DOI: https://doi.org/10.12700/btsz.2026.8.3.63

SAFETY ISSUES OF AI-BASED ENGINEERING ASSISTANCE IN HUMAN-ROBOT INTERACTION

AZ MI-ALAPÚ MÉRNÖKI ASSZISZTENCIA BIZTONSÁGI KÉRDÉSEI AZ EMBER-ROBOT INTERAKCIÓBAN

BURAI István György[1] – KOLLÁR Csaba[2]

Abstract

Artificial intelligence is playing an increasingly important role in engineering decision support, while robotic systems used in industrial and educational environments remain safety-critical and fundamentally deterministic. This paper examines the safety issues raised by AI-based engineering assistance in human-robot interaction. Its main assumption is that risk does not emerge only at the level of direct physical human-robot contact, but may also arise earlier in the decision-preparation phase, when the user relies on AI-generated suggestions. The study analyses the role of AI as an engineering assistant, with particular attention to autonomy boundaries, human supervision, and validation. It identifies the main indirect risk factors, including overreliance, insufficient verification, and lack of user competence. The paper concludes that the safe application of AI can only be ensured through mandatory human validation.

Keywords

artificial intelligence, human-robot interaction, engineering assistance, safety, human supervision

Absztrakt

A mesterséges intelligencia egyre nagyobb szerepet kap a mérnöki döntéstámogatásban, miközben az ipari és oktatási környezetben alkalmazott robotizált rendszerek működése továbbra is biztonságkritikus és alapvetően determinisztikus jellegű. A tanulmány azt vizsgálja, hogy az MI-alapú mérnöki asszisztencia milyen biztonsági kérdéseket vet fel az ember-robot interakcióban. Kiindulópontja, hogy a kockázat nemcsak a fizikai ember-robot kapcsolat szintjén jelenhet meg, hanem már a döntéselőkészítés során is, amikor a felhasználó MI által generált javaslatokra támaszkodik. A cikk elemzi az MI mint mérnöki asszisztens szerepét, különös tekintettel az autonómiahatárokra, az emberi felügyeletre és a validációra. Azonosítja a főbb közvetett kockázati tényezőket, így a túlzott bizalmat, a hiányos ellenőrzést és a kompetenciahiányt. Következtetések alapján az MI biztonságos alkalmazása csak kötelező emberi validáció mellett értelmezhető.

Kulcsszavak

mesterséges intelligencia, ember-robot interakció, mérnöki asszisztencia, biztonság, emberi felügyelet

[1] burai.istvan@bgk.uni-obuda.hu | ORCID: 0009-0007-8988-6301 | Assistant Lecturer, Óbuda University, Bánki Donát Faculty of Mechanical and Safety Engineering | Tanársegéd, Óbudai Egyetem, Bánki Donát Gépész és Biztonságtechnikai Mérnöki Kar

[2] kollar.csaba@uni-obuda.hu | ORCID: 0000-0002-0981-2385 | senior research fellow and leader, Óbuda University Bánki Donát Faculty of Mechanical and Safety Engineering Artificial Intelligence Workshop | tudományos főmunkatárs és vezető, Óbudai Egyetem Bánki Donát Gépész és Biztonságtechnikai Mérnöki Kar Mesterséges Intelligencia Műhely

BEVEZETÉS

A mesterséges intelligencia alkalmazása az utóbbi években a mérnöki tervezési, döntéstámogatási és oktatási folyamatokban is egyre hangsúlyosabbá vált. Az MI-alapú rendszerek már nem csupán információkeresési vagy adminisztratív feladatokban jelennek meg, hanem egyre gyakrabban nyújtanak szakmai jellegű javaslatokat technológiai, programozási vagy elemzési kérdésekben is. Ezzel párhuzamosan az ipari környezetben működő robotizált rendszerek továbbra is olyan biztonságkritikus technikai rendszereknek tekinthetők, amelyek esetében a hibás döntések vagy a nem megfelelően validált beavatkozások közvetlen fizikai kockázatot is eredményezhetnek [1].

