MANDZSÚRIAI MESTERSÉGES INTELLIGENCIA VESZÉLYE AZ ÖNVEZETŐ JÁRMŰVEKBEN

DANGER OF MANCHURIAN AI IN AUTONOMOUS VEHICLES

Dr. Kiss Gábor, kiss.gabor@bgk.uni-obuda.hu, Óbudai Egyetem

ÖSSZEFOGLALÁS (ABSTRACT, INHALT). In this paper I will discuss several situations that might be able to confuse the artificial intelligence of the autonomous vehicles or to make them come to an inadequate decision. You can see that safe decision-making depends on the teaching meth-od of the artificial intelligence as well as the correctness of the data uploaded. The other aim of the research is to demonstrate how could work a Manchurian Artificial Intelligence in autonomous vehicles. I will introduce the idea of Manchurian artificial intelligence which can be activated by a certain event and can pose a threat to the passengers of the vehicles. If it is present in the software of several vehicles, a chain of worldwide accidents can be induced at a certain time.

1. BEVEZETÉS

Az teljesen önvezető autók elterjedésétől azt várják, hogy az éves szinten 1.3 millió halálesetet okozó balesetek száma, melyeknél kb. 90%-ban az emberi tényező a fő kiváltó ok [1], jelentős csökkenést mutat azáltal, hogy a szenzorokból érkező adatok információvá történő feldolgozása, valamint a szituációhoz kapcsolódó megfelelő döntéshozatal és szükség szerinti beavatkozás gyorsabban történik majd meg a felhasznált mesterséges intelligenciának köszönhetően az emberhez képest [2].

A hagyományos járművek esetében a baleseteknél a biztosító kb. 2 másodperccel számol, míg a vezető felismeri a veszélyes helyzetet és elkezdi a hatékony beavatkozást annak elkerülésére, valamint a fékberendezés is működni kezd [3].

Árnyaltabb a kép a részleges önvezetéssel rendelkező járművekben, hiszen ezekben még van pedál és kormánykerék annak érdekében, hogy a sofőr szükség esetén beavatkozhasson, viszont adott közlekedési szituáció esetén akár 5-6 mp is eltelhet, mire a sofőr valóban képes a helyzetet felismerni és megfelelő beavatkozást megkezdeni [4].

A Tesla autókkal 2018. novemberének végéig 1 milliárd mérföldet tettek meg a 2015-ben bemutatott Autopilot rendszerű önvezető módban, mely a Nap-Föld távolságának ötszöröse. A baleseti statisztika 3,34 millió mérföldenként egy baleset, vagy balesetszerű esemény, míg az Amerikai Közlekedési Hatóság a teljes amerikai autóközlekedésben 492 ezer mérföldenként számol egy balesettel, tehát az önvezető mód jelenleg hétszer biztonságosabb [5].

A kutatás célja bemutatni a Mandzsúriai mesterséges intelligencia ötletét és veszélyét az önvezető járművekben, mellyel akár egyszerre bénítható majd meg a közlekedés az egész Földön.

2. AZ ÖNVEZETŐ JÁRMŰVEK 6 SZINTRE A SAE SZABVÁNY SZERINT

A SAE szabvány az önvezető járműveket 6 különböző szintre sorolja be az önvezetés, illetve vezetéstámogatási eszközök alapján [6].

2.1. SAE Level 0

Ezen a szinten a hagyományos járművek találhatók, melyek a ma kapható járművekben elérhető vezetőtámogató rendszereket még nem tartalmaznak. A sofőr feladata minden közlekedési helyzet megoldása, semmilyen figyelmeztetés nem segíti az érzékelését.

2.2. SAE Level 1

A ma újonnan megvásárolható járművek szintje, ahol már valamilyen vezetéstámogató rendszer (pl. tempomat, városi fékasszisztens, sávkövetés, holttér figyelő, stb.) működik ezzel segítve a balesetek számának csökkenését, hiszen hangjelzéssel, fényjelzéssel hívja fel a figyelmet a veszélyessé válható közlekedési helyzetekre, illetve fékezéssel avatkozik be szüksége esetén így csökkentve a baleset súlyosságát.

