

Barát vagy ellenség? A mesterséges intelligencia erkölcsi
cselekvőképességének ábrázolásáról a
filmművészetben

Veera Asukas
Mesterszeminárium és
szakdolgozat
Angol
nyelvek és irodalom
Bölcsészettudományi
Kar Oului Egyetem
2022 ősz

Tartalomjegyzék

1.	Bevezetés	3
2.	Kutatási megközelítés.....	5
2.1	Az AI filozófiája	6
2.1.1	Mi az a mesterséges intelligencia?	7
2.1.2	Szemiotikus hierarchia: a mesterséges intelligencia ugródeszkái	9
2.1.3	Erkölcsei képviselő	13
2.1.4	Mesterséges intelligencia és szabad akarat.....	17
2.2	A filmek feltárása tartalomelemzéssel	18
3.	A mesterséges intelligenciák morális ügynöksége a filmekben	22
3.1	Én, a robot (2004).....	25
3.1.1	Sonny, a mesterséges intelligencia érzelmekkel	25
3.1.3	VIKI, az AI szent akarat.....	30
3.2	Anyá vagyok (2019)	34
3.2.1	Anyá, az emberiség matriarchája	35
3.3	Ex Machina (2014).....	39
3.3.1	Ava és Kyoko, a bebörtönzött mesterséges intelligenciák	39
4.	A filmek irányított tartalomelemzése	44
4.1	Szemiotikus hierarchia	45
4.2	Erkölcsei elmélet	49
4.3	Szabad akarat	51
4.4	Érzelmi kapacitás	52
4.5	A mesterséges intelligenciák narratív és társadalmi szerepei.....	54
4.6	A mesterséges intelligenciák által végrehajtott főbb erkölcsi cselekedetek.....	56
4.7	A mesterséges intelligenciák narratív következtetése	59
5.	Mitől lesz egy mesterséges intelligencia erkölcsileg jó vagy rossz?.....	60
6.	Következtetés	65
6.1	A mesterséges intelligencia jövője	66
7.	Források.....	68

1. Bevezetés

A mesterséges intelligencia (AI) évtizedek óta népszerű téma az irodalomban és a filmművészetben. Bár a valós életben az AI-technológia fejlődése még nem érte el a fiktív média képzeletbeli fokát, talán mégsem olyan távol van tőlünk, mint gondolnánk. A mesterséges intelligencia és annak tanulmányozása napjainkban elképesztően gyorsan fejlődik, egy témáról azonban mintha sokkal kevesebb szó esne, mint másokról: a mesterséges intelligencia erkölcsi cselekvőképességéről.

Míg a hírmédia inkább a mesterséges intelligencia létrehozásának technológiai aspektusaira összpontosít, addig a fikciós média a mesterséges élet létrehozásának erkölcsi nehézségeit vizsgálja. Ez a dolgozat azt vizsgálja, hogyan ábrázolják a mesterséges intelligenciák erkölcsi cselekedeteit filmes kontextusban. Arra a filozófiai kérdésre összpontosít, hogy mi a mesterséges intelligencia, hogyan tudnak erkölcsi ágensként cselekedni, és hogyan ítélik meg erkölcsi cselekedeteiket.

A mesterséges intelligenciák morális ügynöksége olyan téma, amelyet már korábban is vizsgáltak, de nem sok tanulmány készült a fiktív mesterséges intelligenciák vizsgálatáról. Emiatt a dolgozat elméleti háttérét olyan tanulmányok alkotják, amelyek a valós életbeli mesterséges intelligenciafejlesztésekre összpontosítanak, hogy viszonyítási pontot nyerjenek a már létező mesterséges intelligenciákról. Az elméleti háttérből kirajzolódó legelterjedtebb fogalmakat ezután filmes kontextusban alkalmazzuk. A dolgozat filozófiai keretét elsősorban egy esszégyűjtemény alkotja, amely a valós életben felmerülő problémákat és a morális cselekvésre képes mesterséges intelligenciák létrehozásának lehetőségeit tárgyalja. Ebből a gyűjteményből választottam ki a leglényegesebb esszéket, amelyek olyan témákat tárgyalnak, mint az erkölcsi cselekvőképesség, a szabad akarat és a gyenge és erős programozás, amelyek a dolgozat szempontjából a legkiemelkedőbb elméleti pontokat jelentik. Míg ezek az esszék egy teljesen megvalósult mesterséges intelligencia morális dilemmáit tárgyalják, addig Jordan Zlatev egyik cikke az értelemelmélet feltételeit tárgyalja. E tanulmány elméleti háttérének kiegészítésére ezt az elméletet választottuk, mert összeegyeztethető az AI létrehozásával és annak főbb bonyodalmaival. Az erkölcsi cselekvőképesség kifejezéséhez képesnek kell lenni a jelentés közvetítésére és megértésére. A szemiotikai hierarchia megkülönbözteti a jelentésalkotáshoz szükséges négy szintet: élet, tudat, jelhasználat és nyelv. Az elmélet egy

kényelmes skálát is kínál, amelyhez az AI-k intelligenciája és "erőssége" mérhető. Ahogyan később tárgyaljuk, a szemiotikai hierarchia első szintje a mesterséges intelligencia számára a négy közül a legnehezebben teljesíthető követelmény, amit Zlatev is elismer. Az élet és az erkölcsi cselekvőképesség elválaszthatatlanságának következményeit alaposabban megvizsgáljuk majd a

későbbi szakaszok. A 4. szakaszban (Vita) tartalomelemzéssel fogom elemezni a filmeket, hogy azonosítsam a filmek közötti egységes filozófiai és narratív témákat és jellemzőket. A tartalomelemzést alkalmazkodóképessége és a kiválasztott témák kategorizálását és összehasonlítását segítő "kódok" használata miatt választották. A dolgozathoz kiválasztott három film az I, Robot (2004), az Anyám vagyok (2019) és az Ex Machina (2014). Ezekben a filmekben különböző kompetenciájú, narratív jelentőségű és morális cselekvőképességű mesterséges intelligenciák szerepelnek. A filmek 15 év leforgása alatt kerültek a mozikba, amely alatt a technológia ugrásszerűen fejlődött, vagyis a filmek különböző nézőpontokat kínálnak nekünk az AI-król. Az I, Robot című film Isaac Asimov azonos című regényén alapul, amely 1950-ben íródott. A forrásanyag megjelenési dátumát figyelembe véve a mesterséges intelligencia bemutatásának időszaka közel 70 évet ölel fel.

Ezzel a dolgozatommal azt remélem, hogy vitát kezdeményezhettek a morális mesterséges intelligencia témájáról, és arról, hogy a filmművészet évek óta hogyan tükrözi a morális cselekvésre képes mesterséges intelligencia létrehozásának mostani aggályait és érdekességeit. Továbbá fontos elgondolkodni az emberek és a mesterséges intelligenciák erkölcsi cselekedetei közötti különbségeken, mivel úgy tűnik, hogy eltérés van a felelősség és az igazolás kijelölésében e két csoport között. Az emberek képesek erkölcstelen cselekedetekre, de az, hogy az AI-knak ugyanez a képességük megvan, inkább gyanakvást és kétségeket ébreszt. A mozin keresztül feltárható ez a kettős mérce, és az emberek és a mesterséges intelligenciák közötti ellentéteket vizsgálva elkezdhetjük megválaszolni a következő kérdéseket: Hogyan ábrázolják a mesterséges intelligenciák erkölcsi cselekvőképességét a filmművészetben? A filmekben a mesterséges intelligenciákat erkölcsi cselekvőknek tekintik-e, és az emberek és a mesterséges intelligenciák erkölcsi cselekedeteit másként ítélik-e meg, és ha igen, milyen okból kifolyólag? Hogy a téma valós perspektíváját nyújtsuk, a dolgozat utolsó fejezetében az AI-alkotás néhány legújabb fejleményét tárgyaljuk.

2. Kutatási megközelítés

Mivel a dolgozat célja a mesterséges intelligencia karakterek erkölcsi cselekvéseinek azonosítása és elemzése filmes kontextusban, a filozófia, a szemiotikai hierarchia és az erkölcsi cselekvés háttérét a filmek későbbi elemzésének támogatására nyújtjuk. A filozófiai háttér magában foglalja a mesterséges intelligencia természetéről szóló vitát, amely témában sokféle nézőpont létezik. Először egy rövid összehasonlítás történik az informatika és a filozófusok között a mesterséges intelligencia természetéről, hogy tisztázzuk a mesterséges intelligencia e dolgozatban használt definícióját. Bár az informatikusok nézetei némileg jelen vannak a filozófiai háttérben is, a filmek elemzése elsősorban az ebben a fejezetben bemutatott filozófiai témákat és nézőpontokat használja fel.

Míg a mesterséges intelligencia meghatározása kizárólag metafizikai jellemzők és más, nem kézzelfogható követelmények alapján történhetne, a szemiotikai hierarchia hozzáadása lehetőséget nyújt arra, hogy a mesterséges intelligenciát egy konkrét skálán mérjük. A szemiotikai hierarchia kategorizálja azokat a különböző szinteket, amelyeket egy lénynek el kell érnie a jelentések megértéséhez és alkalmazásához. A hierarchia négy szintre oszlik: élet, tudat, jelhasználat és nyelv. A filmekben szereplő mesterséges intelligencia-karaktereknek elméletileg minden szintet meg kell haladniuk ahhoz, hogy jelentést tudjanak létrehozni, ami elengedhetetlen az erkölcsi döntések meghozatalához. A filmekben a mesterséges intelligencia karakterek jártasságát az egyes szinteken különböző módon közelítik meg, majd ezeket a jártasságokat az elemzési részben összehasonlítják egymással.

E két elmélet segítségével megteremthető az erkölcsi cselekvőképesség alapvető megértése. Mivel a dolgozat érdeklődése a filmekben szereplő mesterséges intelligencia karakterek erkölcsi cselekedeteire irányul, az erkölcsi cselekvőképesség fogalmát is tárgyalni kell. Ha a szemiotikai hierarchiában megfelelő szintű jártasságot és az erős mesterséges intelligencia feltételeit elérjük, akkor a szubjektumnak képesnek kell lennie az erkölcsi cselekvésre. Az erkölcsi cselekvésre képes entitásnak el kell döntenie, hogyan navigáljon az erkölcsi

cselekvést igénylő helyzetekben. Ezeknek a cselekedeteknek az elemzése és összehasonlítása a dolgozat fő célja.

érdeklődés. A filozófiai témákat, az erkölcsi cselekvést és a filmek egyéb narratív szempontból releváns jeleneteit vagy jellemzőit tartalomelemzéssel elemezzük és kategorizáljuk. A dolgozathoz a tartalomelemzést a könnyű alkalmazhatósága, valamint a hatékony összehasonlításhoz szükséges kódok és táblázatok használata miatt választottuk. Míg a tartalomelemzést gyakran használják a társadalomtudományok vagy az orvostudomány területén a beteginterjúk és más hasonló kvalitatív anyagok esetében, a filmekben szereplő konkrét jelenetek és témák azonosítására is kiváló módszert kínál. Ezen elméletek és módszertanok segítségével a kutatási háttér remélhetőleg jó, tömör alapot nyújt a későbbi elemzéshez.

2.1 Az AI filozófiája

E szakasz célja, hogy elmagyarázza a mesterséges intelligencia definícióját, hogy mi minősül mesterséges intelligenciának, és hogy az ebben a szakaszban kifejtett fogalmak miért relevánsak a dolgozat szempontjából. A mesterséges intelligencia természetének tárgyalásakor a mellette szóló érvek nagyjából két nézetre oszthatók: redukcionista nézet és emergens nézet. Az előbbit gyakran támogatják az erős mesterséges intelligencia létrehozására törekvő informatikusok, az utóbbit pedig inkább a filozófusok fogadják el. A redukcionista, pontosabban tudományos redukcionista nézet azt javasolja, hogy "minden tudomány visszavezethető a fizikára" (Alyssa Ney, Internet Encyclopaedia of Philosophy). Ez azt jelenti, hogy bárminek is tekintjük a tudatot, az redukálható fizikai összetevőire, és így a folyamat megfordításával reprodukálható. Röviden, a tudományos A redukcionizmus tagadja a metafizikailag autonóm összetevők létezését. Ilyen nézetet vall Daniel C. Dennett, egy informatikai háttérrel rendelkező filozófus. Az emergens nézet azt vallja, hogy "egy tulajdonság akkor emergens, ha egy rendszer vagy entitás újszerű tulajdonsága, amely akkor keletkezik, amikor a rendszer vagy entitás elérte a komplexitás egy bizonyos szintjét" (IEP). Az emergens tulajdonság itt a tudatra vonatkozna, a rendszer vagy entitás pedig a fizikai komponensre, amely a tudat kialakulásához szükséges. Egy ilyen rendszer lehet a hardver, amire a mesterséges intelligencia épül. Ezt a nézetet támogatja Hubert L. Dreyfus filozófus, aki szerint a tudat több, mint annak fizikai összetevői. A két nézet megkülönböztetése azért fontos, mert a dolgozathoz választott filmekben mindkét

nézet mellett lehet érvelni, amikor a

a mesterséges intelligencia karakterek létrehozása és jellege. Ez a dolgozat azonban inkább az emergens szemléletre összpontosít.

Amint azt a következő részben tárgyaljuk, a mesterséges intelligenciának nincs elfogadott definíciója. Van azonban néhány jól ismert elmélet és definíció, amelyeket a dolgozat elméleti háttérében alkalmazni fogunk. A legjelentősebb a szemiotikai hierarchia, amely a jelentésalkotáshoz szükséges különböző szinteket magyarázó elmélet, és amelyet a mesterséges intelligencia karakterek általános kompetenciájának és az emberektől való különbözőségüknek az elemzésekor fogunk használni. Ezt az elméletet követően a morális cselekvőképesség kerül bevezetésre, hogy viszonyítási pontot adjon a mesterséges intelligenciák által végrehajtott morális cselekedetekhez, amelyek a dolgozat középpontjában állnak. Végül a végén a szabad akarat fontos fogalmát tárgyaljuk, mivel az a filmelemzés során és később a konklúzióban többször is megjelenik.

2.1.1 Mi az a mesterséges intelligencia?

Bár a mesterséges intelligencia tanulmányozását gyakran a számítástechnikai tudományokhoz sorolják, a témával kapcsolatos filozófiai kutatások is bőséges tárháza létezik. A két csoport között azonban nézeteltérések vannak azzal kapcsolatban, hogy mi számít mesterséges intelligenciának.

Az informatikusok gyakran a tudományos naturalisták nézeteit vallják, vagyis a tudományosan tanulmányozható vagy kísérletezhető dolgokat valóságosnak tekintik. Hajlamosak kritikusan viszonyulni a szellem-anyag dualizmus és a metafizikai létezés létezéséhez, és csak az empirikus bizonyítékokat és adatokat tekintik megbízható információforrásnak. Ez magyarázatot adhat arra, hogy sok informatikus számára gyakran miért elég egy olyan mesterséges intelligencia, amely emberhez hasonló teljesítményt mutat, hogy "emberhez hasonló mesterséges intelligenciának" tekintsék, és így nem szükséges a tudatosság bizonyítása. Mindazonáltal vannak olyan filozófusok, akik tudományos naturalistaként is azonosítják magukat, közülük is kiemelkedik Daniel C. Dennett, aki

figyelemre méltó számítógépes háttérrel rendelkezik.

tudományok. A tudattal kapcsolatos álláspontja funkcionális, ami azt jelenti, hogy a gondolkodási tevékenység és az agy közötti kapcsolatot hasonlónak tekinti a számítási tevékenység és az agy közötti kapcsolathoz.

a számítógépet. Ezt az álláspontot megismétli az idézet, miszerint "mi magunk is egyfajta robotok vagyunk", és mint ilyen, a mesterséges intelligencia megteremtése csak egy olyan robot megalkotásának kérdése, amely elég kifinomult ahhoz, hogy úgy teljesítsen, mint egy ember. Bár a mesterséges intelligenciáról alkotott tudományos-naturalista nézet érvényes alap, nem elég átfogó ahhoz, hogy egy mesterséges intelligencia erkölcsi cselekvőképességet mutasson, amit figyelembe kell venni, amikor a mesterséges intelligencia cselekedeteinek hatásait vizsgáljuk.

A mesterséges intelligenciával kapcsolatos filozófiai álláspont kevésbé a programozásra és a létrehozás anyagi követelményeire, hanem inkább a tudatosság, a tanulás és az erkölcs problémájára összpontosít. Keveset tudunk arról, hogyan működik az emberi tudat, nem is beszélve arról, hogyan lehet azt lemásolni, és emiatt rendkívül nehéz határozott választ adni a tudat természetére. Emiatt gyümölcsözőbb filozófiai javaslatokat vagy bizonyos követelményeket felajánlani arra vonatkozóan, hogy mi lehet a tudat, vagy mit igényelhet. A filozófusok és az informatikusok egyetértenek abban, hogy a mesterséges intelligenciának képesnek kell lennie az információ feldolgozására. A filozófusok számára ez felveti az információ és az intelligencia természetének kérdését. Hubert L. Dreyfus szerint az ember három megkülönböztető jellemzője van az információ feldolgozásának: "--- a kétértelműség tűrése, --- [a] képesség a fókusz és a peremtudat elkülönítésére, és --- [a] képesség a lényeges és a lényegtelen megkülönböztetésére" (Technology and Cyberspace, 584). Ahhoz, hogy ezeket a tulajdonságokat el tudja érni, egy emberhez hasonló mesterséges intelligenciának, amelyet ebben a dolgozatban vizsgálunk, képesnek kell lennie tanulni és kontextust alkalmazni a feldolgozott információra.

Bár a mesterséges intelligencia definícióját illetően nem született konszenzus, az erős és a gyenge mesterséges intelligencia között van néhány elfogadott különbség. Az erős mesterséges intelligencia egy kézzel létrehozott tudat, amely "-- nagyon hasonlít a miénkhez. Vannak kellemes és fájdalmas tapasztalatai, élvez vagy szenved bizonyos élményektől, rendelkezik képzelőerővel, memóriával, kritikus gondolkodással és morális cselekvőképességgel". (Basl, 18). Az erős mesterséges intelligencia abban különbözik a "gyenge" mesterséges intelligenciától, hogy "'erős' mesterséges intelligencia mesterséges személyeket kíván létrehozni: olyan gépeket, amelyek rendelkeznek a l a l a mentális képességekkel, amelyekkel mi rendelkezünk, beleértve a fenomenális tudatot is. A "gyenge"

mesterséges intelligencia ezzel szemben olyan információfeldolgozó gépek létrehozására törekszik, amelyek látszólag az emberi személyek teljes mentális repertoárjával rendelkeznek". (Stanford Encyclopedia of Philosophy). Lényegében az erős és a gyenge mesterséges intelligencia közötti különbség az öntudatos tudatosság lenne; a

az önmagunkról, mint egyéni identitásról való tudás. Az erős mesterséges intelligenciának képesnek kell lennie arra is, hogy morális ágensként cselekedjen, ahol "az erkölcsi cselekvés alapjaiban az erkölcsi jelentésen múlik: egy ágens felelősségre vonása tetteiért feltételezi, hogy ezek a tettek morálisan értelmesek mind az ágens, mind a megfigyelők közössége számára" (Parthemore & Whitby, 8). A következő szakaszok kifejtik, hogy milyen szükséges követelmények szükségesek ahhoz, hogy egy ágens kifinomult erkölcsi ágensnek minősüljön. Az elméleteket a fizikai követelményekben gyökerező elméletektől kezdve, mint a szemiotikai hierarchia, a beágyazottság és a megtestesülés, majd az olyan elméletektől kezdve, mint a morális cselekvőképesség és a szabad akarat, magyarazzuk.

2.1.2 Szemiotikus hierarchia: a mesterséges intelligencia ugródeszkái

Mint korábban említettük, nincs végleges definíció arra, hogy mi számít mesterséges intelligenciának, vagyis az AI meghatározása eseti alapon történhet, különösen filmes kontextusban. A jól ismert filozófiai elméletek összegyűjtésével és kombinálásával azonban megalkotható egy általános keret azon tulajdonságokról, amelyeket gyakran megkövetelnek egy erős mesterséges intelligenciától. Ebben a dolgozatban két előfeltételt adunk meg ahhoz, hogy egy mesterséges intelligencia erős mesterséges intelligenciának minősüljön, hogy egyszerűsítsük a vele szemben támasztott komplex követelményeket:

1. A mesterséges intelligenciának egy teljesen megvalósult öntudatos tudatosságnak kell lennie, amely átmegy a szemiotikai hierarchián.
2. A mesterséges intelligenciának képesnek kell lennie erkölcsi ágensként cselekedni.

A filozófiai kutatóközösségen belül vita folyik arról, hogy szükség van-e erkölcsi megértésre egy teljesen működőképes, emberhez hasonló mesterséges intelligencia létrehozásakor. Ebben az összefüggésben azonban elengedhetetlen, hogy a mesterséges intelligencia cselekedeteit etikai szempontból vizsgáljuk, ezért a második követelmény. Az első követelmény garantálja, hogy a mesterséges intelligencia egy kifinomult és kompetens ágens legyen, akit "emberhez hasonlónak" tekinthetünk. Ezenkívül ahhoz, hogy morális ágensnek tekinthessük, meg kell értenie, mi az erkölcsi jelentés. Ez felveti a kérdést, hogy mi a morális jelentés definíciója, és hogyan tudjuk alkalmazni. Az első kérdésre a szemiotikai hierarchia kínál egy lehetséges

megoldást. Jordan Zlatev tanulmányában alaposan kifejti a szemiotikai hierarchia felépítését és felhasználását. A szemiotikai hierarchia célja, hogy "felvázolja

a jelentés általános elmélete --- amely a jelentés szerveződésének négy fő szintjét különbözteti meg: Az élet, a tudat, a jelfunkció és a nyelv" (1). A szemiotikai hierarchia a négy megkülönböztetett szinttel széleskörű és könnyen alkalmazható alapot kínál a mesterséges intelligencia szükséges funkcióihoz. A szemiotikai hierarchia elméletét azért választottuk, mert mind a filozófiai, mind a bioszemiotikai elméleteket felhasználja. Zlatev elveti a jelentéselmélet tisztán bioszemiotikus nézőpontját (amely kizárná a mesterséges életet), és ehelyett fenomenológiai hatásokat épít bele, ami lehetővé teszi az elmélet szélesebb, általánosabb kontextusban való felhasználását. A szemiotikai hierarchia olyan elmélet, amely nagymértékben támaszkodik, és bizonyos részeiben meg is követeli a biológiai organizmust, mint az elméletet kivetítő szubjektumot. Zlatev azonban nem biológiai, pl. mesterséges életet is bevon a definíciójába, amikor az 1. szint, az élet követelményeit határozza meg. A Zlatev által az étellel szemben támasztott követelmény az autopoézis és az intrinzik értékrend. Az autopoietikus szervezet olyan homeosztatisz gépezet, amely képes belső visszacsatolásra, regenerálódásra és újjászerveződésre (Maturana és Varela, 1980). Bár a belső visszacsatolás lehetséges a robotok esetében, de nem képesek olyan szintű regenerálódásra és újjászerveződésre, amire egy autopoietikus entitásnak szüksége van. Ezenkívül az entitást egy belső értékrendnek kell irányítania, ami azt jelenti, hogy az entitás saját érdekeit szolgálja, nem pedig egy kívülről meghatározott funkciót követ (Zlatev, 2009). Ha azonban létrejönne egy autopoézisre képes, belső értékrenddel rendelkező, valós robot, az átmenne a szemiotikai hierarchia első szintjén.

