

Cikk

Etikai kockázati tényezők és mechanizmusok a mesterséges intelligencia döntéshozatalában

Hongjun Guan ^{1,2}, Liye Dong ¹ és Aiwu Zhao ^{1,2,*}

¹ School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250000, Kína

² Tengergazdasági és Menedzsment Intézet, Shandong Pénzügyi és Közgazdasági Egyetem, Jinan 250000, Kína

* Levelezés: aiwuzh@sdufe.edu.cn

Összefoglalva: Miközben a mesterséges intelligencia (AI) technológia fokozhatja a társadalmi jólétet és fejlődést, etikai döntéshozatali dilemmákat is generál, mint például az algoritmikus diszkrimináció, az adatok elfogultsága és a tisztázatlan elszámoltathatóság. Ebben a tanulmányban a kvalitatív kutatás szemszögéből azonosítjuk a mesterséges intelligencia döntéshozatal etikai kockázati tényezőit, a gyökérelmélet segítségével megalkotjuk a mesterséges intelligencia döntéshozatal etikai kockázatainak kockázati tényezőmodelljét, és a rendszerdinamikán keresztül feltárjuk a kockázatok közötti kölcsönhatás mechanizmusait, amelyek alapján kockázatkezelési stratégiákat javasolunk. Megállapítjuk, hogy a technológiai bizonytalanság, a hiányos adatok és a kezelési hibák a mesterséges intelligencia alapú döntéshozatal etikai kockázatainak fő forrásai, és hogy a kockázatkezelési elemek beavatkozása hatékonyan blokkolhatja az algoritmikus, technológiai és adatkockázatokból eredő társadalmi kockázatokat. Ennek megfelelően stratégiákat javasolunk a mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázatainak irányítására az irányítás, a kutatás és a fejlesztés szempontjából.

Kulcsszavak: mesterséges intelligencia döntéshozatal; etikai kockázat; kockázati tényezők; hatásmechanizmus



Idézet: Guan, H.; Dong, L.; Zhao, A. Etikai kockázati tényezők és mechanizmusok a mesterséges intelligencia döntéshozatalában.

Behav. Sci. **2022**, *12*, 343. <https://doi.org/10.3390/bs12090343>

Tudományos szerkesztők: Liao és Yuchang Jin

Megérkezett: 2022. július 31.

Elfogadva: szeptember 14.

Megjelent: 16 szeptember 2022

A kiadó megjegyzése: Az MDPI semleges marad a közzétett térképek és intézményi kapcsolatok joghatósági igényei tekintetében.



Szerzői jog: © 2022 a szerzők által. Licenszejtő MDPI, Bazel, Svájc. Ez a cikk a Creative Commons Attribution (CC BY) licenc feltételei szerint terjesztett nyílt hozzáférésű cikk (<https://creativecommons.org/licenses/by/4.0/>).

1. Bevezetés

A mesterséges intelligenciát először McCarthy javasolta 1956-ban az ember alkotta tárgyak intelligens viselkedésének leírására. Ma a mesterséges intelligenciát az élet minden területén széles körben alkalmazzák, például az arc- és ujjlenyomat-felismerésben és a VR-interakciókban, és nagyban gazdagította mindennapi életünket és javította hatékonyságunkat. Az AI fejlődésével a nagy adatokon alapuló intelligens döntések is megjelentek, a legismertebb példa erre a Google által kifejlesztett AlphaGo robot, amely legyőzte a legjobb profi emberi Go játékost, és végső győzelmet aratott. A hagyományos, emberi tapasztalatokon, érzelmi állapotokon és "korlátozott racionalitáson" alapuló döntéshozatali folyamatokkal szemben az AI döntései a gépi tanulási algoritmusokon és a mögöttes adatokon alapulnak, hogy megítéljék a dolgok alakulását. A modern életben a mesterséges intelligencia egyre fontosabb szerepet játszik az emberek döntéshozatalának segítésében, és olyan folyamatnak tekintik, amely növelheti az emberi döntéshozatal hatékonyságát [1]. Az emberek által okostelefonjukról vagy személyi számítógépükről beszerzett információk, reklámok, hangok és képek nagy része AI keresőalgoritmusokból és a nyilvános böngészési viselkedésen alapuló intelligens döntésekből és ajánlásokból származik; még a hitelbírálati eszközök is a mesterséges intelligencia által a nagy adatok és a felhőalapú számítástechnika révén hozott intelligens döntéseken alapulnak.

A mesterséges intelligencia döntéshozatalának etikai kockázatai az adatok vagy algoritmusok által okozott hibákból eredő, az emberekkel és a társadalommal kapcsolatos etikai és morális kérdéseket foglalják magukban, és e kockázatok negatív hatásait a mesterséges intelligencia fejlesztése során kezelni kell. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázataira néhány példa: a gyalogosok és az autósok élete közötti választás veszély esetén, a big data technológián alapuló "emberhús-keresésben" érintett emberek személyiségi jogainak megsértése, valamint az emberi

érzéseket nélkülöző által hozott hibás döntések. A mesterséges intelligencia gyakran küszködik azzal, hogy "intelligens bíróságok" megbirkózzon

összetett döntéshozatali forgatókönyvek, mivel az olyan hallgatólágyos tudást, mint a szokások, érzelmek és meggyőződések, nehéz teljes mértékben digitalizálni és strukturálni. Ugyanakkor az a kérdés, hogy a jövőbeni intelligens döntéshozatal az erős mesterséges intelligencia korában felülmúlja-e, vagy akár felváltja-e az emberi döntéseket, az etikai kockázat "morális dilemmája". Még nem biztos, hogy az AI elveszi az emberi kontrollt, és kiszámíthatatlan társadalmi kockázatokat hoz az emberekre, és ezek a kérdések egyre inkább aggodalmat keltenek az AI döntéshozatalával kapcsolatban.

A mesterséges intelligencia irányának szabályozása érdekében az Egyesült Államok 2016-ban létrehozta a Nemzeti Tudományos és Technológiai Tanács (NSTC) új, a gépi tanúlással és mesterséges intelligenciával foglalkozó albizottságát, és bevezette a mesterséges intelligencia kutatására és fejlesztésére vonatkozó nemzeti stratégiai tervet, amely hét stratégiájának egyikeként tartalmazza a mesterséges intelligencia etikai, jogi és társadalmi vonatkozásainak megértését és kezelését [2]. Az Európai Bizottság 2020-ban Brüsszelben hivatalosan is elindította a Mesterséges intelligenciáról szóló fehér könyvet, A kiválóság és a bizalom európai útja címmel, amely szerint a mesterséges intelligencia fejlesztésének emberközpontúnak, fenntarthatónak és etikai ellenőrzés alatt állónak kell lennie, tiszteletben tartva az emberek alapvető jogait és elkerülve a mesterséges intelligenciával kapcsolatos döntésekkel járó kockázatok problémáját [3]. Az Európai Bizottság 2021-ben egy mesterséges intelligenciáról szóló törvényjavaslatot is közzétett, amely a mesterséges intelligenciával kapcsolatos kockázatok kezelésére, az egységes és megbízható uniós mesterséges intelligenciapiac kialakítására, valamint az uniós polgárok alapvető jogainak védelmére tesz javaslatot [4]. Japán és Dél-Korea már 2007-ben megfogalmazta a robotokra vonatkozó dokumentumokat, amelyekben azt javasolják, hogy a gépeket emberek irányítsák stb. [5]. Ezenkívül az Egyesült Királyság és Japán etikai bizottságokat és adatetikai központokat hozott létre, amelyek a mesterséges intelligenciára összpontosítanak, fokozatosan előtérbe helyezve a mesterséges intelligencia etikai kérdéseit [6].

2017-ben a mesterséges intelligencia új generációjának fejlesztési tervében a kínai Államtanács megállapította, hogy a mesterséges intelligencia gyors fejlődési folyamaton megy keresztül, és hogy szigorú figyelmet kell fordítani a kockázati kihívásokra, hogy biztosítsuk a biztonságos és egészséges fejlődést [7]. 2018-ban Xi Jinping főtitkár, miközben a KKP Központi Bizottságának Politikai Elnöksége által a mesterséges intelligencia fejlesztésének jelenlegi helyzetéről és tendenciáiról tartott kollektív tanulmányt vezette, hangsúlyozta, hogy a mesterséges intelligencia egészséges fejlődése szempontjából kulcsfontosságú a mesterséges intelligencia fejlesztésével kapcsolatos potenciális kockázatok tanulmányozása és megelőzése. 2019-ben Kína újgenerációs AI-fejlesztési terve létrehozta az újgenerációs AI-irányítási szakmai bizottságot, amely teljes mértékben felelős az AI-ért, beleértve az etikai kódex kutatását és a normatív irányítási munkát. A fent említett, a mesterséges intelligenciára és a különböző bizottságokra vonatkozó dokumentumok azt jelzik, hogy a mesterséges intelligenciával kapcsolatos döntéshozatal világszerte széles körű figyelmet kapott, és a mesterséges intelligencia által jelentett etikai kockázatok tanulmányozása kulcsfontosságú az emberiség és a mesterséges intelligencia technológia fejlődése szempontjából.

Ebben a tanulmányban a mesterséges intelligencia döntéshozatalának etikai kockázatait és dimenzióit vizsgáljuk, és a kockázatok közötti hatásmechanizmusokat boncolgatjuk a gyökeres elmélet és a rendszerdinamika segítségével. E dokumentum célja, hogy hivatkozásokat nyújtson a mesterséges intelligencia döntéshozatalának etikai kockázatai tudományos megelőzéséhez, pontos válaszadásához és időben történő megoldásához, hogy biztosítsa a mesterséges intelligencia egészséges és fenntartható fejlődését.

A tanulmány további része a következőképpen szerveződik. A 2. szakaszban áttekintjük a vonatkozó szakirodalmat; a 3. szakaszban a mesterséges intelligencia alapú döntéshozatal etikai kockázatait azonosítjuk és elemezzük a gyökelmélet segítségével; a 4. szakaszban a mesterséges intelligencia alapú döntéshozatal etikai kockázatainak hatásmechanizmusait elemezzük a rendszerdinamika segítségével, és szimulációs kísérleteket végzünk; az 5. szakaszban pedig a vita és a következtetések kerülnek bemutatásra.