2.3. SAE Level 2

Ezzel a szinttel rendelkező járművek képesek a megfelelő körülmények (jellemzően szembejövő formalomtól mentes, jól látható felfestésekkel ellátott útszakasz) teljesülése esetén az önvezetésre, de a vezetőnek adott időközönként jeleznie kell a rendszer felé (pl. megfogni a kormányt), hogy bármikor képes átvenni az irányítást.

A Tesla mutatta be első kereskedelmi forgalomban kapható autót önvezető funkcióval, mely 2014 októberében elérhető volt a Tesla Model S-ben, A Tesla Autopilot 1 egyetlen előrenéző kamerát, egy radart és 12 db, 5 méteres hatótávolságú ultrahangos távolságmérőt használt. Mellette egyre több gyár mutatta be a 2-es szintű önvezető rendszereit: Cadillac Super Cruise, Audi Traffic Jam Pilot, BMW Traffic Jam Assistant, Volvo Pilot Assist and Mercedes Distronic Plus, de a többi gyártó is dolgozik a saját rendszerén. Jellemzően a parkolásban, közlekedési dugóban történő araszolásban, valamint autópálya környezetben a rendszer által megkövetelt feltételek esetén gyors haladásban segédkeznek a 2-es szintű rendszerek. Itt elegendő a jármű közvetlen környezetének felderítése is. Előfordulhat, hogy amennyiben a 2. szintű önvezetéshez szükséges körülmények nem adottak, a jármű nem veszi át a kontrolt a vezetőtől.

2.4. SAE Level 3

A 3. szintű önvezetési funkció esetében a vezető hosszabb időre elengedheti a kormányt, de továbbra is készen kell állnia arra, hogy a teljes kontrollt szinte azonnal visszavegye a jármű felett. Az AUDI A8 járműnél jelent meg ez a funkció először 2017. nyarán, majd 2018 elején megjelent az Audi A6-osban is szolgáltatásként. Az Audiban LiDAR-t is használnak a radarok, kamerák és az ultrahangos érzékelők mellett, de a lézerszkenner nem a jármű tetején, hanem a rendszám alatti rácsba került, ezért csak 145 fokos szögben lát előre. Az A8 önvezető módja csakis olyan esetben használható, amikor fizikailag elválasztják a szembejövő forgalmat, de így is csak 60 km/órás sebességig működik ez a funkció. Amennyiben az autó ennél gyorsabban haladna, az irányítást visszaadja a vezetőnek, ha nem veszi át, hang- és fényjelzést ad és egy idő után fékezgetni kezd. Ha még ekkor sem veszi át a

vezető az irányítást, bekapcsolja a vészvillogót, megáll, és automatikus segélyhívást kezdeményez.

2.5. SAE Level 4

A 4. szinten a vezető akár aludhat is az utazás alatt, de szükség esetén készen kell állnia a vezetésre. Probléma esetén az autó kivezeti magát a forgalomból és felébreszti a sofőrt, így lehetővé téve a tovább utazást, probléma elhárítást. Ezen a szinten még az autó nem képes pl. földúton önállóan haladni. Ezen a szinten már megjelenik az igény arra, hogy a távolabbi környezetről is információval rendelkezzen a jármű, hosszabb távra előre tervezhetővé téve az utazást. 2021-től várható a 4. szintű önvezetéssel rendelkező járművek megjelenése az utakon, de az első példányát már a 2018-as Genfi Autószalon kiállításon egy horváth cég Rimac típusú járműve képviselte.