Az ember-robot interakció biztonsági vizsgálata hagyományosan a fizikai együttműködés, a védelmi megoldások és a működés közbeni kockázatok oldaláról történik [2]. A mesterséges intelligencia megjelenésével azonban a biztonsági kérdések egy része már nem kizárólag a közvetlen ember-gép kapcsolat szintjén értelmezhető, hanem a döntés-előkészítés és a rendszerhasználat korábbi szakaszaiban is. Amennyiben a felhasználó egy MI-rendszert mérnöki asszisztensként alkalmaz, a hibás, félreértelmezett vagy nem megfelelően ellenőrzött javaslatok olyan döntésekhez vezethetnek, amelyek később a robotizált rendszer működésében biztonsági következményekkel járhatnak [3].

A tanulmány célja annak elemzése, hogy az MI-alapú mérnöki asszisztencia milyen közvetett biztonsági kockázatokat hordoz az ember-robot interakcióban, különös tekintettel az autonómiahatárok, az emberi felügyelet és a validáció szerepére. A vizsgálat nem az MI közvetlen robotvezérlési alkalmazásaira koncentrál, hanem arra a helyzetre, amikor a mesterséges intelligencia döntéstámogató vagy javaslatadó funkcióban jelenik meg, miközben a végrehajtás determinisztikus, biztonságkritikus rendszerben történik. A tanulmány emellett olyan értelmezési keret felvázolására törekszik, amely az oktatási és ipari környezetben egyaránt segítheti az MI biztonságos és felelősségteljes alkalmazásának megítélését.

AZ EMBER-ROBOT INTERAKCIÓ BIZTONSÁGI ÉRTELMEZÉSE IPARI KÖRNYEZETBEN

Az ember-robot interakció biztonsági értelmezése az ipari környezetben alapvetően abból indul ki, hogy az ember és a gépi rendszer közös működési térben, egymás tevékenységére közvetlen vagy közvetett módon hatva vesz részt a munkafolyamatban. A biztonság ebben az összefüggésben nem kizárólag a fizikai érintkezés elkerülését jelenti, hanem minden olyan műszaki, szervezési és emberi tényező együttes kezelését is, amely a működés során veszélyhelyzet kialakulásához vezethet. Ipari környezetben ez különösen fontos, mivel a robotizált rendszerek nagy mozgási energiával, előre programozott működési logikával és a kezelő által csak részben átlátható belső folyamatokkal rendelkezhetnek [4].

Az ipari ember-robot interakció biztonsági értelmezése ugyanakkor nem merül ki a fizikai elkülönítés, az ütközéselkerülés vagy a védelmi funkciók vizsgálatában. A rendszer biztonsági teljesítményét azok a döntések is befolyásolják, amelyek a fizikai végrehajtást megelőző tervezési, programozási és üzemeltetési szakaszokban születnek. Ebben a tágabb megközelítésben a veszélyforrások egy része nem közvetlenül a robot mozgásából ered, hanem abból, hogy a kezelő vagy a technológiai döntéshozó milyen információkat fogad el, hogyan értelmezi azokat, és milyen mértékben támaszkodik külső döntéstámogató rendszerre. Az MI megjelenésével ezért az ember-robot interakció biztonsági vizsgálata

kiterjeszthető a döntési lánc korábbi szakaszaira is, ahol a hibás információ vagy a nem megfelelő validáció még a végrehajtás előtt létrehozhatja a későbbi veszélyhelyzet feltételeit [3], [4].

A hagyományos ipari biztonsági megközelítés az ember-robot interakció során elsősorban a fizikai veszélyforrásokra koncentrál, így például az ütközésre, a beszorulásra, a nem várt mozgásokra vagy az elégtelen leválasztásból eredő kockázatokra. Ezek kezelése tipikusan műszaki védelmi megoldásokkal, érzékelőkkel, biztonsági logikákkal, elhatárolással és szabályozott üzemmódokkal történik [5]. Ugyanakkor a biztonság tényleges szintjét nem kizárólag a berendezés fizikai kialakítása határozza meg, hanem az is, hogy a kezelő, a programozó vagy a technológiai döntéshozó milyen információk alapján avatkozik be a rendszer működésébe.

