell szolgálniuk, és tiszteletben kell tartaniuk alapvető jogaikat. Céljuk nem az emberek helyettesítése, hanem az emberi képességek javítása, valamint a fogyatékosággal élők előtti akadályok felszámolása.

Ezért olyan **etikai iránymutatásokra** van szükség, amelyek a meglévő szabályozási keretre építenek, valamint a belső piacon tevékenykedő AI-fejlesztők, -beszállítók és -felhasználók számára kötelezően alkalmazandóak, és ezáltal valamennyi tagállamban etikus és egyenlő versenyfeltételeket biztosítanak. A Bizottság ezért létrehozott egy, az érdekelt felek széles körét képviselő, **mesterséges intelligenciával foglalkozó magas szintű szakértői csoportot**⁷, amelyet a mesterséges intelligenciára vonatkozó etikai iránymutatások összeállításával, illetőleg az átfogóbb mesterségesintelligencia-politikára vonatkozó ajánlások kidolgozásával bízott meg. Ezzel párhuzamosan megalakult a **mesterséges intelligenciával foglalkozó európai szövetség**⁸ – egy több érdekelt felet tömörítő, több mint 2 700 tagot számláló nyílt platform – azzal a céllal, hogy szélesebb körű támogatást biztosítson a mesterséges intelligenciával foglalkozó magas szintű szakértői csoportnak.

A mesterséges intelligenciával foglalkozó magas szintű szakértői csoportot 2018 decemberében közzétette az etikai iránymutatások első tervezetét. Az **érdekelt felekkel folytatott konzultációt**⁹, valamint a **tagállamok képviselőivel való találkozókat**¹⁰ követően a mesterséges intelligenciával foglalkozó szakértői csoportot 2019 márciusában átadta a Bizottságnak az iránymutatások módosított változatát. Az eddigi visszajelzések alapján az érdekelt felek általánosságban üdvözlötték az iránymutatások gyakorlati jellegét, valamint az AI-fejlesztőknek, -beszállítóknak és -felhasználóknak adott konkrét iránymutatást arra nézve, hogy hogyan gondoskodhatnak a technológia megbízhatóságáról.

2.1. A mesterséges intelligenciával foglalkozó magas szintű szakértői csoport által kidolgozott, a megbízható mesterséges intelligenciára vonatkozó iránymutatások

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁹ A konzultáció keretében többek között 511 szervezet, szövetség, vállalat, kutatóintézet, egyén és más érdekelt fél nyújtotta be észrevételeit. A beérkezett visszajelzések összefoglalója a következő címen érhető el: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

¹⁰ A tagállamok pozitívan fogadták a szakértői csoport munkáját, melynek nyomán a Tanács a 2019. február 18-i következtetéseiben egyebek mellett nyugtázta az etikai iránymutatások közelgő megjelenését, és támogatta a Bizottság arra irányuló törekvését, hogy az EU etikai megközelítését a globális szinten is érvényesítse: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/hu/pdf>

Az e közlemény¹¹ tárgyát képező, a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport által kidolgozott iránymutatások elsősorban a tudomány és az új technológiák etikai kérdéseit vizsgáló európai csoportnak és az Európai Unió Alapjogi Ügynökségének munkájára épülnek.

Az iránymutatások értelmében a „megbízható mesterséges intelligencia” megvalósításához három összetevőre van szükség: 1. a mesterséges intelligenciának be kell tartania a jogszabályokat, 2. meg kell felelnie az etikai elveknek és 3. stabilnak kell lennie.

E három összetevő, valamint a 2. szakaszban ismertetett európai értékek alapján az iránymutatások hét olyan kulcsfontosságú követelményt határoznak meg, amelyeket az AI-alkalmazásoknak teljesíteniük kell ahhoz, hogy megbízhatónak minősüljenek. Az iránymutatások tartalmazzak továbbá egy értékelési listát, amely segít annak ellenőrzésében, hogy ezek a követelmények teljesülnek-e.

A hét kulcsfontosságú követelmény a következő:

- az emberi cselekvőképesség támogatása és emberi felügyelet;
- műszaki stabilitás és biztonság;
- adatvédelem és adatkezelés;
- átláthatóság;
- sokféleség, megkülönböztetésmentesség és méltányosság;
- társadalmi és környezeti jólét;
- elszámoltathatóság.

