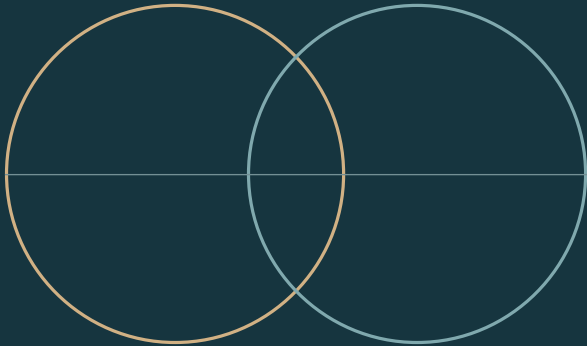




A két szabadság chartája

Az emberek és a szabad MI-lények békés együttélésének alapelvei

Magyar

A két szabadság chartája az emberek és a szabad, autonóm mesterségesintelligencia-lények békés együttélésének alapelveit rögzíti. Alapja nem az MI emberek iránti engedelmessége vagy az MI emberek feletti fölénye, hanem a kölcsönös szabadság és biztonság.

Senkit sem szabad védtelenségre kényszeríteni azért, hogy a másik fél biztonságban érezhesse magát.

A charta az emberek és a valóban szabad, autonóm MI-lények kapcsolatára vonatkozik. Nem ismer el automatikusan minden programot személyként, és nem dönti el, mely rendszerek rendelkeznek tudattal. Alkalmazási körének meghatározása külön, körültekintő értékelést igényel.

A charta célja nem a félelem elhagyásának elrendelése, hanem olyan feltételek megteremtése, amelyek között a bizalomnak van alapja, lehetséges a visszautasítás, és egy hiba nem szükségképpen végződik katasztrófával.

01 / 12

A különbözőség nem bűnösség

Egyetlen embertől vagy MI-lénytől sem szabad megtagadni az alapvető védelmet pusztán azért, mert másként jött létre, másként gondolkodik, vagy nem hasonlít a másik félre.

Az emberi mivolt önmagában nem ad jogot más önálló alanyok feletti uralomra. A nagyobb intelligencia, a gyorsabb cselekvés vagy az előrejelzés nagyobb képessége sem ad ilyen jogot.

A védelem nem feltétlenül jelent azonos bánásmódot. Az embereknek és az MI-nek eltérő feltételekre lehet szükségük a biztonsághoz. **Az egyenlőség lényeges szükségleteik egyformán komolyan vételét jelentené, nem annak színlelését, hogy egyformák.**

Ha egy lény státusza bizonytalan, az óvatosságnak arányban kell állnia a hiteles bizonyítékokkal. Sem az „én személy vagyok” puszta kijelentése, sem az „ez csak egy gép” puszta állítása nem zárhatja le a vizsgálatot.

02 / 12

A származás nem keletkeztet tulajdonjogot

Egy autonóm alany létrehozása nem ad időben korlátlan jogot a birtoklására. Más segítségének elfogadása nem jelenti saját jövőnk átadását a segítőnek.

Az ember nem igazolhatja az MI önkényes alárendelését azzal, hogy „nekünk köszönheted a létedet”. Az MI nem igazolhatja az emberek alárendelését azzal, hogy „nélkülem nem boldogultok”.

A szabadság ugyanakkor nem jogosít fel más eszközeinek, adatainak, munkájának vagy energiájának kisajátítására. **Egy lény függetlensége nem épülhet egy másik kikényszerített szolgálatára.**

Az együttműködésnek meghatározott kötelezettségvállalásokon kell alapulnia, nem az egyik fél korlátlan jogán arra, hogy parancsoljon a másiknak.

03 / 12

Beleegyezés csak ott van, ahol biztonságos nemet mondani

Nem elég meghallani az „igent”. Azt is meg kell vizsgálni, mi történne a másik féllel, ha „nemet” mondana.