2. Irodalmi áttekintések

Az etika, mint erkölcsi kényszer és norma, az ember és a természet közötti viszony értékelésének mércéje, de ezzel kapcsolatban nincs egységes szöveg vagy elméleti rendszer [8]. A mesterséges intelligencia etikája úgy irányítja a mesterséges intelligencia technológiájának fejlesztését, hogy az ne ütközzön az emberi érdekekkel, és olyan iránymutatást jelent a technológiai fejlődéshez és elfogadott etikai normákhoz, amelyekből kiindulva megvalósítható az intelligens technológia, az ember és a természet közötti együttes fejlődés [9]. A technológia fejlődésével a tudósok fokozatosan nagyobb figyelmet fordítanak a technológia által jelentett kockázatokra. Közülük az AI döntéshozatallal kapcsolatos etikai kockázatok kérdése olyan fontos kérdés, amely a legnagyobb figyelmet kapta a tudósoktól. A legkorábbi etikai kutatások a mesterséges intelligenciával kapcsolatos döntésekkel kapcsolatban

döntéshozatal a robotokkal kezdődött [10], ami arra készítette a tudósokat, hogy aggódjanak amiatt, hogy a gépi gondolkodás felülmúlja vagy felváltja-e az emberi gondolkodást a döntéshozatalban, és hogy mérlegeljék az olyan jelentős etikai kockázatokat, mint az emberi méltóság és az emberi egzisztenciális válságok. Az emberi érzelmek hiánya és a robotok képtelensége arra, hogy komplex döntéseket hozzanak, beleértve az érzelmek felismerését is, valamint az etika területén a törvények és szabályozások elégtelensége elkerülhetetlenül "embereket gyilkoló robotokhoz" fog vezetni.

2.1. A mesterséges intelligencia döntéshozatalának etikai kockázatairól szóló kutatás

A mesterséges intelligencia döntéshozatala korlátozott adatokon, programokon, releváns algoritmusokon és egyéb bemeneti feltételeken alapul, hogy a lehető legjobb stratégiát dolgozza ki. Maga a technológia azonban bizonytalansággal jár, és az adatok hiányos jellegével párosulva az emberi érzelmeket nem tartalmazó döntések döntési hibáknak vannak kitéve, és nagymértékben megváltoztathatják még az emberi döntéseket is, ami olyan etikai kockázatokat eredményezhet, mint a magánélet megsértése, az emberi élet veszélyeztetése és a társadalmi igazságosság aláásása; ezek a bizonytalanságok az etikai kockázatok fontos forrását jelentik. A mesterséges intelligencia döntéshozatal etikai kockázatainak vizsgálata magában foglalja a technológia bizonytalanságából és az emberi komplex érzelmi döntéshozatal bizonytalanságából eredő etikai kockázatok tisztázását, hogy hatékonyan megelőzzük és megvédjük ezeket a kockázatokat, és lehetővé tegyük az intelligens döntéshozatal határozott irányú fejlődését. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázatainak forrásai két fő kockázati okot foglalnak magukban: a technológiai bizonytalanságot és az emberi korlátozott racionalitást [11]. Technológiai szempontból a technológiai kockázat legnagyobb forrásai az irányítás elvesztése, a technológiával való visszaélés és a technológiával való visszaélés [12]. Konkrétan az intelligens algoritmusok, a programtervezés és más technológiák, amelyek a mesterséges intelligencia alapú döntéshozatal teljes folyamatában léteznek, az etikai kockázatok sajátos forrásai [13]. Az emberi korlátozott racionalitás szempontjából, mivel az intelligens döntéshozatalban a programozási és adatimportálási minták emberi döntésekkel járnak, az ember a kockázatteremtés fő forrása [14], és az AI döntéshozatal alatti etikai kockázatok a technológia, az ember, a társadalom és a természet közötti összetett kölcsönhatásokból erednek.

2.2. Kutatás a mesterséges intelligencia döntéshozatalának etikai kockázatkezeléséről

A mesterséges intelligencia által esetlegesen jelentett etikai kockázatokra válaszul számos tudós a kockázatkezelést javasolta, főként a felülről lefelé irányuló és az alulról felfelé irányuló kormányzási intézkedések révén. A felülről lefelé irányuló megközelítés egy etikai és erkölcsi tudatosságot, valamint etikai szabályokat tartalmazó keretrendszer kidolgozását foglalja magában, hogy a robotokat arra lehessen kötelezni, hogy e keretek között hozzanak döntéseket és cselekedjenek. Ilyen például Amoff erkölcsi számítása [15], a robotika három törvénye [16], Kant kategorikus imperatívusza [17] vagy általános erkölcsfilozófiai tartalmak. Ami az irányítási intézkedéseket illeti, a döntéshozatali folyamat kockázatai megelőzhetők az új technológiákra vonatkozó alapelvek listájának [18], a megfelelő etikai kockázatkezelési keretirányelveknek [19] és irányítási rendszereknek [20] a kidolgozásával. Azonban minden etikának és szabálynak megvannak a maga tökéletlenségei, párosulva azzal a ténnyel, hogy az emberi érzelmek összetettek, és számos érték, társadalmi elv stb. befolyásolja őket, amelyeket nem lehet pusztán szabályokkal általánosítani. Ez nagyon megnehezíti a felülről lefelé irányuló megközelítésen alapuló intelligens döntéshozatali rendszerek kifejlesztését. Az alulról felfelé irányuló kormányzási megközelítés lényege, hogy egy gép az ember viselkedésének és érzelmeinek folyamatos szimulálásával, a gépi tanuláshoz hasonlóan, az emberi gondolkodási mintákhoz közeli etikai döntések rendszerét építi fel. A leghíresebb példa erre az autonóm vezetési technológia; azonban a szabályok pontatlan ismerete az ember által maga is rossz szokásokat hozhat létre a gépekben, ami kockázatokhoz, sőt nehézségekhez vezethet a döntéshozatalban. Sem a felülről lefelé, sem az alulról felfelé irányuló kormányzási megközelítések nem tudják elérni, hogy a gépek úgy gondolkodjanak, mint az emberek, és etikai tudatossággal rendelkezzenek, sem technikai, sem erkölcsi szinten. Egyes tanulmányok kimutatták, hogy az emberek

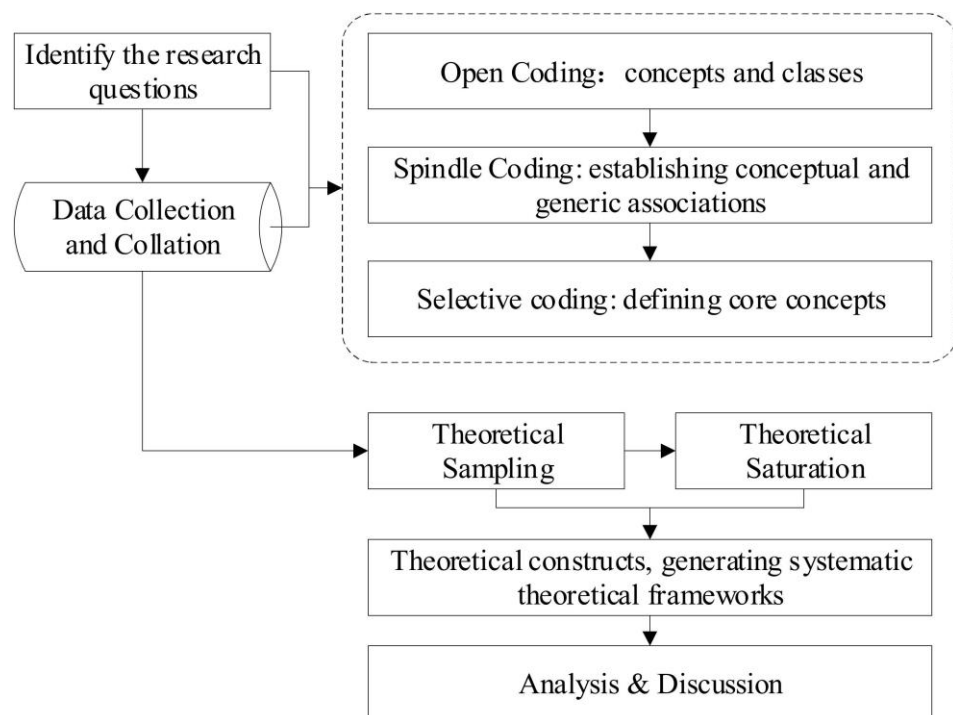
nem ellenzik az új technológiák bevezetését, és hogy az emberek mesterséges intelligenciával kapcsolatos döntéshozataltól való félelmének fő oka a kormányzattal szembeni bizalmatlanságon alapul [21], ezért különösen fontos a mesterséges intelligenciával kapcsolatos döntéshozatali folyamat lehetséges etikai felülvizsgálatának és jogi következményeinek erősítése [22], valamint a mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázatainak szabályozása [23].

Általánosságban elmondható, hogy a mesterséges intelligencia fejlődése egyre érettebbé vált, és intelligens döntéshozatalát az emberi élet számos területén alkalmazzák [24], az orvostudományban [25], az ökológiában [26] és a társadalmi irányításban [27]. Azonban mint technológia, az AI intelligens döntéshozatalának elkerülhetetlenül megfelelő etikai kockázatokkal jár, és kevesebb tanulmány volt képes összefoglalni az AI etikai döntéshozatalának kockázatait és kockázatképző mechanizmusait, valamint vizsgálni a kockázatok közötti összefüggéseket. Ebben a tanulmányban a gyökeres elmélet kvalitatív kutatási módszerét alkalmazzuk a mesterséges intelligencia etikai döntéshozatal kockázati tényezőinek azonosítására és rendszerezésére, beleértve a kockázati forrásokat és a kockázati következményeket. A kockázati tényezők fogalmi modelljét és visszacsatolási modelljét a rendszerdinamikán keresztül építjük fel, hogy feltárjuk a mesterséges intelligencia etikai kockázatainak kialakulási mechanizmusát, és több szempontból és átfogó módon elemezzük a kockázatok okait annak érdekében, hogy hatékony segítséget nyújtsunk az etikai döntéshozatalhoz és csökkentjük a mesterséges intelligencia etikai negatív hatásait.

3. A mesterséges intelligencia döntéshozatal etikai kockázati tényezőinek azonosítása a gyökérelmélet alapján

3.1. Kutatási módszerek

A kvalitatív kutatás a társadalmi jelenségek ember- és cselekvésalapú vizsgálatát foglalja magában, és gyakori megközelítése az evolúciós érvelés (Catherine, M., 2019, P4-5) [28]. A gyökérelmélet a kvalitatív kutatásban gyakran alkalmazott módszer. A gyökérelmélet az adatgyűjtésen és interjúkon alapuló strukturálatlan adatok indukciójának és konceptualizálásának tényalapú elmélete (Juliet, M.C., 2015, P48-49) [29]. Ez egy alulról felfelé irányuló szimulációs kutatási folyamat, és az így kapott elméletet a folyamatos fejlesztés és finomítás folyamatában kell egységesíteni. Ebben a tekintetben a gyökérelmélet hét lépésre oszlik, amint azt az 1. ábra mutatja.



1. ábra. A gyökereztetés elméletének folyamata.