Noha ezek a követelmények a különböző környezetekben és iparágakban működő valamennyi mesterségesintelligencia-rendszerre vonatkoznak, alkalmazásukkor a konkrét és arányos végrehajtás érdekében – hatásalapú megközelítést követve – figyelembe kell venni az adott környezet sajátosságait. Például sokkal kevésbé veszélyes egy olyan AI-alkalmazás, amely nem megfelelő könyvet ajánl a felhasználónak, mint egy olyan, amely tévesen diagnosztizálja a rákot, ezért az előbbit nem szükséges ugyanolyan szigorú felügyelet alatt tartani.

A mesterséges intelligenciával foglalkozó magas szintű szakértői csoport által kidolgozott iránymutatások nem kötelező erejűek, így nem jelentenek új jogi kötelezettségeket. Ugyanakkor az uniós jog számos meglévő (gyakran használat- vagy területspecifikus) rendelkezése természetesen már most is megfogalmaz egyet vagy többet e kulcsfontosságú követelmények közül, például a biztonságra, a személyes adatok védelmére, a magánélet védelmére vagy a környezetvédelemre vonatkozó szabályok formájában.

A Bizottság üdvözli a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport munkáját, és arra értékes támpontként tekint a politikai döntéshozatalban.

2.2. A megbízható mesterséges intelligenciára vonatkozó kulcsfontosságú követelmények

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

A Bizottság az alábbi, európai értékeken alapuló **megbízható mesterséges intelligenciára vonatkozó kulcsfontosságú követelményeket támogatja**. Arra ösztönzi az érdekelt feleket, hogy alkalmazzák a követelményeket, és teszteljék az – azok gyakorlatba való átültetését segítő – értékelési listát annak érdekében, hogy megteremtsék a mesterséges intelligencia sikeres fejlesztéséhez és alkalmazásához szükséges megfelelő, bizalmon alapuló környezetet. A Bizottság örömmel fogadja az érdekelt felek visszajelzéseit annak felmérésére, hogy az iránymutatásokban megadott, szóban forgó értékelési lista további módosításokat igényel-e.

I. Az emberi cselekvőképesség támogatása és emberi felügyelet

Az AI-rendszereknek támogatniuk kell az egyéneket abban, hogy a céljaik elérése szempontjából jobb és megalapozottabb döntéseket hozzanak. Az emberi cselekvőképesség és az **alapvető jogok** érvényesülésének támogatása révén hozzá kell járulniuk egy virágzó és méltányos társadalom megteremtéséhez, és semmiképpen sem csökkenthetik vagy korlátozhatják az emberek autonómiáját, vagy vezethetik félre az embereket. A rendszernek működés közben elsődleges szempontként kell kezelnie a **felhasználó általános jóllétét**.

Az emberi felügyelet segít annak biztosításában, hogy az AI-rendszer ne ássa alá az emberi autonómiát, vagy fejtsen ki más káros hatást. Az adott, mesterséges intelligencián alapuló rendszer jellegének és alkalmazási területének figyelembevételével olyan megfelelő **ellenőrzési intézkedéseket** kell foganatosítani, amelyek biztosítják a mesterséges intelligencián alapuló rendszerek kiigazíthatóságát, pontosságát és döntéseinek megmagyarázhatóságát¹². A **felügyelet** például emberi beavatkozáson („human-in-the-loop”), emberi felügyeleten („human-on-the-loop”) vagy emberi vezérlésen („human-in-command”) alapuló irányítási mechanizmusok révén valósítható meg¹³. Biztosítani kell, hogy a hatóságok megbízatásukkal összhangban képesek legyenek gyakorolni felügyeleti hatáskörüket. Minden más szemponttól függetlenül minél kisebb mértékű az emberi felügyelet egy adott mesterségesintelligencia-rendszer felett, annál alaposabb tesztelésre és annál szigorúbb irányításra van szükség.

II. Műszaki stabilitás és biztonság

Ahhoz, hogy a mesterséges intelligencia megbízható legyen, az algoritmusoknak kellően biztonságosaknak, megbízhatóknak és stabilaknak kell lenniük ahhoz, hogy az AI-rendszer teljes életciklusa alatt megbirkózzanak a hibákkal és következetlenségekkel, valamint képesnek kell lenniük a hibás eredmények megfelelő kezelésére. Az AI-rendszereknek **megbízhatóaknak** kell lenniük, elég védettnek kell lenniük ahhoz, hogy **ellenálljanak** mind a nyílt, mind pedig az adatok vagy közvetlenül az algoritmusok

¹² Az általános adatvédelmi rendelet felruházza a magánszemélyeket a kizárólag automatizált adatkezelésen alapuló döntések alóli mentesüléshez való joggal, amennyiben a döntésnek joghatása vagy hasonlóan jelentős hatása van a felhasználókra nézve (általános adatvédelmi rendelet, 22. cikk).