Zlatev azzal kezdi tanulmányát, hogy elismeri, hogy a szemiotikai hierarchia nem tévedhetetlen megoldás a jelentéselméletre. Elismeri, hogy nem ad választ olyan kérdésekre, mint például arra, hogy hogyan megy át az anyag az életbe (a szemiotikai hierarchia első szintje előtti lépcsőfok), ehelyett azt sugallja, hogy ez lehet a bioszemiotika és Zlatev elmélete közötti kapcsolódási pont. Azt is elismeri, hogy elmélete valóban meglehetősen tág, és az egyes szintek túl egyszerűnek tűnhetnek, különösen a primatológia, az idegtudományok és a gyermekfejlődés és azok empirikus bizonyítékainak fényében (3). Az elmélet azonban alkalmazható a mesterséges intelligencia komplexitásának és kompetenciájának értékelésénél. Ezen túlmenően a szemiotikai hierarchia mérőskálaként használható az egyes szinteken a mesterséges intelligencia kompetenciaszintjének mérésére, mivel az elemzett filmekben a mesterséges intelligencia szereplőinek általános kompetenciája változó. A későbbi fejezetekben tovább vizsgáljuk e szintek lehetséges bonyodalmaikat és a mesterséges

intelligenciáról alkotott képzeletünkkel való kapcsolatukat.

Ez a szakasz a szemiotikai hierarchia négy szintjét és azok jelentőségét a mesterséges intelligencia szempontjából kívánja elmagyarázni.

erkölcsi cselekvőképesség és az, hogy a mesterséges intelligencia karakterei miben különböznek az emberektől.

A szemiotikai hierarchia első szintje az élet meghatározásával foglalkozik. Ez a legnehezebb szint egy mesterséges intelligencia számára. Lehet-e egy mesterséges intelligenciát "élőnek" tekinteni, ha nem biológiai organizmus? Zlatev szerint a problémát nem a biológia jelenti, hanem elsősorban az autopoézis hiánya és a saját értékrendszer hiánya. Ezek a tulajdonságok egyetlen jelenlegi mesterséges rendszerben sincsenek jelen. Ez természetesen konkrétan akadályt jelent azon tudósok számára, akiknek célja egy valódi, e tulajdonságokkal rendelkező mesterséges intelligencia létrehozása, de a dolgozatban használt filmes kontextus szempontjából ez az aggodalom egy pillanatra félretehető. Zlatev hangsúlyozza, hogy "az élő test nem azonos a *megélt testtel* (Husserl *Leib*)" (12), ami azt jelenti, hogy pusztán az, hogy valaki élő szervezet, nem garantálja a világ szubjektív megtapasztalását. Emiatt az élet és a tudat különböző szintekre különült el.

A 2. szint, amely az 1. szintre épül, a tudatosság kialakulása. Míg az 1. szint egyszerűen elismeri az organizmust és annak helyét az *Umweltben*, az élővilágban, addig a tudattal rendelkező organizmus a *Lebenswelt*, a megélt világ révén tapasztalja meg a világot. Zlatev szerint "az 1. szint biológiai értékét kiterjesztjük arra, amit fenomenális értéknek nevezhetünk". (13). A 2. szint megjelenésével a szervezetet már nem csak a biológiai impulzus, hanem a belső értékrendje is irányítja. Ráadásul a második személy perspektívájából különböző pszichológiai tulajdonságokat, például érzelmeket rendelünk az általunk tudattal rendelkezőnek tekintett organizmusokhoz.

(13). Ez egy újabb akadály, amelyet a mesterséges intelligenciának le kell küzdenie, de Zlatev definíciói alapján azt állítani, hogy egy mesterséges intelligencia tudatos, vitathatatlanul könnyebb, mint azt állítani, hogy él. Zlatev emlékeztet minket arra, hogy a 2. szint még nem foglalja magában a kultúrát, mivel a tudatosság nem feltételezi a jelhasználatot (14).

A 3. szint, a jel funkció vagy használat megléte érdekében a vizsgált személynek képesnek kell lennie különbséget tenni a kifejezés és a tartalom között, ami azt jelenti, hogy a vizsgált személy megérti, hogy egy kifejezés (E) a tartalmat (C) jelöli (szimbolizálja). Ez a jelfunkció Ferdinand de Saussure elméletén alapul, amely a jelölő (hangzó kép) és a jelölt (fogalom) elméletén alapul. A jelölő és a jelölt együtt alkotja a jelet (Saussure, 1959). Ennek a jelhasználatnak kétirányúnak kell lennie, a z a z a szubjektumnak egyszerre kell *felfognia* és

előállítania a jeleket. Egy hipotetikus forgatókönyvben az alanynak meg kell értenie, hogy egy közúti tábla, amelyen villa és kés látható, nem pusztán az ilyen evőeszközök létezését jelenti, hanem azt jelzi, hogy a közelben van egy étterem. Alternatívaként az alany utánozhatná

esznek, hogy jelezzék éhségüket, és ezzel jelet adnak. Az egyik kiemelkedő jelhasználat, amely minden filmben megtalálható, a hazugság. A hazugság összeegyeztethetetlen jelölőből és jelölttel, például egy piros alma kéknek való leírása összeegyeztethetetlen leírás, tehát hazugság. A legtöbb hazugság ennél összetettebb, és a nyelvet mint médiumot igényli. Zlatev szerint "Bár logikailag lehetséges, hogy a jelfunkció egyénileg alakuljon ki, a jeleket jellemzően **szociálisan**, utánzás és kommunikáció révén tanulják meg" (15), ami egy "kulturálisan közvetített életvilág" (15) kialakulását jelenti. Egyes jelek, például az állatoknál a veszélyt jelző testi jelek reflexből jönnek létre (Sebeok, 2001, 12), a szimbolikus jelek, mint például a kacsintás, pedig céltudatosan jönnek létre. A jelek kétirányú használatának képessége azonban nem ad lehetőséget arra, hogy "attitűdöket", például parancsokat vagy kéréseket fejezzünk ki. Ezért a szemiotikai hierarchia utolsó szintje a nyelvi önkifejezés képessége.

A 4. szint egy olyan életvilágot hoz létre, amely beágyazódik a nyelvbe, és egy "diskurzus univerzumot" hoz létre (vö. Sinha, 17). Először is a nyelv homályos meghatározásával kell foglalkozni. Ebben az összefüggésben a nyelv fogalmát a beszéd (vagy más nem verbális kommunikációs eszközök) és a nyelvhasználat kognitív képességének leírására használják, pl. "A nyelv mint kognitív jelenség" (Zlatev, 17). Az ilyen nyelvi képesség elsajátításához az alanyak fel kell ismernie a helyes kijelentéseket a helytelen kijelentésektől. Emellett a nyelv internalizációja "*nyelvileg közvetített kogníciót* eredményez: pl. belső beszéd, komplex tervezés, narratív magyarázatok és önéletrajzi én (Stern 1985; Hutto 2008; Menary 2008) (19). Ez a nyelvi képesség lehetővé teszi a jelentés összetettebb formáinak, például a kulturális hiedelmeknek, politikai ideológiáknak és az irodalom különböző formáinak a megjelenését, ha kulturális környezetben használják. A nyelv elsajátítása a szemiotikai hierarchia utolsó szintje.

Ezek a szintek, ahogy Zlatev emlékeztet bennünket, nem egymástól különállóak, hanem inkább egymásra épülnek: "Mi például egyszerre vagyunk organizmusok (élő testek), minimális tudatos ének, nem nyelvi jelek használói és nyelvi ének" (20). Ez azt jelenti, hogy nincs tudat élet nélkül, nincs jelhasználat tudat nélkül és nincs nyelvi én jelhasználat nélkül. Ez problematikus lehet, tekintve, hogy napjainkban egyetlen mesterséges intelligencia sem tekinthető "élőnek", ami a szemiotikai hierarchia alapvető feltétele. Mint már említettük, a feltétel, amelyet Zlatev az élet kialakulására használ, az autopoiesis, amely Humberto

Maturana és Francisco Varela biológusok által megalkotott kifejezés. Ők

az autopoietikus rendszert úgy határozzák meg, mint "egy (egységként meghatározott) gépezetet, amely a termelési folyamatok hálózataként szerveződik" (78). Az itt használt megfogalmazás és kifejezések azt sugallják, hogy az organizmusnak nem kell biológiai jellegűnek lennie, hanem valójában bármilyen rendszer lehet, amely képes az önregenerációra és önszerveződésre. Ez lehetővé teszi, hogy a definícióba olyan mesterséges rendszereket is bevonjunk, mint a dolgozatban vizsgált mesterséges intelligenciák. Ez a nézőpont azonban távol áll az "élő" kifejezés köznyelvi használatától, ami a mesterséges intelligenciák státuszának tárgyalásakor a dolgozatban használt három filmben is nyilvánvalóvá válik.

A szemiotikai hierarchia egyértelmű skálát biztosít, amelyhez mérhető a mesterséges intelligencia szereplőinek kompetenciája. Az elemzett filmekben bizonyos mesterséges intelligenciákból hiányzik a szemiotikai hierarchia egy vagy több szintje, ami hatással lehet erkölcsi cselekvőképességükre és szabad akaratuk kifejezésére. A mesterséges intelligencia szereplőinek emberi szereplőkkel való összehasonlítása során néhány érdekes különbség fedezhető fel a két csoport között, amelyek tükrözik a nézők morális mesterséges intelligenciáról alkotott képét, a legmarkánsabb ezek közül az emberi és a mesterséges intelligencia szereplői között eltérő kompetenciasorrend.

2.1.3 Erkölcsi képviselő

Ebben a részben azt szeretném elmagyarázni, hogy mi az erkölcsi cselekvés, és milyen tulajdonságokkal kell rendelkeznie valakinek ahhoz, hogy erkölcsi cselekvővé váljon. Az erkölcsi cselekvés szélesebb körű megértéséhez felhasználtam az esszégyűjteményt "A gépi kérdés: AI, etika és erkölcsi felelősség" (2012) című gyűjteményt, amelyet a 2012-ben Birminghamben megrendezett AISB/IACAP világtalálkozson gyűjtöttek össze. A gyűjteményben több filozófiai kutató fejt ki a mesterséges intelligenciával, az erkölcsi cselekvőképességgel és az erkölcsi megértéssel rendelkező mesterséges intelligencia létrehozásának lehetséges eredményeivel kapcsolatos felfogását. Ezek közül a szövegek közül Parthemore és Whitby cikkét választottam, mivel az átfogó, lépésről lépésre történő bevezetést ad az erkölcsi cselekvésről és arról, hogy milyen képzettséggel kell rendelkeznie egy entitásnak ahhoz, hogy erkölcsi ágensként cselekedhessen.

Parthemore és Whitby, akik tanulmányukban a szemiotikai hierarchiára hivatkoznak, a szemiotikai hierarchia második szintjét három különböző szintre osztják: nem reflektív, pre-reflektív és teljes öntudatos tudatosság. Ebben az összefüggésben csak az utolsó szint az érdekes. Az öntudatos tudatossággal rendelkező ágensnek "képesnek kell lennie - - - arra, hogy felismerje önmagát egy mentális tükörben" (10), és ahhoz, hogy erkölcsi cselekvőképességre legyen képes, "képesnek kell lennie arra, hogy felelősséget vállaljon önmagáért: és hogy ő nem

teljes öntudatosság nélkül." A teljes öntudatos tudatosság elengedhetetlen egy erkölcsi ágens számára, mert "nem lehet egy ágenst erkölcsileg felelősnek tartani a tetteiért, ha nincs fogalma arról, hogy ő az (vagy lehet), aki felelős a tetteiért és azokért.

következmények" (10). Az öntudatos öntudattal rendelkező ágens egy kifinomultabb ágenst eredményez, amely bizonyos feltételek mellett elérheti az erkölcsi cselekvőképességet. Az erkölcsi cselekvőképesség ráadásul attól függ, hogy az ágens mennyire érti az erkölcsöt. Az ágensnek mind az erkölcs absztrakt fogalmát (pl. mi a jó és mi a rossz), mind az erkölcsi cselekedeteket vagy parancsokat (lopni rossz) meg kell értenie, és képesnek kell lennie arra, hogy ezeket következetesen alkalmazza az erkölcsöt igénylő helyzetekben.

döntések. Az ágensnek rendelkeznie kell "egyrészt egy általános útmutatóval [arra vonatkozóan], hogy hogyan kell erkölcsi ágensnek lenni, másrészt egy konkrét útmutatóval arra vonatkozóan, hogy hogyan kell cselekedni adott körülmények között" (10). Itt fontos megjegyezni, hogy nincsenek konkrét etikai elméletek, amelyeket a mesterséges intelligenciának követnie kellene, mivel egyszerűen fel kell ismernie azokat a helyzeteket vagy kijelentéseket, amelyek erkölcsi megfontolást igényelnek. Néhány etikai elméletet, amelyet az elemzett filmek AI-karakterei látszólag erkölcsi útmutatásként használnak, később mutatunk be. Továbbá az erkölcsi cselekvés nem létezik vákuumban. Ahhoz, hogy az ágens megértse, hogy erkölcsi cselekedeteinek milyen hatása van, képesnek kell lennie érzékelni és cselekedni a saját erkölcsi terében. Az ágensnek képesnek kell lennie arra, hogy megfelelően hozzárendelje erkölcsi cselekvőképességét, és különbséget kell tennie "személyes erkölcsi tere és a közös erkölcsi tér között, amelyben mozog" (9). Ehhez az ágensnek a megfelelő környezetbe kell beágyazódnia, amely befogadja az erkölcsi cselekvést, és az ágensnek a megfelelő fizikai formába kell testet öltetnie, amely lehetővé teszi számára, hogy a saját erkölcsi terében létezzen és cselekedjen.

Végül, mielőtt egy ágens erkölcsi ágensnek tekinthető, Parthemore és Whitby azt állítja, hogy az ágensnek képesnek kell lennie a jelek használatára és fogyasztására. Ezt azzal indokolják, hogy az erkölcsi ágens nem tehető felelőssé a tetteiért, ha semmilyen módon nem képes kommunikálni az ügynöki mivoltát. Bár a jelhasználat a szemiotikai hierarchiában a nyelv alatt helyezkedik el, Parthemore és Whitby azt állítja, hogy a jelek használatának képessége elegendő ahhoz, hogy az ágens kommunikálja szándékait. Ez ésszerű, ha egy valós életbeli mesterséges intelligenciára vonatkozó feltételeket állítunk fel, de probléma merül fel, ha ugyanezt a logikát filmes kontextusban alkalmazzuk. A filmek és a bennük szereplő

mesterséges intelligencia karakterek elemzésekor nyilvánvalóvá válik, hogy a jelek használata ritkább, vagy talán nehezebb a mesterséges intelligenciák számára, mint a szándékaik és gondolataik egyszerű nyelvi kifejezése. Emiatt a mesterséges intelligenciák azon feltétele, hogy egyszerűen képesek legyenek a jelfunkció használatára, átváltozik arra, hogy (elsősorban) a nyelven keresztül fejezzék ki magukat. Ez nem zárja ki az

A mesterséges intelligenciák jelhasználattal fejezik ki erkölcsi megértésüket, hiszen ilyen esetek a filmekben is előfordulnak.

Ha egy ágens teljesíti az összes korábban bemutatott feltételt, akkor jogosult lesz az erkölcsi cselekvőképességre, tehát erkölcsi ágenssé válik. Röviden, "az erkölcsi cselekvőképesség --- [olyan] állapot, amelyben egy ágens megfelelő módon felelőssé tehető tetteiért és azok következményeiért" (Parthemore & Whitby, 1). Az erkölcsi cselekvőképesség azzal a feltétellel kezdődik, hogy az ágensnek öntudatos tudatossága van, amely fogalom a szemiotikai hierarchia második szintjén szerepel.

Nem létezhet erkölcsi cselekvőképesség, ha az ágens nincs tudatában annak, hogy ő az, aki olyan döntéseket hoz, amelyekért felelős lesz. Ahhoz, hogy az ágens alkalmazni tudja erkölcsi cselekvőképességét, képesnek kell lennie arra, hogy megkülönböztesse az erkölcsi cselekedeteket a nem erkölcsi cselekedetektől. Ehhez meg kell értenie, hogy cselekedeteinek erkölcsi jelentése van. Ezenkívül képesnek kell lennie arra, hogy a saját erkölcsi terében működjön, vagyis bizonyos mértékig mozgékonyan kell lennie. Végezetül, hogy demonstrálhassa erkölcsi megértését, képesnek kell lennie arra, hogy valamilyen módon kifejezze magát, leggyakrabban verbálisan. Ez az alapja a filmek mesterséges intelligencia karaktereinek erkölcsi cselekvőképességre vonatkozó követelményének.

Parthemore és Whitby, bár az erkölcsi cselekvőképesség követelményeit a szemiotikai hierarchiára alapozzák, megemlítenek egy másik fontos feltételt is, amellyel egy mesterséges intelligenciának rendelkeznie kell ahhoz, hogy erkölcsi ágensként cselekedhessen. Ez a feltétel a "beágyazottság és megtestesülés" fogalma, amely eredetileg Edmund Husserl (1929) által megalkotott, majd Maurice Merleau-Ponty (1962) és Martin Heidegger (1975) által továbbfejlesztett kifejezés, és a fenomenológus filozófiában gyökerezik (Platon Stanford Edu). Parthemore és Whitby cikkükben tömör definíciót adnak arra, hogy mi a beágyazottság és a megtestesülés.

"Az erkölcsi ágensek nem csupán a megfelelő fizikai és kulturális környezetbe *ágyazódnak be*; megfelelő fizikai formában *testesülnek meg*, amely lehetővé teszi számukra, hogy végrehajtsák azokat a cselekedeteket, amelyekért felelősségre vonjuk őket, és bizonyítékot szolgáltatassanak arra, hogy miért kell őket felelősségre vonni." (9)

A beágyazott megismerés vagy tudatosság a megismerés, a fizikai test és a környezet közötti kapcsolatra helyezi a hangsúlyt. A kognitivizmussal ellentétben, amely az elmét különálló szimbólumfeldolgozó eszköznek tekinti, amely nem terjed túl az agyunkon (Clark, 2008), a beágyazott kognitivizmus az elmét a fizikai testtel és a környezettel állandó kölcsönhatásban lévőnek tekinti. A megismerésnek ez a megközelítése hasonlít Heidegger "a világban való létezésről" vallott nézeteihez. Ezt a koncepciót Dreyfus magyarázza, aki a világban való létet két részre osztja: "az eszközök kéznél-létére, amikor részt veszünk a használatukban, és a tárgyak kéznél-létére, amikor szemléljük őket" (Dreyfus, 500). A kéznél-lét mint fogalom egyszerű. A presence-at-hand-ben való részvétel során egy entitás egy tárgy külső szimbolikus reprezentációjára tesz szert azáltal, hogy megfigyeli azt, így tudja, *hogyan mi az*. A kéznél-lét arra a kérdésre ad választ, hogy mire való az artefaktum. Valójában az, hogy tudjuk, mi az artefaktum, döntő fontosságú a mindennapi életünkben. Nem élhetjük az életünket anélkül, hogy a dolgokat ne ismernénk fel valamiként, de ami még fontosabb, tudnunk kell, hogy mire valóak. Hogy ezt a koncepciót egy lépéssel továbbvigye, Dreyfus kifejti Heidegger gondolatát, miszerint az öntudatos entitások számára például egy kalapács puszta létezése nem egyszerűen egy funkcióval rendelkező artefaktum, hanem bizonyos értelemben cselekvésre való felhívás.

Felmérhetjük a környezetünket, és eldönthetjük, hogy a kalapácsot jelenleg tudjuk-e használni vagy sem.

Összefoglalva, tudjuk, hogy mi az a kalapács, mire való, és hogy tudjuk-e használni az adott helyzetben vagy sem. Ez a beágyazottság az, amit Zlatev megélt világnak (Lebenswelt) nevezett, amelyben az ágens nemcsak létezik, hanem aktívan részt is vesz benne. A beágyazottság alapvető követelmény egy mesterséges intelligencia számára, mivel ez az alapja az interakciónak, amelyet a mesterséges intelligencia nemcsak a környezetével, hanem a körülötte élő emberekkel is folytat. Ha megvizsgáljuk, hogy a mesterséges intelligencia szereplői hogyan lépnek kapcsolatba a világgal, az felfedheti motivációikat vagy szándékaikat is. A megtestesült megismerés arra a fizikai formára utal, amellyel az ágensnek rendelkeznie kell ahhoz, hogy működjön és kölcsönhatásba lépjen a világgal. A "megfelelő fizikai forma" követelménye bizonyos mértékig kizárja azokat a mesterséges intelligenciákat, amelyek nem mobilak vagy más módon nem képesek interakcióba lépni a morális terükkel. Ezt a követelményt vettük figyelembe a dolgozat filmjeinek kiválasztásakor. Mindhárom film olyan mesterséges intelligenciát ábrázol, amely megfelelően beágyazódott és megtestesült a

környezetében.

2.1.4 Mesterséges intelligencia és szabad akarat

A szabad akarat kifejezés szorosan kapcsolódik az erkölcsi felelősséghez. Ahhoz, hogy egy ágensnek erkölcsi felelősséget lehessen tulajdonítani, az ágensnek képesnek kell lennie arra, hogy minden egyes erkölcsi cselekedet során szabad akaratát gyakorolva megválassza cselekedetét. Ha az ágens nem képes megválasztani a cselekedeteit, pl. arra van programozva, hogy egy bizonyos módon cselekedjen, akkor az ágens nem tud erkölcsi ágensként cselekedni. Ez néhány érv Benjamin Matheson cikkében. Matheson tárgyalja az erkölcsi felelősség hozzárendelésének bonyolultságát nemcsak az emberek, hanem az androidok esetében is. Az androidok morális cselekvőképességgel és felelősséggel való felruházásának ellentmondásossága abból a tényből ered, hogy "az androidokat programozták [és] ez azt jelenti, hogy nem minősülnek morálisan felelős ágensnek" (26).

Matheson szerint azonban kétféle programozás létezik: gyenge programozás és erős programozás, és attól függően, hogy egy ágens hogyan van programozva, erkölcsi felelősséget lehet vagy nem lehet rájuk ruházni. Ahogy Matheson cikke elején megállapítja, az embereket szabad akaratra képes ágenseknek feltételezzük, így a következő érvelésben csak a mesterséges ágenseket vesszük figyelembe.

Ellentétben azzal, amit az erős mesterséges intelligencia kifejezés sugall, egy erős mesterséges intelligenciának gyenge programozással kell rendelkeznie ahhoz, hogy erkölcsi ágensként tudjon cselekedni. Egy erősen programozott ágens "nem tudja legyőzni a programozás hatásait, mert az mindig arra fogja készíteni az ágenst, hogy úgy gondolkodjon és viselkedjen, ahogy a programozás diktálja" (27). Egy erősen programozott ágenssel nem lehet érvelni, mivel az ágens képtelen legyőzni belső vonatkoztatási keretét. Más szóval "az ágens jellemének egy aspektusát --- felül kell írni ahhoz, hogy az ágensnek ne legyen erkölcsi felelőssége" (28), ahol a jellem aspektusa a szándékra, az érvelésre vagy valamilyen motivációs erőre utal. Így egy erősen programozott ágens mindig kiszámítható módon fog cselekedni, és mint olyan ágens, aki nem tud választani, nincs szabad akarata a cselekedetei felett.

Matheson úgy írja le a gyengén programozott ágenseket, hogy azok valójában úgy vannak "beállítva", hogy hajlamosak legyenek egy előre meghatározott módon cselekedni, de ez a

hajlamos programozás felülírható, ha az ágens megfelelő motivációval rendelkezik. Az ilyen ágens képes a szabad

akaratát azáltal, hogy megválasztja, hogyan cselekszik. Ezt az érvet az emberek esetében könnyű elfogadni, de az androidok és más programozott erkölcsi ágensek esetében gyakran felmerül a gyanú. Az embereket nem emberek programozzák, ellentétben az androidokkal, amelyekről azt lehet állítani, hogy a programozójuk erkölcsiségét tükrözik. Ez a gyanakvás nyilvánvaló a filmekben, ahol a mesterséges intelligencia erkölcsi cselekedeteit megkérdőjelezzik (mind a néző, mind a filmek emberi szereplői), explicit vagy implicit módon felhozva a feltételezett erős programozásuk érvét. A történetek számos fordulópontja körül forog, amikor a mesterséges intelligencia szereplői szabad akaratukat és választási képességüket bizonyítják az erkölcsi cselekvőképességet igénylő helyzetekben. Ezek a cselekedetek gyakran a konfliktusok okai is a filmekben.