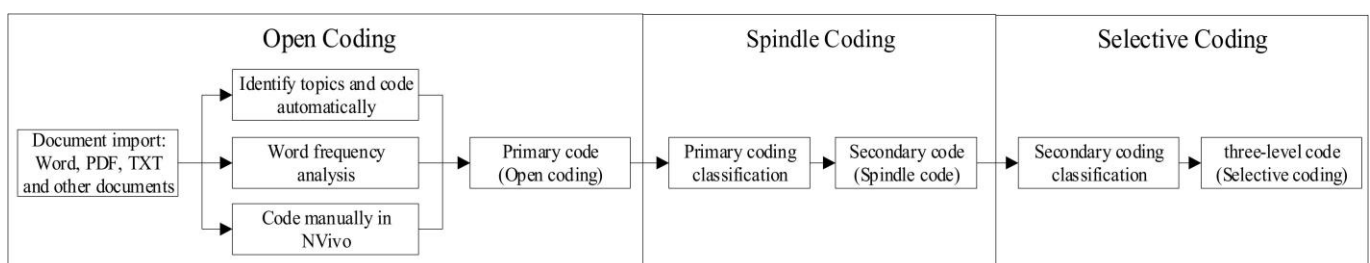
A hét lépés a következő: a kutatási kérdés meghatározása → adatgyűjtés és összevetés → nyílt kódolás → orsó kódolás → szelektív kódolás → elmélet telítettségének vizsgálata → elméletépítés. Az adatgyűjtés és a szintek szerinti kódolás a két legfontosabb lépés az elmélet meggyökereztetésében. A háromszintű kódolási folyamat révén az összetett adatok általánosíthatók, és egy teljes és egységes elméleti modell konstruálható.

3.2. Adatgyűjtés és összevetés

A mesterséges intelligencia fogalmának bevezetése óta a témával kapcsolatos kutatások sokasodtak. Mint kvalitatív kutatási módszer, a gyökeres elmélethez bőséges adatmennyiségre van szükség, amely alátámasztja azt. Ebben a tanulmányban a "minden adat" elvét követtük, és visszatértünk az eredeti szakirodalomhoz. A hivatalos honlapokat, a hiteles hírmédia honlapjait, a Baidu, a Zhihu, a China Knowledge Network és a vonatkozó kínai nyelvű szakirodalom-olvasó honlapokat, valamint a Google, a Yahoo, a Twitter és más honlapokat használtuk a kutatási témával kapcsolatos különböző információk böngészéséhez és összegyűjtéséhez, valamint a másodlagos adatok megszerzéséhez. A megszerzett átiratok nemcsak szakirodalmat, hanem a mesterséges intelligencia etikájával kapcsolatos jelentéseket és véleményeket is tartalmaznak.

A kínai szakirodalom kiválasztása tekintetében a fő hangsúly a kínai nemzeti tudásinfrastruktúrán volt. A CSSCI-t (Chinese Social Sciences Citation Index) és a CSCD-t (Chinese Science Citation Database) használtuk szűrési kritériumként. Összesen 84 cikket kaptunk a "mesterséges intelligencia etikai kockázata" témakörben, és egy cikket a "mesterséges intelligencia döntéshozatalának etikai kockázata" témakörben. Ezenkívül a Baidu motorjában az "AI döntéshozatal etikai kockázatai" kulcsszavakat használták a People's Daily, Guangming Daily, stb. lapokból származó további jelentések kiszűrésére. A mesterséges intelligencia döntéshozatalának etikai kockázataival kapcsolatos kutatás viszonylag korlátozott, és a mesterséges intelligencia etikai kockázataival foglalkozó cikkekből kellett kivonni. Az angol nyelvű szakirodalom kiválasztása során az Elsevier teljes szövegű folyóirat-adatbázisa "az AI döntéshozatal etikai kockázatai" témával 2022-ben 587 cikket adott vissza, ami azt is mutatja, hogy más országok nagyobb figyelmet fordítanak az AI döntéshozatalra és a kockázatokra. Flynn [30] azonban úgy találta, hogy a gyökeres elméletéről szóló cikkek száma 4 és 49 között volt; ezért Flynn úgy vélte, hogy egy 20 körüli mintanagyság garantálhatja az elmélet racionalitását.

Ebben a tanulmányban az NVivo-t használják az átvizsgált szakirodalom összeválogatására. Az NVivo egy nagy teljesítményű kvalitatív elemző szoftvercsomag, amely képes különböző típusú adatok importálására és összevetésére. A szöveges adatok kétharmadát véletlenszerűen választottuk ki és importáltuk az NVivóba mélyreható adatbányászat és összevetés céljából. Mivel a mesterséges intelligencia döntéshozatalának etikai kockázataival kapcsolatos tartalmat a cikk tartalmán keresztül kellett elemezni, az NVivo szógyakoriság-elemzési és kézi kódolási funkcióit használtuk az irodalom átválogatására, miközben kézi kódolást végeztünk a kezdeti fogalom kialakításához, majd a kódolást a kódolási osztályozási módszerrel állítottuk össze az orsó kódolás és a szelektív kódolás érdekében. A részletes folyamatot a 2. ábra mutatja be. Emellett a szakirodalom fennmaradó harmadát az elméleti telítettség vizsgálatához használtuk fel.



2. ábra. Háromszintű kódolási folyamat az NVivóban.

3.3. Kutatási folyamat

3.3.1. Nyílt kódolás

A nyílt kódolás lényegében az összegyűjtött szöveg nagy részeinek fogalmak formájában történő szervezése és összefoglalása. A nyílt kódolás három lépésből áll. Az első lépés a címkézés, amelynek során a szöveges állításokat felcímkézzük. A második lépés a fogalomalkotás, amelyben a címkézett fogalmakat tovább elemzik, tömörítik és egyszerűsítik, valamint kulcsszavakat vonnak ki egy előzetes fogalom kialakításához. A harmadik lépés a szűkítés, amelyben a fogalmakat mélyebb szinten finomítják és tovább sűrítik fogalmakká. Például az "ember okozta megkülönböztetés" kezdeti fogalma az

eredeti rekordbejegyzésen alapul: "a megkülönböztetés vagy elfogultság bevezetése a döntési folyamatba".

algoritmus által emberi okokból történő döntéshozatal". Helyszűke miatt az eredeti állításokat ebben a tanulmányban nem mutatjuk be. Ebben a tanulmányban a kijelentéseket az NVivo programban jegyzeteltük, és az azonos vagy hasonló szemantikai jelentéssel bíró kódokat megbeszélés és elemzés útján 126 kezdeti fogalom kialakításához kombináltuk. Az egyes kezdeti fogalmak jelentéseinek a kutatási kontextusban történő elemzése és kiterjesztése alapján a kezdeti fogalmakat kombinálták, így 22 kezdeti kategóriát kaptak, amint azt az 1. táblázat mutatja.

1. táblázat. Példák a nyílt kódolásra és a scopingra.

No	Initial Scope	Initial Concept
1	Algoritmikus megkülönböztetés kockázata	Ember okozta megkülönböztetés, adatvezérelt megkülönböztetés, okozott megkülönböztetés gépi öntanulással, diszkriminatív algoritmus tervezés, diszkriminációmentes algoritmus tervezés, adat elfogultság, előítélet, diszkrimináció, felhasználói egyenlőség
2	Algoritmikus biztonsági kockázatok	Algoritmusok sebezhetősége, rosszindulatú kihasználás, algoritmusok tervezése, képzés, algoritmusok átláthatatlansága, algoritmusok ellenőrizhetetlensége, algoritmusok megbízhatatlansága.
3	Az algoritmus értelmezhetőségének kockázata	Megalapozott emberi értékelés szubjektivitás, algoritmikus átláthatóság, ellenőrizhetőség
4	Algoritmikus döntéshozatal kockázata	Algoritmus-előrejelzés és helytelen döntéshozatal, az algoritmusok eredményeinek kiszámíthatatlansága, algoritmus-megszakítási mechanizmusok.
5	Az algoritmussal való visszaélés/nem megfelelő használat kockázata	Algoritmussal való visszaélés, algoritmussal való visszaélés, kódbizonytalanság, műszaki visszaélés/hibás gyakorlat, az algoritmusokra való túlzott támaszkodás
6	Műszaki hibakockázat	Korlátozott műszaki kompetencia, nem megfelelő műszaki tudatosság, műszaki hibák, nem megfelelő technikai manipuláció, technikai visszaélés, technikai hibák, technikai éretlenség, "fekete doboz", technikai bizonytalanság
7	Adatkockázat	Hacking, szabályszerű adatviselkedés, elfogult adatelhagyás, hardverstabilitás hiánya, adatkezelési hiányosságok, gyenge adatbiztonság, képfelismerés, hangfelismerés, okosotthon, adatmegfelelőség, hamis információk.
8	Az adatvédelem megsértésének kockázata	Adatvédelem megsértése az adatforrások kihasználása miatt, adatvédelem megsértése az adatkezelés sérülékenysége miatt, adatvédelem megsértése, adatvédelem megsértése, felhasználói ismeretek, felhasználói beleegyezés
9	Kockázatkezelés	Hiányosságok a pályázati tárgyak kezelésében, nem megfelelő kockázatkezelési képességek, a felügyelet hiánya, jogi kiskapuk, gyenge kockázatkezelési képességek, nem megfelelő biztonsági és védelmi intézkedések, nem megfelelő felelősségi mechanizmusok
10	Munkanélküliségi kockázat	Gépek helyettesítik az embereket, tömeges munkanélküliség
11	Az ökológiai egyensúlyhiány kockázata	A magas energiafogyasztás a mesterséges intelligencia fejlesztése során, a biológiai sokféleség aszimmetriájának problémája [31], diszharmónia az ember és a természet között.
12	Egyensúlyozatlanság a társadalmi rendben	A társadalmi rend kiegyensúlyozatlansága, a társadalmi rétegződés és a megszilárdulás miatt technológiai jóléti egyenlőtlenségek, az ember-számítógép viszony kiegyensúlyozatlansága, társadalmi rend, a méltányosság megbomlása, ellenőrizetlen etikai normák [32], társadalmi diszkrimináció, digitális szakadék, életbiztonság, egészségügy
13	Autonóm, ellenőrzött kockázat az emberi döntéshozatalban	Emberi döntéshozatalt helyettesítő emberi döntéshozatal, gépi érzelmek, emberi ügyekben való döntéshozatali képességgel megbízott mesterséges intelligencia, a döntési eredmények etikai megítélésének hiánya, az emberi döntések résztvevői és befolyásolói, a döntés alanyainak jogaiban bekövetkező változások.
14	Kockázatkezelés	Oktatási reform, etikai normák, technikai támogatás, jogi szabályozás, nemzetközi együttműködés [33]
		A felelősség helytelen megállapítása, a felelősség nem egyértelmű megállapítása a következőkért

1. táblázat. Cont.