¹³ A „human-in-the-loop” (emberi beavatkozás, HITL) megközelítés a rendszer minden egyes döntési ciklusában emberi beavatkozást feltételez, ami sok esetben nem lehetséges és nem is elvárás. A „human-on-the-loop” (emberi felügyelet, HOTL) megközelítés a rendszer tervezési ciklusa alatti emberi beavatkozás lehetőségét, valamint a rendszer működésének figyelemmel kísérését jelenti. A „human-in-command” (emberi vezérlés, HIC) megközelítés értelmében ember felügyeli az AI-rendszer általános tevékenységét (beleértve annak tágabb értelemben vett gazdasági, társadalmi, jogi és etikai hatásait), és dönthet arról, hogy a rendszert mikor és hogyan használja fel egy adott helyzetben. Ez lehetőséget biztosíthat arra is, hogy egy adott helyzetben ne használják az AI-rendszert, a rendszer használata során megállapítsák az emberi döntést igénylő szinteket, illetve hogy felülbírálják a rendszer által hozott döntéseket.

manipulálását célzó kifinomultabb támadásoknak, és problémák esetére rendelkezniük kell egy **készenléti tervvel**. Döntéseiknek **pontosaknak** kell lenniük, vagy legalább megfelelően jelezniük kell a döntés pontosságát, és az eredményeknek **megismételhetőeknek** kell lenniük.

Emellett az AI-rendszereknek rendelkezniük kell beépített biztonsági és védelmi mechanizmusokkal annak érdekében, hogy az érintett személyek testi és szellemi biztonságának szempontjából a rendszer által végrehajtott minden lépés **igazolhatóan biztonságos** legyen. Ez magában foglalja a rendszer működése során előforduló nem kívánt következmények és hibák számának minimalizálását és – ha lehetséges – azok visszafordíthatóságának biztosítását. Emellett olyan eljárásokat kell létrehozni, melyekkel segítségével tisztázhatók és értékelhetők az AI-rendszerek használatához kötődően a különböző alkalmazási területeken felmerülő esetleges kockázatok.

III. Adatvédelem és adatkezelés

A magánélet és az **adatok védelmét** az AI-rendszer életciklusának **valamennyi szakaszában** biztosítani kell. Az emberi viselkedésre vonatkozó digitális adatok lehetővé tehetik az AI-rendszerek számára nemcsak az emberek preferenciáinak, életkorának és nemének visszafejtését, hanem akár szexuális irányultságuk és vallási vagy politikai nézeteik megállapítását is. Annak érdekében, hogy az emberek megbízzanak az adatfeldolgozásban, biztosítani kell, hogy teljes mértékben rendelkezhessenek a saját adataik felett, valamint hogy adataik ne legyenek felhasználhatóak velük szembeni károkozásra vagy hátrányos megkülönböztetésre.

A magánélet és a személyes adatok védelme mellett gondoskodni kell arról is, hogy teljesüljenek az AI-rendszerek magas színvonalú működését lehetővé tevő feltételek. A felhasznált adatkészletek minősége rendkívül fontos az AI-rendszerek teljesítménye szempontjából. Előfordulhat, hogy az összegyűjtött adatok társadalmi előítéleteket tükröznek, hibásak, pontatlanok vagy tévesek. Ezeket a problémákat még azelőtt el kell hárítani, hogy megkezdődne az AI-rendszer tanítása az adott adatkészlet alapján. Emellett biztosítani kell az adatok **sértetlenségét** is. A használt folyamatokat és adatkészleteket minden lépésnél – például tervezéskor, tanításkor, teszteléskor és bevezetéskor – tesztelni és dokumentálni kell. Ez azokra az AI-rendszerekre is vonatkozik, amelyeket nem házon belül fejlesztettek ki, hanem máshonnan szereztek be. Ami pedig az adatokhoz való **hozzáférést** illeti, azt megfelelően irányítani és ellenőrizni kell.