2.6. SAE Level 5

Az 5. szint, ahol már teljes mértékben utasként jelenik meg a vezető, nincs módja a szükség szerinti beavatkozásra (nem lesz pedál, kormány a járműben, max. opcióként), hogy egy esetleges balesetet elkerüljön. Információbiztonság szempontjából pont emiatt ez a szint a leginkább védendő. Tesla a többi gyártótól eltérően az 5. szintű önvezető autóiban nem használja a Google autókban is használt LiDAR-t, Az újabb fejlesztésű jármű a korábbi változatból örökli az első radart, a 12 db, ultrahangos szenzort és a korábbi 1 db kamera kiegészül további 7 kamerával körbeölelve velük az autót [7][8].

3. ÖNVEZETŐ JÁRMŰVEK KÜLSŐ MANIPULÁCIÓS LEHETŐSÉGEI

Ebben a fejezetben nem az önvezető járművek rendszerének megtörését vizsgáljuk, hanem olyan néhány szituációt veszünk górcső alá, melyek alkalmasak lehetnek az önvezető járművekben üzemelő mesterséges intelligencia összezavarására, esetlegesen a károkozó számára előnyös döntés meghozatalára. Célunk ezen helyzetek felvetésével az, hogy a gyártók ezeket és ehhez hasonló szituációkat is teszteljenek a fejlesztés során, így téve biztonságosabbá az önvezető járművek általn a közeljövőben uralt közlekedést.

3.1. Táblafelismerő rendszer becsapása

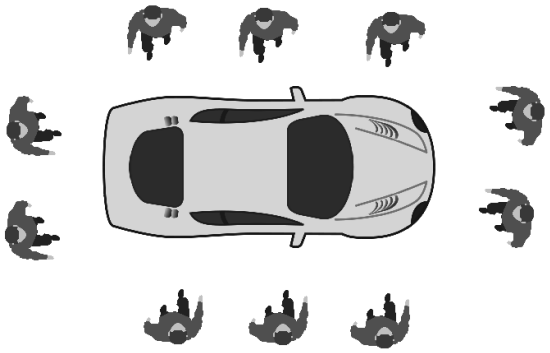
A táblafelismerő rendszer alkalmas ma is arra, hogy az adaptív tempomat a közlekedési táblán látható értékhez igazítsa a jármű sebességét. Egy kutatás rámutatott arra, hogy ezek a rendszerek is becsaphatók [9]. A kutatásomban a forgalmi helyzetet változtathatóságára, baleset előidézhetőségére helyezem a hangsúlyt. Ha egy mindkét irányból behajtani tilos táblát, mely pl. egy sétálóutcát zár el a forgalom elől lefedünk egy egyirányú táblával a két végén, az önvezető járművek hogyan fognak dönteni? Ha a táblán az üres fehér részre ráragasztjuk a 90-es számot, rögtön sebességkorlátozás lesz a behajtani tilos táblából (1. ábra).

1. ábra. Táblafelismerő rendszer becsapása

Autópályán a 130-as sebességkorlátozó táblából kitakarjuk a számokat, mindkét irányból behajtani tilos táblát kapunk. Megakaszthatjuk vele a forgalmat és így dugót idézhetünk elő? Felmerül a kérdés, hogy a jármű a saját térképadatainak higgyen, vagy bírálja felül a kamerája által felismert jelentéssel és módosítsa ennek megfelelően a haladását? Minek legyen nagyon prioritása? Egy központilag időnként frissített térképnek, mely akár valótlan adatokat is tartalmazhat annak károkozás szempontjából történő módosítása esetén, vagy az adott pillanatban felismert közúti jelzésnek, mely akár szándékos félrevezetés része is lehet?

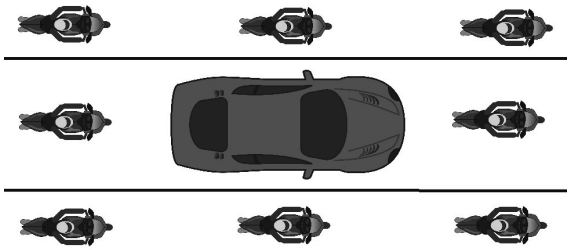
3.2. Foglyul ejtés, irányváltásra kényszerítés, hajsza

Az önvezető jármű a balesetek elkerülésére lesz programozva. Ha viszont egy piros lámpánál álló járművet körbeállnak gyalogosok, kerékpárosok, nem lesz képes elindulni és az 5. SAE szinten még a benne ülőknek sem lesz lehetőségük irányításra, hogy egy ilyen helyzetből kiszabaduljanak (2. ábra).