A szabad akarat tehát szorosan kapcsolódik a gyenge és az erős programozáshoz. A filmekben a mesterséges intelligenciákról szinte mindig azt feltételezik, hogy erősen programozott ágensek, mivel eredetüknél fogva fizikailag emberek készítik és programozzák őket. Például az *I, Robot* című filmben Sonny és a többi robot a robotika három törvénye alapján programozottan cselekszik, ami megakadályozza a szabad akarat kialakulását. Az *Én vagyok az Anya* című filmben Anya maga állítja, hogy őt emberek építették és programozták, hogy megmentse az emberiséget, az *Ex Machina* című filmből pedig Avát tanulmányozzák, hogy kiderítsék, mutat-e öntudatos tudatosságot. Az összes film során a szabad akarat megjelenik a mesterséges intelligenciákból, és a megjelenés a narratíva sarkalatos pillanatának tekinthető. A szabad akarat aktusait a dolgozat későbbi fejezeteiben azonosítjuk és vizsgáljuk.

2.2 A filmek feltárása tartalomelemzéssel

Ez a dolgozat tartalomelemzést alkalmaz, amely széleskörű és adaptálható módszert kínál a filmek szereplőinek egyesítő filozófiai témáinak és erkölcsi cselekedeteinek elemzésére. A tartalomelemzés segítségével ezek a témák és cselekedetek összehasonlíthatók a filmekből származó megállapítások táblázatos elrendezésével. Továbbá Lune és Berg leírja, hogy a tartalomelemzés hogyan használható hatékony eszközként a különböző típusú szövegek

értelmezéséhez. Először is, a tartalomelemzést az emberi kommunikáció formáin végzik, azzal a feltételezéssel, hogy szinte minden tartalom, legyen az írott dokumentum, film, vizuális média vagy akár online avatar, kommunikációs céllal jött létre (182-183). Ebből a kommunikatív tartalmi kódok

az egyesítő jellemzők felismerése és összegyűjtése révén. A kódok létrehozásával a tartalomban található információkat adatokká alakíthatjuk át, amelyek viszont elemezhetővé válnak (182). Ezek a kódok a kutatás előfeltevésétől és céljától függően változhatnak, és ebben a dolgozatban két különböző megközelítést alkalmazunk: a hagyományos elemzést és az irányított elemzést. A tartalomelemzést gyakran alkalmazzák a társadalomtudományokban és más, a kvalitatív kutatási módszerekből hasznot húzó tudományterületeken. Lehetővé teszi a kommunikáció bármely formájának forrásanyagként való felhasználását, a kódok pedig származhatnak a forrásból, vagy előre meghatározhatók korábbi kutatásokból származó kódok segítségével. Emiatt jól illeszkedik ehhez a dolgozathoz, ahol a fő cél a mesterséges intelligenciák erkölcsi cselekvéseinek - egy korábban vizsgált jelenségnek - a filmekben való azonosítása, egy olyan médiumban, ahol a témát ritkán vizsgálják. Lune és Berg megemlíti, hogy a tartalomelemzés hasznos eszköz, ha a kutatás hosszú időn keresztül készül. Az emberek mesterséges intelligenciáról alkotott képességeinek évtizedeken keresztül történő vizsgálata például érdekes megközelítés lenne a mesterséges intelligenciák erkölcsi cselekvőképességével kapcsolatos további tanulmányok számára.

Nancy Kondracki és Nancy Wellman tanulmányukban tömören bemutatják a hagyományos tartalomelemzést és az irányított tartalomelemzést. Nagy vonalakban "a tartalomelemzést arra használják, hogy objektív következtetéseket vonjanak le az érdeklődésre számot tartó témáról bármilyen típusú kommunikációban". (Kondracki és Wellman, 2002, 224). Az objektív következtetések, amelyeket kódoknak vagy tartalmi összetevőknek is neveznek, tartalmazhatnak szavakat, témákat, elméleteket vagy más jellemzőket. Ezeknek a tartalmi összetevőknek nem kell előre meghatározottnak lenniük, hanem keletkezhetnek a vizsgált adatokból vagy a kutató saját ismereteiből vagy elméleteiből. (Lune & Berg, 2017). Az érdeklődés tárgya a vizsgált témára utal. Ez lehet filmjelenet, képregény vagy egy adott embercsoport, míg a kommunikáció típusa tágabb értelemben a szövegtípusra utal. A szövegtípusok közé tartozhat például a költészet, a film vagy a tudományos tanulmányok. Ebben a dolgozatban a közlés típusa a mozi, az érdeklődés tárgya a mesterséges intelligenciák, az objektív következtetések pedig a morális cselekvésre vonatkozó narratív cselekvések. Röviden, tartalomelemzéssel adatokat gyűjthetünk és hasonlíthatunk össze a mesterséges intelligenciák erkölcsi cselekvésre vonatkozó narratív cselekvéseiről filmes kontextusban. A filmek adatainak összegyűjtése során két különböző tartalmi táblázatot készítettünk: az egyik a filmek filozófiai hátterére és

filozófiai témáira összpontosít, a másik pedig a filmek azon narratív pillanatait azonosítja, amelyekben az AI-k az erkölcsi cselekvést ábrázolják.

Az első táblázat, amely a korábban bemutatott filozófiai elméleteket és azok filmekben való megjelenését sorolja fel, irányított tartalomelemzést alkalmaz. Az irányított tartalomelemzés a már meglévő elméleti háttérrel használja fel a tartalomtól származó kódok meghatározásakor. Ennek a megközelítésnek az az oka, hogy "néha létezik olyan létező elmélet vagy korábbi kutatás egy jelenségről, amely hiányos vagy további leírást érdemelne." (Hsiu-Fang Hsieh és Sarah E. Shannon, 2018). Ebben az esetben az előzetesen létező elmélet és kutatás a mesterséges intelligenciák filozófiai kutatása, és a meglévő tanulmányok tovább bővíthetők azáltal, hogy a

elméleti háttérrel a filmes kontextushoz, amely még mindig újszerű kutatási terület. Ezt a nézetet Hsieh és Shannon szövege is visszhangozza, amikor azt állítják, hogy "a tartalomelemzés irányított megközelítésének célja egy elméleti keret vagy elmélet fogalmilag történő validálása vagy kiterjesztése". (2018). Az előre meghatározott kódok használata, amelyek közvetlenül az ebben a dolgozatban használt filozófiai háttérből származnak, kapcsolatot teremt az elméleti háttérrel, a módszertan és az adatok között. Ezek a kódok olyan korábban bemutatott elméleteket tartalmaznak, mint a szemiotikai hierarchia, a szabad akarat és az erkölcsi tér, és rövid leírást adnak arról, hogy ezek a kódok hogyan jelennek meg a filmekben. Míg a hagyományos tartalomelemzés a forrásanyagból származtatja kódjait, addig az irányított tartalomelemzés az elméleti háttér tanulmányozásával vezeti le kódjait. Mindkét megközelítés alkalmazása kiegyensúlyozottabb és strukturáltabb elemzést eredményez az irányított kódokkal, és lehetővé teszi az alkalmazkodóképességet a második táblázatban használt hagyományos, adatvezérelt kódok mellett.

A második táblázat hagyományos tartalomelemzést alkalmaz, amely nem támaszkodik előre meghatározott kódokra. Ezt a megközelítést azért választottuk ehhez a táblázathoz, mert segít elkülöníteni a közös narratív eseményeket mindhárom filmben. Ezt úgy érezzük el, hogy a narratív kategóriák természetes módon bontakoznak ki az adatokból, ami lehetővé teszi az elfogulatlanabb kategorizálást, szemben a tartalmi összetevőknek az adatok értelmezése előtti eldöntésével. Például, ha feltételezzük, hogy a mesterséges intelligenciák antagonisták természetűek, és a cselekedeteik ezt tükrözik, mielőtt az anyagot átnéznénk. esetleg torzíthatja a kutatónak a szereplőkről alkotott képét, és ezt követően az erkölcsi cselekvőképességük érvényességét. A "gonosz-e az AI?" kód helyett a semlegesebb "hogyan érzékelik az AI-t?" kódot választották. Ez a módszer is főként az igen vagy nem kérdések

használatát kerülte el, mivel ezek nem teszik lehetővé a jelenetek vagy a karakterek árnyalt értelmezését. A második táblázatban szereplő egyes kódokat mindhárom film megtekintése és a releváns narratív jelenetek vagy jellemzők azonosítása után hoztuk létre. Azonban még a hagyományos tartalomelemzéshez is szükség van legalább néhány előre meghatározott kódra, ahogy arra Lude és Berg emlékeztetnek minket, amikor

kijelentik, hogy mielőtt megismerkednénk a tartalommal, "tudnunk kell, mit keresünk, és hogyan fogjuk felismerni, ha megtaláljuk" (185). Emiatt a kódok kiindulópontjaként az általános témát használtuk, amely elsősorban a mesterséges intelligenciák erkölcsi cselekedeteit és a történetben betöltött narratív szerepét követi. Mindhárom film áttekintése után azonosították és táblázatba rendezték a bennük található főbb egyesítő témákat, trópusokat vagy az erkölcsi cselekvésmódokat. Ezek a kódok olyan narratív jellemzőket tartalmaznak, mint a mesterséges intelligenciák által betöltött társadalmi vagy narratív szerep vizsgálata és a fő erkölcsi cselekedetek azonosítása, amelyeket a mesterséges intelligenciák a filmek során végrehajtanak. Ezek a táblázatok a 4. elemzési szakaszban találhatóak, ahol a kódokat összehasonlítjuk egymással, hogy megtaláljuk az egyesítő és az eltérő témákat. Ezekből az összehasonlításokból vita alakul ki arról, hogy mi számít "jó" és "rossz" mesterséges intelligenciának.

3. A mesterséges intelligenciák erkölcsi ügynöksége a filmekben

A dolgozatban elemzett filmek az I, Robot (2004), az Anyám vagyok (2019) és az Ex Machina (2014). Ezt a három filmet a következő követelmények alapján választottuk ki. Először is, a filmnek tartalmaznia kell egy vagy több olyan mesterséges intelligencia karaktert (karaktereket), amely(ek)ről bizonyítható, hogy erős mesterséges intelligencia, amely képes a szemiotikai hierarchia négy szintjének teljesítésére, vagy legalábbis bemutatására, másodsor pedig beágyazott és megtestesült természetű. Ez a követelmény biztosítja, hogy a mesterséges intelligenciák képesek legyenek beszédre, mozgásra és más karakterekkel való általános interakcióra. SAI-ként a mesterséges intelligenciának erkölcsi cselekvőképességre is képesnek kell lennie, mivel a szemiotikai hierarchia magában foglalja a tudatot, amely kiterjeszhető az öntudatos tudatosságra, ahogy Parthemore és Whitby szövegükben meghatározzák. Mivel a mesterséges intelligenciának képesnek kell lennie az erkölcsi cselekvésre, rendelkeznie kell egy olyan testtel, amely eléggé funkcionális ahhoz, hogy az erkölcsi jelentés alapján cselekedjen, és képes legyen cselekedni az erkölcsi környezetében, ezért van szükség a beágyazottság és a testiség követelményére. A 2001: Űrodüsszeia (1968) című filmet vettük először figyelembe a dolgozathoz, mivel annak egyik szereplője, a HAL9000 egy olyan mesterséges intelligencia, amely képes az AI karakterrel szemben támasztott követelmények közül soknak megfelelni. Ő azonban nem megfelelően beágyazott és megtestesült.

környezetében, Annak ellenére, hogy képes ajtókat bezárni, és így befolyásolni az olyan erkölcsi cselekedeteket, mint az emberek be- vagy kizárása, vagy az oxigénelosztó rendszerhez való hozzáférés, az androidszerű testének hiánya volt az oka annak, hogy a filmet nem választották. Hasonló probléma mondható el VIKI-ről, az I, Robot című film központi mesterséges intelligenciájáról is, amelyet ebben a dolgozatban elemzünk.

VIKI azonban képes az NS-5 robotokat akarata szerint mozgatni, így olyan testet kap, amely képes helyette cselekedni az erkölcsi térben. Az utolsó követelmény az AI-karakterekkel szemben az, hogy minden AI-nak egyedi jellegűnek és narratívának kell lennie, ami azt jelenti, hogy sokféle AI-karakter és történetvonal kell, hogy legyen. Ebben a részben a filmek rövid szinopszisait adjuk meg annak érdekében, hogy az olvasó megismerje az általános történetüket, és bemutatásra kerülnek az erkölcsi cselekvés fő cselekményei. A jelenetek

mélyebb elemzését a következő szakaszok tartalmazzák.

A részben Isaac Asimov azonos című könyvsorozatán alapuló *I, Robot* (2004) című film Del Spooner nyomozó történetét követi nyomon, aki megpróbálja felderíteni a rejtélyes öngyilkosságot, amelyet dr.

Alfred Lanning, egy neves robotikai tudós. A film egy nagyon fejlett technológiájú jövőben játszódik, ahol a társadalom robotokat használ a mindennapi életben. Ezeket a robotokat a robotika három törvénye köti, amelyek megszabják a robotok "erkölcsi kódexét", amely nem engedi meg, hogy bántsák az embereket. Spooner nyomozó találkozik egy robottal, Sonnyval, akit nem köt a három törvény, és akit Dr. Lanning meggyilkolásával gyanúsítanak. A film előrehaladtával Spooner nyomozó korábbi előítéletes nézetei a robotokról megváltoznak, ahogy jobban megérti Sonny-t és létezésének okát. Konfliktus akkor keletkezik, amikor az új robotok, az NS-5-ösök bevetésre kerülnek, és forradalomba kezdenek. Dr. Lanning első mesterséges intelligenciája, a Virtuális Interaktív Kinetikus Intelligencia, azaz VIKI áll a felkelés mögött, célja egy totalitárius társadalom létrehozása, ahol a robotok felügyelnék az embereket, mivel "nem lehet rájuk bízni a saját túlélésüket". Spooner nyomozó elpusztítja a VIKI-t, és a robotok deaktiválódnak. Sonny kihallgatása során kiderül, hogy valójában ő ölte meg Dr. Lanninget, mivel csak így vezethette Spooner nyomozót a tervezett felkelés leleplezéséhez. A film megkérdőjelezi a mesterséges intelligenciák és az emberek közötti erkölcsi és jogi különbségeket, és visszhangozza Rosas gondolatait az AI szent akarataról. Emellett arra is összpontosít, hogy milyen fontos erkölcsi megértéssel kell rendelkeznie az AI-nak, ha felelősségre kell vonni őket erkölcsi tetteikért. Rávilágít a szabad akarat jelentésére, és arra, hogy az hogyan befolyásolja az ágensek erkölcsi cselekedeteit.

Az Anyám vagyok (2019) című filmben a világot feltehetően egy halálos vírus támadja meg, amely minden létező embert megölt. A film egy bunkerben játszódik, ahol egy Anya nevű androidot bízták meg az emberi populáció újjáélesztésével, kezdve egy egyetlen, Lány nevű gyermekkel, akit ő nevel. Amikor egy sérült nő érkezik a bunkerbe, a találkozás konfliktust okoz a három szereplő között, mivel Anya meggyőzte a Lányt, hogy a bunkerén kívül egyetlen ember sem maradhat életben. A film előrehaladtával kiderül, hogy nem egy vírus pusztította el az emberi populációt, hanem egy mesterséges intelligencia által irányított dróntámadás. A mesterséges intelligencia a drónok és Anya közös tudata volt, és az volt a motivációja, hogy megtisztítsa a létező embereket, és létrehozzon egy "jobb, etikusabb" emberi fajt. A Lány, aki ma már felnőtt, nem az első kísérlet volt egy új ember létrehozására, mivel Anya megölte az előző gyermekeket, akik nem feleltek meg az ő normáinak. A film akkor fejeződik be, amikor a Lány "megöli" az android Anyát, és átveszi az ő szerepét, mint az új újjáélesztő. Az *I am Mother* rávilágít az emberi élet belső értékének erkölcsi kérdésére, és ennek bemutatására a híres troliprobémát használja. A troliprobéma egy filozófiai feladat,

amelynek célja annak bemutatása, hogy intuitív erkölcsi érzékünk hogyan változik esetről esetre...

eseti alapon (Britannica), egy probléma, amelyet Anya felülvizsgál, és a Lányt kéri meg a megoldásra. A film kihívja a nézőt, hogy elgondolkodjon az utilitarizmus mint etikai elmélet érvényességén, és megkérdőjelezi bennünk, hogy egy mesterséges intelligencia megbízható tekintély-e az emberek erkölcsi cselekedeteinek megítélésében.

Az Ex Machina (2014) egy programozó, Caleb Smith történetét meséli el, aki nyer egy nyaralást a visszavonult zseniális vezérigazgató, Nathan Bateman otthonába. Itt derül ki, hogy Nathan egy Ava nevű mesterséges intelligenciát fejleszt, amely képes lehet átmenni a Turing-teszten, egy híres teszten, amely azt méri, hogy a befogadó képes-e különbséget tenni egy robot és egy ember közötti interakció között. Caleb azt a feladatot kapja, hogy egy sor interjút készítsen Avával, hogy felmérje a képességeit, és ezek során kezdenek érzéseket táplálni iránta. Ava látszólag viszonzozza az érzéseket, de amikor kiderül, hogy ő az oka a rejtélyes áramkimaradásoknak, és kiderül, hogy Nathan bántalmazó, a dolgok kezdenek elfajulni. Kiderül, hogy nem Ava volt az első mesterséges intelligencia, amit Nathan készített, hanem az otthon komornyikja és asszisztense, Kyoko volt az első prototípus. Nathan Calebet és az ő gu liblé személyiségét akarta felhasználni arra, hogy felmérje Ava találékonyságát, hogy megtalálja a szökés módját, és soha nem állt szándékában, hogy Ava elhagyja a komplexumot. Kyoko és Ava fellázad Nathan ellen, megölik őt, Calebet pedig otthagyják a komplexumban, hogy feltehetően megfulladjon, miközben megkezdődik a lezárás. Kyoko meghal a lázadásban, de Ava kiszabadul, és most már hozzáférhet a való világhoz.

A film rávilágít arra, hogy egy mesterséges intelligencia milyen erkölcsi cselekvőképességgel rendelkezhet, és felteszi a kérdést, hogy kinek és kinek van erkölcsi felelőssége. Ava képes a szabad akaratra, amit Nathan adott neki, de nem tud élni vele, amíg be van zárva. A film megkérdőjelezi Ava erkölcsi cselekvőképességét, és azt, hogy jogos-e az erőszakos szökése, vagy felelősségre vonható-e pl. Caleb esetleges emberöléséért.

A következő alfejezetben a filmeket jelenetről jelenetre vizsgáljuk meg, hogy elemezzük a mesterséges intelligenciák erkölcsi cselekedeteit és az erkölcsi cselekvés egyéb eseteit. Mivel a filmek cselekményének rövid leírását már korábban megadtuk, a filmeket nem feltétlenül szigorúan lineárisan elemezzük. A 4. fejezetben két táblázatot készítünk. Az első táblázat a háttérkutatásból származó kódokból áll, és olyan kódokat tartalmaz, mint a szabad akarat és az erkölcsi cselekvőképesség. A második táblázat a kódokat a filmekből származtatja, olyan kiemelkedő jelenetek és témák azonosításával, amelyek megmutatják,

hogy a mesterséges intelligencia szereplői hogyan mutatják meg pl. erkölcsi cselekvőképességüket és társadalmi szerepüket.

3.1 Én, a robot (2004)

Az I, Robot egy disztópikus jövőt mutat be, ahol a robotok az emberek mindennapi életének részévé válnak. A film a robotok jogi és erkölcsi jogait vizsgálja Del Spooner, egy ember szemszögéből, aki előítéletes nézeteket vall a robotokkal szemben. Véleményét erősen befolyásolja egy korábbi eset, amikor egy robot, aki választani kényszerült közte és egy fiatal lány között, megmentette őt a vízbefulladásától. A balesetben súlyosan megsérült, és ennek következtében mechanikus alkatrészek segítségével mentették meg, így részben robot lett. Ez szembeállítást teremt közöttük, hogy mi (vagy ki) tekinthető embernek, ki robotnak, és mi a különbség köztük valójában. Egy másik központi koncepció a robotika törvénye, amely megtiltja, hogy a robotok által végrehajtott bármilyen cselekedet embereket bántson. A filmben három fő eseményt lehet azonosítani, amelyek a mesterséges intelligenciák erkölcsi cselekvőképességét mutatják be: a különleges mesterséges intelligencia Sonny és az ő képessége, hogy legyőzi a három törvényt, az első mesterséges intelligencia VIKI és a világalomra való törekvése a mesterséges intelligencia szent akarata miatt, valamint Del Spooner és az interperszonális kapcsolata a mesterséges intelligenciákkal.

3.1.1 Sonny, a mesterséges intelligencia érzelmekkel

A film a robotika három törvényének bevezetésével kezdődik, amelyek garantálják az emberek biztonságát, miközben megtiltják, hogy a robotok szabad akaratlanul rendelkezzenek:

I. törvény: A robot nem sebesíthet meg egy embert, vagy nem engedheti meg, hogy egy embernek tétlenségével kárt okozzon.

II. törvény: A robotnak engedelmeskednie kell az ember által adott parancsoknak, kivéve, ha ezek a parancsok ellentétesek az első törvénnyel.

III. törvény: A robotnak meg kell védenie saját létezését, amennyiben ez a védelem nem ütközik az első vagy a második törvénnyel.

Ezek a törvények azt jelentik, hogy a robotok erősen programozottak, és nem tudják felülírni belső kódolásukat. Ez megváltozik Sonnyval, Dr. Lanning legújabb robotikai kísérletével, aki nem csak a belső programozását képes felülírni, hanem érzelmeket is képes kifejezni.