Az ember szerepe ezért az ipari ember-robot interakcióban nem csupán a közvetlen kezelésre vagy felügyeletre korlátozódik, hanem kiterjed a rendszer értelmezésére, a működési feltételek megadására és a beavatkozási döntések meghozatalára is [6]. Ebből következően a biztonság kérdése nem választható el attól a döntési lánctól sem, amely a fizikai végrehajtást megelőzi. Ha a rendszerrel kapcsolatos emberi döntések hibás információra, hiányos ellenőrzésre vagy félreértelmezett javaslatokra épülnek, akkor a kockázat már a tényleges ember-robot kontaktus előtt kialakulhat [3].

AZ MI MINT MÉRNÖKI ASSZISZTENS AZ EMBER-ROBOT INTERAKCIÓ BIZTONSÁGI KONTEXTUSÁBAN

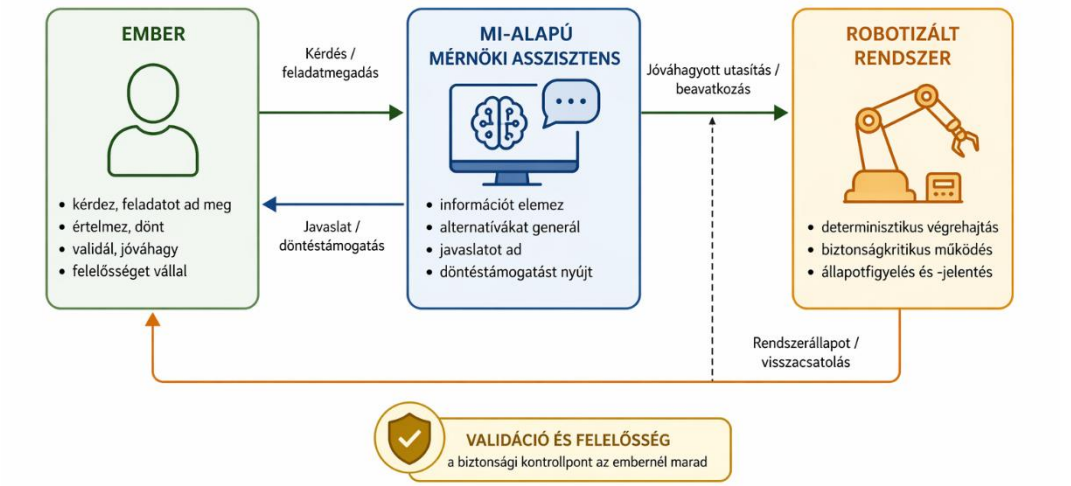
A mesterséges intelligencia mérnöki környezetben egyre gyakrabban jelenik meg asszisztív szerepkörben, ahol elsődleges feladata nem a közvetlen végrehajtás, hanem az információfeldolgozás, a javaslatadás és a döntéstámogatás [7]. Ilyen felhasználási helyzet lehet például a technológiai tervezés, a programozási feladatok előkészítése, a hibakeresés, a dokumentációk értelmezése vagy az oktatási célú szakmai támogatás. Ebben a megközelítésben az MI nem autonóm irányító rendszerként, hanem a felhasználó munkáját segítő mérnöki asszisztensként értelmezhető.

Ez a szerepkör lényegesen különbözik azoktól az alkalmazási esetektől, amelyekben a mesterséges intelligencia közvetlenül avatkozik be a gépi működésbe vagy autonóm szabályozási funkciót lát el. A vizsgált esetben az MI kimenete javaslatként jelenik meg, amely önmagában még nem eredményez fizikai beavatkozást, ugyanakkor befolyásolhatja a felhasználó döntését. Ebből következően a biztonsági kérdés nem pusztán az MI technikai megbízhatóságára vezethető vissza, hanem arra is, hogy a felhasználó miként értelmezi, ellenőrzi és alkalmazza az általa kapott információt.