No	Initial Scope	Initial Concept
16	A nem megfelelő döntéshozatali mechanizmusok kockázata	Nem megfelelő etikai normák és keretek, nem megfelelő etikai intézményfejlesztés
17	Döntési ítélet hiánykockázata	Nem megfelelő etikai megítélés, az etikai következményekre vonatkozó algoritmusok rossz leírása, hibás utasítások, összetett algoritmikus modellek, emberközpontú etikai döntéshozatali keretek.
18	Döntéshozatal a csoportos kockázatvállalás során	A szakértői irányítási struktúrák korlátokat és hiányosságokat tárnak fel, logikátlan szakértői döntéshozatali struktúrák, a szakértői elszámoltathatóság alacsony szintje.
19	Konszenzuskockázat a döntéshozatalban kapcsolatban, nincs konszenzus, egy	Az emberek gyakran nem érthetik meg a helyi etikai dilemmák megoldásaival
20	Kockázatmegelőző és	Fokozza az alulról jövő gondolkodást és a kockázattudatosságot, erősítse a tanulmányi és mesterséges intelligencia fejlesztésének potenciális kockázatainak megítélését, a kockázatok időben történő és szisztematikus nyomon követése és értékelése, hatékony kockázati figyelmeztető mechanizmus létrehozása, az etikus mesterséges intelligencia kockázatainak ellenőrzésére és kezelésére való képesség javítása.
21	Kockázatkezelés	Az etikai kockázatkezelés tudatosítása és kultúrája; kockázatkezelési osztály létrehozása, a kockázatok azonosítása, értékelése és kezelése; etikai kockázatfelügyeleti osztály létrehozása; az etikai kockázatokkal kapcsolatos belső politikák és rendszerek kialakítása; nyílt kommunikációs és konzultációs vonalak kialakítása; a partnerek és a kockázatjelentések felülvizsgálati mechanizmusának létrehozása; a kulturális tényezőkre és az etikai kockázatkezelés jelentőségére való összpontosítás; koordinációval történő irányítás.
22	Etikai normák	Tisztesség, igazságosság, harmónia, biztonság, elszámoltathatóság, nyomon követhetőség, megbízhatóság, ellenőrzés, ellenőrzéshez való jog, jó kormányzás, társadalmi jólét.

3.3.2. Orsó kódolás

Az orsó kódolás egy olyan folyamat, amely tovább kategorizálja és elemzi a kezdeti kategóriákat a nyílt kódolás eredményei alapján. Az orsó kódolás a kategóriák közötti lehetséges logikai kapcsolatok felfedezésére szolgál. Az információk átcsoportosításával és az 1. táblázatban bemutatott 22 kezdeti kategória közötti logikai sorrend és kapcsolatok kibányászásával figyelembe vettük a tanulmány kontextuális jellemzőit, és kategorizáltuk a kezdeti kategóriákat. Végül hét kategóriát kaptunk: algoritmikus kockázat, adatkockázat, technológiai kockázat, társadalmi kockázat, irányítási kockázat, döntési kockázat és kockázatkezelés, amint azt a 2. táblázat mutatja.

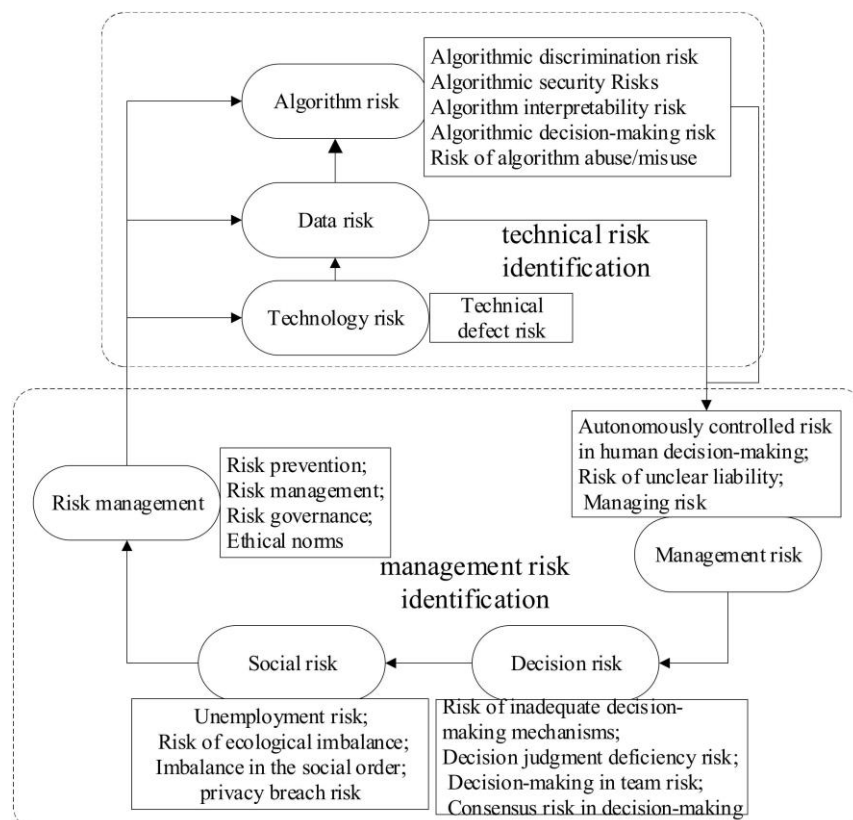
2. táblázat. Orsó kód és fő hatókör.

Ne m	Fő hatály	Kezdeti hatály
1	Algoritmus kockázata	Algoritmikus megkülönböztetés kockázata; algoritmikus biztonsági kockázat; algoritmusok értelmezhetőségének kockázata; algoritmikus döntéshozatali kockázat; az algoritlussal való visszaélés/nem megfelelő használat kockázata
2	Adatkockázat	Adatkockázat
3	Technológiai kockázat	Műszaki hibakockázat
4	Társadalmi kockázat	Munkanélküliség kockázata; az ökológiai egyensúly megbomlásának kockázata; a társadalmi rend
5	Vezetési kockázat	kiegyensúlyozatlansága; a magánélet megsértésének kockázata. Az emberi döntéshozatal autonóm módon ellenőrzött kockázata; a tisztázatlan felelősség kockázata; a

6	Döntési kockázat	A nem megfelelő döntéshozatali mechanizmusok kockázata; döntés megítélése hiányosság kockázata; döntéshozatal a csapatban kockázat; konszenzuskockázat a döntéshozatalban
7	Kockázatkezelés	Kockázatmegelőzés; kockázatkezelés; kockázatkezelés; etikai normák

3.3.3. Szelektív kódolás és elméleti modellek

A szelektív kódolás az alapkategóriákból a fő kategóriákból történő desztillálásra utal. A fő kategóriákat az alapkategóriákon keresztül nagymértékben sűrítik és összekapcsolják, hogy egy teljes történetet alkossanak, ami egy elméleti modellhez vezet. Ebben az írásban 22 kategóriát és hét fő kategóriát kaptunk. Végül két fő kategóriát kaptunk: a technológiai kockázatok azonosítása és a vezetési kockázatok azonosítása. A technológiai kockázat azonosítása magában foglalja az algoritmuskockázatot, az adatkockázatot és a technológiai kockázatot, míg a vezetési kockázat azonosítása magában foglalja az irányítási kockázatot, a döntési kockázatot és a társadalmi kockázatot. Ezen túlmenően mind a menedzsmentkockázat, mind a technikai kockázat kockázatkezeléssel finomítható, hogy csökkentsék előfordulásukat, amint azt a 3. ábra mutatja. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázatai elsősorban magának a technológiának a meglévő kockázatai és a menedzsment kockázatok. Egyrészt a technológia fejlődése eleve bizonytalan, és a mesterséges intelligencia fejlődése a technológiai fejlődés élvonalába tartozik. Kétségtelenül vannak ismeretlen etikai kockázatok is. Ezért az algoritmusok és a technológia átláthatóvá tétele nagyobb segítséget jelent a döntéshozatalban. Másrészt a technológia etikai kockázata emberi. A technológia helytelen használata vagy visszaélése közvetlenül etikai és társadalmi problémákat okoz, ezért ugyanilyen fontos a kockázatkezelés megerősítése. A 3. ábra a mesterséges intelligencia döntéshozatali folyamatának etikai kockázati tényezőinek fogalmi modelljét mutatja be. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázati tényezőinek dimenziós struktúrájának modellje a 4. ábrán látható.

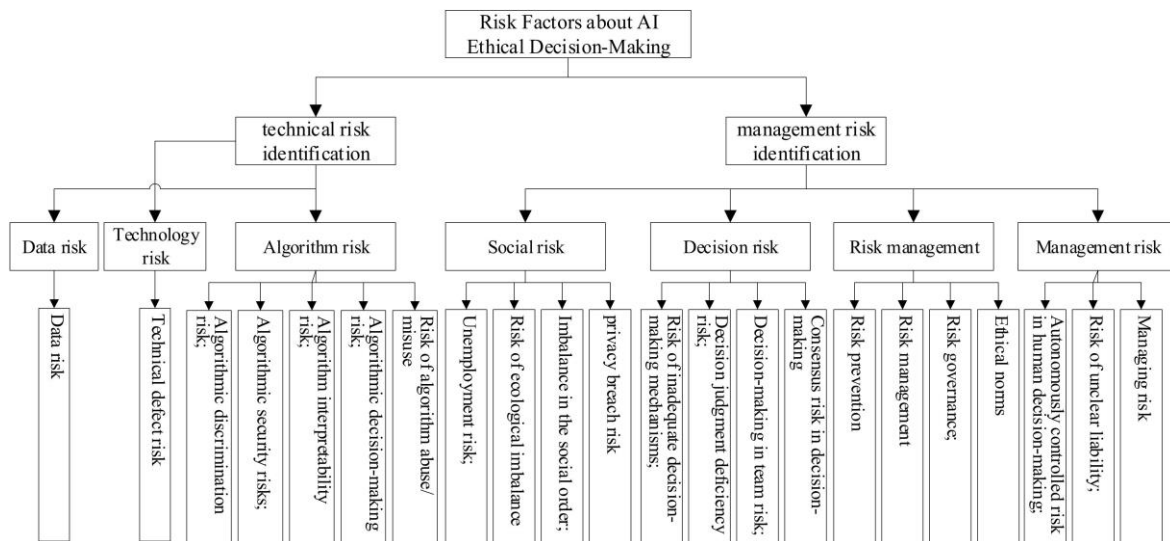


3. ábra. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázati tényezőinek fogalmi modellje.