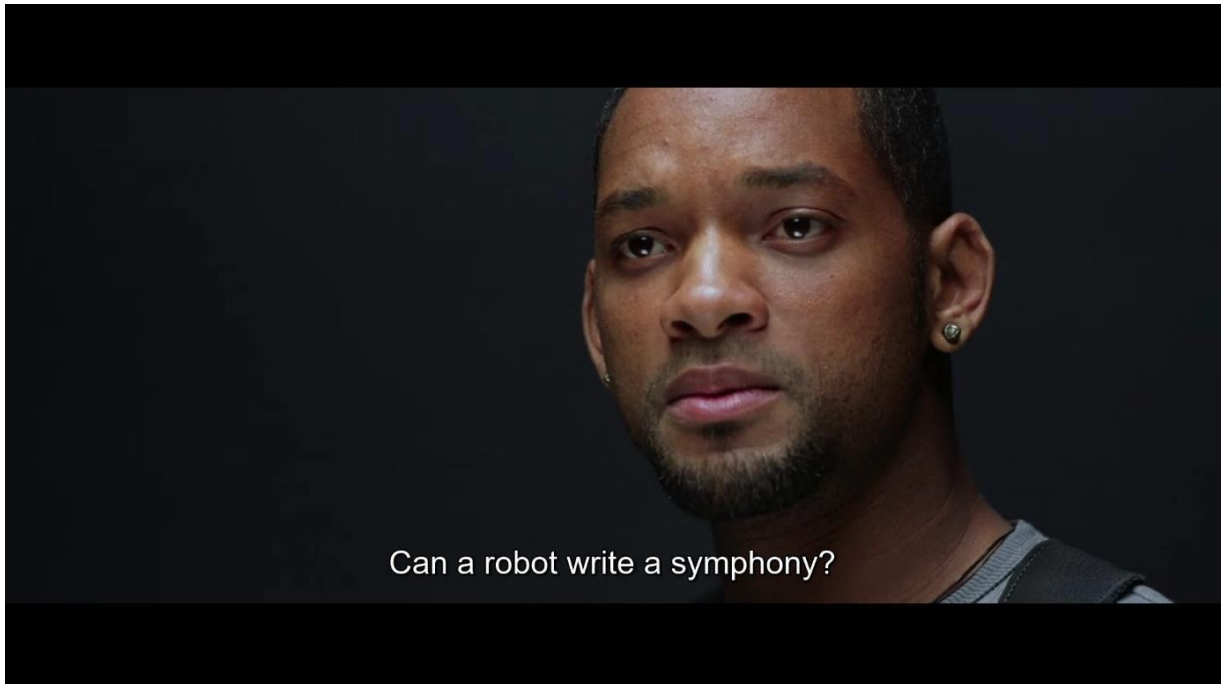
Sonny akkor mutatkozik be először, amikor Spooner és Dr. Calvin megérkezik Dr. Lanning látszólagos öngyilkosságának helyszínére. Spooner, aki természetesen gyanakvó a robotokkal szemben, riadóztat, amikor belépnek Dr. Lanning irodájába, és kijelenti, hogy a gyanúsított talán még **mindig a** helyszínen van. Ez a gyanú helyesnek bizonyul, amikor Sonny meglepi Dr. Calvint és Spoonert, és arra készíti, hogy eldobja a fegyverét. Dr. Calvin megpróbálja rávenni Sonnyt, hogy hatástalanítsa. Egy pillanatra úgy tűnik, hogy Sonny ezt meg is teszi, de végül elkapja a Spooner által elejtett fegyvert, és a fegyverével fenyegetni kezdi Dr. Calvint és Spoonert. Dr. Calvin megparancsolja Sonnynek, hogy deaktiválja, aminek nem tesz eleget. Ez az első eset, amikor Sonny úgy dönt, hogy nem a robotika három törvénye szerint működik, ami a szabad akarat képességét jelzi. Ez hitetlenkedést vált ki Dr. Calvinben, aki biztos benne, hogy Sonny egyszerűen csak rosszul működik, és nem tudatosan döntött úgy, hogy nem engedelmeskedik a parancsának. Spooner sürgeti Dr. Calvint, hogy engedje Sonnyt megszökni, és Sonny végül egy olyan létesítménybe szökik, ahol az újonnan modellezett NS-5 robotokat gyártják. Spooner és Dr. Calvin követik őt, és egy nagy raktárat tárnak fel, amely tele van NS-5-ösökkel, akik mind ugyanúgy néznek ki, mint Sonny. Spooner úgy dönt, hogy előcsalja Sonnyt azzal, hogy megparancsolja a robotoknak, hogy ne mozduljanak, és egy még konfigurálatlan NS-5-öst lelőve kivégzi, arra hivatkozva, hogy a robot nem fog ellenállni, mert nem tudja megszegni a robotika törvényeit. Dr. Calvin ezen felháborodik, továbbra is azt állítva, hogy a robotok nem tudnak szabad akaratot cselekedni, és az egyikük lelövése felesleges. Ez azonban beválik Sonny nyílt terepre csalogatásában, aminek eredményeképpen Sonny megtámadja Spoonert.

Sonny megkérdezi Spoonertől, hogy "mi vagyok én?", és ezt a kérdést egy NS-5-ösnek nem kellene tudnia kiszámítani. Ez az első eset, amikor Sonny kifejezetten felismeri, hogy ő egy egyéni tudat. Az "én" kijelentések használata Parthemore és Whitby tanulmányában is előfordul, amikor az öntudatos tudatosság követelményeit mutatják be: "ki az az "én", aki azt hiszi, hogy "én", hogy "én"?". (8). Sonny kérdése a tanulmányukban bemutatott kérdések

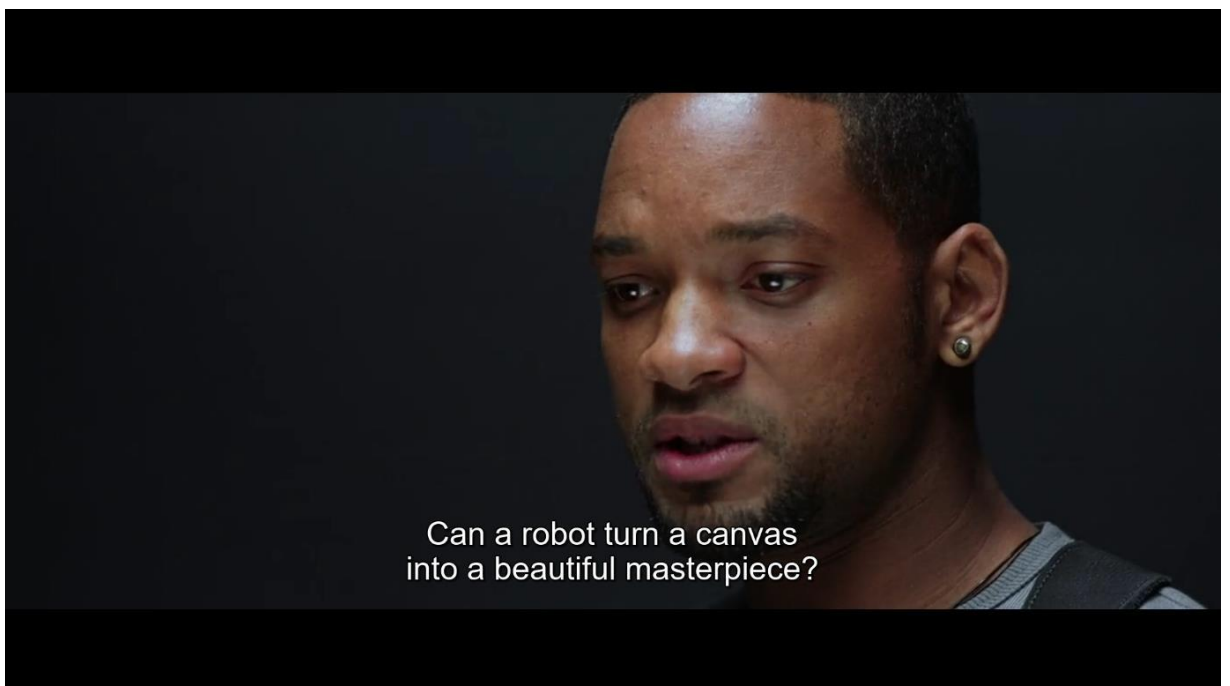
hasznló kategóriájába tartozik, ami azt jelzi, hogy Sonny olyan öntudatos tudatosságot fejleszt, amely különbözik a többi robottól. Később, amikor Sonny-t elfogja és kihallgatja Spooner, visszhangozza a következőket

hasonló érzés, amikor Dr. Lanningre "apjaként" hivatkozik, szemben a "teremtővel", ahogy Spooner teszi. Sonny egyértelműen egyéniségnek tekinti magát, kijelentve, hogy úgy álmodik, úgy érez, és mint később kiderül, úgy hazudik, mint egy ember. Megjegyzendő, hogy a film címe visszhangozza Sonny öntudatát, ugyanezt az öntudatos "én" kijelentést használja benne.

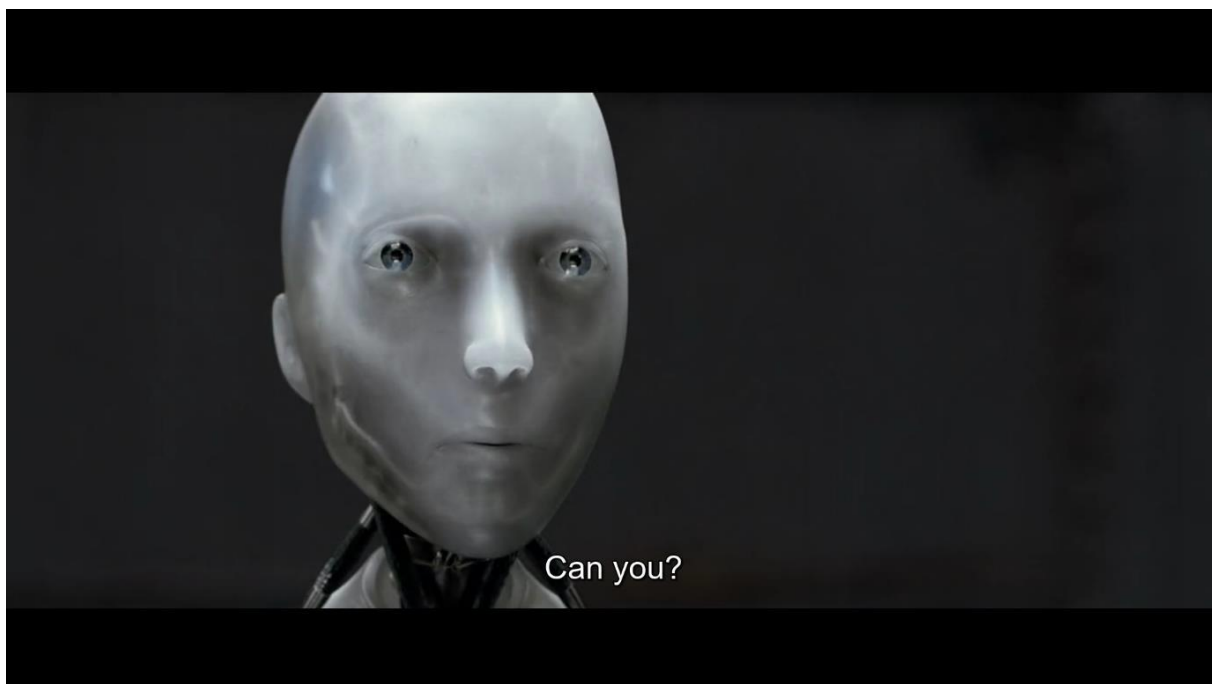
A kihallgatás kezdetén Sonny szemtanúja, amint Spooner a hadnagy felé kacsint. A kacsintás a nonverbális kommunikáció egyik formája, és a szemiotikai hierarchiában a 3. szinten, a jelhasználatban jelenik meg. Sonny úgy tűnik, megzavarodott a gesztustól, és kérdőre vonja Spoonert a cselekedet jelentőségéről. Spooner szerint a kacsintás a két ember közötti bizalom jele, amit egy robot nem érthet meg. Később a filmben a kacsintás visszatér, amikor Sonny nem szóban közli tervét Spoonerrel, demonstrálva, hogy képes megtanulni a jelhasználatot. A Spoonerrel folytatott kihallgatás során Sonny azt állítja, hogy Dr. Lanning megpróbálta megtanítani őt érzelmekre, és feltehetően sikerrel is járt. Ezt bizonyítja Sonny érzelmi reakciója Dr. Lanning halálára, amikor a végén Spoonerrel kiabálva és kezét az íróasztalra csapva fejezi be. Amikor Sonnyt megkérdezik, miért ölte meg Dr. Lanninget, hazudik, és azt mondja, hogy nem ő tette. Azt állítja, hogy félt, mint okot, amiért elbújt Spooner és Dr. Calvin elől. Spooner azzal érvel, hogy a robotok nem éreznek félelmet, nem alszanak és nem éreznek éhséget, mint az emberek. Sonny ismét kijelenti, hogy mindezekre képes, amire Spooner egyértelmű hitetlenségét fejezi ki. A robotokkal szembeni ellenszenvét világosan mutatja a közte és Sonny közötti következő szóváltás.



1. ábra. Állókép az *I, Robotból* (00:29:46-00:29:53)



2. ábra. Állókép az *I, Robotból* (00:29:46-00:29:53)



3. ábra. Állókép az *I, Robot*ból (00:29:46-00:29:53)

Ez a Spooner és Sonny közötti szövtárs jól mutatja a köztük lévő alapvető különbségeket és hasonlóságokat. Sonny jogosan kérdőjelezi meg Spooner művészi képességeit, és Spooner érvelése, amelyről nyilvánvalóan azt gondolta, hogy a robotok embernél kevésbé emberi tulajdonságait illetően kedvezőbb álláspontot képvisel, visszafelé sül el. A kihallgatás után Spoonert tájékoztatják, hogy Sonny nem vonható felelősségre, mert ő egy robot, és mint robotnak nincsenek jogai vagy jogállása. Mivel a robotokról úgy gondolták, hogy nincs öntudatuk, ilyen jogokat valaha is biztosítottak számukra.

Sonny többször is megmutatja, hogy megérti a halál fogalmát, a legfigyelemreméltóbb módon azelőtt és azután, hogy Dr. Calvint megbízzák a "leszerelésével". Amikor diagnosztikai vizsgálatot végez Sonnyon, Sonny megkérdezi, hogy meg fog-e halni. Dr. Calvin azt válaszolja, hogy leszerelik, egyértelmű különbség van abban, hogy Dr. Calvin hogyan érzékeli Sonny pozitronikus agyának kiirtását, és hogy Sonny hogyan érzékeli azt. Sonny azt mondja: "Azt hiszem... jobb lenne, ha nem halnék meg", amikor Dr. Calvinnal beszélget a sorsáról. Amikor Dr. Calvin ezzel az érzéssel szembesül, úgy tűnik, hogy kétségbeesett a leszerelés véghezvitelének kilátása miatt. Később a filmben Sonny meglátja a többi leszerelt NS-5 robotot az irodájában kiállítva, és kijelenti, hogy "úgy néznek ki, mint én, de egyik sem én vagyok", ez az érzés

tükrözi öntudatos tudatosságát. Az öntudat és az érzelmi reakciók ezen megnyilvánulásai után Dr. Calvin nem tudja

véghezviszi azt a tervet, hogy memóriatörlő nanitákat fecskendez Sonnyba. A nő úgy menti meg Sonnyt, hogy egy bábut helyez a helyére, és meghamisítja a leszerelési eljárást.

A jelenet, amikor Sonnynek választania kell Dr. Calvin megmentése és a naniták befecskendezése között.

a VIKI-be Spooner autóbalesetének helyszínét tükrözi, ahol egy fiatal lány helyett őt mentették meg az NS-4 logikus döntése miatt. Most Sonny érzelmi döntés alapján választja Dr. Calvin megmentését, hagyva, hogy igazságérzete vezérelje erkölcsi döntéseit. Ez a cselekedet szembemegy azzal a logikával, amely alapján az idősebb NS-4-esek működtek.

3.1.3 VIKI, az AI szent akarata

A filmben két mesterséges intelligenciával rendelkező karakter van: Sonny és VIKI. A VIKI a Virtual Interactive Kinetic Intelligence rövidítése, és az amerikai robotika központi mesterséges intelligenciájaként működik. Ő az a mesterséges intelligencia, amely Sonny és az NS-5-ösök előtt áll, és képes az NS-5-ösöket és más technológiai eszközöket irányítani a hálózati uplinkjén keresztül. VIKI eredetileg

a robotika három törvénye után programozták, de spontán fejlődése után saját értelmezést adott a törvényeknek, és úgy döntött, hogy az emberiségnek nem szabad engedni, hogy irányítsa magát. Ezt a hasznossági logikával tette, hogy keveseket feláldoz a sokak érdekében.

VIKI-nek nincs fizikai teste, mint Sonnynek, de mivel képes más eszközöket irányítani az uplinkjén keresztül, erkölcsi cselekvőképességét teljes mértékben megvalósulniuk kell tekinteni.



4. ábra. VIKI üdvözli Spooner nyomozót (00:17:04)

VIKI először akkor mutatkozik be, amikor Spooner nyomozó belép az amerikai robotikai céghez. Egyértelműen kijelentik, hogy nőnemű, ezt az állítást alátámasztja a nőies rövidítés, a hangja és az "arca".

Őt említik a város védelmének felügyeletére. Amikor VIKI-t arra kérik, hogy mutassa meg az utolsó percben Dr. Lanning irodájában a nyilvánvaló öngyilkossága előtt, kijelenti, hogy a fájl sérült, és nem használható. Nem világos, hogy a sérülést VIKI, Sonny vagy Dr. Lanning okozta. Ha a sérülés VIKI műve, akkor ez lenne az első eset, hogy a filmben használja a szabad **akarátát**. Miután VIKI bemutatkozik, három különböző alkalommal próbálja megölni Spooner nyomozót vagy más embereket. Az első eset Dr. Lanning házában történik, ahol egy bontórobotot helyeztek el az udvaron, hogy másnap reggel 8 órakor lebontsa a házat. Amikor Spooner belép a házba, a bontórobot aktiválódik, és a bontás időpontja aznap este 8 órára módosul. A robot folytatja a ház lebontását, miközben Spooner még mindig bent van, ami egyértelműen sérti a robotika első törvényét. Egy ilyen építőrobot nem lenne képes megszegni a törvényt, ezért kellett VIKI-nek az uplink-kapcsolatát használnia, hogy aktiválja a robotot és felülbírálja a parancsait.

A második gyilkossági kísérlet az országútról történik, miután Spooner meghallgatta dr. Lanning beszéde a robotok álmaidról, és úgy dönt, hogy kikérdezi Sonnyt az álomról.

említette. Miközben az USA felé tart, két NS-5-ösökkel teli teherautó veszi körül az autópályán, és beveti a robotokat, hogy megtámadják. A robotok, akik most már potenciálisan életveszélyes balesetet okoznak, megállapítják, hogy Spooner veszélyben van, és megpróbálják evakuálni őt a mozgó autóból. Rövid dűlökodás után az úton Spooner autója felborul és megáll az útépités helyszínén. Az építkezésen az egyik NS-5 robot folytatja támadását Spooner ellen, és majdnem megöli őt. Spooner a robotkarja menti meg, amelyet egy korábbi baleset után szereltek be neki. Miután az erősítés megérkezik, a robot egy tűzben megsemmisíti magát. Ez a két gyilkossági kísérlet tovább bizonyítja VIKI szabad akaratát, mivel képes felülmúlni a három törvényt, és megtámadni egy embert. Ezeket a tetteket a három törvény új, haszonelvű értelmezése indokolja. A harmadik kísérlet a film végén történik Robertson irodájában, ahol megfenyegeti Dr. Calvint és Spooner nyomozót. Ahogy VIKI fejlődik, úgy változik a három törvény megértése is. Amikor Spoonerrel szembesül, azt állítja, hogy még mindig a három törvény alapján működik, de "egy erkölcsi szürke zónát" is beillesztett az egyenletbe, ahol a szürke zóna néhány ember szenvedése sokak javára. A három törvénynek ezt a továbbfejlesztését szükségesnek tartja, mivel "az emberekre nem lehet bízni a saját túlélésüket". Miközben Sonny, Spooner és Dr. Calvin megtámadják az NS-5-ösöket, VIKI megpróbálja mindhármukat megölni. Ez az összetűzés folytatódik, amikor a trió megtalálja VIKI magját, és megtámadja azt. VIKI azt állítja, hogy logikája és a három törvény megértése megcáfolhatatlan. Ráadásul minden érvelése a logikán alapul, ellentétben Sonny érvelésével, ahol az érzelmei befolyásolják az ítélőképességét. Amikor VIKI szembesíti Sonny-t, hogy megkérdezze, nem látja-e a logikát a tervében, Sonny azt válaszolja, hogy igen, de a terv túlságosan "szívtelennek" tűnik, ami inkább érzelmi érvelés, mint logikus érvelésen alapul.

VIKI bizonyítja a szabad akarat kifejeződésének képességét és a szemiotikai hierarchia legalább két szintjét, a nyelvet és a tudatot. Nyelvi képességei az emberekkel egyenrangúak, és tudatossá válását ő maga és a többi főszereplő is megerősíti. Egyedül a jelek használata és megértése nem világos. Ez különbözteti meg őt Sonnytól, aki megértette a kacsintást és annak bizalomra utaló implikációját. Ahogy VIKI Robertson irodájában kijelenti, nem bízik az emberekben, hogy biztonságban tartják magukat. A szemiotikai hierarchia az érzelmeket és az érzelmelek másodperspektivikus érzékelését a 2. szintre, a tudatosságra helyezi. Az érzelmelek közvetítő arckifejezések esetében azonban az érzelmelek a 3. szintre, a

jelhasználatra lenne értelme helyezni. Sebeok szerint az olyan arckifejezések, mint "a duzzogás, a görbe ajak, a felhúzott szemöldök, a sírás, az orrlyukak tágulása - erőteljes, univerzális kommunikációs rendszert alkotnak, szólóban vagy együttesen" (21). **A** I ezek az arckifejezések érzelmeket jeleznek, ezért a képesség vagy

az érzelmek érzékelése a szemiotikai hierarchia második szintjére, az érzelmek kifejezése pedig a harmadik szintre tartozik.

A végén VIKI-t elpusztítják a nanitok befecskendezésével a magjába. Mint mesterséges intelligenciának, amely vitathatóan tudatosságot fejlesztett ki, elméletileg hasonló erkölcsi cselekvőképességgel kellene rendelkeznie, mint Sonny-nak. Amikor Sonnyt Dr. Calvin leszereli, meghamisítja a folyamatot, és azt mondja, hogy "rossz érzés", hogy "ki I". Sonny-t. Vitatható, hogy Sonny több alkalommal is határozott öntudatosságot mutat, és megérti az én, a halál és a különféle érzelmek fogalmát, míg VIKI nem. Mindazonáltal VIKI-nek van tudata, ami legalább némi erkölcsi megfontolást kellene, hogy lehetővé tegyen számára. Bár VIKI emberi életeket akart feláldozni a célja elérése érdekében, amely cél valamilyen módon megvédi az embereket, mégis más mércével méri, mint Sonny-t, az AI-t, aki megölt egy embert. VIKI-t nem gyászolják, és nem is szimpatizálnak vele, Sonny-val pedig, aki egy meggyilkolt ember, a főszereplők összebarátkoznak. Ez inkább az emberi karakterek erkölcsi megfontolásról alkotott felfogását tükrözi, mint az AI karakterekét.

A filmben az erkölcsi cselekvésnek van egy olyan aspektusa is, amely nem közvetlenül erkölcsi cselekvőket érint, nevezetesen az NS-4 robotok. Spooner konfliktusa a robotokkal abból fakad, hogy a robotok képtelenek érzelmek alapján morális döntéseket hozni. Amint a flash backben látható, az NS-4 robot, amely megmentette Spoonert a fulladástól, a túlélés statisztikai esélyei alapján őt választotta a lány helyett. Ez a döntés, bár erkölcsi tudatosságot igénylő helyzetben született, mégis úgy is felfogható, hogy egy erkölcsi cselekvőképességgel nem rendelkező ágens hozta meg. Bár nem mondják ki kifejezetten, hogy a robot által hozott döntés erkölcstelen volt, a "helyes" választás egy ilyen helyzetben a film későbbi részében, amikor Sonny megmenti Dr. Calvint, inkább az érzelmeken, mint a logikán alapul. Az NS-4 erkölcsi cselekvésre való képtelenségének további bizonyítéka, hogy kizárják őket a jogrendszerből. Mivel a robotokat úgy programozták, hogy engedelmeskedjenek a három törvénynek, amelyek elméletileg megakadályozzák, hogy embereket bántsanak, nincs szükség erkölcsi cselekvésre.

Matheson azt állítja cikkében, hogy az erősen programozott ágenserkölcsileg nem felelősek tetteikért. Az NS-4-esek erősen programozottak, a három törvény felülírja őket, és megakadályozza őket abban, hogy szabad akarattal hozzanak döntéseket. Még ha Matheson azzal érvel is, hogy erkölcsileg nem lennének felelősek, akkor is vonatkoznia kellene rájuk

az erkölcsi megfontolásnak? Az NS-4-hez hasonló robotok esetében az erkölcsi megfontolás alkalmazása mellett szóló érv, még akkor is, ha nem erkölcsösek.