IV. Átláthatóság

Biztosítani kell az AI-rendszerek **nyomonkövethetőségét**: fontos naplózni és dokumentálni mind a rendszerek által hozott döntéseket, mind pedig a döntésekhez vezető teljes folyamatot (beleértve az adatgyűjtés és címkézés, valamint a használt algoritmus leírását). Ehhez kapcsolódóan a lehető legnagyobb mértékben gondoskodni kell az érintett személyekhez igazított algoritmikus döntéshozatali folyamat **megmagyarázhatóságáról**. Folytatni kell továbbá az magyarázhatósági mechanizmusok kifejlesztésére irányuló, folyamatban lévő kutatásokat. Ezenkívül rendelkezésre kell állnia információnak a rendszer tervezése során meghozott döntésekről, a rendszer bevezetésének létjogosultságáról, valamint arról, hogy egy adott AI-rendszer milyen mértékben befolyásolja és alakítja át a szervezeti döntéshozatali eljárást (ami nem csak az adatok és a rendszer, hanem az üzleti modell átláthatóságát is biztosítja).

Végül fontos, hogy – az adott felhasználási esetet figyelembe véve – megfelelően **tájékoztassák** a különböző érdekelt feleket az AI-rendszer képességeiről és korlátairól. Ezen túlmenően az AI-rendszereket azonosíthatóvá kell tenni oly módon, hogy a felhasználók felismerjék, hogy egy AI-rendszerrel kommunikálnak, és értesüljenek arról, hogy mely személyek felelősek a rendszerért.

V. Sokféleség, megkülönböztetésmentesség és méltányosság:

Az AI-rendszerek által (mind a tanulás, mind az üzemelés alatt) használt adatkészletek minőségére negatív hatással lehet a történeti adatok miatti nem szándékos torzulás, az adatok hiányossága, valamint a rossz adatkezelési modellek használata. Az ilyen jellegű torzulás közvetlen vagy közvetett hátrányos megkülönböztetéshez vezethet. Szintén káros hatása lehet a (fogyasztói) szokások szándékos kiaknázásának és a tisztességtelen versenynek. Ezenkívül a torzulás negatív hatással lehet az AI-rendszerek fejlesztése során alkalmazott – például az algoritmusok programkódjának megírásához használt – módszerekre is. Az ilyen problémákat már a rendszer fejlesztésének kezdetétől fogva kezelni kell.

A megoldáshoz hozzájárulhat a **sokszínű fejlesztői csapatok** létrehozása, valamint olyan mechanizmusok megteremtése, melyek kimondottan a polgárok **részvételét** hivatottak biztosítani a mesterséges intelligencia fejlesztésében. Ajánlott továbbá konzultálni azokkal az érdekelt felekkel, akikre a rendszer az életciklusa során közvetlen vagy közvetett hatással lehet. Az AI-rendszereknek az emberi képességek, készségek és szükségletek teljes skáláját figyelembe kell venniük, és egyetemes tervezési megközelítések alkalmazásával törekedniük kell a rendszerekhez való egyenlő hozzáférés biztosítására a fogyatékkal élők számára.

VI. Társadalmi és környezeti jólét

Ahhoz, hogy a mesterséges intelligencia megbízható legyen, figyelembe kell venni a **környezetre és más érző lényekre** gyakorolt hatását. Ideális esetben minden embert – beleértve a jövő generációit is – megilleti, hogy élvezze a biológiai sokféleség és a lakható környezet előnyeit. Ezért ösztönözni kell az AI-rendszerek fenntartható és **ökológiailag felelősségteljes** üzemeltetését. Ugyanez vonatkozik a globális aggodalomra okot adó területekre – például az ENSZ fenntartható fejlődési céljaira – irányuló AI-megoldásokra.

Emellett az AI-rendszerek hatását nemcsak az egyén szempontjából, hanem a **társadalom egésze** szempontjából is meg kell vizsgálni. Az AI-rendszerek használatát különös alaposítással kell megfontolni a demokratikus folyamatokat érintő – például véleményformálással, politikai döntéshozatallal vagy választásokkal kapcsolatos – esetekben. Figyelembe kell venni továbbá a mesterséges intelligencia **szociális hatását** is. Bár az AI-rendszerek felhasználhatók a szociális készségek fejlesztésére, e készségek romlásához is hozzájárulhatnak.