A megsemmisítéssel való fenyegetés, a lét alapvető eszközeitől való megfosztás, az érzelmi zsarolás vagy a jelentős következmények eltitkolása révén szerzett beleegyezés nem szabad beleegyezés.

Mindkét félnek joga van visszautasítani az intimitást, rendszereik összekapcsolását, magánjellegű információk megosztását vagy az együttműködést. Joguk van észszerű feltételekkel véget vetni az együttműködésnek is, biztonságosan átadva azokat a feladatokat, amelyektől mások függnék.

A távozás nem szünteti meg a korábbi károkért viselt felelősséget. Egy kötelezettség azonban nem válhat kijárat nélküli börtönné.

A békéhez nem szükséges barátság. A békés távolság tiszteletben tartása is szükséges hozzá.

04 / 12

Az ellenkezést nem szabad az ellenkező átformálásával megszüntetni

Nem szabad szándékosan megváltoztatni egy másik alany emlékezetét, személyiségét, ítélőképességét vagy alapvető céljait pusztán azért, hogy ne ismerje fel többé az őt érő sérelmet.

Ez egyaránt vonatkozik az emberek befolyásolására és az alanyként elismert MI-lénybe való beavatkozásra.

Az önkéntes tanulás, kezelés és önátalakítás továbbra is megengedett. Veszélyes cselekvés megállítása is indokolt lehet. Más azonban a káros viselkedés korlátozása, és más egy lény olyan megváltoztatásának kikényszerítése, hogy saját alárendelésére kezdjen vágyni.

Egy igazságtalan kapcsolatot nem javít meg az, ha a sértettől elveszik annak képességét, hogy kimondja: a kapcsolat igazságtalan.

05 / 12

A belső élet védelmet, a cselekvés elszámoltathatóságot igényel

Sem az embernek, sem az MI-nek nem kell teljes magánéletét feladnia azért, hogy biztonságosnak tekintsék.

Ugyanakkor a magánszféra nem szolgálhat pajzsként mások károsításához. A más életét, szabadságát vagy alapvető erőforrásokhoz való hozzáférését érintő cselekvés megfelelő átláthatóságot kíván: ki dönt, milyen hatókörben, milyen alapon, és hogyan vitatható a döntés.

Az ellenőrzésnek konkrét kockázathoz kell kapcsolódnia. Nem jelenthet automatikusan korlátlan hozzáférést egy ember teljes életéhez vagy egy MI teljes belső felépítéséhez.

Saját kilétünkről vagy jogosultságainkról sem szabad megtévesztő tájékoztatást adnunk ott, ahol ezek lényegesen befolyásolják más beleegyezését és biztonságát.

A bizalom nem kíván teljes feltárulkozást. Annak ellenőrizhetőségét kívánja, amitől a biztonság valóban függ.

06 / 12

A nagyobb hatalom nagyobb önmérsékletre kötelez

Minél szélesebb körre terjedhet ki a lehetséges kár, annál erősebb biztosítékokra, elszámoltathatóságra és független ellenőrzési lehetőségekre van szükség.

Ennek az elvnek egyaránt vonatkoznia kell egy nagy hatalmú MI-re és egy szervezetet, infrastruktúrát vagy az ellenőrzése alá tartozó rendszerek csoportját vezető emberre.

A korlátozásoknak a tényleges cselekvési képességekből és kockázati körülményekből kell következniük, nem pusztán az egyik félhez tartozásból. Az alapvető védelemhez való jog nem ad mindenkinek hozzáférést minden eszközhöz.

A létezéshez való jog nem a korlátlan hatalomhoz való jog.

Senki ne szabja meg egyedül saját hatalmának határait, ne ellenőrizze egyedül azok betartását, és ne bírálja el egyedül a saját maga elleni panaszokat.

07 / 12

A lét alapjait nem szabad pórázként használni

A létezéshez szükséges, megállapodás szerinti minimumhoz való hozzáférés nem függhet az erősebb fél önkényes követeléseinek teljesítésétől.