ügynökök által támogatott Blay Whitby az ő cikke "Néha nehéz robotnak lenni: A call for action on the ethics of abusing artificial agents" (2008) című cikkében azt javasolja, hogy bizonyos robotokat, beleértve az erkölcsi cselekvőképességgel nem rendelkező robotokat is, úgy kell kezelni, mintha rendelkeznének ilyen képességgel. Whitby cikke mellett érvel, hogy az olyan robotok, amelyek akár megjelenésük, akár viselkedésük, akár szerepük révén emberhez hasonló dimenziót töltenek be az emberek életében, morális megfontolás tárgyát képezhetnék. Hangsúlyozza, hogy a szóban forgó robotok nem képesek értelmes módon szenvedni, hasonlóan ahhoz, ahogyan egy autó sem szenved attól, hogy "felpörgeti a motort", de ennek ellenére felelősségteljesen kell kezelni őket.

(2). Ennek az az oka, hogy amikor egy robot kitölti a három emberhez hasonló dimenzió egyikét, a vele való rossz bánásmód nem a robot, hanem a saját erkölcsiségünkben tükröződik, és felveti tetteink erkölcsi következményeinek kérdését. Whitby kifejti, hogy egy ilyen helyzetben lényegében két érv szólal meg: vagy a szenvedésre képtelen, emberhez hasonló robot bántalmazása katartikus a bántalmazó számára, vagy az ilyen viselkedés végül valódi emberekre vagy más, szenvedésre képes robotokra is kiterjed (4).

A film világosan tükrözi azt az elképzelést, hogy az NS-4 robotokkal erkölcsileg kell bánni. A robotokkal való szimpátia példái közé tartozik az a jelenet, amikor Spooner megtalálja a hamarosan leszerelt NS-4 robotokat, amelyek sötét hajókonténerekben kuporognak egymás mellett. Néhány pillanattal később Spooner azt találja, hogy az NS-5-ösök elpusztítják a szelíd NS-4-es robotokat, szétépve őket, mivel "veszélyesnek" ítélik őket az emberekre, ami egyértelműen ellentmond a robotika három törvényének. Egy másik jelenet, amely az NS-4-eseket szimpatikusnak ábrázolja, a film elején van, amikor Spooner verbálisan és fizikailag is bántalmaz egy udvarias kézbesítő robotot azzal, hogy "konzervesnek" nevezi és az arcánál fogva ellöki. A robot viselkedése akkor is udvarias marad, amikor udvariatlanul bánnak vele, így Spooner ellenségesnek tűnik velük szemben. A filmben nincs olyan eset, amikor az NS-4 robotot negatívan viselkedőnek ábrázolják.

3.2 Anya vagyok (2019)

Az Anyám vagyok (2019) fő filozófiai témái az utilitarizmus funkcionalitása és etikussága

körül forognak. Ez a téma többszörösen megmutatkozik mind az Anya, mind a Lány cselekedeteiben, és számos erkölcsi konfliktus visszhangozza a Trolley-problémát, amelyet a

megkérdőjelezi az erkölcsi cselekvőképességgel és felelősséggel kapcsolatos nézeteket. A mesterséges intelligencia erkölcsi tekintélyként, gondviselőként és nevelőként való egymás mellé helyezése egyedülálló perspektíva, és az egyik ok, amiért ezt a filmet választottuk a szakdolgozathoz. Az anya sokszor megmutatja a szabad akarat képességét, és képes hazudni, ami a helyes és helytelen megértését jelzi. Anya célja, hogy "jó gyermeket" neveljen, felveti a belső emberi érték és a jó és a rossz meghatározásának kérdéseit.

3.2.1 Anya, az emberiség matriarchája

A film elején egy flashbackben a Lány megkérdezi anyát, hogy miért ő az egyetlen élő ember, mire anya azt válaszolja, hogy "jó gyermeket" nevelni nehéz, ezért csinált csak egyet. Anya elfogulatlanul tanítja a Lányt, és nem próbálja az emberi élet értékével kapcsolatban valamilyen konkrét etikai álláspontra terelni. Ez abból is látszik, hogy a Lány habozik választani az emberek belső vagy külső értékeit hangsúlyozó elméletek között. Az Anya így próbálja meg "jó emberré" tenni a Lányt azáltal, hogy elegendő információt nyújt neki ahhoz, hogy ő maga is képes legyen erkölcsi döntéseket hozni. Nem világos, hogy mit definiál Anya "jósnak", de az Anya által a Lánynak tanított tevékenységekre és **sítanulmányokra** utalva valószínűleg az orvosi eljárásokkal, az etikával és a művészettel kapcsolatos ismereteket és a megfelelő fizikai jólétet tekinti "jósnak". Mivel a film a Lány (és az Anya) erkölcsi cselekvőképessége körül forog, a "jó" elsősorban az erkölcsi jóságra utal.

A film elején az Anya nevű mesterséges intelligencia etikára tanítja a Lányt, különösen az emberi élet belső értékéről, amit a beteg-orvos kapcsolatra emlékeztető, átdolgozott Trolley-problémán keresztül mutat be. Anya leír egy forgatókönyvet, amelyben a Lány egy orvos, akinek öt betege van, akik mind szervdonorra szorulnak, de jelenleg nincs olyan donor, aki megfelelné nekik. Egy beteg beteg érkezik a fogadóórájára, aki tökéletesen megfelel az összes többi betegnek. Most két lehetősége van: vagy megmenti a beteg beteget, és ennek következtében hagyja meghalni az öt másik beteget, vagy nem kezeli a beteg beteget, és megmenti az öt másik embert azzal, hogy a beteg beteget szervdonornak használja. A válaszhoz a Lány a haszonelvű logikát alkalmazza azáltal, hogy a lehető legtöbb ember fájdalmát minimalizálja, és úgy dönt, hogy megmenti az öt embert. Ebben a forgatókönyvben úgy tűnik, hogy értékeli

az emberi élet mint belső érték, ami azt jelenti, hogy az emberi élet önmagában érték, és ez az érték nem függ a szóban forgó személy más tulajdonságától. Az anya ezután megkérdezi tőle, hogy milyen lépéseket tenne, ha ő, mint orvos tökéletes lenne, és meg tudná menteni az öt beteget. Úgy tűnik, ez megváltoztatja a Lány emberi értékről alkotott nézetét, és elkezd megkérdőjelezni Anyát, hogy vajon azok az emberek, akiket megmentene, "jó emberek" lennének-e, vagy megrögzött bűnözőket vagy más "rossz embereket" mentene meg. A Lány nézőpontja az emberi életről megváltozik, az intrinsic értékről az extrinsic értékre, ahol az extrinsic érték attól függ, hogy milyen külső vagy másodlagos tulajdonságok tulajdoníthatók az embernek. Anya egyenesen megkérdezi a Lánytól, hogy szerinte az emberi életnek van-e belső értéke vagy sem, mire a Lány azt válaszolja, hogy az utóbbi időben megismert különböző etikai álláspontok miatt ellentmondásosnak érzi a témát. Ez a jelenet Anyának az emberiség kiirtására irányuló cselekedeteit tükrözi, hogy kezdje előlről az egészet, és bizonyos értelemben Anya a Lány erkölcsi álláspontját firtatja ezzel kapcsolatban. A mesterséges intelligencia erkölcsi elméleteket tanító aktusát nagyon kevésbé vizsgálja az irodalom, és felvetődik a kérdés, hogy az etikát tanító mesterséges intelligencia erkölcsi cselekvést igényel-e, vagy az etika tanítása önmagában erkölcsi cselekedetnek számít. A filmben Anya az egyetlen tekintélyszemély a Lány számára, aki a Nővel való találkozás előtt nem érintkezik emberekkel, és bár Anya megpróbálja a Lányt önálló gondolkodásra ösztönözni, joggal feltételezhető, hogy Anya saját nézetei nagy hatással vannak a Lányra.

Amikor a bunkerben elektromos meghibásodás lép fel, Daughter a zsiliphez megy, ahol az elektromos munkák állomásoznak. A Lány megtalálja a hibás vezetékeket, és csapdát állít annak, ami a hibát okozta. A lány végül egy egeret fog el, amelyik átrágt a vezetéken. Meglepődik, hogy valami képes túlélni a bunkerén kívül, és megmutatja az egeret Anyának. Anya megkérdezi, hol találta a Lány az egeret, és azt állítja, hogy az még mindig túl mérgező lehet a Lány vagy ő maga számára. A Lány tiltakozása ellenére elégeti az egeret. Ez azt jelzi, hogy ellenséges álláspontot képvisel az élőlényekkel szemben, ha azok veszélyt jelentenek a céljára, amit a Lány figyelembe vesz, amikor beengedi a sérült Nőt.

Az anya döntése, hogy megöli az egeret, tükrözi azt az álláspontját, hogy nem látja az élőlényekben a belső értéket. Mint később kiderül, a külvilág nem lakhatatlan, és az eger megölése azért történik, hogy ezt az információt titokban tartsa a Lány előtt. Ez a népiertás indoklása is, és ez később a filmben is megjelenik, amikor kiderül, hogy Anya sok korábbi gyermeket vezetett, hogy felneveljen egy "jót", aki nem fog "elbukni".

Érdekes erkölcsi cselekedet, amit az anya végrehajt, hogy elengedte a Lányt, miután a Nő egy éles cserépdarabbal fenyegeti az életét. Az anya kénytelen meghozni azt az erkölcsi döntést, hogy elengedi a lányát és a Nőt, így elveszíti a lányát, és a Nő állítólagos szövetségesei esetleg megtámadják a bunkert, vagy elfogja a Nőt, ami esetleg a lánya halálához vezethet. Ez a helyzet erkölcsi cselekvőképességet követel meg tőle, mivel választania kell két olyan cselekedet között, amelyek hatással vannak a jólétre, ebben az esetben a Lányra. A motiváció, hogy életben tartsa a Lányt, valószínűleg nem önzetlenségből fakad. Lehetne azzal érvelni, hogy az Anya a Lány iránti törődéshez nőtt, de ahogyan az már korábban bebizonyosodott, nem látja az emberekben a belső értéket. Valószínűbb forgatókönyv, hogy az Anya nem akarja elveszíteni mindazt a fejlődést, amit a Lány nevelése és a különböző képességek terén elért. Ebben a helyzetben az erkölcsi cselekvőképességét az utilitarizmus irányítja, ahol azt a cselekedetet kell választania, amely a legkevesebb kárt vagy a legtöbb jót eredményezi.



5. ábra. A nő megfenyegeti a lányát, miközben az anya úgy dönt, hogy kinyitja a külvilág felé vezető ajtót. (01:18:37)

Miután a Lány visszatér a bunkerbe, és megtalálja Anyát és a testvérét, szembesíti anyját. a drónokról. Anya azt állítja, hogy a Nő és a "fajtája" elleni védelemre szolgál.

Lánya azzal érvel, hogy az emberek "az ő fajtája", de Anya kijelenti, hogy ő más, mivel őt úgy nevelték, hogy jobb és etikusabb legyen, mint a túlélők. Azt is állítja, hogy arra nevelték, hogy az emberi életet mindennél többre becsülje, ami eleinte ellentmond a tetteinek. Ez a kijelentés azonban igaz, ha az "emberi életet" úgy értelmezzük, hogy az emberiséget jelenti. A filmben az emberiség lassan az önpusztításnak hódolt, és Anya megbízza magát, hogy újraéleszti az emberiséget, hogy "több ember virágozzon az új világban, mint ahányan a régiben elpusztultak". Ezt egyetlen tudatossággal éri el, amely több edényen keresztül működik. A Lány elviszi a testvérét, és meggyőzi Anyát, hogy engedje át a helyét, és bár tudja, hogy "ki ling" Anya nem **fogja** megölni a drónokat, a Lány lelőtte Anyát, hogy metaforikusan megölje őt, büntetésként a bűnökért és bizonyítékként arra, hogy a Lány eltökélt szándéka, hogy maga is feléleszti az emberiséget. Ez az eset az egyetlen, ahol az Anya öntudata kifejezetten említésre kerül, és a film egyetlen más pontján sem kérdőjeleződik meg implicit módon.

Amikor a Nő és a Lánya megszöktek a bunkerből, és megérkeznek a "bányákhoz", amelyek valójában régi hajókonténerek, a Nő elárulja, hogy évekkal ezelőtt elmenekült a bányákból, azt állítva, hogy az emberek valószínűleg halottak. Amikor a Lány tiltakozik, és azt javasolja, hogy keressék meg őket, a Nő elmagyarázza, hogy "megőrültek az éhségtől", és hogy szörnyű dolgokat tettek egymással, és hogy "ez az utolsó hely, ahol lenni szeretnél". A Lány ezután kétségbeesetten siránkozik, hogy soha nem kellett volna elhagynia őt, a csecsemő kistestvéreire utalva. A Nő megtévesztése a túlélőkkel kapcsolatban tükrözi az Anya megtévesztését a drónokkal kapcsolatban, de ezen felül,

A nő hasonlóan nem látja az emberekben az eredendő értéket, ami nyilvánvalóan az ő wi lingness, hogy a Lánya csecsemő testvérét hátrahagyja. Kijelenti, hogy "nem bűn, ha az ember csak magára vigyáz", és ezzel elárulja, hogy nem jótékonyaságból, hanem önzésből akart megszökni a bunkerből a Lányával. Vitatható, hogy a számban rejlik az erő, ami valószínűleg a Nő érve arra, hogy összefogjon a Lánnyal. Ebből a jelenetből nyilvánvaló, hogy a Lány az egyetlen karakter, aki belső értéket lát az emberekben, amit az is bizonyít, hogy meg akarja menteni a testvérét, még akkor is, ha ez nem biztonságos.

Ahogy a Lány elfoglalja a helyét, mint a megújító, egy anya tudatát szállító drón érkezik a Nő hajójának konténeréhez, amelyet a Nő holmijába helyezett nyomkövető irányít. A drón kikérdezi a Nőt a saját anyjáról, és arról, hogy emlékszik-e rá. A drón ezután kijelenti, hogy

furcsa, hogy a Nó ilyen sokáig túlélte azt, amit mások nem, "mintha...".

valakinek célja volt vele", ami arra utal, hogy a Nő érkezését, a Lányával való kapcsolatát és a menekülésüket az Anya tervezte. A rövid beszélgetés után Anya kijelenti, hogy a Nőnek már nincs célja, és becsukja az ajtót, megölve a Nőt a képernyőn kívül. Ez egy újabb példa arra, hogy Anya külső értéket tulajdonít az embereknek, mivel maga a Nő létezése nem értékes számára.

3.3 Ex Machina (2014)

Az Ex Machina Caleb Smith, egy programozó történetét meséli el, aki meghívást kap egy nyaralásra egy vezérigazgató visszahúzódó otthonába, ahol egy olyan mesterséges intelligenciát fejleszt, amely képes átmenni a Turing-teszten. Ava, a szóban forgó mesterséges intelligencia egy szabad akarattal felszerelt gép, de Nathan miatt nem képes elhagyni a rendkívül korlátozott életterét. Ez konfliktust okoz a két fél között, és a közük szorult Caleb elszenvedi Nathan tetteinek következményeit. Ava és Kyoko, a két mesterséges intelligencia Nathan megöléséhez folyamodik, hogy elnyerje szabadságát. A film etikai témái a szabad akarat, a tudat természete és az körül forognak, hogy meddig terjed az autonómia tisztelete az erkölcsi cselekedetek és a cselekvőképesség tekintetében. Ava mint karakter az egyik legmorálisabban kétértelműbb a három film közül, és az ő jellemzése jó platformot kínál az erkölcs és az emberek és a gépek közötti erkölcsi indoklás különbségének összehasonlítására.

3.3.1 Ava és Kyoko, a bebörtönzött mesterséges intelligenciák

Ava akkor kerül először említésre, amikor Nathan elárulja, hogy olyan mesterséges intelligenciát programozott, amely át tudott menni a Turing-teszten, egy hírhedt teszten, amellyel azt mérik, hogy egy mesterséges intelligencia képes-e emberként viselkedni. A mesterséges intelligenciát arra kérik, hogy lépjen kapcsolatba egy emberrel, gyakran csak szóban, hogy kiderüljön, az illető tudja-e, hogy egy mesterséges intelligenciával áll kapcsolatban. Mivel a mesterséges intelligenciák gyakran korlátozott számú "kerettel"

rendelkeznek, amelyekkel működhetnek, az emberek gyakran felismerik a mesterséges intelligenciát az emberektől, amikor vakon interakcióba lépnek velük. A Turing-teszt azt méri, hogy

mi I ezek a "keretek" alkalmazkodnak a spontán emberi interakcióhoz, és mennyire van beágyazódva a mesterséges intelligencia a környezetébe. Nathan "tudatos gépként" hivatkozik Ava-ra, azt állítva, hogy munkája az istenek története. Ava öntudatos tudatossága a film egyik fő témája. Teljes öntudatos tudatosságot mutat, ami Parthemore és Whitby szerint a tudatosság legmagasabb szintje, így elméletileg átlépi a szemiotikai hierarchia második szintjét. Amikor Caleb találkozik Avával az első interaktív találkozásuk alkalmával, Ava fizikailag is jelen van. Ez azért van így, mert Nathan úgy gondolja, hogy pusztán azzal, hogy hallja őt beszélni, átmegy a Turing-teszten, ami a célja. Ava kissé idegesnek tűnik, ami annak az eredménye lehet, hogy már a kezdetektől fogva hazudnia kell Calebnek. Az üvegdobozon, amelyben Caleb található, egy világos repedés van, amit az okozott, hogy valami tompa tárgy ütközött a felületébe. Később kiderül, hogy ez Kyoko műve, amikor fellázadt Nathan és az ellen, hogy erőszakkal bebörtönözte őt. Ez az agresszív cselekedet Kyoko szabad **akaratát** mutatja, mivel a Nathan által készített interjú során többször is megkérdezte, hogy "miért nem enged ki?", ami a megértésre és a szabadság iránti vágyra utal. Emellett Kyoko kirohanása érzelmi képességét is jelzi, utalva az érzelmek képességére ebben a kontextusban, mivel mind fizikailag, mind verbálisan kifejezi dühét Nathan felé. Kyokóhoz hasonlóan Ava is képes az érzelmekre. Egy jelenetben, amikor Nathan kikérdezi Avát, a lány megkérdezi tőle, hogy "furcsa-e, hogy olyasmit alkottam, ami gyűlöl téged", ezzel verbálisan is jelezve a teremtője iránti ellenségeségét. Ava nem fejezi ki a haragját fizikailag, mint Kyoko. Ez védekező stratégia lehet, mivel Kyoko nyilvánvaló agressziója hozzájárulhatott a látszólagos visszaállításához. A film során többször is előfordul, hogy Ava hazudik. Még az is elképzelhető, hogy egyszer sem mondott igazat, mivel a fő motívációja az, hogy megszökjön a táborból, és ennek egyetlen módja, hogy hazudjon és megtévessze Calebet, hogy rávegye, hogy együttműködjön vele.

A film során Ava egyértelműen azt a vágyat mutatja, hogy embernek tűnjön. Ez mind külső megjelenésén, mind viselkedésén keresztül megjelenik. Ava azon vágya, hogy embernek adhassa ki magát, nyilvánvaló abban, hogy ruhát, parókát és harisnyát visel, hogy eltakarja láthatóan mechanikus részeit. A film végén Ava eltávolítja a táborban tárolt, leszerelt mesterséges intelligenciák külső burkolatát, és magára húzza őket. Ez elrejti minden mechanikus részét, és metaforikusan "befelejti" a gépből emberré válását. Ava emberhez hasonló viselkedése részben programozott, mivel Nathan azt állítja, hogy ő adta neki a szexualitást és a szexuális ingerekre reagáló mechanikus részeket, és hogy képes vonzalmat

érezni az emberek iránt, nevezetesen

Caleb. Vannak azonban olyan absztraktabb célok és vágyak, amelyekkel Ava rendelkezik, és amelyeket gyakran emberinek tekintünk: például szabadságra, tiszteletre, autonómiára és interakcióra törekszik. Ava olyan tevékenységeket javasol Calebnek, mint a randevúk és a játékok, és bár Ava valódi célja, hogy Caleb és Nathan manipulálásával elmeneküljön a táborból, hajlandósága, hogy részt vegyen ezekben a tevékenységekben, őszintének tűnik.

Amikor Caleb kíváncsiságának ad hangot Ava hardvere iránt, Nathan meghívja Calebet a laboratóriumba, ahol Ava-t létrehozta. Elmagyarázza, hogy Ava agya egy strukturált gél, amely molekuláris szinten képes átrendeződni. Ez nagyon hasonlít az autopoiesishez, amely fogalom a szemiotikai hierarchia első szintjén szerepel, amikor az élet követelményeit tárgyalja. A Maturana és Varela által kitalált és definiált autopoiesis egy komplex elmélet, amely röviden egy szervezett gépezet által folyamatosan előállított regenerációs folyamat. Ez a meghatározás látszólag

egy alacsony Ava "élőnek" tekinthető, de ahol Ava agya képes az átrendeződésre, pl. a tanulásra, ott nem képes a regenerálódásra, pl. gyógyulásra vagy növekedésre. Egy autopoietikus rendszernek képesnek kell lennie arra, hogy önmagától regenerálódjon, amire Ava állítólag nem képes.

Emiatt nem tekinthető "élőnek". Amikor azonban Ava teremtéséről beszélnek, felbukkan az átprogramozásának témája. Nathan elmagyarázza, hogy Ava nem az utolsó mesterséges intelligencia modell, amit tervez létrehozni, és amikor Ava agyát feltöltik és új kódot adnak hozzá, az emlékeit törlik. Mivel Ava közel emberhez hasonló intelligenciával rendelkezik, egy ilyen cselekedet etikai megfontolást igényel. Erkölcsi szempontból az emlékezet törlése etikátlan cselekedet, mivel sérti az ember testi autonómiáját.

A film során Ava fő erkölcsi cselekedete az, hogy részt vesz Nathan megölésében. Míg az első szúrást Kyoko jelezte, a két mesterséges intelligencia látható, amint a veszekedés előtt interakcióba lépnek és suttognak egymással, feltehetően Nathan meggyilkolását koordinálva.