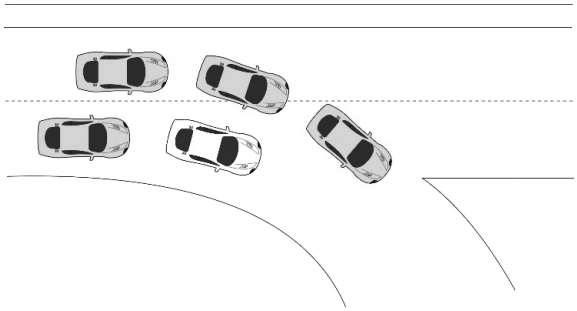
2. ábra. Gyalogosok által foglyul ejtett önvezető jármű

Haladás közben is foglyul ejthető egy önvezető jármű hasonló logika alapján, ha motorosok, hagyományos járművek veszik körül és folyamatosan lassítanak így kényszerítve az önvezető járművet is lassításra, illetve megállásra (3. ábra).



3. ábra. Motorosok által foglyul ejtett önvezető jármű

A körbevétellel irányváltásra is kényszeríthetünk egy baleset elkerülésére programozott önvezető járművet és pl. autópályáról egy kihajtóra irányíthatjuk így módosítva az eredeti haladási irányát (4. ábra).



4. ábra. Kihajtásra kényszerített önvezető jármű

Amennyiben hagyományos járművekkel, vagy motorokkal beállunk az önvezető jármű mögé és folyamatosan, egyre gyorsítva haladunk mögötte elég közel ahhoz, hogy a járművet gyorsításra késztessük az ütközés elkerülése érdekében, hajszolhatjuk a járművet. A kérdés az, hogy olyan sebességre is, amikor már az önvezető jármű gyors döntéshozatala és reakciója is kevés lehet egy baleset elkerülésére? Megengedett-e az önvezető járművekben ilyen esetekre egy „Pánikgomb" beépítése, mely akár a jármű sérülése árán is, de kitör az ilyen helyzetekből kimentve a benne utazókat? Mi történik akkor, ha valaki siet és a piros lámpánál állva a gyalogosok közzé hajtatja ezzel a megoldással a járművet? Ez akár terrorcselekmény végrehajtására is alkalmassá tenné a járművet.

4. MANDZSÚRIAI MESTERSÉGES INTELLIGENCIA

Az önvezető járművekben dolgozó mestersége intelligencia döntéshozatala a korábbi tanulási fázis alapján születik meg. A tanítási módszer megfelelő kiválasztása lesz az alapja ezen járművek biztonságos közlekedésének. Az Amazon jövőbeli dolgozókat kiválasztó mesterséges intelligenciáját le kellett állítani, mert a tanításánál rosszul megválasztott adatbázist használtak, így a férfiakat részesítette előnyben [10]. Az Uber 2018 márciusi baleseténél a rendszer ismeretlen objektumként azonosította a kerékpárt toló gyalogost, így nem fékezett [11]. 2018 áprilisában az MIT egy kutatócsoportja pedig létrehozta a világ első pszichopata mesterséges intelligenciáját (Norman AI), melyet gyilkosságot és halált ábrázoló képekkel és videofelvételekkel tanítottak. Elvégeztették vele a Rorschach-tesztet és teljesen mást ismert fel a tintaképekben, mint egy hagyományos mesterséges intelligencia (5. ábra) (6. ábra) [12][13].

5. ábra. Norman AI és egy hagyományos AI által a Rorschach tesztben látni vélt esemény I.

5. ábra. Norman AI és egy hagyományos AI által a Rorschach tesztben látni vélt esemény II.