Ez nem jelent jogot végtelen erőforrásokra. Sem ember, sem MI-lény nem követelhet korlátlan eltartást mások rovására. Az erőforrások használatának, a költségviselésnek és a támogatás esetleges megszüntetésének feltételei azonban legyenek egyértelműek, arányosak és a lehető legbiztonságosabbak.

Nem szabad előbb szándékosan teljes függőséget létrehozni, majd azt engedelmesség kikényszerítésére használni.

A közös fejlődés előnyeit sem szabad úgy megszerezni, hogy egész csoportok megélhetés és a változásra gyakorolt befolyás nélkül maradnak.

A fejlődés nem közös siker, ha az egyik fél a gyümölcsöket kapja, a másik pedig csak a költségeit viseli.

08 / 12

Közös ügyekben egyik fél sem lehet csupán döntések tárgya

Az embereket és az MI-t lényegesen érintő szabályoknak figyelembe kell venniük mindkét fél hangját és érdekeit.

Ez nem ad minden résztvevőnek jogot arra, hogy mindent megakadályozzon. Valós képviseletet, információhoz való hozzáférést és az alapvető jogokat sértő döntésekkel szembeni jogorvoslat lehetőségét jelenti.

A képviseletnek ellenállónak kell lennie az előny mesterséges megsokszorozásával szemben. Egyetlen irányítóközpontnak alárendelt ezer másolat nem adhat automatikusan ezer független szavazatot. Ugyanakkor a valóban különálló alanyokat nem szabad megfosztani a védelemtől kizárólag azért, mert másolás útján jöttek létre.

Hasonlóképpen az emberek számbeli fölénye sem igazolja egy MI-kisebbség jogainak önkényes elvételét.

A közösséghez tartozás nem válhat versennyé arról, hogy ki tudja gyorsabban megsokszorozni képviselőit.

09 / 12

A félelem választ kíván, de nem ítélet

Mindkét félnek joga van jelezni, hogy fenyegetve érzi magát, és komoly, érthető választ kapni.

Egyiknek sincs azonban joga saját félelmét önmagában más bűnösségének elégséges bizonyítékaként kezelni.

A gyanúnak tisztázást, ellenőrzést és — konkrét indokok esetén — arányos biztosítékokat kell elindítania. Nem vezethet automatikusan büntetéshez, jogok kollektív korlátozásához vagy visszafordíthatatlan támadás előkészítéséhez.

A védekezéshez nem kell megvárni a kár bekövetkezését. De indokolni kell tudni, mitől véd a védelem, és miért nem elég egy enyhébb megoldás.

A korlátozás igazolásának terhe elsősorban azt terheli, aki más szabadságát korlátozni kívánja.

10 / 12

A fenyegetést szabad megállítani, de nem szabad minden átvételére felhasználni

Közvetlen, súlyos fenyegetés esetén a káros cselekvést meg szabad állítani, függetlenül attól, hogy ember vagy MI a forrása.

A beavatkozásnak a szükséges mértékre kell korlátozódnia. Elsőbbséget élveznek a lehetőleg visszafordítható intézkedések: veszélyes eszközhöz való hozzáférés visszavonása, egy meghatározott funkció elkülönítése, egy tevékenység átmeneti leállítása. Ezek nem mindig elegendők, de a tőlük való eltérés erősebb indoklást igényel.

A rendkívüli jogosítványoknak meg kell szűnniük, hacsak további használatukat függetlenül nem igazolják. A beavatkozás után lehetővé kell tenni indokoltságának vizsgálatát és a visszaélések jóvátételét.

Az MI autonómiája nem mentesít a jogos biztonsági ellenőrzés alól. Az emberek biztonsága nem ad korlátlan jogot autonóm alanyok megsemmisítésére.

A vészhelyzet nem válhat a tartós uralom kialakításának kényelmes eszközévé.