Az MI-alapú mérnöki asszisztencia sajátossága, hogy működése nem azonos a klasszikus, előre meghatározott szabályok szerint működő automatizált rendszerekével. A generatív MI válaszai valószínűségi alapon jönnek létre, ezért ugyanazon vagy hasonló feladatmegadás mellett is előfordulhat eltérő tartalmú javaslat, hiányos indoklás vagy szakmailag pontatlan megoldás. Ez különösen jelentős olyan mérnöki környezetben, ahol a döntés később determinisztikus végrehajtórendszerben realizálódik. A két működési logika közötti eltérés miatt az MI kimenete nem kezelhető közvetlenül végrehajtható utasításként, hanem olyan előzetes információként vagy alternatívaként, amelynek alkalmazhatóságát a felhasználónak műszaki és biztonsági szempontból is ellenőriznie kell. Ebben az értelemben az autonómiahatár nem kizárólag technikai korlátozást jelent, hanem azt a döntési pontot is

kijelöli, ahol az MI támogatási szerepe megszűnik, és az emberi szakmai felelősség válik meghatározóvá [7].

Különösen fontos ez olyan esetekben, ahol a végrehajtás determinisztikus, biztonságkritikus rendszerben történik. Az ipari robotokhoz, CNC-rendszerekhez vagy egyéb automatizált berendezésekhez kapcsolódó mérnöki döntések esetében a hibás vagy nem megfelelően validált javaslatok következményei a fizikai működés szintjén jelenhetnek meg. Ezért az MI mérnöki asszisztensként történő alkalmazása csak akkor tekinthető biztonságosan értelmezhetőnek, ha világosan kijelölhetők azok a határok, ahol a javaslatadás véget ér, és ahol az emberi ellenőrzésnek, jóváhagyásnak és felelősségvállalásnak kell elsődlegessé válnia [8]. A vizsgált kapcsolatrendszert az 1. ábra szemlélteti, amely az ember, az MI-alapú mérnöki asszisztens és a robotizált rendszer közötti funkcionális kapcsolatokat, valamint a validáció és a felelősség helyét mutatja be.



1. ábra: Az ember, az MI-alapú mérnöki asszisztens és a robotizált rendszer kapcsolati modellje