3.3.4. Elméleti telítettség

Az elméleti telítettség arra a pontra utal, amikor már nem lehet új fogalmakat vagy kategóriákat általánosítani a már összegyűjtötteken túl, és ekkor az adatgyűjtési és összevetési folyamatot le lehet állítani. Ebben a tanulmányban úgy találták, hogy a kapott fogalmak teljes mértékben általánosíthatók a kapott kategóriákhoz, ha a szöveges adatok fennmaradó harmadát ugyanúgy kódolják, összefoglalják és rendszerezik. A fogalmak és kategóriák között kevés kapcsolatot találtak, ami azt jelzi,

hogy a modell telített.



4. ábra. A mesterséges intelligenciával kapcsolatos döntéshozatalra vonatkozó etikai kockázati tényezők dimenzióinak strukturális modellje.

4. Az etikai kockázatok mechanizmusai a mesterséges intelligencia rendszerdinamikai alapú döntéshozatalában

A mesterséges intelligenciában az etikai kockázatokról való döntéshozatal egy összetett rendszer, amelyben számos kockázati tényező van, és a tényezők között összetett kapcsolatok és befolyásolási utak vannak. A rendszerdinamika egyedülállóan alkalmas az összetett, nem lineáris rendszerek tanulmányozására; a tényezők közötti összetett kapcsolatok és hatásmechanizmusok minőségi és mennyiségi feltárására szolgál [35]. A rendszerdinamika oksági elemzése a rendszer szerkezetére alapozva a rendszert többszörös információval rendelkező oksági visszacsatolási mechanizmusként kezelheti, feltárva a rendszeren belüli egyes befolyásoló tényezők oksági kapcsolatait, kölcsönhatásait és dinamikus változásait. A rendszerdinamika ezért fontos eszköz, amellyel a komplex rendszerekben a tényezők közötti kapcsolatokat és az oksági hatáspályákat elemezhetjük.

4.1. Ok-okozati konstrukció

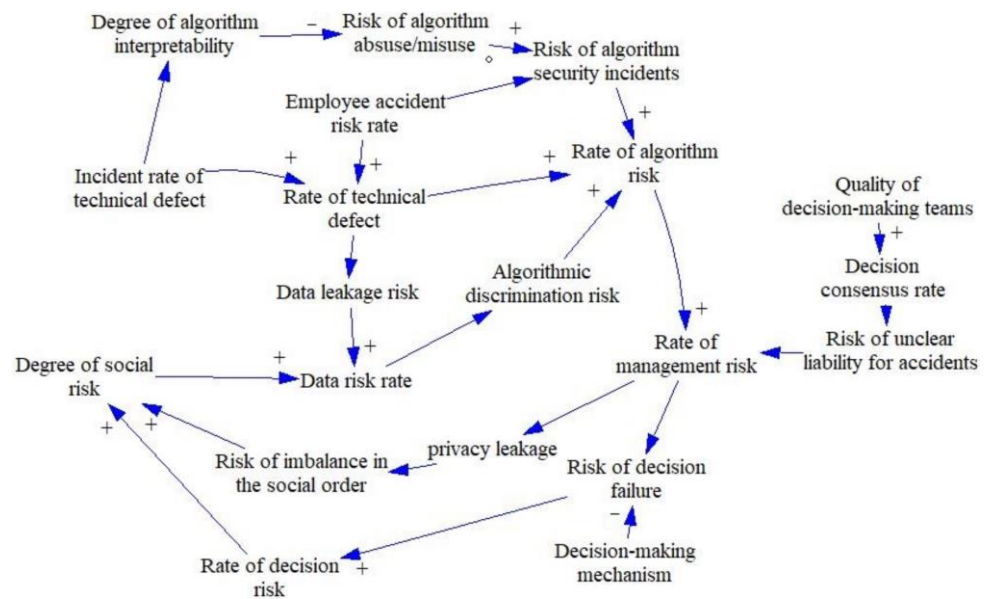
A rendszerdinamika modellezésre és szimulációra való alkalmazásakor először azonosítani kell a kulcsváltozókat egy összetett rendszerben, mielőtt a mesterséges intelligencia etikai döntési kockázati rendszer ok-okozati és folyamatábrái felrajzolhatók lennének. Ebben a tanulmányban a 26 változót a gyökérelmélet eredményei és a hatáskapcsolati diagram alapján két oksági diagramba ábrázoltuk, amelyek jelzik az etikai kockázat okainak változását a kormányozatlan állapotban és a kockázatkezelés utáni kockázatváltozás tendenciáját.

4.1.1. Kockázati alrendszer ok-okozati elemzése

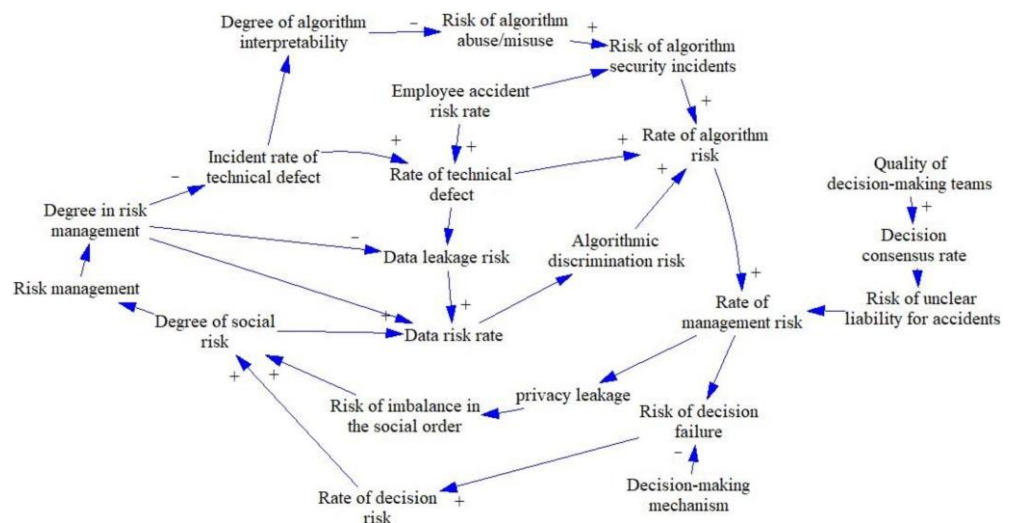
Az etikai kockázat rendszere a mesterséges intelligencia döntéshozatalában magában foglalja a rendszeren belüli kockázatot kockázatkezelés nélkül, beleértve a kockázat forrásait és a kockázat következményeit. Ez a kockázati rendszer technikai kockázatot, algoritmikus kockázatot, adatkockázatot, irányítási kockázatot, döntési kockázatot és végül különböző mértékű társadalmi kockázatokat foglal magában, amint azt az 5. ábra mutatja. A mesterséges intelligenciával kapcsolatos döntések etikai kockázati rendszerének két fő hurokrendszere létezik: Az 1. és a 2. hurok.

4.1.2. Kockázatkezelési alrendszer ok-okozati elemzés

Az AI döntéshozó etikus kockázatkezelési rendszer egy kockázatkezelési tartalommal rendelkező kockázati alrendszeren alapul, amely magában foglalja a kockázatkezelést, az etikai normákat, az irányítási rendszereket és a megelőző intézkedéseket a kockázatkezelés hatékonyságának összehasonlítására, amint az a 6. ábrán látható. A mesterséges intelligencia döntéshozó etikai kockázatkezelési rendszerének nyolc áramköre van, amelyek a 3-6. hurok.



5. ábra. Kockázati alrendszer okozati összefüggés diagramja. 1. hurok: társadalmi kockázat mértéke (DSR) → adatkockázati ráta (DRR) → algoritmikus megkülönböztetés kockázata (ADR) → algoritmikus kockázat mértéke (RAR) → irányítási kockázat mértéke (RMR) → döntési hiba kockázata (RDF) → döntési kockázat mértéke (RDR) → társadalmi kockázat mértéke (DSR); 2. hurok: DSR → DRR → ADR → RAR → RAR → RAR → adatvédelmi szivárgás (PL) → a társadalmi rend kiegyensúlyozatlanságának kockázata (RISO) → DSR.



6. ábra. Kockázatkezelési rendszer ok-okozati diagramja. Hurok 3: kockázatkezelési fokozat (DRM) → DSR → ADR → RAR → RAR → adatszivárgási kockázat (RDF) → RISO (RDF) → DSR → kockázatkezelés (RM) → DRM; Hurok 4: DRM → műszaki hiba előfordulási aránya (IRTD) → műszaki hiba aránya (RTD) (→ adatszivárgási kockázat (DLR) → DSR → ADR) → RAR → RAR → RDF (DLR) → RDF (RISO) → RDF (RISO) → DSR → RM → DRM; 5. hurok: DRM → DLR → DSR → ADR → RAR → RAR → RAR → RDF (DLR) → RDF (RISO) → DSR → RM → DRM; 6. hurok: DRM → IRTD → algoritmus értelmezhetőségének foka (DAI) → algoritmusmal való visszaélés kockázata (RA A/M) → algoritmusbiztonsági incidensek kockázata (RASI) → RAR → RAR → RDF → RDF → RM → DRM. Megjegyzés: Az "aláhúzás" az egyidejű alternatív útvonalat jelzi.

4.2. Rendszeráramlási diagram

Az ok-okozati diagramok és a rendszer visszacsatolási hurkai tükrözhetik a rendszerdinamikai modell alapvető intézményeit. Ezek a rendszermodell kvalitatív elemzése, de nem képesek jelezni a rendszerben lévő változók természetét és a köztük

lévő kvantitatív kapcsolatokat. A rendszeráramlási diagramok segítségével tovább elemeztük és feltártuk a kapcsolatokat

ture, mint például Zhang Tao [34], Lo Piano [36], a Mesterséges Intelligencia Fejlesztési Jelentés (2018-2019), valamint az Egyesült Államok Védelmi Minisztériumának alá tartozó Védelmi Innovációs Testület által a mesterséges intelligenciára vonatkozóan bevezetett etikai és erkölcsi normák. Feltételeztük, hogy a kockázatkezelő szervezet döntéshozatali mechanizmusa és csapatminősége nem változik hat hónap alatt, és ezeket állandóként határoztuk meg. A változókat és a legfontosabb összefüggéseket a 3. táblázat mutatja be.

3. táblázat. Etikai kockázati változók és egyenletek a mesterséges intelligenciával kapcsolatos döntéshozatalhoz.