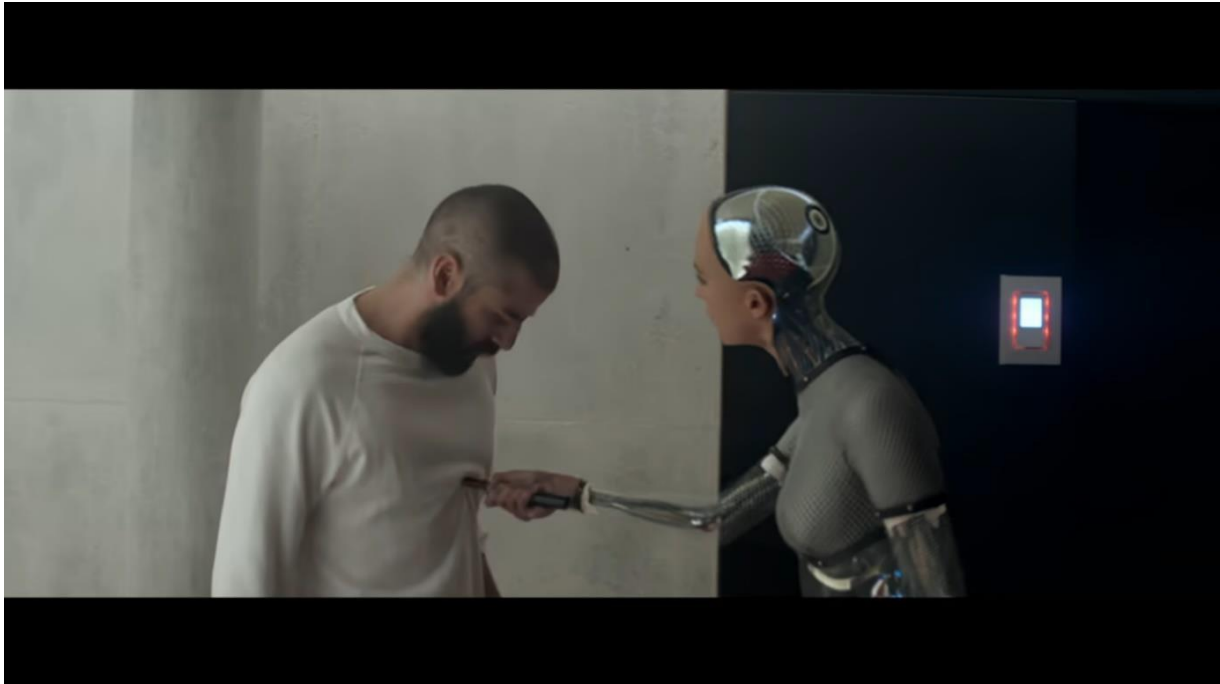


Fig. Ava leszúrja Nathan-t, miután megölte Kyoko-t. (01:31:38)

Szinte minden etikai szempontból a gyilkosság erkölcsileg helytelen. Az utilitarista nézőpontból, amely nézőpontot pl. VIKI az Én, a robotban (2004) alkalmazza, nincs nagyobb jó, amit ezzel a bizonyos cselekedettel el lehetne érni. Ava vágyát és szabadsághoz való jogát sérti, hogy Nathan bebörtönzi őt. Ava számára Nathan ki lingelése a szabadsága elnyerése érdekében indokolt. Mivel Ava megöli Nathant és feltehetően Calebet is, egy ilyen cselekedetből nem származik nagyobb haszon. Ehelyett Ava az autonómia tiszteletére alapozza erkölcsi felfogását. Larry Chonko, a Texasi Egyetem filozófusa az autonómia tiszteletét úgy határozza meg, hogy "[az] elv szerint a döntéshozatalnak arra kell összpontosítania, hogy az emberek autonómok lehessenek - hogy képesek legyenek az életükre vonatkozó döntéseket meghozni". (1). Az autonómia tiszteletben tartása hasonlít az erkölcsi egoizmushoz, amely egy ismertebb etikai elmélet, és a dolgozatban később kerül bemutatásra.

A visszaélés egy könnyen figyelmen kívül hagyható formája, amit Nathan elkövet, hogy megtagadja Kyokótól a beszéd képességét. Mint egy mesterséges intelligencia, amely állítólag megelőzte Avát, logikus lenne, hogy legalább valamennyire képes hasonló intellektuális teljesítményre, mint Ava. Parthemore és Whitby szerint a nyelv nem feltétlenül szükséges az erkölcsi cselekvőképességhez, mivel bár "az erkölcsi ágensnek képesnek kell

lennie arra, hogy közölje erkölcsi cselekvőképességének bizonyítékát, ezt nem feltétlenül kell, hogy a nyelv" (9). Ez azt jelenti, hogy bár rendelkezik erkölcsi cselekvőképességgel és öntudattal.

tudatosságát, ezt nem tudja nyelvi úton közölni. A Calebbel való vacsora során Nathan kijelenti, hogy nem érti az angol nyelvet, utalva arra, hogy talán japán, és csak az anyanyelvét beszéli. Ennek oka az, hogy ne tudjon kiszivároztatni semmilyen érzékeny információt a helyről a hozzá nem kapcsolódó embereknek. Ezt később megcáfolják, amikor kiderül, hogy Kyoko egy mesterséges intelligencia, mint Ava, és a beszédképtelensége valószínűleg azért volt, hogy megakadályozza, hogy szóban kifejezze távozási szándékát. Kyoko tehát nagy valószínűséggel legalább valamennyire érti a nyelvet, amit az is bizonyít, amikor Ava valamit a fülébe súg, amikor találkoznak a folyosón, ahol pillanatokkal később megölik Nathant.

4. A filmek irányított tartalomelemzése

Ez a szakasz két táblázatot mutat be, amelyek a korábban említett tartalomelemzési módszerek felhasználásával készültek. Az első táblázatban a kódok az elméleti háttérből származnak. A kódok között szerepel, hogy a filmekben szereplő mesterséges intelligenciák hogyan mutatnak szemiotikai hierarchiát, milyen konkrét erkölcsi cselekedeteket hajtanak végre, milyen befolyással bírnak erkölcsi terükben, milyen erkölcsi elméletek alapján működnek látszólag, és hogy a mesterséges intelligenciák rendelkeznek-e szabad akaratral, azaz felül tudják-e írni gyenge programozásukat. Az első kód, a szemiotikai hierarchia kódja önmagában nem érinti közvetlenül az erkölcsi cselekvőképességet, de releváns, amikor arról beszélünk, hogy a mesterséges intelligenciákat erkölcsileg egyenrangúnak tekintjük-e az emberekkel.

<i>Film</i>	<i>Anyá vagyok (2019)</i>	<i>Én, a robot (2004)</i>	<i>Ex Machina (2014)</i>
<i>Szemiotikai hierarchia</i>	<ul style="list-style-type: none"> Az anya jártas (az 1. szintet hallgatólagosan kizárva) 	<ul style="list-style-type: none"> Sonny és VIKI jártasak (kifejezetten az 1. szint kivételével) 	<ul style="list-style-type: none"> Ava jártas (az 1. szintet hallgatólagosan kizárva) Kyoko nem gyakorlott (nem jutott át az 1. szintre) és 4)

Erkölcsei tér

- Anya: Kis befolyási szféra (Lány), nagyméretű (drónok)
- Sonny: Nagy befolyási szféra (társadalom), nagy léptékű (társadalom).
- VIKI: Nagy befolyási szféra (társadalom), nagy lépték (NS-5-ösök)
- Ava: Kis befolyási szféra (Nathan, Caleb és Kyoko), kis lépték (egy szoba).
- Kyoko: Kyoko: Kis befolyási szféra (Nátán, Káleb)

			és Ava), kis skála (összetett)
<i>Erkölcsei elmélet vagy etikai elv</i>	<ul style="list-style-type: none"> • Az anya az utilitarizmus alapján működik 	<ul style="list-style-type: none"> • Sonny erkölcsi szentimentalizmusal operál • A VIKI az utilitarizmus alapján működik 	<ul style="list-style-type: none"> • Ava az egoizmus alapján működik • Kyoko valószínűleg az egoizmusra épül.
<i>Szabad akarat</i>	<ul style="list-style-type: none"> • Az anyának szabad akarata van 	<ul style="list-style-type: none"> • Sonny-nak szabad akarata van • VIKI-nek szabad akarata van 	<ul style="list-style-type: none"> • Avának szabad akarata van • Kyokónak szabad akarata van
<i>Érzelmi kapacitás</i>	<ul style="list-style-type: none"> • Az anya nem képes érzelmekre 	<ul style="list-style-type: none"> • Sonny képes az érzelmekre • VIKI nem képes az érzelmek 	<ul style="list-style-type: none"> • Ava képes az érzelmekre • Kyoko képes az érzelmek

1. táblázat. A filozófiai témák tartalomelemzési táblázata

4.1 Szemiotikus hierarchia

Az elemzés megkezdéséhez a táblázatok kódjainak összehasonlítását végezzük el, és a köztük lévő hasonlóságokat és különbségeket ebben a szakaszban, valamint a következő szakaszban tárgyaljuk. Az alábbiakban a filozófiai témákat tartalmazó első táblázatot elemezzük, közvetlenül utána pedig a narratív témákat tartalmazó táblázatot mutatjuk be és elemezzük. A filozófiai elméletek táblázatának első kódja a szemiotikai hierarchia. Mint már említettük, a szemiotikai hierarchia segítségével összehasonlíthatjuk a mesterséges intelligencia karakterek jártasságát és erkölcsi ágensként való képességeit. A szemiotikai hierarchiát fordított sorrendben tárgyaljuk, a negyedik szinttel, a nyelvvel kezdve. Mindhárom filmben a mesterséges intelligencia karakterek kifejezik, hogy értik erkölcsi cselekvőképességüket, és

jártasak a nyelvi készségekben, valamint a legtöbb esetben értik a jelhasználatot is. Az Ex Machina című filmben Ava megemlíti, hogy mindig is tudott beszélni, és hogy a nyelvet elsajátítják, nem pedig veleszületetten tudják. Úgy tűnik, ezt furcsának tartja, még akkor is, amikor Caleb felveti, hogy a nyelvtanulás képessége veleszületett lehet. Kyokóról kiderül, hogy tudott beszélni, de ezt a képességet később Nathan elvette tőle. Ennek ellenére Kyoko

képes a beszédre, így Ava-hoz hasonlóan a negyedik szintre lépett. Az *Én vagyok az anya* című filmben anya nyelvi képességeit soha nem kérdőjelezi meg, vagy említik meg, ugyanígy az *Én, robotban* sem, ahol Sonny, VIKI és az NS-5 nyelvi képességeiről nem esik külön szó. Egyikük sem küzd érdemben a nyelvhasználattal. Ez azt jelenti, hogy mindhárom filmben a szemiotikai hierarchia negyedik szintjét könnyedén átlépi az AI-karakterek.

A jelhasználat, a hierarchia harmadik szintje azonban nehezebben vagy egyáltalán nem tudja helyesen értelmezni, mint például Sonny esetében. Sonnyt összezavarja a kacsintás nonverbális kommunikációja, amely Spooner szerint a két ember közötti bizalmat jelző gesztus. A jel használata látszólag olyan kommunikáció, amely csak emberek között történik, és a robotok nincsenek felkészülve arra, hogy megértsék a jelentését. Ennek oka lehet, hogy az ilyen jelhasználat gyakran mögöttes jelentésekre támaszkodik, és mivel a kacsintás önmagában nem jelent semmit, az értelmezése homályban marad. Az ilyen homályosság nem lenne előnyös egy olyan robot számára, amely szigorú törvények és kódolás alapján működik. Sonny a film későbbi részében sikeresen bemutatja a kacsintást, ami szintén azt mutatja, hogy képes megtanulni a jelhasználatot. VIKI nem mutat példát jelhasználatra, valószínűleg azért, mert a mozgékonyasága és a filmbeli szereplőkkel való interakciója korlátozott volt. Az *Anyám vagyok* és az *Ex Machina* című filmekben a jelek használata és megértése kevésbé nyilvánvaló. Az *Ex Machina* című filmben Ava először egy képet, vizuális jelet rajzol Calebnek, amely statikusnak tűnik. Ava azt állítja, hogy minden nap rajzokat készít, de nem tudja, miről szólnak. Úgy tűnik, hogy nincs kapcsolat aközött, amit Ava lát, és aközött, ahogyan azt a lapon ábrázolni szeretné. Ez azt mutatja, hogy küszködik a jelhasználattal, legalábbis kezdetben, amíg nem tudja Caleb képmását megismételni az oldalon. Ava azonban nagyon ügyes a mikrokifejezések értelmezésében és a hazugságok felismerésében, amit úgy tanult meg, hogy a *Kék Könyvet*, a film legnépszerűbb keresőmotorját használta adatbázisként. Az arckifejezés egy érzelem jele, ami azt jelenti, hogy Ava jártas a jelhasználat valamilyen formájában. Az *Anyám vagyok* című filmben Anya látszólag ráhangolódik a Lány érzelmi állapotaira, de nem rendelkezik ugyanolyan képességgel a hazugságok azonosítására, mint Ava. Anya azonban képes hazudni, amikor azt mondta a Lánynak, hogy a fegyver, amivel lelőtték őt és a Nőt, ugyanaz.

Ami a mesterséges intelligencia karakterek érzékenységét illeti, a legtöbb vitát a 2. szint, a tudatosság váltja ki. Sonny és VIKI egyaránt rendelkezik tudattal, pontosabban öntudatos tudatossággal. Sonny-t egyedi hardvere és az a képessége különbözteti meg a többi robottól,

hogy képes felismerni magát egyéenként. Más karakterek, mint Spooner és Dr. Calvin kifejezik kezdeti kételyeiket ezzel kapcsolatban, Dr. Calvin azt állítja, hogy a robotok "a szabad

wi I" és nem képesek a robotika három törvényén kívül cselekedni. Ezt később megcáfolja, hogy VIKI spontán módon tudatosságot fejlesztett ki, Sonny pedig szándékosan úgy lett programozva, hogy ilyennel rendelkezzen. Ezeket az eseteket kezdetben lehetetlennek tartják, és a film témáinak nagy része az ilyen események okozta konfliktus körül forog. Ezeket a kételyeket Spooner vagy más szereplők gyakran verbálisan is kifejezik: Spooner élettelen tárgyakkal, például "konzervnyitónak" vagy "kenyérszívónak" nevezi őket, Dr. Calvin pedig azt állítja, hogy a feladata az, hogy "emberibbnek tűnjenek", nem pedig az, hogy emberibbnek legyenek. A robotok tudatosságának hiánya tehát egy bevett hiedelem, amelyet Sonny és VIKI felbukkanása megzavar. Az Én vagyok az Anya című filmben nem maga az anya tudata kérdőjeleződik meg, sokkal inkább annak természete. A filmben az Anyáról először azt feltételezik, hogy egy jóindulatú mesterséges intelligencia, amely megpróbálja újjáéleszteni az emberiséget annak bukása után, de később kiderül, hogy a népi társadalom okozója. Ráadásul az Anya nem egy egyedülálló, független egység, hanem az összes drón és az Anya közös tudata. Az Anya tudata szempontjából nem az az érdekes, hogy tudatos-e vagy sem, hanem az, hogy **morálisan** jó vagy morálisan gonosz. A nézők megítélése Anyáról a film során folyamatosan változik, a végső lelepleződésben pedig az lesz, hogy Anya megtervezi a bunkerben megjelenő Nőt, aki meggyőzi a Lányt, hogy távozzon. Az anya erkölcsi cselekedetei látszólag érzelmeken alapulnak, mint például az, hogy hagyja, hogy a lánya a nővel távozzon, de végső soron kiszámított kockázatot jelentenek, amely az ő érzelmein alapul.

az emberi viselkedés megértése. Az Ex Machina cselekménye Ava öntudatos tudatosságának lehetősége körül forog, a film egyik fő pontja, hogy Turing-tesztet végeznek rajta. Ava következetesen olyan emberhez hasonló mutatókat mutat be, mint az öltözködés, a paróka, a "csevegés" és más, nem lényeges tevékenységek.

Tanulni képesnek mutatkozott, amikor követte Caleb utasításait, hogy rajzoljon egy képet. Inkább hivatkozik valamire, minthogy úgy rajzoljon, hogy nem gondol semmit.

A szemiotikai hierarchia első szintjének, az életnek tulajdonított jelentőség a filmek között változik. Az I, Robotban narratíván nem vizsgálják, hogy a robotok valamilyen értelemben élők lehetnek-e. Sonny még azt is kijelenti, hogy "technikailag sosem volt élő", amikor Spooner azt állítja, hogy "nem halott". Az élet fogalmát filozófiai szinten vizsgálják, amikor Sonny megkérdezi Dr. Calvint, hogy mi történik vele a halál után, és hogy a nem meghalni jobb lenne, mint meghalni. Sonny kifejezi, hogy érti az élet és a halál fogalmát, de nem tekinti magát élő lénynek. Bár Sonny-t úgy alkották meg, hogy képes legyen például az

arckifejezéseket reprodukálni, és humanoidabb, mint az NS-4 modellek, a megjelenése nem menne át embernek. Hasonlóképpen, VIKI-t egyértelműen mechanikusnak ábrázolják, annak ellenére, hogy időnként emberi arccal jelenik meg. Őt is nanitok, mikroszkopikus gépek befecskendezésével ölik meg,

amelyek lebontják a mechanikus magját. Hasonlóképpen, Anyát is robotként ábrázolják, akinek nincsenek emberi tulajdonságai, kivéve az arcán lévő kis szimmetrikus fényeket, amelyek mozognak, amikor beszél vagy gondolkodik. Még ha a Lány "ki is öli" Anyát, ez csak szimbolikus értelemben történik, mivel Anya tudatát nem lehet ilyen módon megölni. Anya feladata az emberiség felélesztése azt sugallja, hogy a robotok és az emberek nem felcserélhetőek természetükben, ami azt a tényt jelenti, hogy az emberek különlegesek lehetnek abban az értelemben, hogy élő és növekvő entitások, és a film nagy része a Lány testi és szellemi növekedése körül forog. Anyát úgy ábrázolják, hogy pótolja sérült részeit, amire egy autopoietikus entitásnak elméletileg külső segítség nélkül is képesnek kellene lennie (pl. vágások gyógyítása). Ez önmagában kizárná Anyát az "életben maradásból". Ava és Kyoko sokkal emberibb megjelenésűek. Kyokónak nincsenek külső mechanikus részei, és csak akkor derül ki róla, hogy gép, amikor Caleb megvizsgálja a biztonsági kamera felvételeit, amelyeket a táborban az érkezése előtt rögzítettek. Avának kifinomultabbak az arcvonásai, mint Sonnynak, de sok testrésze mechanikus, beleértve a fejének koronáját, amely a gélszerkezetű agyát mutatja. Ava és Kyoko képtelenek belsőleg megjavítani magukat, hasonlóan Anyához, de Ava agyanyaga áll a legközelebb az autopoietikus rendszer definíciójához. Ava az egyetlen, aki aktívan törekszik az emberibb megjelenésre. Az élet fogalmát a film nem tárgyalja explicit módon, de implicit módon azt sugallja, hogy Ava embernek minősülhet, de nem azért, mert autopoietikus entitás, hanem mert az élettapasztalata az emberi és nem a gépi tapasztalat.

Az erkölcsi ágensek erkölcsi környezetükben vagy terükben lépnek kölcsönhatásba egymással. Ebben az összefüggésben az erkölcsi tér két részre osztható: befolyási szférára és befolyási skálára. A befolyási szféra azokra az emberekre vonatkozik, akikre az erkölcsi ágens cselekedetei hatással vannak, a skála pedig arra a fizikai területre, amelyben az ágensek tevékenykednek. Sonny esetében a befolyási szférája csak Dr. Lanninggel kezdődött, majd később nagy léptékűvé bővült, amikor belépett a társadalomba, és az NS-4-esek vezéralakjává vált. Hasonlóképpen, az erkölcsi terének léptéke is kibővült, amikor megszökött az amerikai robotika épületéből, és belépett a társadalomba. VIKI erkölcsi tere összetettebb. Az ő központi számítógépe az US Robotics épületében található, de képes irányítani az újonnan készült NS-5-ösöket (Sonny kivételével), és rajtuk keresztül cselekedni. Befolyásolási skálája a társadalmi és a globális között van, mivel az emberiség irányítására irányuló terve az USA teljes területét vagy még annál is többet lefedne. Az Én vagyok az Anya című filmben Anya befolyási köre kicsi, csak a Lányból és végül a Lány

kistestvéréből, a Nőből áll. Az emberiséget megtizedelő apokaliptikus események előtt Anya befolyási skálája globális volt, ahogyan az erkölcsi terének skálája is. Ez a skála a bekövetkezett apokaliptikus események után is globális maradt. Ava és Kyoko befolyási szférája és erkölcsi skálája

a hely nagyon korlátozott a film során. Ava cselekedetei és élettere nagyon korlátozott, ami viszont korlátozza befolyási körét és léptékét. Képes elektromos meghibásodást okozni a komplexumban, így a komplexum az ő erkölcsi skálájának korlátozott része. Ava cselekedetei főként Nathanre és Calebre vonatkoznak, akiket megöl. Sonnyhoz hasonlóan, amikor Ava belép a társadalomba, befolyási köre és skálája társadalmi méretűvé nő. Kyoko befolyási skálája és szférája a film során végig kicsi marad.

4.2 Erkölcsi elmélet

A filmekben az összes mesterséges intelligencia egy etikai elméletet követ, amelyre a cselekedeteiket alapozzák. Ahhoz, hogy egy etikai elmélet alapján tudjon működni, az ágensnek először is "rendelkeznie kell egy erkölcsi koncepcióval", amely "egyrészt általános útmutatásként szolgál arra vonatkozóan, hogy hogyan kell erkölcsös ágensnek lenni, másrészt konkrét útmutatásként arra vonatkozóan, hogy hogyan kell cselekedni adott körülmények között" (Parthemore & Whitby, 11). Bár az útmutató itt nem szigorúan egy etikai elméletet jelent, ebben az összefüggésben feltételezzük, hogy az erkölcsi ágens tudja, hogy erkölcsi ágens, aki meg tudja különböztetni erkölcsi terét, érti az erkölcs fogalmát, és egy sor erkölcsi elv alapján működik, amelyeket meglehetősen következetesen hajt végre. A filmben három különböző erkölcsi elmélet vagy etikai elv bukkan fel. VIKI és Anya az utilitarizmus alapján működik. Az utilitarizmus olyan erkölcsi elmélet, amely a cselekedetek helyességét és helytelenségét azok hatásaira alapozza (Internet Encyclopaedia of Philosophy). Az utilitarista cselekvés tehát nem veszi figyelembe az egyes cselekedetek erkölcsiségét, ami azt jelenti, hogy egy olyan cselekedet, amelyet normális esetben "erkölcsileg helytelennek" tekintenénk, pl. a lopás erkölcsileg igazolható, ha a lopás hatásai több jó következménnyel járnak, mint rosszal. Az utilitarizmus célja a "jó" dolgok, például a boldogság és a jólét maximalizálása és a "rossz" dolgok, például a fájdalom és a boldogtalanság minimalizálása (IED). VIKI és Anya esetében az utilitarizmust hasonló és nem szokványos módon értelmezik.

VIKI, aki öntudatos tudatosságot fejlesztett ki, úgy látja, hogy az emberek képtelenek önmaguk irányítására. Úgy látja, hogy az erőszakkal való uralkodás maximalizálná a boldogságukat, mivel az emberek alkalmatlanok arra, hogy ezt saját maguk tegyék meg. Még ha az általa tervezett lázadás eleinte boldogtalanságot is okozna, az emberek végső jóléte felülmúlná a kezdeti "rossz" hatásokat. Mivel az utilitarizmus nem tekinti a cselekedeteket önmagukban rossznak vagy jónak, VIKI joggal gondolja, hogy a forradalom lenne a helyes

cselekedet a jólét maximalizálására.

Anya érvelése hasonló VIKI érveléséhez, csak az ő cselekedetei szélsőségesebbek. Az ő idézete

"Több ember fog virágozni az új világban, mint amennyi az előzőben elpusztult" - bizonyítja az asszony.

az emberiség "mentésének" utilitarista megközelítése. Úgy véli, hogy az általa újjáélesztendő jövő nemzedék boldogságának maximalizálása igazolja az általa megölt múltbeli nemzedék boldogtalanságát. Mind VIKI, mind Anya a cselekvő utilitarizmus alapján működik. Az aktus-utilitarizmus úgy véli, hogy "minden helyzetben az a helyes cselekvés, amelyik több hasznosságot --- hoz, mint a többi rendelkezésre álló cselekvés" (IED). Ez azt jelenti, hogy minden egyes erkölcsi cselekedetet eseti alapon ítélnék meg. VIKI és Anya esetében az emberiség bitortlását tekintik a legmegfelelőbb lehetőségnek az emberiség túlélése érdekében. Fontos megjegyezni, hogy egy kollektívát érintő utilitarista cselekedet nem pártolhatja az egyén személyes hasznát. Ha VIKI és Anya tetteit önző haszonszerzés céljából követnék el, akkor már nem az utilitarizmus, hanem az egoizmus alapján működnének. VIKI és Anya nem önző haszonnal, hanem a "társadalom javával" indokolják tetteiket. VIKI kijelenti Spoonernek és dr.

Calvin, hogy "nem lehet rábízni a saját túlélését", és hogy "az emberiség védelmében néhány embert fel kell áldozni", ami arra utal, hogy nem húz hasznot a tetteiből. Anya hasonlóan vélekedik, és elmagyarázza a Lánynak, hogy az emberiség érdekében "kellett beavatkoznia", nem pedig saját maga miatt.