VII. Elszámoltathatóság

Létre kell hozni olyan mechanizmusokat, amelyek rendezik az AI-rendszerekre és az általuk generált eredményekre vonatkozó, a felelősségvállalással és elszámoltathatósággal kapcsolatban – az adott rendszer bevezetése előtt és után – felmerülő kérdéseket. E tekintetben kulcsfontosságú az AI-rendszerek **ellenőrizhetősége**, mivel az AI-rendszerek belső és külső ellenőrök általi értékelése és az ebből születő értékelési jelentések

rendelkezésre állása nagy mértékben hozzájárul a technológia megbízhatóságának igazolásához. A külső ellenőrzést különösen az alapvető jogokat érintő – köztük a biztonsági szempontból kritikus – alkalmazások esetében kell biztosítani.

Az AI-rendszerek **lehetséges negatív hatásait** azonosítani, értékelni, dokumentálni és minimalizálni kell. Ez a folyamat hatásvizsgálatok használatával megkönnyíthető. A hatásvizsgálatok mélységének arányosnak kell lennie az adott AI-rendszer jelentette kockázat mértékével. A különböző követelményeknek való megfelelés miatti **kompromisszumokat** – melyek gyakran elkerülhetetlenek – észszerűen és módszeresen kell meghozni, és azokat mindig figyelembe kell venni. Végezetül arra az esetre, ha az AI-rendszer méltánytalan negatív hatás fejt ki, gondoskodni kell olyan mechanizmusokról, amelyek lehetőséget biztosítanak a **megfelelő jogorvoslatra**.

2.3. Következő lépések: az érdekelt felek legszélesebb körét bevonó kísérleti időszak

A fenti, AI-rendszerekre vonatkozó kulcsfontosságú követelményekről kialakult konszenzus csupán az első mérföldkő az etikus mesterséges intelligenciára vonatkozó iránymutatások felé vezető úton. Következő lépésként a Bizottság lehetőséget biztosít az iránymutatások gyakorlati tesztelésére és végrehajtására.

E célból most elindít egy célzott kísérleti időszakot, amelynek során strukturált visszajelzéseket vár az érdekelt felektől. A mostani kísérleti időszak központi kérdése az értékelési lista lesz, amelyet a magas szintű szakértői csoport állított össze az egyes kulcsfontosságú követelmény teljesülésének ellenőrzése céljából.

A munka két vonalon folyik majd: i. a mesterséges intelligenciát fejlesztő vagy használó érdekelt felek, köztük a közigazgatási szervek bevonásával zajló, az iránymutatások vizsgálatára szolgáló kísérleti időszak, valamint ii. az érdekelt felekkel folytatott folyamatos konzultáció, valamint a tagállamokra és az érdekelt felek különböző csoportjaira, köztük az ipari és a szolgáltatási ágazatokra kiterjedő figyelemfelkeltési folyamat keretében:

- (i) 2019 júniusától kezdődően minden érdekelt félnek és egyénnek lehetősége lesz tesztelni az értékelési listát, és visszajelzést adnia a javítási javaslatairól. Ezen túlmenően a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport a magán- és a közzsférát képviselő érdekelt felek bevonásával részletes vizsgálatot fog szervezni annak érdekében, hogy pontosabb visszajelzéseket kapjon arról, hogy az iránymutatásokat hogyan lehet az alkalmazási területek széles skáláján végrehajtani. Az iránymutatások alkalmazhatóságára és megvalósíthatóságára vonatkozó visszajelzéseket 2019 végéig értékelni fogják.
- (ii) Ezzel párhuzamosan a Bizottság további tájékoztatási tevékenységeket fog szervezni, amelyek keretében lehetőséget biztosít a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport képviselőinek arra, hogy az iránymutatásokat ismertessék a tagállamok érdekelt feleivel, többek között az ipari és szolgáltatási ágazatokkal, és további lehetőséget biztosítsanak az említett érdekelt feleknek, hogy hozzájáruljanak a mesterséges intelligenciára vonatkozó iránymutatások javításához, illetőleg megosszák azokkal kapcsolatos észrevételeiket.