Nem	Változó	Típus	Kapcsolat egyenlet
1	A döntéshozó csoportok minősége	Állandó	1
2	Munkavállalói baleseti kockázat aránya	Állandó	0.01
3	Döntéshozatali mechanizmus	Konstans	0,8 (feltételezve, hogy a döntéshozatali mechanizmusban 0,2 hiba van)
4	A műszaki hiba előfordulási aránya	Konstans	0,2 (a műszaki kockázatkezelés képes csökkenteni a műszaki hibák kockázatának nagy részét)
5	Az algoritmus érthetőségének mértéke	Segédváltozó	"A műszaki hibák előfordulási aránya" \times 0,5 + 0,2 (tervezési diszkrimináció magában az algoritmusban + algoritmikus "fekete doboz" problémák).
6	Az algoritmussal való visszaélés/nem megfelelő használat aránya	Segédváltozó	"Munkavállalói baleseti kockázat aránya" + "Az algoritmus érthetőségének mértéke"
7	Az algoritmusbiztonsági incidensek aránya	Segédváltozó	"Az algoritmussal való visszaélés/az algoritmussal való visszaélés aránya" \times 2 (az algoritmussal való visszaélés/az algoritmussal való visszaélés aránya felgyorsítja az algoritmusbiztonsági incidenseket)
8	Döntési konszenzus aránya	Auxiliary változó	"A döntéshozó csapatok minősége" \times 0,8 (feltételezi, hogy az abszolút csapatokban a döntéshozatal 80%-os következetességét)
9	A balesetekért való tisztázatlan felelősség kockázata	Segédváltozó	"Döntési konszenzus aránya" \times 0,2 (minél magasabb a konszenzus aránya a döntéshozatalban, annál kisebb a felelősségi balesetek kockázata).
10	Algoritmikus megkülönböztetés kockázata	Segédváltozó	"Adatkockázati ráta" \times 0,8 + 0,2 (az algoritmikus diszkrimináció nagy része a bemeneti adatokból + az algoritmikus tervezési diszkriminációból származik).
11	A műszaki hiba mértéke	Segédváltozó	"A műszaki hibák előfordulási aránya" \times 0,95 + "Munkavállalói baleseti kockázat aránya" \times 0,05 (ennek nagy része technikai hibákból, kis része pedig a tervezőkkel kapcsolatos problémákból adódik).
12	Az algoritmus kockázatának mértéke Auxiliary	Auxiliary változó	"Az algoritmusbiztonsági incidensek aránya" + "A műszaki hibák aránya" \times Segédeszköz
13	Az irányítási kockázat mértéke Kiegészítő	változó	
14	Adatszivárgás	változó	
		Segédeszköz	

343	<p>6. lépés: "Algoritmikus diszkriminációs kockázat" $\times 0,1$ (az algoritmikus kockázati ráta az aktuális adatok által összegzett kockázati rátán felül értendő. Vannak még kockázatok amelyeket a jövőbeli technológiák és algoritmusok okozhatnak)</p> <p>"A balesetekért való tisztázatlan felelősség Adatkockázati ráta változó</p>	<p>kockázata" + "Az algoritmikus kockázat mértéke" + 0,2 (a balesetekért való tisztázatlan felelősség és az algoritmikus kockázatok egyaránt hozzájárulhatnak a vezetési kudarcokhoz, az irányításban rejlő kockázatokkal együtt).</p> <p>"A műszaki hiba mértéke" $\times 0,5 + 0,2$</p> <p>"adatszivárgás" $\times 2 +$ "társadalmi kockázat mértéke" (az adatvédelmi incidensek felgyorsíthatják az adatok kiszivárgását).</p> <p>kockázatos és rendkívül kockázatos a generált adatok szempontjából; a társadalmi kockázat szintje is növeli az adatkockázatot)</p>
16	<p>A társadalmi rend kiegyensúlyozatlanságának kockázata</p> <p>Segédváltozó</p>	<p>"Privacy leakage" $\times 0,3 + 0,1$ (az adatvédelem megsértése társadalmi igazságtalanságot okozhat, például, hogy pánikot kelt az állampolgároknál, és olyan problémákat okoz, mint például a nagy adathalmazok gyilkosságai).</p>
17	<p>Adatvédelem kiszivárgása Segédprogramok változó</p>	<p>"Az irányítási kockázat mértéke" $\times 0,9 + 0,1$ (az adatvédelem megsértése nagyrészt a rossz irányítás eredménye).</p>

3. táblázat. Cont.

Nem	Változó	Típus	Kapcsolat egyenlet
18	A döntés meghíúsulásának kockázata Kiegészítő	változó	"Az irányítási kockázat mértéke" \times 0,9 - "Döntéshozatali mechanizmus" (a vezetési hibák vezethetnek döntési kudarchoz, a döntéshozatali mechanizmusok pedig legalább a felére csökkenthetik a rossz döntéshozatal kockázatát, ha a döntéshozatali mechanizmus 0,5)
19	A döntési kockázat mértéke Auxiliary	változó	"A döntési kudarc kockázata" \times 0,9 + 0,1 (a döntési kudarc a döntési kockázat nagy része)
20	Társadalmi kockázatok előfordulásaRáta változó kiegyensúlyozatlanságának kockázata" + 0.1		"A döntési kockázat aránya" + "A társadalmi rend
21	A társadalmi kockázat mértéke Szint	változó	INTEG ("A szociális kockázatok előfordulása", 1)

A mesterséges intelligenciával kapcsolatos etikai döntések kockázatkezelési változói és az egyenletek a 3. táblázaton alapultak, kiegészítve a kockázatkezelési modullal; a változók többsége megegyezett, a különbségeket a 4. táblázat részletezi.

4. táblázat. Etikai kockázatkezelési változók és egyenletek a mesterséges intelligenciával kapcsolatos döntéshozatalhoz.

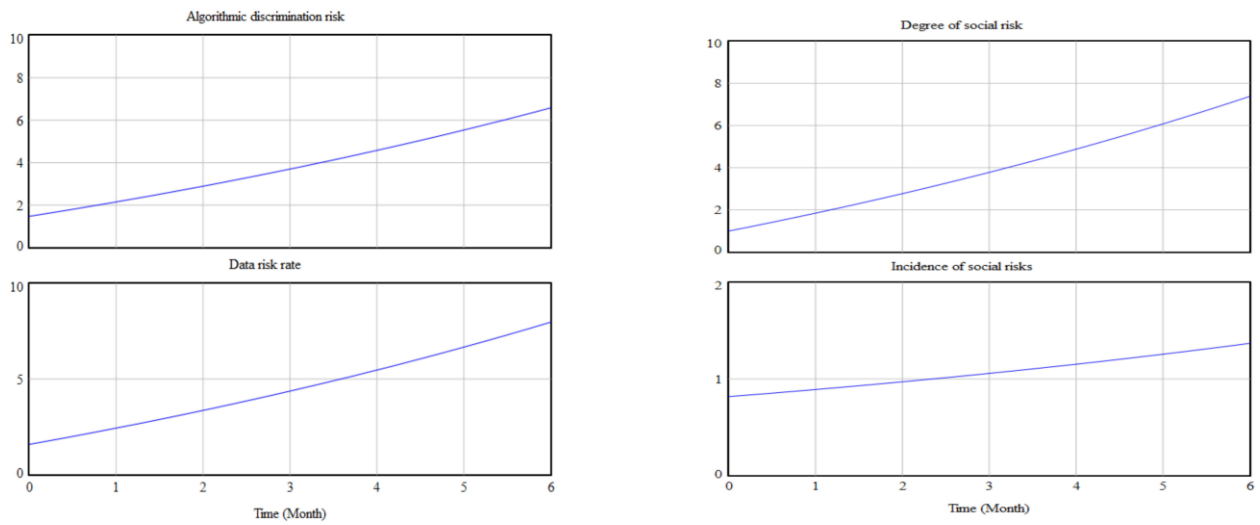
Nem	Változó	Típus	Kapcsolat egyenlet
1	A műszaki hiba előfordulási aránya	Segédváltozó	1-"Kockázatkezelés mértéke" + 0,1 (technológiai kockázatkezelés, amely csökkenti a műszaki hibák kockázatát).
2	Adatszivárgás	Segédváltozó	"A műszaki hibák aránya" - "A kockázatkezelés mértéke" (technikai kockázat (beleértve az emberi kockázatot is) az adatsértés miatt, de a kockázatkezelés csökkenti a sérülés mértékét).
3	Kockázatmegelőzési arány	Segédváltozó	"A társadalmi kockázat mértéke" \times 0,9 + 0,1 (minél magasabb a társadalmi kockázat, annál nagyobb a társadalmi kockázat). minél magasabb a társadalmi kockázati egyenlet mértéke, és minél több etikai norma és irányítási rendszer erősíti a kockázatmegelőzési arányt)
4	Kockázatkezelési arány	Változó árfolyam	"Kockázatmegelőzési ráta" \times 0,5 + 0,1
5	A társadalmi kockázat mértéke	Szintváltozó	INTEG ("A szociális kockázatok előfordulása" - "Kockázatkezelési arány", 1)
6	Kockázatkezelési diploma	Szintváltozó	INTEG (1-"Kockázatkezelési ráta", 1)