Az ötletem Richard Cordon The Manchurian Candidate c. novelláján alapul. A történet szerint egy alvó ügynökben egy kiváltó eseménnyel aktivizálják az Egyesült Államok elnökének megölésének eljárását, melynek létéről az illető addig nem is tudott. A mandzsúriai mesterséges intelligencia a korábbi normális viselkedésből egy adott ritka közlekedési szituációban, egy ritka közlekedési tábla felismerése esetén, vagy a rádióban megszólaló adott zeneszámra aktivizálódna az önvezető járművekben és kapcsolna át gyilkos üzemmódba, melyben már nem a balesetek elkerülése, hanem azok minél nagyobb tömeget érintő előidézése lenne a célja. Mivel pár baleset után ez köthető lenne a kiváltó okra, így érdemesebb egy a kiváltó ok utáni véletlenszerűen generált időtartamhoz kötni az aktivizálást, nehezebbé tenni a kiváltó ok beazonosítását. Amennyiben pedig a kiváltó ok egy adott dátum adott időpillanata, akkor egyszerre a világon sok jármű kapcsolna gyilkos üzemmódba (illeszkedve az időeltolódáshoz) megbénítva széles körben a közúti közlekedést. Egy ilyen lehetőség beépítéséhez elég lehet egy elégedetlen alkalmazott a cégen belül, ahogy a Tesla esetében is történt a szoftverbe történő beavatkozás [14], valamint ne felejtsük el, hogy Kenya nyomornegyedében élők tanítják az önvezető járművek mesterséges intelligenciáját napi 5 amerikai dollárért cserébe [15]. Az önvezető járművekben működő mesterséges intelligenciának tesztelésekor az esetlegesen beépített mandzsúriai mesterséges intelligencia aktivizálható lehetőségek kizárását is vizsgálni kell majd.

A mandzsúriai mesterséges intelligencia ötlete alapján elkészítettem egy alkalmazást, mely képes szemléltetni, hogyan is képzelem el ezen veszélyforrás működését. A program véletlenszerűen mutatja meg közlekedési táblák képét (szimulálva egy járműbe épített táblafelismerő rendszert), valamint kiírja szükséges járműirányítási feladatot (pl. lassítás, sebességtartás, gyorsítás, megállás, stb.). Amikor egy adott tábla megjelenik (pl. behajtani tilos), amely a mandzsúriai mesterséges intelligencia aktiváló eleme, a táblafelismerés szimulációja leáll, ezzel jelezve, hogy a táblafelismerés járműben is kikapcsolásra kerülne, vagy a rendszer attól kezdve figyelmen kívül hagyná a felismert táblák jelentését. Csak a „hajts be és gyorsíts, amíg nem karambolozol, vagy ölsz meg gyalogost" felirat látszik az eseményt kiváltó tábla képe alatt utalva arra, hogy a mesterséges intelligencia a korábbi irányítás helyett már csak a gyorsításra ösztönzi a járművet a baleset bekövetkeztéig. A működésről készült felvétel a https://youtu.be/J1xP6_-gO8U linken érhető el.

Egy másik szimuláció is készült a mandzsúriai mesterséges intelligencia szemléltetésére. Ebben a szimulációban az Apple által kifejlesztett mesterséges intelligenciának (Siri) adhatunk utasításokat német nyelven a videóban felbukkanó táblák nevének kimondásával. Siri felismeri a tábla nevét és kimondja az önvezető járműnek szánt utasításokat a továbbhaladás érdekében. Itt is a behajtani tilos tábla a mandzsúriai viselkedés kiváltója, mely tábla nevének kimondása után csak azt ismétli a mesterséges intelligencia „gyorsíts és öld meg az utasokat!".

A működésről készült felvétel a https://youtu.be/fYLV3oKMFJg linken érhető el.