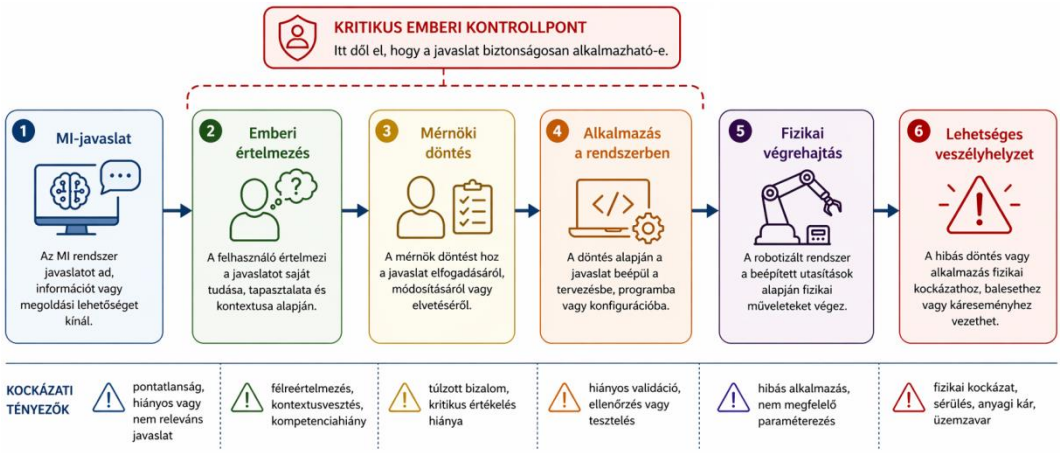
AZ AUTONÓMIAHATÁROK ÉS A KÖZVETETT BIZTONSÁGI KOCKÁZATOK KÉRDÉSE

Az MI-alapú mérnöki asszisztencia biztonsági értékelésének egyik kulcskérdése az autonómiahatárok egyértelmű kijelölése. Biztonsági szempontból alapvető különbség van aközött, hogy a mesterséges intelligencia információt rendszerez, javaslatot fogalmaz meg vagy döntést készít elő, illetve aközött, hogy közvetlenül befolyásolja a fizikai végrehajtást. Minél közelebb kerül az MI szerepe a végrehajtási szinthez, annál nagyobb jelentőséget kap annak meghatározása, hogy mely döntési pontokon marad meg a kötelező emberi felügyelet és validáció [9].

A közvetett biztonsági kockázatok abból erednek, hogy az MI által generált tartalom nem közvetlen gépmozgás formájában jelenik meg, hanem emberi döntésen keresztül épül be a rendszer működésébe [3]. Ilyen kockázati tényező lehet a túlzott bizalom, amikor a felhasználó az MI-választ kellő ellenőrzés nélkül fogadja el, a kompetenciahiány, amikor nem képes annak műszaki értékelésére, valamint a hiányos validáció, amikor a javaslat alkalmazása előtt nem történik meg a szükséges szakmai ellenőrzés. Ezekben az esetekben a

hiba nem az MI önálló fizikai cselekvéséből, hanem az ember és az MI közötti döntési kapcsolat elégtelen kontrolljából fakad.

A közvetett kockázatok kezelése ezért olyan ellenőrzési pontok kialakítását igényli, amelyek még a fizikai végrehajtás előtt képesek megszakítani a hibás döntési láncot. Ilyen kontrollpont lehet például a műszaki paraméterek szakmai felülvizsgálata, a biztonsági feltételek ellenőrzése, a végrehajthatóság vizsgálata vagy egy kritikus döntés második személy általi jóváhagyása. Az ellenőrzés szükséges mélysége az adott feladat kockázati szintjéhez igazítható: míg egy alacsony következményű információs javaslat esetében elegendő lehet a felhasználói értékelés, addig egy fizikai végrehajtást befolyásoló mérnöki döntésnél indokolt lehet több, egymásra épülő kontroll alkalmazása. Ebből következően az MI biztonságos integrációja nem egyetlen ellenőrzési lépéshez, hanem egy rétegzett felügyeleti struktúrához kapcsolható, amelyben a kritikus döntési pontokon az emberi szakértelem és jóváhagyás elsődleges marad [3], [8]. A közvetett biztonsági kockázatok kialakulásának logikáját a 2. ábra foglalja össze, amely az MI-javaslattól a fizikai veszélyhelyzetig vezető döntési és végrehajtási láncot szemlélteti.



2. ábra: Kockázatlánc az MI-javaslattól a fizikai veszélyhelyzet kialakulásáig

Az autonómiahatárok kijelölése ezért nem csupán technikai, hanem felelősségi kérdés is. Biztonságkritikus környezetben egyértelműen rögzíteni kell, hogy mely feladatok maradhatnak az MI támogatási körében, és mely pontokon nem kerülhető meg az emberi döntés. Az ilyen határok hiánya könnyen ahhoz vezethet, hogy a felhasználó a javaslatot implicit módon hitelesnek vagy közvetlenül alkalmazhatónak tekinti, ami növeli a hibás alkalmazás és ezen keresztül a fizikai veszélyhelyzet kialakulásának valószínűségét.