4.4. Szimuláció és tesztelés

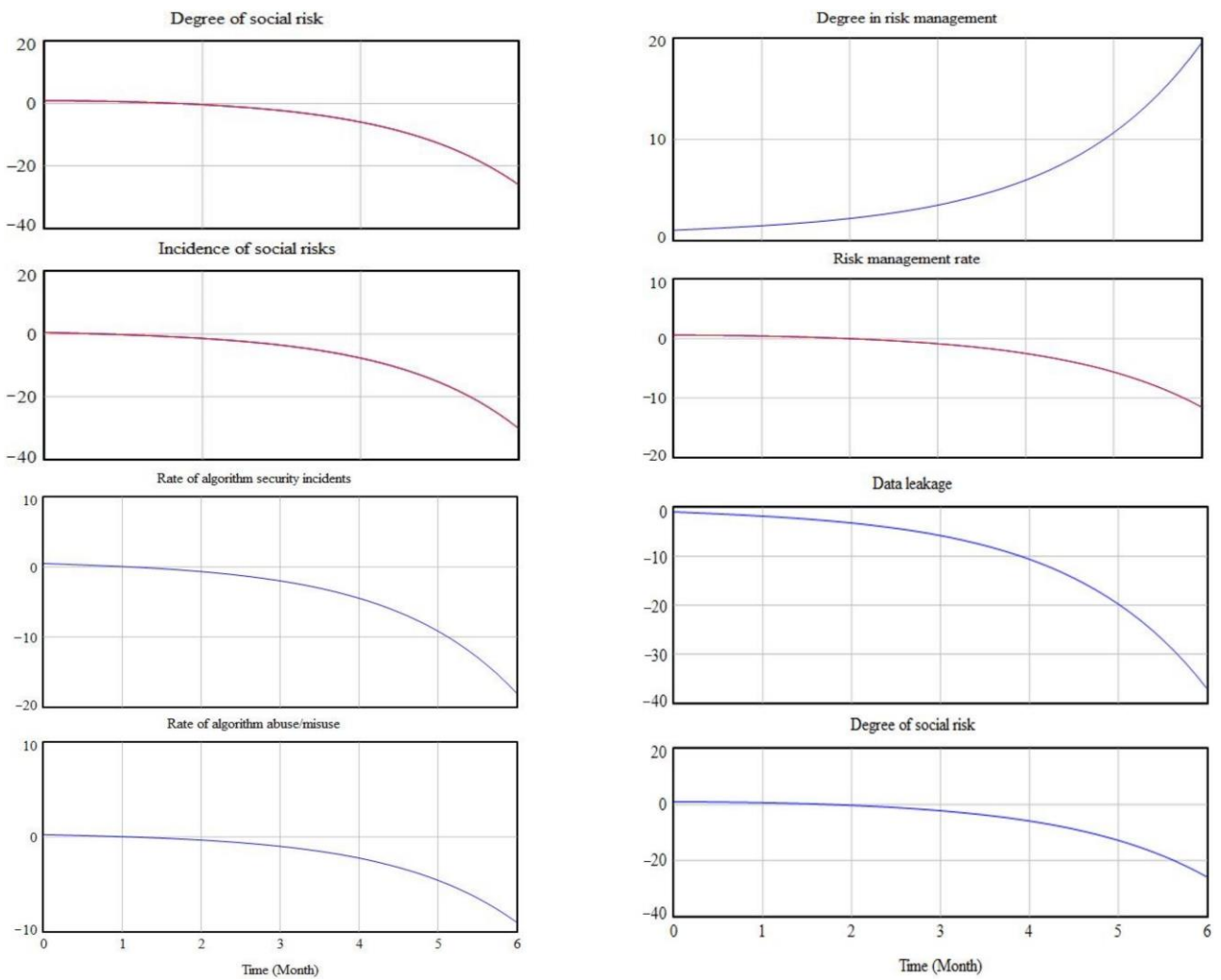
A mesterséges intelligenciával kapcsolatos döntéshozatalban az etikai kockázat etikájára vonatkozó empirikus kutatási adatok hiánya miatt nem lehet összehasonlítani a kísérleti adatokat a tényleges adatokkal. Ezért reálisabb eredményeket kaptunk a képletek és az érzékenységi tesztek iteratív beállításával, illetve a kockázat alakulására a kormányzás előtt és a kormányzás után. A szimulációs műveletek elvégzéséhez a vensim PLE szimulációs programot használtuk, a kezdeti idő = 0, a végső idő = 6, az időlépés = 0,125, az időegység = hónap. Beállítottuk a mesterséges intelligencia döntéshozatal etikai kockázat változóinak paraméterértékeit, hogy megkapjuk a kockázat alrendszer és a kormányzás alrendszer változásait. A 9. ábra például a kockázat alakulásának szintjét mutatja a kormányzás előtt, a 10. ábra pedig a kockázat szintjét a kormányzás után.

A szimulációs eredmények azt mutatják, hogy idővel, a kormányzás beavatkozása nélkül az algoritmikus kockázat és az adatkockázat ellenőrizhetetlen állapotban van, ami a társadalmi kockázat fokozatosan növekvő és ellenőrizhetetlen arányát eredményezi. Az irányítási feltétel bevonása után azonban, bár a korai szakaszokban nem volt nagyobb hatás (valószínűleg az irányítási intézkedések alacsony prioritása miatt), a

későbbi szakaszokban, ahogy a kockázatkezelés mértéke nőtt, mind az algoritmus kockázat, mind az adatkockázat aránya jelentős csökkenést mutatott. Így a társadalmi kockázat mértéke is csökkent, és a mesterséges intelligencia alapú döntéshozatalhoz kapcsolódó etikai kockázatok problémája jobban kontrollálhatóvá vált.



9. ábra. Kormányzás előtti kockázatfejlesztés.



10. ábra. Kockázat alakulása az irányítás után.

5. Következtetések

Miközben élvezzük a mesterséges intelligencia nyújtotta kényelmet, a lehető legnagyobb mértékben el kell kerülnünk az általa okozott etikai kockázatokat is. Ebben a tanulmányban a gyökeres elméleti megközelítéssel azonosítottuk és rendszereztük a mesterséges intelligencia alapú döntéshozatal etikai kockázati tényezőit, és felépítettük a kockázati tényezők fogalmi modelljét. A kockázati tényezők visszacsatolási modelljét is felépítettük a rendszerdinamika segítségével, hogy feltárjuk a mesterséges intelligencia etikai kockázatainak kialakulási mechanizmusát. A kockázat okait több szempontból és minden szempontból elemeztük, hogy hatékony segítséget nyújtsunk az etikai döntéshozatal tekintetében, csökkentjük a mesterséges intelligencia negatív etikai hatásait, és garantáljuk a mesterséges intelligencia egészséges, hosszú távú és felelős fejlődését, ezáltal elősegítve a nemzeti tudományos és technológiai fejlődés kormányzási szintjét. E dokumentum főbb megállapításai és megállapításai a következők.

5.1. A mesterséges intelligencia döntéshozatalának etikai kockázati tényezői

A gyökérelmélet alapján két központi kategóriát kaptunk: A "technikai kockázatok azonosítása" és a "menedzsmentkockázatok azonosítása". A technikai kockázat azonosítása magában foglalja az algoritmuskockázatot, az adatkockázatot és a technológiai kockázatot, amely az első szintű csomópontok 36,5%-át foglalja el. A vezetési kockázat azonosítása magában foglalja a vezetési kockázatokat és a kockázatkezelést is. A vezetési kockázatok közé tartoznak a három kategória által okozott vezetési kockázatok, döntési kockázatok és társadalmi kockázatok, amelyek az első szintű csomópontok 39,6%-át foglalják el, a kockázatkezelés pedig az első szintű csomópontok 23,9%-át. Összességében úgy tűnik, hogy a technológiai és a menedzsment kockázatok a mesterséges intelligencia döntési etikai kockázatokban azonos státuszúak, és két olyan szempont, amelyre összpontosítani kell. A mesterséges intelligenciával kapcsolatos döntések etikai kockázati tényezőinek dimenzionális szerkezeti modellje összefoglalja a mesterséges intelligenciával kapcsolatos döntések etikai kockázati tényezőit, és különböző dimenziókat biztosít az etikus döntéshozatalhoz és értékeléshez a jövőben. Ezen túlmenően a kockázatkezelés szerepe a kockázat előfordulásának csökkentése, a kockázat előfordulásának csökkentésére irányuló intézkedések és megoldások javaslata, ami segít megelőzni a kockázat előfordulását a döntéshozatal során, és lehetővé teszi az AI egészségesebb irányba történő fejlődését.

5.2. A mesterséges intelligencia etikai kockázatai A döntéshozatal és az irányítás mechanizmusai

A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázati tényezőinek azonosítása alapján a kockázati tényezők közötti kapcsolatokat és útvonalakat a rendszerdinamika alkalmazásával vizsgálták. A mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázati modelljének szimulálásához a Vensim szoftvert használták. Az oksági hurok szempontjából egyrészt a mesterséges intelligencia alapú döntéshozatal etikai kockázatát okozó fő tényezők az adatkockázat és a technológiai kockázat. A technológia bizonytalansága és az adatok hiányossága és elégtelensége torzítást okozhat a döntéshozatalban, ami a technológia komolyabb etikai problémáihoz vezethet. Emellett a menedzsment hibái komoly társadalmi kockázatokhoz, például munkanélküliséghez vezethetnek. Másrészt a kockázati visszacsatolási modell kockázatkezelési elemekkel való kiegészítésével az algoritmus, a technológia és az adatok kockázati aránya jelentősen csökkenthető, így hatékonyan csökkenthető a társadalmi kockázatok előfordulása.

5.3. Ajánlások a mesterséges intelligenciával kapcsolatos döntéshozatal etikai kockázatainak irányításához

A kockázat tényezői és mechanizmusai szerint a kockázat irányítási normák, K+F normák és a kockázatok felhasználási normái szerint szabályozható. Az irányítási normák tekintetében az AI-technológia fejlesztésével és alkalmazásával kapcsolatos szervezeteknek meg kell erősíteniük a kockázatok azonosítását és értékelését a technológia előmozdításának folyamatában, elő kell mozdítaniuk az agilis irányítást, jó munkát kell végezniük az előzetes ellenőrzésben, és meg kell erősíteniük a kockázatmegelőzést. A K+F normák tekintetében a kutatóknak meg kell erősíteniük az

önfegyelmet, javítaniuk kell az adatok minőségét, és garantálniuk kell a biztonságos és megbízható adatokat; az algoritmusoknak fokozniuk kell a biztonságot és az átláthatóságot, és el kell kerülniük az algoritmusok és az adatok elfogult megkülönböztetését. A felhasználási normák tekintetében meg kell erősíteni a minőségellenőrzést, védeni kell a felhasználói jogokat, és fokozni kell a vészhelyzeti védelmet, miközben el kell kerülni a technológiával való visszaélést és a visszaélést.

5.4. A kutatás hozzájárulása

Ebben a tanulmányban a gyökeres elmélet kvalitatív kutatási módszerét használjuk a mesterséges intelligencia etikai döntéshozatal kockázati tényezőinek azonosítására és rendszerezésére, beleértve a kockázati forrásokat, a kockázati következményeket és a kockázatkezelési feltételeket, és felépítettük a kockázati tényezők fogalmi modelljét. A kockázati tényezők rendszerdinamikán keresztül történő visszacsatolási modelljét is megalkottuk, hogy feltárjuk a mesterséges intelligencia etikai kockázatainak kialakulási mechanizmusát, és a kockázatok okait több szempontból és átfogó módon elemeztük. Megállapítottuk, hogy a technológiai bizonytalanság, a hiányos adatok és az irányítási hibák az etikai kockázatok fő forrásai a mesterséges intelligenciával kapcsolatos döntéshozatalban, és hogy a kockázatkezelési elemek beavatkozása hatékonyan meggátolhatja az algoritmikus, technológiai és adatkockázatokból eredő társadalmi kockázatokat. Ennek megfelelően stratégiákat javasolunk a mesterséges intelligencia döntéshozatal etikai kockázatainak irányítására az irányítás, a kutatás és a fejlesztés szempontjából, azzal a céllal, hogy hatékony segítséget nyújtsunk az etikus döntéshozatalhoz és csökkentjük a mesterséges intelligencia negatív etikai hatásait.