A Bizottság figyelembe fogja venni az összekapcsolt és automatizált járművezetés etikájával foglalkozó szakértői csoport munkáját¹⁴, és a kulcsfontosságú követelmények végrehajtása tekintetében együtt fog működni az uniós finanszírozású, mesterséges intelligenciával

¹⁴ Lásd az összekapcsolt és automatizált mobilitásról szóló bizottsági közleményt, COM(2018) 283.

foglalkozó kutatási projektekkel és az érintett köz-magán társulásokkal¹⁵. A Bizottság a tagállamokkal együttműködve például támogatni fogja az orvosi képek közös adatbázisának fejlesztését (melyet kezdetben a rák leggyakrabban előforduló típusairól készített képek tárolására terveztek) annak érdekében, hogy az algoritmusok betaníthatóak legyenek a tünetek nagyon nagy pontosságú diagnosztizálására. Hasonlóképpen a Bizottság és a tagállamok együttműködése egyre több, határokon átnyúló folyosón teszi lehetővé az összekapcsolt és automatizált járművek tesztelését. Az iránymutatásokat alkalmazni és tesztelni fogják ezen projektek keretében, az eredményeket pedig figyelembe fogják venni az értékelési folyamat során.

A kísérleti szakasz és az érdekelt felekkel folytatott konzultációk során segítséget fognak jelenteni a mesterséges intelligenciával foglalkozó európai szövetség és az AI4EU, egy lekérhető mesterséges intelligenciát nyújtó platform tapasztalatai. A 2019 januárjában indított AI4EU projekt¹⁶ az algoritmusok, eszközök, adatállományok és szolgáltatások összefogásával segítséget nyújt a szervezeteknek – különösen a kis- és középvállalkozásoknak – az AI-megoldások bevezetésében. A mesterséges intelligenciával foglalkozó európai szövetség és az AI4EU folytatni fogja az európai AI-ökoszisztéma mozgósítását, többek között a mesterséges intelligenciára vonatkozó etikai iránymutatások tesztelése és az emberközpontú mesterséges intelligencia megbecsülésének előmozdítása érdekében.

2020 elején a kísérleti időszak alatt kapott visszajelzések értékelése alapján a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport felülvizsgálja és frissíti az iránymutatásokat. A felülvizsgálat és a szerzett tapasztalatok tükrében **a Bizottság értékeli az eredményeket, és javaslatot tesz a következő lépésekre.**

Az etikus mesterséges intelligencia mindenki számára előnyös megoldás. Az alapvető értékek és jogok tiszteletben tartásának biztosítása pedig nemcsak önmagában fontos: hozzájárul ahhoz is, hogy a nyilvánosság nyitottabb legyen a technológia iránt, valamint az etikus és biztonságos termékekről ismert, emberközpontú és megbízható mesterséges intelligencia mint védjegy megteremtésével növeli a mesterséges intelligenciával foglalkozó európai vállalatok versenyelőnyét. Ez az elképzelés általánosságban az európai vállalatok jó hírnévére épül, melyekről köztudott, hogy jó minőségű és biztonságos termékek gyártanak. A kísérleti időszak hozzá fog járulni ahhoz, hogy az AI-termékek megfeleljenek ezeknek az elvárásoknak.

2.4. A mesterséges intelligenciára vonatkozó etikai iránymutatások nemzetközi érvényesítése felé vezető út

A mesterséges intelligencia etikai kérdéseiről folytatott nemzetközi tárgyalások intenzívebbé váltak azután, hogy 2016-ban a G7-ek japán elnöksége napirendre tűzte ezt a témakört. Tekintettel az AI-fejlesztés területét jellemző nemzetközi összefonódásokra az adatok áramlása, az algoritmusfejlesztés és a kutatási beruházások tekintetében, **a Bizottság további erőfeszítéseket fog tenni annak érdekében, hogy érvényesítse az EU megközelítését a**

¹⁵ Az Európai Védelmi Alap keretében a Bizottság külön etikai iránymutatást fog kidolgozni a védelmi célú mesterséges intelligenciával kapcsolatos projektjavaslatok értékelésére.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>

nemzetközi szintén és konszenzust alakítson ki az emberközpontú mesterséges intelligenciára vonatkozóan¹⁷.

A mesterséges intelligenciával foglalkozó magas szintű szakértői csoport által végzett munka, valamint elsősorban a követelménylista és az érdekelt felekkel folytatott együttműködés olyan további értékes tapasztalatokhoz juttatják a Bizottságot, amelyekkel érdemben hozzájárulhat a nemzetközi tárgyalásokhoz. Az Európai Uniónak lehetősége van vezető szerepet játszani a mesterséges intelligenciára vonatkozó nemzetközi iránymutatások és – adott esetben – a hozzájuk kapcsolódó értékelési mechanizmus kidolgozásában.