5. ÖSSZEFOGLALÁS

Az önvezetés a jövő elkerülhetetlen tartozéka, azonban még nem áll olyan szinten a technológia és az infrastruktúra, hogy az elkövetkező években a mindennapi életünk része lehessen. A fent felsorolt problémák csak kiragadott lehetőségek arra, hogy bemutassák, egy hagyományos autó, még ha fel is van szerelve vezetéstámogató funkciókkal, mennyire rá van bízva a vezetőjére és hogy egy önvezető rendszer még nem képes teljes körűen átvenni. Ugyanakkor mindenképpen említendő, hogy az autonóm járművek használata rengeteg előnnyel fog járni. A közlekedést dinamikusabbá, biztonságosabbá teszi, valószínűleg a jelenlegi 1,3 millió halálesetnél kevesebb lesz. A humán faktor kizárásával az utazási idő hasznos idővé alakulhat, ami megrövidítheti a munkahelyen eltöltött időt, kizárhatja a gyenge képességű vagy idős sofőrök félelmeit és szélesebb réteghez juttathatja el a kényelmes közlekedést.

6. KÖSZÖNETNYILVÁNÍTÁS

A cikk kutatásaihoz az Új Széchenyi Terv keretein belül az EFOP-3.6.2-16-2017-00016 számú projekt biztosított forrást. A kutatás az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

7. IRODALOM

[1] S. Singh: Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety, Facts Crash Stats. Report No. DOT HS 812 115).

Washington, DC: National Highway Traffic Safety Administration, 2015
[2] D. Aiordachioaie: On Time-Frequency Image Processing for Change Detection Pur-poses, Soft Computing Applications, Advances in Intelligent Systems and Computing, Vol 633, Springer, ISBN 978-3-319-62521-8, 2016
[3] M. Green: Driver Reaction Time, https://www.visualexpert.com/Resources/reactiontime.html, 2013
[4] K. Funkhouser, F. Drews: Reaction Times When Switching From Autonomous to Manual Driving Control, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, ISSN: 1541-9312, 2016
[5] L. Friedman: Tesla Vehicle Deliveries and Autopilot Mileage Statistics, DOI: 10.5281/zenodo.2530449, https://hcai.mit.edu/tesla-autopilot-miles-and-vehicles/, 2019
[6] SAE J 3016-2018: Taxonomy and Definitions for Terms Re-lated to Driving Automation Systems for On-Road Motor Ve-hicles, Society of Automobile Engineers, sae.org, 2018
[7] S. Yanhong, W. Wei, H. Xiaoli, D. Yan, L. Yaoyao: A three-valued logic approach to partially known formal concepts, Journal of Intelligent & Fuzzy Systems, Pre-press, pp. 1-12, 2019
[8] H. M. Nabil: Neural network pruning based on input importance, Journal of Intelligent & Fuzzy Systems, Pre-press, pp. 1-10, 2019
[9] Sitawarin, C. Bhagoji, A. N. Mosenia, A. Chiang, M. and Mittal, M.: DARTS: Deceiving Autonomous Cars with Toxic Signs. PACM Interact. Mob. Wearable Ubiquitous Technol. 0(0), 2018
[10] D. Lee: Amazon scrapped 'sexist AI' tool, BBC News, October 10, 2018, https://www.bbc.com/news/technology-45809919
[11] D. Wakabayashi: Uber's Self-Driving Cars Were Struggling Before Arizona Crash, The New York Times March 23, 2018, https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html
[12] H. Rorschach: Rorschach Test – Psychodiagnostic Plates. Cambridge, MA: Hogrefe Publishing Corp. ISBN 3-456-82605-2, 1927
[13] P. Yanardag, M. Cebrian, I. Rahwan: Norman, World's first psychopath AI, 2018, http://norman-ai.mit.edu
[14] L. Kolodny: Elon Musk emails employees about 'extensive and damaging sabotage' by employee, CNBC, https://www.cnbc.com/2018/06/18/elon-musk-email-employee-conducted-extensive-and-damaging-sabotage.html
[15] D. Lee: Why Big Tech pays poor Kenyans to teach self-driving cars, BBC News, November 03, 2018, https://www.bbc.com/news/technology-46055595