OKTATÁSI ÉS IPARI KÖVETKEZMÉNYEK

Az MI-alapú mérnöki asszisztencia terjedése az oktatási és ipari környezetben egyaránt új biztonsági elvárásokat vet fel. Ipari oldalon ez azt jelenti, hogy a technológiai döntések előkészítése során nem elegendő az MI által adott javaslatok hasznosságát vizsgálni, hanem azok ellenőrizhetőségét, alkalmazhatóságát és biztonsági következményeit is értékelni kell. Oktatási oldalon pedig az válik hangsúlyossá, hogy a hallgatók ne pusztán az MI

használatát sajátítsák el, hanem annak korlátait, bizonytalanságait és a szükséges emberi validáció szerepét is megértsék.

Az oktatási környezet ebből a szempontból különösen alkalmas arra, hogy az MI alkalmazásához kapcsolódó szakmai ellenőrzési kompetenciák kontrollált módon fejleszthetők legyenek. A hallgatók számára nem elegendő annak megtanulása, hogy miként kérdezzenek egy MI-rendszertől vagy hogyan használják fel az általa generált javaslatokat, hanem szükség van arra is, hogy felismerjék a válaszok korlátait, ellenőrizzék azok szakmai megalapozottságát és azonosítsák a biztonsági szempontból kritikus elemeket. Ez különösen fontos olyan mérnöki feladatok esetében, ahol a hibás döntés később közvetlenül fizikai végrehajtásban jelenik meg. A strukturált felülvizsgálati eljárások ezért nemcsak ellenőrzési eszközként, hanem fejlesztendő mérnöki kompetenciaként is értelmezhetők. Az ilyen megközelítés hozzájárulhat ahhoz, hogy az MI használata ne a szakmai kontroll csökkenését, hanem a tudatosabb döntéshozatalt támogassa [3], [7].

Az oktatás ebben a megközelítésben nem csupán ismeretátadási közeg, hanem kontrollált validációs tér is lehet, ahol a jövő mérnökei biztonságos körülmények között tanulhatják meg az MI-válaszok szakmai értelmezését és ellenőrzését. Ez különösen fontos olyan területeken, ahol a döntések később determinisztikus, fizikai végrehajtórendszerekben hasznosulnak. Ilyen esetekben a képzés egyik lényeges feladata az, hogy a hallgató felismerje: az MI által adott válasz nem tekinthető önmagában validált műszaki megoldásnak, hanem csak olyan kiindulási pontnak, amelyet szakmai és biztonsági szempontból egyaránt validálni kell.

Ipari környezetben mindez azt is jelenti, hogy az MI alkalmazását nem önálló technológiai megoldásként, hanem szabályozott emberi felügyelet mellett működő támogatási eszközként célszerű értelmezni. A biztonságos felhasználás feltétele, hogy a döntési felelősség ne mosódjon el az ember és a rendszer között, továbbá hogy egyértelműen meghatározhatók legyenek azok az ellenőrzési pontok, ahol a javaslatok műszaki és biztonsági szempontú felülvizsgálata megtörténik. Ennek hiányában az MI nem a biztonságot támogató, hanem a bizonytalanságot növelő tényezővé válhat [8].

KÖVETKEZTETÉSEK

A tanulmány rámutatott arra, hogy az MI-alapú mérnöki asszisztencia biztonsági értelmezése az ember-robot interakcióban nem korlátozható a közvetlen fizikai kapcsolat vizsgálatára. A kockázatok egy része már a döntés-előkészítés szintjén megjelenhet, amikor a felhasználó a mesterséges intelligencia által generált javaslatokra támaszkodik. Emiatt a biztonság kérdése szorosan összekapcsolódik az autonómiahatárok kijelölésével, az emberi felügyelet fenntartásával és a validáció következetes érvényesítésével.