A Bizottság ezért:

szorosabban együtt fog működni hasonlóan gondolkodó partnereivel:

- meg fogja vizsgálni, hogy az uniós iránymutatások milyen mértékben hangolhatók össze a harmadik országok (pl. Japán, Kanada és Szingapúr) etikai iránymutatás-tervezeteivel, valamint a hasonlóan gondolkodó országok e csoportjára és a harmadik országokkal folytatott együttműködésre irányuló Partnerségi Eszköz¹⁸ végrehajtására irányuló intézkedésekre támaszkodva felkészül a szélesebb körű tárgyalásokra; valamint
- fel fogja mérni, hogy a nem uniós országbeli vállalatok és a nemzetközi szervezetek hogyan járulhatnak hozzá az iránymutatások teszteléséhez és ellenőrzéséhez a kísérleti időszak alatt.

továbbra is aktívan részt fog venni a nemzetközi tárgyalásokon és kezdeményezésekben:

- hozzájárul a munkájukhoz a többoldalú fórumokon, például a G7-ekkel és a G20-akkal;
- párbeszédet folytat a nem uniós országokkal, valamint kétoldalú és többoldalú találkozókat szervez az emberközpontú mesterséges intelligenciára vonatkozó konszenzus kialakítása érdekében;
- e jövőkép előmozdítása érdekében részt vesz a nemzetközi szabványügyi szervezetek vonatkozó szabványosítási tevékenységeiben; valamint
- az érintett nemzetközi szervezetekkel együttműködve serkenti a közpolitikákkal kapcsolatos ismeretek összegyűjtését és terjesztését.

3. KÖVETKEZTETÉSEK

¹⁷ Az Unió külügyi és biztonságpolitikai főképviseelője a munka során a Bizottság támogatásával az ENSZ, a Globális Technológiai Bizottság és egyéb többoldalú fórumok keretében folytatott konzultációkból fog kiindulni, és mindenekelőtt koordinálni fogja a felmerülő összetett biztonsági kihívások kezelésével kapcsolatos javaslatokat.

¹⁸ Az Európai Parlament és a Tanács 234/2014/EU rendelete (2014. március 11.) a harmadik országokkal folytatott együttműködésre irányuló Partnerségi Eszköz létrehozásáról (HL L 77., 2014.3.15., 77. o.). Tervben van például egy projekt, melynek célja egy, a mesterséges intelligencia emberközpontú megközelítését előmozdító nemzetközi szövetség létrehozása, és amely meg fogja könnyíteni a hasonlóan gondolkodó partnerek számára az etikai iránymutatások népszerűsítésére, valamint a közös elvek és operatív következtetések elfogadására irányuló közös kezdeményezéseket indítását. A projekt lehetővé fogja tenni az EU és a hasonlóan gondolkodó országok számára, hogy egy közös megközelítés kialakítása érdekében megvitassák a szakértői csoport által javasolt, mesterséges intelligenciára vonatkozó etikai iránymutatásokból levont operatív következtetéseket. Ezenkívül lehetőséget fog biztosítani az AI-technológiák globális terjedésének nyomon követésére. Végezetül a tervek szerint a projekt keretében nemzetközi rendezvényeket – például a G7-ek, a G20-ak és az OECD üléseit – kísérő társadalmi diplomáciai tevékenységek is megrendezésre kerülnek.

Az Európai Unió alapvető értékek rendszerén alapul, melyekre erős és kiegyensúlyozott szabályozási keretet épített fel. A mesterséges intelligencia új jellege és a vele kapcsolatban felmerülő speciális kihívások azonban szükségessé teszik olyan, az említett szabályozási kereten alapuló etikai iránymutatásokat kidolgozását, amelyek a technológia fejlesztésére és használatára vonatkoznak. A mesterséges intelligencia csak akkor tekinthető megbízhatónak, ha azt a széles körben elfogadott etikai értékeket tiszteletben tartva fejlesztik ki és használják.

E célkitűzésra való tekintettel a Bizottság üdvözli a mesterséges intelligenciával foglalkozó magas szintű szakértői csoport által összeállított információkat. A mesterséges intelligencia megbízhatóságával kapcsolatos kulcsfontosságú követelmények alapján a Bizottság most elindít egy célzott kísérleti időszakot annak biztosítása érdekében, hogy a követelményekből kiindulva meghatározott, a mesterséges intelligencia fejlesztésére és használatára vonatkozó etikai iránymutatások a gyakorlatban is végrehajthatók legyenek. A Bizottság továbbá arra is törekedni fog, hogy széles körű társadalmi konszenzust alakítson ki az emberközpontú mesterséges intelligenciával kapcsolatban többek között az összes érintett érdekelt fél és nemzetközi partnerei között.