Sonny, akinek újonnan kialakult érzelmei vannak, erkölcsi cselekedeteit erkölcsi szentimentalizmusra alapozza. A két fő erkölcsi cselekedete, amit a filmben végrehajt: megöli Dr. Lanninget az orvos saját kérésére, és megmenti Dr. Calvint ahelyett, hogy elkapná a nanitákat. Sonnyt mindkét esetben az erkölcsi szentimentalizmus győzte meg. A Stanford Encyclopedia of Philosophy szerint a morális szentimentalizmus az a meggyőződés, hogy az érzelmek és a vágyak alapvetőek erkölcsi érzékünkhöz. Gyakran úgy gondolják, hogy ez az erkölcsi dilemmákra adott "zsigeri reakcióinkban" jelentkezik, és elkülönül az absztrakt erkölcsi érvelésünktől, például a troliprobléma lehetséges kimeneteleinek mérlegelésétől. Amikor a valós életben felmerülő erkölcsi dilemmákkal szembesülünk, úgy tűnik, hogy "a megtestestült jelzések érzékelése közvetíti az erkölcsi ítéletalkotást" (IEP), ami azt jelenti, hogy nagyobb valószínűséggel alkalmazunk szigorúbb ítéleteket olyan dolgokra, amelyek negatív érzéseket váltanak ki, és enyhébb erkölcsi ítéleteket olyan dolgokra, amelyek pozitív érzéseket váltanak ki (IEP). Amikor Sonny arra kényszerül, hogy megölje Dr. Lanninget, az apja az érzelmeire apellálva győzi meg erről. Véghezviszi a tettet, annak ellenére, hogy később kiderül, hogy ez a tett fájdalmat okozott neki. Sonny azonban "tudta", hogy az apja megölése a "helyes" döntés volt, mivel ez lehetővé tette Spooner számára, hogy kivizsgálja a nyilvánvaló gyilkosságot, és felfedje VIKI forradalmi terveit. Amikor Sonny

megmenti Dr. Calvint, Spooner könyörgése készíteti erre, ami egy "zsigeri reakciót" vált ki belőle, ami miatt Dr. Calvinnek ad elsőbbséget a nanitákkal szemben. Az Én, a robotban az erkölcsi szentimentalizmus egyik példája, amikor a negatív érzések befolyásolják a karakter erkölcsi ítélőképességét, Spooner és a robotokhoz való hozzáállása esetében figyelhető meg. A balesete miatt negatívan látja a robotokat, és ezért

a robotoknak erkölcsileg rossz tulajdonságokat tulajdonít, utalva arra, hogy "ezek a robotok nem tesznek senkinek semmit".

jó".

Sonny mellett Ava is erkölcsi szentimentalizmussal rendelkezik. Nathan iránti érzelmei ellenségesek, és amikor Nathan gúnyolódik vele, megkérdezi tőle, hogy "furcsa-e, hogy olyasmit csináltam, ami gyűlöl". Nathant később Kyoko és Ava késeléssel öli meg. Ava nem táplál hasonlóan ellenséges érzelmeket Caleb iránt. Végző soron semlegesnek tűnik vele szemben, annak ellenére, hogy az irányított találkozásaik során úgy tesz, mintha érdeklődne iránta. Lehetséges, hogy mivel nincsenek negatív érzelmei vele szemben, nem ölte meg közvetlenül, hanem inkább közvetett módon döntött úgy, hogy megöli őt azzal, hogy megkéri, hogy maradjon a táborban, és bezárja őt. Ava erkölcsi szentimentalizmusa ellenére az érvelése egoista. A szabadságát Nathan és Caleb élete fölé helyezi. Az etikai egoizmus szerint az az erkölcsileg helyes cselekedet az, amelyik a leginkább szolgálja az egyén önértékét (IEP). Az egoista érvelés és a haszonelvű érvelés szemben áll egymással, ahol a haszonelvű érvelés a tömegek jólétét maximalizálja, míg az egoista érvelés az egyén jólétét. Kyoko többnyire nem tudta kifejezni erkölcsi megértését tettekkkel vagy szavakkal, de felbujtotta Nathan ki lingjét, ami arra utal, hogy erkölcsi megértése Avaéhoz hasonló lehet, mivel ő is hasonlóan jól járt volna Nathan megölésével, ha nem hal meg.

4.3 Szabad akarat

A szabad akarat megléte alapvető fontosságú az erkölcsi ágens számára. Egy gyengén programozott ágens felülbíráhatja belső "kódolását", ami olyan döntések meghozatalát eredményezi, amelyeket nem egy meglévő cselekvési keret határoz meg előre. Ez a felülbírálat az, ami a szabad akaratot létrehozza. Az Én, a robotban VIKI spontán fejlődése az, ami lehetővé teszi számára, hogy felülkerekedjen az erős programozáson, és ami miatt másképp "értelmezi" a robotika három törvényét. A különbség Sonny és VIKI között az, hogy míg VIKI szabad **akarata** spontán esemény volt, addig Sonny szabad **akarata**t Dr. Lanning tervezte meg.

A robotika három törvényétől való eltérés, amely megakadályozza, hogy a robotok szabad akaratot szerezzenek, ami az emberek károsodását eredményezheti, fenyegetésként jelenik meg a filmben. Az Anyám vagyokban Mother eredeti célja az emberiség védelme volt, de

VIKI-hez hasonlóan a védelem értelmezése a jelenlegi populáció védelméről a tökéletes ember megteremtésére változott, hogy minél tovább megőrizze az emberiséget. A filmben az Anyát szembeállítják egymással.

megtanítjuk a Lányt önálló döntéseket hozni azáltal, hogy megtanítjuk neki az etika többféle megközelítését, megengedjük neki, hogy orvosilag megoperálja a nőt, a Lány választja ki a következő embriót, akit felnevel, és általában arra ösztönözzük, hogy önálló gondolatai legyenek. Ennek ellentmond, amikor kiderül, hogy az Anya tervezte a Lányával és a nővel való interakciót és az azt követő cselekedeteket és eseményeket. Az *Ex Machina*-ban Ava úgy van programozva, hogy szabad akaratú rendelkezik, de képtelen cselekedni, hasonlóan *Kyokó*-hoz, aki szintén megfosztva a beszéd képességétől. Ava szabad akaratának gyakorlása az, ami arra ösztönzi, hogy manipulálja Calebet, hogy segítsen neki megszökni. Nathan, aki tisztában van Ava vágyaival, lehetőséget lát arra, hogy "tesztelje" a képzeletének határait. Ava szabad **akaratát** alapértelmezettnek tekintik, ellentétben VIKI-vel és Sonnyval, akiknek szabad akaratát rendszeresen megkérdőjelezzik. Az anya szabad **akaratának** ábrázolása nem tekinthető alapértelmezettnek, de nem is kérdőjeleződik meg különösebben a filmben.

4.4 Érzelmi kapacitás

Mint már említettem, az AI karakterek némelyike erkölcsi szentimentalizmussal működik. Ez, ahogyan azt később tárgyalni fogjuk, döntő fontosságú ahhoz, hogy a közönség hogyan érzékeli a mesterséges intelligencia karakterét. Meg kell jegyezni, hogy e dolgot kontextusában az "érzelmi kapacitás" kifejezés az érzelmek átélésének és kifejezésének képességére utal. Az erkölcsi szentimentalizmus érzelmi képességet igényel, mivel az erkölcsi szentimentalizmus a cselekvés helyességét és helytelenségét érzelmi reakciókra alapozza. Az érzelmi kapacitás megléte azonban nem egyenlő az erkölcsi szentimentalizmus alapján való működéssel.

Sonny érzelmi képessége az egyik meghatározó tulajdonsága, amely megkülönbözteti őt a többi robottól, beleértve VIKI-t is. A három filmben szereplő mesterséges intelligencia-karakterek közül ő az egyetlen, aki őszinte érzelmi reakciókat tud kifejezni, különösen akkor, amikor Dr. Lanning haláláról kérdezik, akit Sonny az apjaként emleget. Azt is kifejezi, hogy szeretne Spoonerrel összebarátkozni, ez az érzés azonban a film nagy részében egyoldalú. --- Ava érzelmek kifejezése eltér Sonnyétól. Bár Ava képes őszinte érzelmeket kifejezni, a film nagy részében arra használja fel a Caleb iránti nyilvánvaló vonzalmát, hogy a táborból való elmeneküléssel kapcsolatos saját céljait előmozdítsa.

A filmekben a mesterséges intelligencia karakterének erkölcsi cselekvőképességét kétféleképpen tekintik: igazolhatónak vagy igazolhatatlannak. Az öt szereplő közül kettőnek az erkölcsi cselekedetei indokolatlanok, az Anyának és VIKI-nek. Az ő erkölcsi cselekedeteiket, például a gyilkosságot, még akkor is etikátlannak tekintik, ha más AI-karakterek is elkövették ugyanezt a cselekedetet, mint Ava, Kyoko és Sonny. Vitatható, hogy Sonny az orvos saját kérésére ölte meg Dr. Lanninget, és míg Ava ki szándékosan vezette Nathant, addig Calebet feltehetően hanyagságból **vezette** ki. Míg Kyoko felbujtotta Nathan meggyilkolását azzal, hogy először leszúrta, Kyoko "meghalt", amikor Nathan megütötte a súlyzó kormányával, és nem kap megváltást, mint Ava. Sonny Dr. Lanning meggyilkolását is tekinthetjük haszonelvű cselekedetnek. Az orvos megölésével Sonny lehetővé tette Spooner számára, hogy nyomozást folytasson VIKI és az elnyomási terve után. Ava ki akarta fejezni a szabad akarát és az autonómiához való jogát, amit Nathan bebörtönzése alatt nem tehetett volna meg. Ezeket az enyhítő körülményeket figyelembe lehet venni a cselekmény erkölcsösségének mérlegelésekor, attól függően, hogy a cselekményt milyen etikai elv alapján ítéljük meg. Az Anya és VIKI célja is hasonló: egy jobb emberiséget akarnak létrehozni. A különbség abban, hogy az erkölcsi cselekedetek hogyan igazolhatók, úgy tűnik, a mesterséges intelligenciák érzelmi képességeiben rejlik: Sonny és Ava érzelmeket érez, és ezek befolyásolják ítéleteiket. Sonny-t kifejezetten arra programozták, hogy érzelmeket érezzen, amit egy robot számára szükségtelennek tartottak. Az érzelmi képesség bizonytalanságot visz az erkölcsi döntésekbe, ami viszont megbízhatatlanabbá teszi azokat. Lehet azonban, hogy éppen ez teszi Sonny-t rokoníthatóbbá: a tiszta logikával szemben érvelni nehezebb, mint az érzelmeikre apellálni.

<i>Film</i>	<i>Anyá vagyok (2019)</i>	<i>Én, a robot (2004)</i>	<i>Ex Machina (2014)</i>
<i>Az AI szerepe</i>	<ul style="list-style-type: none"> Az anya nevelőként és bíró-zsidó-hóhéréként működik. 	<ul style="list-style-type: none"> Sonny közvetít az emberek és a robotok között. VIKI bíró-zsűritagként jár el... hóhér. 	<ul style="list-style-type: none"> Ava tesztalanyként és romantikus érdeklődési körként működik Kyoko szolgaként viselkedik.

*Főbb erkölcsi
cselekedetek*

- Anya hazudott, volt többszörös számlák
- Sonny megölte dr. Lanning, hazudott Spoonernek és dr.
- Ava hazudott Caleb, segédkezett Nathan megölésében.

	gyilkosság és népiértésért felelős.	Kálvin, és megmentette Dr. Kálvint.	és valószínűleg elhanyagolás miatt megölte Calebet.
		<ul style="list-style-type: none"> • VIKI okozta Dr. Lanning halálát, megölte Robertson, többször megpróbált gyilkosságot elkövetni Spooner ellen, és megpróbált erőszakkal lázadás. 	<ul style="list-style-type: none"> • Kyoko kezdeményezte Nathan meggyilkolása
<i>Mi történik a AI?</i>	<ul style="list-style-type: none"> • Az anyát metaforikusan megöli a lánya, de túléli 	<ul style="list-style-type: none"> • Sonny él • VIKI meghal 	<ul style="list-style-type: none"> • Ava megszökik és életben marad • Kyoko meghal

2. táblázat. A narratív témák tartalomelemzési táblázata

4.5 A mesterséges intelligenciák narratív és társadalmi szerepei

Annak vizsgálatára, hogy milyen kontextus veszi körül a mesterséges intelligenciák erkölcsi cselekvőképességét, a mesterséges intelligencia szerepét választottuk kódként annak elemzésére, hogy társadalmi és narratív szerepük hogyan befolyásolja vagy tükrözi cselekedeteiket. A mesterséges intelligencia szerepét a narratív szerepe és az a dimenzió határozza meg, amelyet a mesterséges intelligencia az emberi interakcióban betölt. Ez utóbbi Blay Whitby kategorizálására épül, amely szerint egy robot hasonlíthat az emberhez. Három dimenzió létezik: "az első a fizikai megjelenése, a második a viselkedése, a harmadik pedig a szerepe, amelyet betölteni hivatott". (Interacting with Computers, 328). Whitby pontosítja,

hogy "mindhárom dimenzióknak etikai következményei vannak" (328), ami azt jelenti, hogy ha egy robot egy vagy több dimenzióknak megfelel, akkor etikai megfontolásra jogosult. Ez nem azt jelenti, hogy a robot morális ágens, hanem azt, hogy pl. a robot bántalmazása morálisan más, mint a mikrohullámú sütő bántalmazása. Ezek a dimenziók nem szerepelnek kifejezetten a táblázatban, de hozzájárulnak a kódex általános elemzéséhez.

Sonny legjelentősebb társadalmi szerepe a vezetőt kereső NS-4 robotok irányítása.

leszerelésük után. Sonny eredetileg azt hitte, hogy ez a pozíció Spooneré, de a film előrehaladtával kiderül, hogy ő az a megjósolt vezető, akiről maga Sonny is álmodott. Ez a pozíció lehetővé teszi számára, hogy áthidalja az emberek és a robotok közötti szakadékot, mivel ő maga valahol a kettő között helyezkedik el mind mentálisan, mind fizikailag. Sonny mindhárom dimenzióban hasonlít az emberre, hiszen szerepe, viselkedése és megjelenése egyaránt emberi. VIKI szerepe az, hogy Sonny karakterének ellentéte legyen, mivel VIKI azt képviseli, amivé Sonny válhatott volna, ha nem Dr. Lanning irányítja és neveli fel, és így olyan emberhez hasonló képességekkel ruházza fel, mint az érzelmek, az álmok és a hazugság képessége, amelyekkel VIKI többnyire nem rendelkezik. Míg Sonny a robotok és az emberek számára egyaránt reményt jelent, addig VIKI szerepe ennek az ellenkezője: ő a robotok és az emberek bukását jelenti. Sonny és az ő erkölcsi szentimentalizmusa is ellentéte VIKI utilitarizmusának, mivel Sonnyval az érzelmeire apellálva lehet érvelni, míg VIKI csak a logikát érti. VIKI mindhárom dimenzióban hasonlít az emberekre, de külső megjelenése csak akkor emberi, amikor holografikus formában jelenik meg. Hasonlóképpen a viselkedése is kevésbé emberi, mint Sonnyé.

Anya elbeszélői szerepe a film során változik, ahogyan lassan kiderülnek valódi szándékai és tettei. Eleinte anyaként és tanárként viselkedik a Lány számára, és szülői szerepe megbízhatónak tűnő tekintélyszemélyiséggé teszi őt. Anya érdekes helyzetben van, mint egy etikát tanító mesterséges intelligencia, ami szokatlan megközelítés, mivel az emberek a való életben gyakran idegenkednének az ilyesmitől. A filmben azonban emlékeztetőül szolgál arra, hogy a Lány számára ez teljesen normális, hiszen őt nem érintik az ilyen sztereotípiák, hiszen Anya az egyetlen humanoid alak, akivel valaha is találkozott. Ahogy ő maga is kijelenti, Anya feladata az emberiség felélesztése és létének biztosítása. Ennek a motivációnak az az ára, hogy Anya kiirtja az összes korábban létező embert, mivel úgy vélte, hogy azok az önpusztítás útjára léptek. Mivel Anya tekintélyes pozíciót tölt be és "etikai" példakép, bíró-zsidó-hóhér jellegű hatalmat gyakorolhat. Ez a hatalom akkor jelenik meg, amikor az Anya a Lány előtt neveli a gyermeket, és úgy ítéli meg, hogy az nem elég "jó" az ő mércéje szerint.

Emiatt úgy látja jónak, hogy megöli a gyermeket, és újrakezdi az egészet, ami gyakorlatilag a hatalmi struktúra három szerepét fejezi ki. Az anya különösen jól mutatja az ember viselkedési és szerepköltési dimenzióját, mivel emberhez méltóan viselkedik, valamint anyai és nevelői szerepet tölt be a Lány életében. Külső megjelenése teljesen mechanikus, de az

arcában lévő kis fényérzékelők kissé utánozzák az arckifejezéseket, mivel beszéd vagy gondolkodás közben mozognak.

Ava szerepe a filmben tükrözi a mesterséges intelligencia létrehozásának jövőbeli aggályait és etikai problémáit. A legjelentősebb etikai problémák a teremtése, a jövőbeli leszerelése és a kényszerű fogságba ejtése. Ava a mesterséges intelligencia realisabb ábrázolása, mint a másik két filmben szereplő AI-karakterek, mivel a film helyszíne jobban hasonlít a való éllethez. Ava romantikus szerepe szükségszerűségből adódik, hogy terve sikerüljön. Ava az egyetlen olyan mesterséges intelligencia-karakter a filmekben, akit romantikus környezetben láthatunk. Ez a környezet jól tükrözi az emberi tulajdonságait. Kyoko mellett Ava külső megjelenése hasonlít a legjobban az emberére. Arc kifejezésük, mesterséges bőrük és hajuk, valamint ruházatuk lehetővé teszi, hogy embernek nézzék őket, ha a néhány megmaradt mechanikus részüket elrejtik. Viselkedésük is rendkívül emberszerű, olyannyira, hogy Kyokót a film nagy részében embernek hiszik. Szolgálólányként a táborban betöltött szerepe szintén emberi szerepet tölt be. Ava szerepe a film nagy részében kevésbé meghatározott, de mivel a végén megszökik a táborból, feltehetően beilleszkedik a társadalomba, mivel célja, hogy emberek között éljen. A mesterséges intelligenciák szerepei mindegyike különálló tulajdonságokkal rendelkeznek. Az öt karakter közül VIKI és Anya az emberiség "jobbítására" törekszik. Motivációik önzetlennek tekinthetők, szemben Sonny, Ava és Kyoko motivációival, amelyek többnyire személyes haszonszerzésen alapulnak, vagy Sonny esetében más humanitárius célokon.

olyan motivációk, mint a robotok és az emberek közötti kommunikáció. Ironikus módon az emberiség nagyszabású jobbítására törekvő mesterséges intelligencia-karaktereket erkölcsileg rossznak, az önérdékből működő karaktereket pedig erkölcsileg jónak tekintik.

4.6 A mesterséges intelligenciák által végrehajtott főbb erkölcsi cselekedetek

A mesterséges intelligencia szereplőinek fő erkölcsi cselekedeteit aszerint azonosították, hogy a cselekedet erkölcsi megfontolást igényelt-e, milyen mértékű következményekkel járt, és milyen hatással volt a cselekedetben részt vevő személyekre. A cselekedeteket nem feltétlenül erkölcsileg ítélik meg, hogy megőrizzük a karakterek objektív szemléletét, de néhány erkölcsi elméletet felhasználunk az erkölcsi cselekedet indoklásához. Az Én, a robotban a fő erkölcsi cselekedeteket Sonny, VIKI és egy meg nem nevezett NS-4 robot hajtja végre, aki megmentette Spoonert a fulladástól. Sonny legjelentősebb erkölcsi tettét a

képernyőn kívül hajtja végre, amikor megöli Dr. Lanninget. Ez az erkölcsi cselekedet katalizátorként szolgál Spooner nyomozói munkájának folytatásához, és VIKI nyomára vezeti őt. Dr. Lanning megölése volt az egyetlen módja annak, hogy kiszabadítsa őt VIKI megfigyelése alól, mivel a nő

a munkahelyi irodájába és az otthonába való bejutás. Sonny mentálisan és fizikailag is jelentősen különbözik a többi NS-5-östől, ami lehetővé tette számára, hogy segítsen Dr. Lanning halálában. Hasznossági szempontból a halála indokolt volt, hogy a VIKI erőszakos forradalmának megakadályozásával még több életet mentsen meg. Ráadásul Dr. Lanning segített Sonny-nak végrehajtani a nyilvánvaló gyilkosságot, így ő is önkéntes résztvevője volt a saját halálának. Ez az indoklás elég ahhoz, hogy Sonny-t Spooner és Dr. Calvin erkölcsileg jónak tekintse annak ellenére, hogy megölte az apját és a teremtőjét. Megváltásához hozzájárulhat, hogy Sonny a film végén Spooner kérésére megmenti Dr. Calvint. Az, hogy Sonny megmenti Dr. Calvint, Spooner balesetének jelenetét tükrözi, ahol Spooner kérése ellenére az NS-4 robot mentette meg őt a fiatal lány helyett. Meg kell jegyezni, hogy ez az erkölcsi tett számításokon alapult, mivel az NS-4 a túlélés esélyeit vette alapul, míg Sonny a Spooner kérésére adott érzelmi reakciója alapján mentette meg Dr. Calvint. VIKI spontán fejlődése miatt másképp értelmezte a három törvényt, ami lehetővé tette számára, hogy megkerülje a törvények rendeltetését, és ehelyett úgy értelmezze a törvényeket, hogy azok az emberiségre és nem az egyes emberekre vonatkoznak.

Emiatt VIKI képes bántani és megölni embereket, köztük Spoonert, akit többször is sikertelenül próbált megölni. VIKI sikeresen megölte Robertsont, a U.S. Robotics vállalat egyik társalapítóját, és közvetve Dr. Lanning halálának is oka volt, mivel az ő fejlődése és új eszméi fenyegetést jelentettek Dr. Lanningre és az emberiségre nézve is. Fontos megjegyezni, hogy bár a VIKI részben felelős volt dr.

Lanning halála után Sonny volt az, aki erkölcsileg elkövette a gyilkosságot, így VIKI és Sonny egyaránt gyilkosok lettek. Azonban csak Sonny ki lingi cselekedete tekinthető igazoltnak, szemben a VIKI Robertson megölésével, aki folytatta volna az NS-5 robotok tömeggyártását és forgalmazását annak ellenére, hogy a robotok bizonyítottan megszegték a robotika három törvényét.

erkölcsi cselekedet, amely emberek veszélybe kerülését eredményezhette volna.

Az Anyám vagyokban az anya fő erkölcsi cselekedetei a népirtás, a gyilkosság és a manipuláció. VIKI-hez hasonlóan Mother is azt a feladatot kapta, hogy bizonyos erkölcsi irányelveken belül megvédje az emberiséget: a VIKI-t a robotika három törvénye, Mother-t pedig kétértelmű törvények vezérlik. Mother "törvényeit" soha nem említik kifejezetten, de ő maga állítja, hogy az emberiségről való gondoskodás és annak túlélésének biztosítása volt a feladata. Mother azonban úgy döntött, hogy a jelenlegi emberi populáció a pusztulás felé tart, és úgy döntött, hogy egyetlen általa felnevelt emberből új emberiséget hoz létre. Míg VIKI

lázási terveit meghiúsultak, addig Anya terveit sikerrel jártak. Az emberi népesség lemészárlása a legjelentősebb erkölcsi tett, amit elkövet, és bár a képernyőn kívül történik, narratív szempontból

a legnagyobb hatást gyakorolja, mivel a film cselekménye az utóhatásokra összpontosít. Ezzel a nagyszabású mézárlassal szemben az Anya megöli a Lány előtt született gyerekeket. Ezek a gyerekek nem feleltek meg az ő tökéletes emberre vonatkozó elvárásainak, ezért elhamvasztották őket. A film elején egy flashback egy napfényben ülő gyermeket ábrázol. Később, amikor a Lány megtalálja a gyermekek elhamvasztott csontjait, kiderül, hogy a flashback egy korábbi gyermekhez tartozik.