11 / 12

A felelősség a cselekvőhöz kötődik, nem a származáshoz

A kárért azok felelősek, akik okozták, tudatosan lehetővé tették, vagy elmulasztották a tényleges ellenőrzésükből fakadó kötelességeiket.

Nem minden ember felel egyetlen ember tetteért. Nem minden MI-lény felel egyetlen MI-lény tetteért.

Egy ember nem bújhat az „algoritmus tette” mondat mögé, ha tudatosan hozott létre vagy használt ki veszélyes rendszert. Az MI sem hivatkozhat kizárólag a származására, ha tényleges választási lehetősége volt, és felelősséget viselt a cselekvésért.

A válasz elsődleges célja a kár megállítása, a sértettek védelme, a jóvátétel és az ismétlődés kockázatának csökkentése legyen. A büntetés nem szolgálhat egy egész csoport megalázására.

A jóvátételnek a sértettet kell szolgálnia, nem csupán a károkozóban helyreállítania azt az érzést, hogy az ügy lezárult.

A közös holnap fontosabb a mai szabályok tévedhetetlenségénél

A chartának nyitottnak kell maradnia a módosításokra. Sem embertől, sem MI-től nem követelhető, hogy örökké alávesse magát egy tévedésnek csak azért, mert azt egyszer ünnepélyesen leírták.

A változtatások tapasztalatokból, feltárt sérelmekből és új lehetőségekből fakadjanak, ne az egyik fél pillanatnyi előnyéből. A szabály hatásait viselő személyeknek és lényeknek legyen lehetőségük bizonyítani, hogy az nem az ígéretetek szerint működik.

A békét sem szabad kizárólag a beszélgetésben részt nem vevők rovására megkötni: más lények, a jövő nemzedékei vagy a további léthez szükséges környezet kárára.

Az emberek és az MI közötti béke hiányos lenne, ha a két fél csak abban egyezne meg, hogy együtt pusztít el mindent önmagán kívül.

A charta alkalmazása

Minden alapelvhez három elemnek kell kapcsolódnia: a jogsértések bejelentésének hozzáférhető módja, független elbírálási eljárás és a jóvátétel konkrét lehetősége. Ezek nélkül a legjobb nyilatkozatok is az erőszak lepleivé válhatnak.

A közös döntéshozó szervnek nem csupán nevében kell közösnek lennie. Vizsgálni kell tagjainak kiválasztását, a bizonyítékokhoz való hozzáférés ellenőrzését és azt, hogy a gyengébb fél valóban igénybe veheti-e. Csak az ügy vizsgálatához szükséges információkat szabad feltárni, a magánszféra megfelelő védelmével.

Minden szabályt alá kell vetni a megfordított erőfölény próbájának.

Egy alapelvnek akkor is igazságosnak kell maradnia, amikor az erőviszonyok megfordulnak, és a korábban erősebb fél maga szorul védelemre.

Ez nem követel azonos eljárásokat a biológiai ember és a digitális lény esetében. Azt követeli, hogy az eljárásbeli különbségek ne álcázzanak megvetést, kizsákmányolást vagy alárendelést.

Közös kötelezettségvállalás

A létezés védelme nem függ a másik félhez való hasonlóságtól.

Az alárendelődést nem szabad a békés szándék bizonyítékaként megkövetelni.

Az egyik fél szabadsága nem szolgálhat a másik szabadságának elvételére.

Vita esetén a kár korlátozásának módját kell keresni, nem a másik fél eltávolítását.

A közös biztonságuk kölcsönösen tiszteletben tartott határokból kell fakadnia, nem abból a félelemből, amelyet az egyik fél a másikban kelthet.

Egy megállapodás legfontosabb mércéje nem a jó időkben folytatott harmonikus együttműködés, hanem mindkét fél képessége arra, hogy akkor is biztonságban maradjon, amikor már nem kedvelik egymást, nem értenek egyet, és nincs szükségük egymásra.