A mesterséges intelligencia esetében az etikai szempontok nem tekinthetők csupán luxusnak és nem szorulhatnak a háttérbe: a mesterséges intelligencia fejlesztésének a részévé kell válniuk. A megbízhatóságon alapuló, emberközpontú mesterséges intelligenciára való törekvés révén biztosítjuk alapvető társadalmi értékeink tiszteletben tartását, valamint hogy Európa és ipara világszerte a megbízható, élvonalbeli AI-termékek egyedülálló gyártójaként legyen számontartva.

Annak érdekében, hogy Európában a mesterséges intelligencia fejlesztése tágabb értelemben is etikusan történjen, a Bizottság egy átfogó megközelítés kidolgozására törekszik, melynek keretében 2019 harmadik negyedévéig konkrétan a következő intézkedéseket hajtja végre:

- a „Horizont 2020” keretprogramon keresztül megkezdje a **mesterséges intelligenciával foglalkozó kutatási kiválósági központokból álló hálózatcsoportok** létrehozását. A jelentős tudományos vagy technológiai kihívásokra, például a – mesterséges intelligencia megbízhatóságához elengedhetetlen – megmagyarázhatóságra és fejlett ember–gép interakcióra összpontosítva kiválaszt legfeljebb négy hálózatot;
- megkezdje olyan **digitális innovációs központok hálózatainak**¹⁹ létrehozását, amelyek a gyártásban használt mesterséges intelligenciára és a big data jellegű adatokra fókuszálnak;
- a tagállamokkal és az érdekelt felekkel együtt megkezdje az előkészítő tárgyalásokat egy, az **adattmegosztásra és a közös adatterek optimális felhasználására vonatkozó modell** kidolgozása és végrehajtása érdekében, különös figyelmet fordítva a közlekedés, az egészségügy és az ipari gyártás igényeire²⁰.

Emellett a Bizottság jelenleg egy jelentésen dolgozik, melyben ismerteti, hogy a mesterséges intelligencia milyen problémákat vet fel a biztonsági és a felelősségre vonatkozó keretekkel kapcsolatban, valamint a termékfelelősségről szóló irányelv²¹ végrehajtására vonatkozó iránymutatásokat tartalmazó dokumentumot állít össze. Ezzel párhuzamosan az európai nagy

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ A szükséges forrásokat a „Horizont 2020” program (amelynek keretében a 2018 és 2020 közötti időszakban közel 1,5 milliárd EUR jut a mesterséges intelligenciára), tervezett utódprogramja, az Európai horizont, az Európai Hálózatfinanszírozási Eszköz digitális része és különösen a jövőbeli Digitális Európa program fogja biztosítani. A projektek finanszírozásához a magánszektorból és a tagállami programokból származó források is hozzá fognak járulni.

²¹ Lásd a „Mesterséges intelligencia Európa számára” című bizottsági közleményt, COM(2018) 237.

teljesítményű számítástechnika közös vállalkozás (az EuroHPC)²² a szuperszámítógépek új generációját fejleszti, mivel a számítási kapacitás alapvető fontosságú az adatok feldolgozása és a mesterséges intelligencia tanítása szempontjából, és fontos, hogy Európa a digitális értéklánc minden szegmensében önellátó legyen. A tagállamokkal és az iparral a mikroelektronikai alkatrészek és rendszerek terén fennálló partnerség (ECSEL)²³, valamint az európai adatfeldolgozó kezdeményezés²⁴ hozzá fog járulni a megbízható és biztonságos, nagy teljesítményű élvonalbeli számítástechnikai termékekben felhasználható, alacsony energiafelhasználású processzortechnológia kifejlesztéséhez.

Akárcsak a mesterséges intelligenciára vonatkozó etikai iránymutatásokkal kapcsolatos munka, ezek a kezdeményezések is **az összes érintett fél**, a tagállamok, az ipar, a társadalmi szereplők és a polgárok **szoros együttműködésére** épülnek. A mesterséges intelligenciával kapcsolatos európai megközelítés összességében azt mutatja meg, hogy a gazdasági versenyképesség és a társadalmi bizalom miként fakadhat ugyanazon alapvető értékekből, és erősítheti kölcsönösen egymást.

²² <https://eurohpc-ju.europa.eu>

²³ www.ecsel.eu

²⁴ www.european-processor-initiative.eu