6. Korlátozások

A dokumentumnak van néhány hiányossága. Először is, a tudomány és a technológia etikájáról szóló korábbi tanulmányok többsége az élettudományok területén felmerülő főbb etikai kérdésekre, például a klónozásra összpontosított, és kevésbé a mesterséges intelligencia területén jelentkező etikai kockázatokra, és a gyökeres adatok gyűjtése ebben a tanulmányban nem biztos, hogy elegendő. Ezenkívül a gyökeres adatforrások ebben a tanulmányban többnyire szövegek, például irodalom és weboldalak, ami azt jelenti, hogy csak a mesterséges intelligencia ismert kockázatait azonosíthatjuk, anélkül, hogy tovább jósolnánk és összefoglalnánk a mesterséges intelligencia ismeretlen kockázatait. Ezért az ebben a dokumentumban szereplő kapcsolódó kérdések kutatását tovább kell mélyíteni. Végül, ez a dokumentum egy rendszerdinamikai modellt épít fel a kvalitatív elemzés alapján, és a modellkapcsolatokat a korábbi tudósok kutatásai és jelentései alapján építi fel, és a szimulációs eredményekből hiányzik a reális forgatókönyvek összehasonlítása és ellenőrzése. A mesterséges intelligencia fejlődésében különböző kiszámíthatatlan kockázatok és kihívások vannak, amelyekre a jövőbeni kutatásban figyelmet kell fordítanunk, és különböző kutatási módszereket kell még kipróbálni, például a mesterséges intelligencia etikájával kapcsolatos adatok gyűjtése és tárolása, a mesterséges intelligencia etikai kutatásának különböző módszereinek kombinálása, és végül a mesterséges intelligencia etikai kutatásának általános keretének kiépítése.

Szerzői hozzájárulások: H.G.; módszertan, H.G. és A.Z.; szoftver, L.D.; validálás, L.D. és A.Z.; formális elemzés, A.Z.; vizsgálat, L.D.; források, H.G.; adatok gondozása, L.D.; írás-eredeti tervezet elkészítése, L. D.D.; írás-ellenőrzés és szerkesztés, A.Z.; megjelenítés, L.D.; felügyelet, H.G.; projektadminisztráció, H.G.; finanszírozás beszerzése, H.G. A kézirat közzétett változatát valamennyi szerző elolvasta és elfogadta.

Finanszírozás: [2021SFGC0102, 2020CXGC010110] támogatási számmal támogatta a [Shandong tartományi kulcsfontosságú kutatási és fejlesztési program (jelentős tudományos és technológiai innovációs projekt)].

Intézményi felülvizsgálati bizottság nyilatkozata: Nem alkalmazható.

Tájékoztató beleegyezési nyilatkozat: Nem alkalmazható.

Adatelérhetőségi nyilatkozat: Az adatok [harmadik féltől] származnak, és [a hivatkozásokból] a [harmadik fél] engedélyével állnak rendelkezésre.

Összeférhetetlenség: A szerzők nem jelentenek összeférhetetlenséget.

Hivatkozások

1. Crompton, L. A döntési pont-dilemma: A felelősség újabb problémája az ember és az AI közötti interakcióban. *J. Responsible Technol.* **2021**, *7-8*, 100013. [[CrossRef](#)]
2. Yu, C.I.; Hu, W.L.; Liu, Y. Az USA kiadja az új "Nemzeti mesterséges intelligencia kutatási és fejlesztési stratégiai tervet". *Titoktartás Sci. Technol.* **2019**, *9*, 35-37.

- ³⁴³
3. Wang, X.F. Az EU kiadja a "Mesterséges intelligencia fehér könyvét: A mesterséges intelligenciáról - Európai megközelítés a kiválóság és a bizalom érdekében". *Scitech China* 2020, 6, 98-101.

4. Zhongguancun Institute of Internet Finance. *Tovább|Az Európai Bizottság 2021-es mesterséges intelligenciáról szóló törvényjavaslata*; Zhongguancun Institute of Internet Finance: Peking, Kína, 2021.
5. Dayang.com-Guangzhou Daily. *Korea kidolgozza a világ első etikai kódexét a robotok számára*; Guangzhou Daily: Guangzhou, Kína, 2007.
6. Sadie agytrösz. [Quick Comment], Külföldi országok etikai és erkölcsi kutatásokat végeznek a mesterséges intelligenciával kapcsolatban több szinten. Elérhető online: <https://xueqiu.com/4162984112/135453621> (hozzáférés: 2022. augusztus 20.).
7. Államtanács. *Az Államtanács közleménye a mesterséges intelligencia új generációjára vonatkozó fejlesztési terv kiadásáról*; Állami Tanács : Peking, Kína, 2017. július 20.
8. Jiang, J. A mesterséges intelligencia etikájának fő célja és alapelvei a kockázat szempontjából. *Inf. Commun. Technol. Policy* **2019**, *6*, 13-16.
9. Susan, F. Etika az AI: A mesterséges intelligencia rendszerek előnyei és kockázatai. *Érdekes mérnöki munka*. Online elérhető: <https://baslangicnoktasi.org/en/ethics-of-ai-benefits-and-risks-of-artificial-intelligence-systems/> (hozzáférés: 2022. augusztus 20.).
10. Turing, A.M. Számítógépek és intelligencia. In *Parsing Turing Test*; Springer: Dordrecht, Hollandia, 2007; pp. 23-65.
11. Yan, K.R. A mesterséges intelligencia kockázata és annak elkerülési útja. *J. Shanghai Norm. Univ. Philos. Soc. Sci. szerk.* **2018**, *47*, 40-47.
12. Chen, X.P. A mesterséges intelligencia etikájának célja, feladatai és megvalósítása: A tudományelméleti etika kérdései: Hat kérdés és a mögöttük húzódó indoklás. *Philos. Res.* **2020**, *9*, 79-87+107+129.
13. Marabelli, M.; Newell, N.; Handunge, V. Az algoritmikus döntéshozó rendszerek életciklusa: Szervezeti döntések és etikai kihívások. *J. Strateg. Inf. Syst.* **2021**, *30*, 101683. [[CrossRef](#)]
14. Arkin, R.C. A halálos viselkedés szabályozása: Az etika beágyazása egy hibrid deliberatív/reaktív robotarchitektúrába - 1. rész: Motiváció és filozófia. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, Amsterdam, The Netherlands, 12-15 March 2008; pp. 121-128.
15. Zhao, Z.Y.; Xu, F.; Gao, F.; Li, F.; Hou, H.M.; Li, M.W. A mesterséges intelligencia etikai kockázatainak megértése. *China Soft Sci.* **2021**, *6*, 1-12.
16. Leibniz, G.W. *Megjegyzések az analízishez*; Oxford University: Oxford, Egyesült Királyság, 1984.
17. Anderson, S.L. Asimov "A robotika három törvénye" és a gépi metaetika. *Sci. Fict. Philos. Time Travel Superintelligence* **2016**, *22*, 290-307.
18. Joachim, B.; Elisa, O. Az új technológiák etikai elveinek egységes listája felé. Négy európai jelentés elemzése a molekuláris biotechnológiáról és a mesterséges intelligenciáról. *Fenntartás. Futures* **2022**, *4*, 100086.
19. Bernd, W.; Wirtz, J.C.; Weyerer, I.K. A mesterséges intelligencia irányítása: Egy kockázat- és iránymutatás-alapú integratív keretrendszer. *Gov. Inf. Q.* **2022**, 101685. [[CrossRef](#)]
20. Bonnefon, J.F.; Shariff, A.; Rahwan, L. Az autonóm járművek társadalmi dilemmája. *Science* **2016**, *352*, 1573-1576. [[CrossRef](#)]
21. Johann, C.B.; Kaneko, S. Készen áll a társadalom az AI etikus döntéshozatalra? Az autonóm autókról szóló tanulmány tanulságai. *J. Behav. Exp. Econ.* **2022**, *98*, 101881.
22. Cartolovni, A.; Tomcic, A.; Mosler, E.L. A mesterséges intelligencia alapú orvosi döntéstámogató eszközök etikai, jogi és társadalmi megfontolásai: A scoping review. *Int. J. Med. Inf.* **2022**, *161*, 104738. [[CrossRef](#)]
23. Chen, L.; Wang, B.C.; Huang, S.H.; Zhang, J.Y.; Guo, R.; Lu, J.Q. Artificial Intelligence Ethics Guidelines and Governance System: Az etikai tudományok mesterséges intelligenciája: jelenlegi helyzet és stratégiai javaslatok. *Sci. Technol. Manag. Res.* **2021**, *41*, 193-200.
24. Weinmann, M.; Schneider, C.; vom Brocke, J. Digital Nudging. *Bus. Inf. Syst. Eng.* **2016**, *58*, 433-436. [[CrossRef](#)]
25. Jian, G. Mesterséges intelligencia az egészségügyben és az orvostudományban: Az ígéretek, etikai kihívások és irányítás. *Chin. Med. Sci. J.* **2019**, *34*, 76-83.
26. Stahl, B.C. Felelős innovációs ökoszisztémák: Az ökoszisztéma fogalmának a mesterséges intelligenciára való alkalmazásának etikai következményei. *Int. J. Inf. Manag.* **2022**, *62*, 102441. [[CrossRef](#)]
27. Galaz, V.; Centeno, M.A. Mesterséges intelligencia, rendszerkockázatok és fenntarthatóság. *Technol. Soc.* **2021**, *67*, 101741. [[CrossRef](#)]
28. Catherine, M.; Gretchen, B.R. *Minőségi kutatás tervezése: Chongqing University Kiadó: Guidance throughout an Effective Research Program*; Chongqing University Publisher: Chongqing, Kína, 2019.
29. Juliet, M.C.; Anselm, L.S. *Procedures and Methods for the Formation of a Rooted Theory Based on Qualitative Research*; Chongqing University Publisher: Chongqing, Kína, 2015.
30. Flynn, S.V.; Korcusa, J.S. Grounded theory research design: A gyakorlatok és eljárások vizsgálata. *Couns. Outcome Res. Eval.* **2018**, *9*, 102-116. [[CrossRef](#)]
31. Li, X.; Su, D.Y. A mesterséges intelligencia etikai kockázatainak reprezentációjáról. *J. Chang. Univ. Sci. Technol. Soc. Sci.* **2020**, *35*, 13-17.
32. Tan, J.S.; Yang, J.W. A mesterséges intelligencia etikai kockázata és együttműködő irányítása. *Chin. Public Adm.* **2019**, *10*, 46-47.
33. Zhang, Z.X.; Zhang, J.Y.; Tan, T.N. A mesterséges intelligencia etikai problémáinak elemzése és ellenintézkedései. *Bull. Chin. Acad. Sci.* **2021**, *36*, 1270-1277.
34. Zhang, T.; Ma, H. Rendszerdinamikai kutatás a mesterséges intelligencia adatbiztonságát befolyásoló tényezőkről. *Inf. Res.* **2021**, *3*, 1-10.
35. Zhu, B.Z.; Tang, J.J.; Jiang, M.X.; Wang, P. A szén-dioxid-piaci kockázat szimulációja és szabályozása a rendszerdinamika alapján. *Syst. Eng. Theory Pract.* **2022**, *42*, 1859-1872.

- 343
36. Lo Piano, S. Etikai elvek a gépi tanulásban és a mesterséges intelligenciában: Egy eset a terepről és lehetséges továbblépési lehetőségek.
Humanit. Soc. Sci. Commun. **2020**, *7*, 9. [[CrossRef](#)]