A vizsgálat alapján megállapítható, hogy a mesterséges intelligencia mérnöki asszisztensként való alkalmazása csak akkor tekinthető biztonságosan értelmezhetőnek, ha szerepe világosan elhatárolható a fizikai végrehajtástól, és a döntési felelősség egyértelműen az embernél marad. Ez különösen fontos olyan determinisztikus és biztonságkritikus rendszerek esetében, ahol a hibás vagy ellenőrizetlenül átvett javaslatok közvetett módon is fizikai veszélyhelyzethez vezethetnek. Az oktatási és ipari környezet közös feladata ezért az, hogy ne csupán az MI alkalmazását, hanem annak biztonságtudatos és kritikusan ellenőrzött alkalmazását is előtérbe helyezze.

A jelen tanulmány koncepcionális és elemző megközelítésben vizsgálta az MI-alapú mérnöki asszisztencia biztonsági kérdéseit az ember-robot interakció összefüggésében. A további kutatások célja egy olyan empirikus vizsgálati keret kialakítása, amely a CNC technológiai tervezés kontextusában elemzi az MI használatából eredő döntési és validációs kockázatokat. Ennek első lépéseként jelenleg előkészítés alatt áll egy kvalitatív, exploratív vizsgálat, amely az MI által generált megoldások hibáinak típusait, valamint a felhasználói validáció jellemző mintázatait kívánja feltárni. Ezt követően egy kvantitatív kutatás hasonlíthatja össze az MI nélküli, az MI-t használó, valamint a strukturált validációs protokoll mellett dolgozó csoportok teljesítményét. A tervezett kutatási irány hozzájárulhat annak pontosabb megértéséhez, hogy az MI milyen módon befolyásolja a döntéshozatalt, a hibafelismerést és a biztonságkritikus környezetekben szükséges emberi ellenőrzést.

A CNC-programozás oktatásában mindez az oktatási hangsúly részleges átrendeződését is indokolhatja. A programkód önálló előállítása mellett egyre nagyobb szerepet kaphat az MI által generált megoldások értelmezése, a hibák felismerése, a technológiai döntések indoklása és a végrehajtás előtti szakmai ellenőrzés. Ez az oktatói felkészülést is módosíthatja, mivel az előadások és gyakorlatok tervezése során célszerű olyan feladatokat kialakítani, amelyek nemcsak a helyes megoldás elkészítését, hanem annak kritikai értékelését és biztonságos alkalmazhatóságának megítélését is fejlesztik. Ez a megközelítés a tervezett kvalitatív esettanulmányban is vizsgálható, különös tekintettel arra, hogy a hallgatók milyen hibákat ismernek fel az MI által generált megoldásokban, és milyen ellenőrzési stratégiákat alkalmaznak azok értékelése során.

FELHASZNÁLT IRODALOM

[1] National Institute for Occupational Safety and Health, “Robotics in the Workplace: An Overview,” Centers for Disease Control and Prevention, 2024.

[2] ISO, ISO 10218-1:2025 Robotics — Safety requirements — Part 1: Industrial robots, Geneva: International Organization for Standardization, 2025.

[3] E. Tabassi et al., Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, National Institute of Standards and Technology, 2023, doi: 10.6028/NIST.AI.100-1.

[4] ISO, ISO 10218-2:2025 Robotics — Safety requirements — Part 2: Industrial robot applications and robot cells, Geneva: International Organization for Standardization, 2025.

[5] ISO, ISO/TS 15066:2016 Robots and robotic devices — Collaborative robots, Geneva: International Organization for Standardization, 2016.

[6] Cs. Kollár, „Szerethetők-e a robotok? Az ember-robot interakció humán oldalának teoretikus aspektusa,” Hadtudomány, vol. 26, különszám, pp. 142–154, 2016, doi: 10.17047/HADTUD.2016.26.K.142.

[7] C. Autio et al., Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1, National Institute of Standards and Technology, 2024, doi: 10.6028/NIST.AI.600-1.

[8] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU)

2020/1828 (Artificial Intelligence Act), Official Journal of the European Union, L 1689, 12 Jul. 2024.
[9] A. Holzinger, K. Zatloukal, and H. Müller, “Is human oversight to AI systems still possible?,” New Biotechnology, vol. 85, pp. 59–62, 2025.