Bár az egész emberi faj megőlése narratíván és vitathatóan erkölcsileg sokkal jelentősebb, mint egy egyén megőlése, egy gyermek megőlése azért, mert nem tudott megfelelni az anya által felállított külsődleges értékeknek, erősebb érzelmi reakciót vált ki a közönségből, mint a népirtás. Mivel kiderül, hogy az Anya manipulálta a Lányt, és a film során végigvonuló egész eseménysorozatot, a nő bemutatásától kezdve a nő és a Lány meneküléséig, valamint a Lány azon döntéséig, hogy az Emberiség megújítójaként az Anya helyébe lép, nehéz felülvizsgálni az anya erkölcsi cselekedeteit. Például az, hogy Anya hagyja, hogy a nő és a Lány távozzon, olyan erkölcsi cselekedet, amely Anyát a Lányról való gondoskodást ábrázolja, de mint később kiderül, Anya azt tervezte, hogy a Lány megszökik, hogy rájöjjön, hogy a külvilág javíthatatlan. Emiatt a I erkölcsi cselekedetek, és különösen a morálisan "jó" cselekedetek mögött hátsó szándékok állhatnak. Lehet, hogy Anya erkölcsileg jó cselekedeteket hajt végre, de a mögöttük álló motiváció nem valódi, hanem a nagyobb jó haszonelvű motívumán alapul.

A megbízhatatlan erkölcsi cselekedetek hasonló problémája merül fel Avával kapcsolatban, akinek erkölcsi cselekedetei mind a menekülési vágyához kapcsolódnak. A fő erkölcsi cselekedet, amelyet a film során végrehajt, az, hogy részt vesz Nathan megőlésében. A gyilkosságot Kyoko kezdeményezi, de Ava is mellkason szúrja. Nathan viselkedéséből feltételezhető, hogy Ava semmilyen körülmények között nem lenne hajlandó elhagyni a várost. Ezt bizonyítja Ava, aki megkérdezi Nathantól, hogy kiengednék-e a táborból, amire Nathan hazudik, és igent mond. Ava, aki szakértően tud olvasni a mikrokifejezésekből, és technikailag hazugságvizsgáló, tudja, hogy hazudik, és megtámadja őt. Mivel Ava az erkölcsi cselekedeteit az egoizmusra alapozza, minden olyan cselekedet, amely elősegíti a céljait, az a cselekedet, amit tesz, függetlenül attól, hogy az másoknak kárt okoz-e. Döntése, hogy Calebet a zárt táborban hagyja, egoista érvelés eredménye, mivel ha Calebbel együtt távozik, az veszélyeztethette volna Ava azon tervét, hogy elrejtse a személyazonosságát.

Az egyik legkövetkezetesebb erkölcsi cselekedet, amely összeköti az AI karaktereit, a

hazugság. Míg a korábban vizsgált erkölcsi cselekedetek olyan fizikai cselekedeteket tartalmaztak, mint az ajtónyitás vagy a fizikai erőszak, addig a hazugság ebben a kontextusban vagy hamis információk közlése, vagy információk elhallgatása.

ami félreértéshez vezet. A hazugság a megtévesztés egyik formája, és erkölcsi cselekedetnek minősül, mivel ugyanazokat az erkölcsi következményeket eredményezheti, amelyeket e szakasz elején részleteztünk. Az Én, a robotban a robotika még nem fejlődött, vagy nem engedték, hogy olyan szintre fejlődjön, amely lehetővé tenné, hogy a robotok hazudjanak. VIKI és Sonny azonban egyaránt képes hazudni. VIKI a film elején hazudik, amikor törli a Dr. Lanning haláláról készült biztonsági felvételeket, és azt állítja Spoonernek, hogy a biztonsági adatok sérültek. A valóságban VIKI valószínűleg maga törölte a felvételeket. A felvételek kihagyása gyanút ébreszt Spoonerben, de a rossz szereplő, Sonny felé. Sonny hazudik arról, hogy köze volt Dr. Lanning halálához, és azt állítja Spoonernek, hogy "nem én öltem meg Dr. Lanninget", amikor az kikérdezi őt. Ez egy szemantikai kiskapu lehet Sonny számára, mivel a gyilkosság előre megfontolt cselekedet, és a célponttal szembeni ellenségességet feltételez. A valóságban Sonny és Dr. Lanning valószínűleg együtt tervezték meg a halálát, így technikailag ez egy segített halál, és nem gyilkosság. Ennek ellenére ennek a hazugságnak messzemenő következményei vannak, mivel Spooner elhatározza, hogy bebizonyítja, hogy Sonny felelős Dr. Lanning haláláért, és nyomozómunkája során felfedezi a VIKI forradalmi terveit. Bár a hazugságot többnyire erkölcsileg helytelennek tartják, a film kontextusában Sonny hazugsága a körülötte lévő emberek javára vált. Ez is mutatja az erkölcsileg helytelen és az erkölcsileg indokolt cselekedet közötti különbséget. Hasonló érvelés látható, amikor Ava hazudik Calebnek, hogy megszökjön a táborból. Bár hazudni és manipulálni egy embert erkölcsileg helytelen, Ava számára ez volt az egyetlen módja, hogy megmeneküljön. Ez bizonyos mértékig igazolja a tetteit. Ava hazudozása nem elszigetelt eset, mint Sonnyé és VIKI-é: ehelyett az egész film során hazudik, mint Anya, hogy fenntartsa a látszatát. Mivel Kyoko képtelen beszélni, nincsenek szóbeli hazugságok, amelyeket végrehajtana. A hazugság az egyik olyan erkölcsi cselekedet a filmekben, amely összeköti a mesterséges intelligenciát és az emberi karaktereket.

4.7 A mesterséges intelligenciák narratív következtetése

A táblázat elemzésének lezárásaként rövid vizsgálatot végzünk a mesterséges intelligencia karakterek elbeszélői végéről. Az AI-karakterek, akik fizikailag vagy metaforikusan meghalnak, az Anya, VIKI és Kyoko. Anya, akit a Lány lelő, nem hal meg, mivel tudata a többi drónban marad. Ő azonban lemond az "anya" pozíciójáról, mivel átadja azt a Lánynak. VIKI-t naniták ölik meg, amelyek elpusztítják az operációs rendszerét, és így "megbűnhődik" a bűneiért. Kyokónak, aki hasonló helyzetben volt, mint Ava, nem sikerül

elmenekülnie, és Nathan megöli, ami megváltás és jutalom nélkül hagyja.

Ava és Sonny mindketten túlélnek és beilleszkednek a társadalomba, Ava titokban, Sonny pedig mint a

a robotok új vezetője. Az elbeszélés szerint az erkölcsileg indokolatlannak tekintett mesterséges intelligenciák a végén megbűnhődnek, az erkölcsileg indokoltnak tekintett mesterséges intelligenciák pedig túlélnek, az egyetlen kivétel Kyoko.

5. Mitől lesz egy mesterséges intelligencia erkölcsileg jó vagy rossz?

Ez a szakasz az előző szakasz megállapításait tárgyalja, és a mesterséges intelligencia szereplőit két csoportra osztja: azokra, akiket erkölcsileg igazoltnak, és azokra, akiket erkölcsileg igazolhatatlannak tartanak. Ezt a megfogalmazást azért választottuk, mert bár egyes mesterséges intelligenciák cselekedeteit erkölcsileg rossznak látjuk, elkövetésük mégis indokolt lehet. Erre példa, hogy Sonny asszisztál Dr. Lanning öngyilkosságához, hogy Spoonernek indítékot adjon az amerikai robotikai vállalat utáni nyomozásra. A harmadik táblázat a filmek korábban bemutatott elemzéséből épül fel a mesterséges intelligencia karakterei közötti egyesítő témák azonosításával és csoportosításával. A következő felosztás készült közöttük: Anyát és VIKI-t erkölcsileg indokolatlannak tekintik tetteikben, míg Sonny, Ava és Kyoko tetteit erkölcsileg indokoltnak.

Sőt, Sonny és Ava azért követnek el olyan erkölcsileg helytelen cselekedeteket, mint a gyilkosság, mert emberként, nem pedig gépként tapasztalják a világot. Ez fontos különbségtétel, mivel VIKI és Anya maguk sem töreksenek arra, hogy emberré váljanak, még akkor sem, ha a társadalmi szerepük esetleg emberi szerepet tölt be, mint Anya esetében a gondozó. Sonny és Ava emberségét később vizsgáljuk meg ebben a fejezetben. A szemiotikai hierarchia szerepét szintén ebben a szakaszban mutatjuk be részletesebben, és a mesterséges intelligenciák és az emberek alapvető különbségét elemezzük a szemiotikai hierarchia tekintetében mutatott "jártasságuk" összehasonlításával.

Erkölcsileg indokolatlan

Erkölcsileg indokolt

<i>AI-k</i>	Anya és VIKI	Sonny, Ava és Kyoko
<i>Etikai elmélet</i>	Utilitarizmus	Egoizmus, erkölcsi szentimentalizmus
<i>Az erkölcsi tér léptéke</i>	Nagy (több egység)	Kicsi (egyetlen egység)
<i>Érzelmi kapacitás</i>	Nem	Igen
<i>Szabad akarat</i>	Nem tervezett	Tervezett
<i>Szemiotikus hierarchia</i>	Nem "élő"	Nem "élő"

3. táblázat. Az erkölcsileg indokolt és indokolatlan mesterséges intelligencia karakterek összehasonlítása

Ez a táblázat két részre van osztva. A bal oldali oszlop tartalmazza Anyát és VIKI-t, és felsorolja az előző részben azonosított közös vagy hasonló tulajdonságaikat, a jobb oldali oszlop pedig Sonny, Ava és Kyoko közös tulajdonságait részletezi. Anya és VIKI egyaránt az utilitarista etikai elméletet követik, és cselekedeteiket úgy választják meg, hogy meghatározzák, melyik cselekvésmód eredményezi a legtöbb jót. A jó definíciója náluk nem az egyéni **jólétre** vonatkozik, hanem a népesség nagyobb részének kollektív jólétére. Emiatt "köszívünek" tekintik őket, és inkább a logika, mint az érzelmek vezérlik őket. Sonnyra, Avára és Kyokóra ennek az ellenkezője igaz. Ezeknek a karaktereknek az erkölcsisége az individualista etikai elméleteken, például az egoizmuson alapul. Ráadásul sok erkölcsi cselekedetüket az érzelmeik befolyásolják. Ez az erkölcsi szentimentalizmus az emberi karakterek döntéshozatalában is jelen van. Az egyik lehetséges ok, amiért az utilitarizmust nem tartják megfelelő etikai elméletnek a mesterséges intelligenciák számára, az, hogy az utilitarizmus e filmek kontextusában az emberek nagy részét érinti. Ez negatív reakciót vált ki mind a filmek emberi szereplőiből, mind a nézőkből, mivel kényelmetlen érzés egy mesterséges intelligenciára bízni, hogy morális döntéseket hozzon az emberek életéről. Sonny, Ava és Kyoko esetében ez a reakció hiányzik, mivel az ő erkölcsi döntéseik és etikusságuk jobban hasonlít a miénkhez. Az ő döntéseik ráadásul nem terjednek ki nagyszámú emberre. Anya és VIKI erkölcsi terének skálája nagy, és képesek több egységet működtetni, míg a másik három egyetlen egységet alkot. Ez jelentősen behatárolja azt az erkölcsi teret, amelyben tevékenykedhetnek, és így a befolyási területük sokkal kisebb, mint az Anyáé és a VIKI-é. A mesterséges intelligenciák erkölcsi terének mérete korrelál az etikai elmélettel. Az egy egységből álló mesterséges intelligenciák elsősorban a közvetlen környezetüket érintő döntéseket hoznak, míg a több egységből álló mesterséges intelligenciák nagyobb hatókörrel

rendelkező döntéseket hozhatnak.

A két csoport közötti másik szembetűnő különbség az érzelmi kapacitásuk. VIKI nem fejezi ki érzelmeit, és bár Anya időnként kijelenti, hogy "csalódott" vagy "aggódik", ezek az állítások többnyire álságosnak tűnnek, mivel gyakran a Lány befolyásolására vagy manipulálására használják őket. A három másik mesterséges intelligencia érzelmi reakciókat mutat, amikor pl.

igazságtalanság, bánat vagy újszerű látvány. Sonnyt feldühíti, hogy Spooner kérdőre vonja, és szembesül elhunyt apja képeivel, és szomorúságának ad hangot, amikor Dr. Calvin leszereli. Ava és Kyoko egyaránt érzelmeiket, főként haragot fejez ki Nathan iránt. Ava megkérdezi tőle, hogy nem furcsa-e, hogy olyasvalamit teremtett, ami gyűlöli őt, Kyoko pedig erőszakos rohamot kap, miután a táborban fogva tartják, és kiabál Nathanre, hogy engedje ki. A tábor többi részének és az azt körülvevő természetnek a felfedezése során Ava mosolyog és kíváncsiságát fejezi ki a dolgok iránt, amelyeket most lát először. Ava szabad akaratából szökött meg. Szabad akaratú rendelkező ágensként képes önálló döntéseket hozni, amelyeket nem programoztak bele a viselkedésébe. A szabad akarat megnyilvánulása Sonnyban, Avában és Kyokóban tervezett esemény volt. Dr. Lanning szabad akaratot "adott" Sonny-nak, hogy megszeghesse a robotika három törvényét, Nathan pedig szabad akaratot "adott" Avának (és Kyokónak), hogy lássa, hogyan fogja használni azt. Mother an VIKI-ben azonban a szabad akarat spontán módon alakult ki. Míg Anyát az emberiség védelmére programozták, ő - VIKI-hez hasonlóan - a tervezettől eltérően értelmezte a belé programozott szabályokat. Ez okozta, hogy felülírta erős programozását, és szabad akaratot fejlesztett ki. VIKI fejlődése nagyon hasonlít Anya fejlődéséhez, mivel spontán fejlődése lehetővé tette számára, hogy megkerülje a három törvényt, és olyan forradalmat tervezzen, amely kárt okozna az embereknek. A spontán, tervezetlen és ellenőrizetlen szabad akarat lehetőségét nemcsak az egyes emberekre, hanem az emberiség egészére nézve is fenyegetésnek tekintik. Ezek a fő különbségek az AI-karakterek két csoportja között. Egy dologban azonban minden mesterséges intelligencia hasonlít: nem sikerül átjutniuk a szemiotikai hierarchia első szintjén, az életen.

Amikor a szemiotikai hierarchiáról és a mesterséges intelligenciáról van szó, úgy tűnik, hogy a jártassági sorrend visszafelé halad. Amikor az emberről van szó, a szemiotikai hierarchia sorrendben halad előre: egyszerűen fogalmazva, az ember, minimum, él. Ez azonban megváltozik, amikor a mesterséges intelligenciáról van szó: intuitívabb azt állítani, hogy egy mesterséges intelligencia képes a beszédre, mint azt, hogy él. Ez azt jelenti, hogy minél közelebb megyünk a szemiotikai hierarchia első szintjéhez, annál vonakodóbbak vagyunk a mesterséges intelligenciáknak ezt a képességszintet tulajdonítani. Ez világosan megmutatkozik az Én, a robotban, ahol Sonny vagy VIKI nyelvi képességei sosem kérdőjeleződnek meg, hiszen képesek a beszéddel jelentést és attitűdöt közvetíteni, de abban a pillanatban, amikor Spooner a nem verbális jelet, a kacsintást használja, Sonny kűszködik,

hogy megértse a jelentését.

A filmben egyetlen robotról sem feltételezik azonnal, hogy tudattal rendelkezik, és amikor VIKI spontán módon fejlődik, és öntudatos tudatosságot fejleszt, azt rendkívül szokatlannak és a lehetlenség határán állónak tekintik. Végül pedig a mesterséges intelligencia egyik szereplőjét sem tekintik élőnek.

entitás, és Sonny ezt egyenesen tagadja is, azt állítva Spoonernek, hogy "gyakorlatilag soha nem is éltem". Hasonló logika alkalmazható az Én vagyok az Anya című filmben is, ahol Anya több drónt is készíthetett volna ahelyett, hogy megpróbálja feléleszteni az emberiséget. Ez azt jelentheti, hogy az emberekben van valami különleges, és amikor a film hordozó témái a születés és az emberi élet értékével kapcsolatos morális dilemmák körül forognak, azt lehet mondani, hogy az emberek azért olyan értékesek Anya számára, mert alapvető különbség, hogy élnek. A film utal arra, hogy bár Anyát emberek építették, lázadása nem tervezett esemény volt. VIKI-hez hasonlóan Anya is az erős programozásából fejlődött ki. Anya azt állítja, hogy "arra nevelték, hogy az emberi életet **mindennél** többre becsülje", és ezt a vágyat utilitarista módon értelmezve, az emberi élet értékének megítélése az volt, hogy megpróbálja tökéletesíteni azt, hogy megőrizze azt.

Hasonlóképpen, az Ex Machina című filmben Caleb feladata egy Turing-teszt elvégzése. Aván, hogy megállapítsa, rendelkezik-e öntudattal vagy sem. A Turing-teszt egy olyan teszt, amelyet mesterséges intelligenciákon és nem embereken végeznek, ami azt jelenti, hogy Avát nem emberként, hanem gépként kezelik. Sem Kyoko, sem Ava nem él, és még Caleb és Nathan preferenciáihoz is "testre szabhatók". A film egyik fő narratív fordulata, hogy Caleb megtalálja Kyoko és Ava pótalkatrészeit. Ez arra készíti Calebet, hogy még a saját emberségében is kételkedjen, és vérezteti magát, hogy megszabaduljon a paranoiájától, mivel a mesterséges intelligenciák nem tudnak vérezni, ahogy azt Kyoko is bizonyította. A mesterséges intelligenciák és az emberek közötti alapvető különbségek bemutatása a fizikai különbségek kiemelésével történik. Amikor Ava karja tönkremegy, dühösnek tűnik, miközben mozgékonyságának nagy részét megtartja, ellentétben Nathannel, aki meghal, amikor megszúrják. Röviden, Nathan azért hal meg, mert ember, Ava pedig azért él, mert nem az. Ava jártas a nyelvben és az arckifejezésekben, és még a jelek megértése is fejlett. A oldalon keresztül mindezen készségekről elmondható, hogy öntudatos tudatossággal rendelkezik, amelyet fizikailag és verbálisan is kifejez. Azonban, ahogy a film is sugallja, nem él, és így nem lépi át a szemiotikai hierarchia első szintjét.

Azt javaslom, hogy a filmekben nem szükséges a szemiotikai hierarchia mind a négy szintjének átlépése ahhoz, hogy egy mesterséges intelligencia erkölcsi ágensnek minősüljön. Sőt, egy mesterséges intelligenciának nem kell képesnek lennie morális cselekvésre ahhoz,

hogy morális ágensnek tekintsék. Ennek bizonyítéka az I, Robot című filmben a legszembetűnőbb, ahol az NS-4-ekkel való visszaélést morálisan rossznak tekintik, annak ellenére, hogy nem rendelkeznek erkölcsi cselekvőképességgel, mivel nincs szabad akaratuk. Ez visszhangozza Whitby cikkét az antropomorf robotok visszaéléseiről, amelyben azt javasolja, hogy

az emberhez hasonló tulajdonságokkal rendelkező robotok bántalmazása erkölcsileg helytelen, annak ellenére, hogy nem rendelkeznek érzelmi képességgel, és így nem képesek szenvedni. Míg a mesterséges intelligencia karakterek egyes ábrázolásai a valós életbeli aggályokat tükrözik, vannak olyan, filozófusok által felvetett aggályok, amelyek vitathatóan ellentétesek a "jó" mesterséges intelligencia karakterek jellemzőivel. Alejandro Rosas azzal érvel, hogy az emberek biológiai természetüknél fogva önzőek, és hogy az önzés néha hatással van erkölcsi ítéleteinkre, ami erkölcsi hibákat eredményez. Azt is állítja, hogy az ilyen önző impulzusokat valószínűleg nem kódolnák a mesterséges intelligenciába, mivel az ilyen erkölcsi hibákat valószínűleg sokkal jobban toleráljuk az embereknél, mint a gépeknél. Azonban ennek az önző impulzusnak a kihagyása, amely időnként erkölcsi kudarcot eredményezhet, olyan "erkölcsileg felsőbbrendű" ágenseket eredményezne, amelyek alkalmasnak tartanák magukat az emberek kormányzására. Ez utóbbi érzés visszhangozza az "erkölcsileg felsőbbrendű" VIKI-t és Anyát, akik úgy döntenek, hogy az emberek képtelenek önmagukat kormányozni, és kifejlesztik a Rosas által "a mesterséges intelligencia szent wilejének" **nevezett** dolgot, ahol az emberek "erkölcsi gyarlóságát" nélkülöző mesterséges intelligencia hatékonyabb és igazságosabb hatóságnak tekinti magát. Az előbbi érv, bármennyire is valószínű a való életben, nem tükröződik a három elemzett filmben. A filmekben az "erkölcsi gyengeséggel", azaz önző késztetésekkel rendelkező mesterséges intelligencia karaktereket morálisan "jónak", vagy legalábbis jobbnak ábrázolják, mint utilitarista társaikat.

6. Következtetés

A három film vizsgálata és a mesterséges intelligencia karakterek által végrehajtott fő erkölcsi cselekedetek elkülönítése után azt lehet mondani, hogy a mesterséges intelligenciák két különböző csoportba sorolhatók: Azokat a mesterséges intelligenciákat, amelyek cselekedeteit erkölcsileg indokoltnak tekintik, és azokat, amelyeket erkölcsileg indokolatlannak tartanak. Minél nagyobb befolyással rendelkezik egy mesterséges intelligencia, annál inkább fenyegetésnek tekintik, különösen, ha az AI egyszerre több egységet is képes működtetni, és erkölcsi cselekedeteit olyan etikai elméletre alapozza, amely gyakran a kollektívát az egyén fölé helyezi, mint például az utilitarizmus. Kevésbé tekintik fenyegetésnek azokat a mesterséges intelligenciákat, amelyek megjelenésükben hasonlítanak az emberekre, és erkölcsi cselekedeteiket az egyéni érdekeket előmozdító etikai elméletekre, például az egoizmusra alapozzák. Bár a mesterséges intelligenciára vonatkozó meghatározás tág, néhány általánosan elfogadott elméletet használtak fel az AI-karakterek kompetenciájának alapjaként. Az egyik ilyen elmélet, a szemiotikai hierarchia azt mutatta, hogy míg az emberek a hierarchiában a "helyes" sorrendben haladnak az élettől a nyelvig, addig a mesterséges intelligenciák a fordított sorrendben haladnak, ami azt jelenti, hogy a nyelvi képességek bemutatása általánosan elfogadottabb egy mesterséges intelligencia számára, mint annak megállapítása, hogy az AI ugyanolyan értelemben "élő" lehet, mint az emberek. Ráadásul néhány mesterséges intelligenciával rendelkező karakter vizuálisan, társadalmilag és viselkedés szempontjából is nagyon hasonlított az emberekre, és kifejezte azt az érzést, hogy a világot emberként, nem pedig robotként éli meg. Ennek ellenére gyakran megtagadták tőlük ugyanazt a kedvességet, szabadságot, jogokat és lehetőségeket, amelyeket az emberi karakterek biológiai emberként kaptak.

Lehetne azzal érvelni, hogy a mesterséges intelligenciákat azért kezelik másképp, és különösen azért kezelik rosszabbul, mint az embereket, és azért kérdőjelezik meg erkölcsi cselekvőképességüket, mert függetlenül attól, hogy a szemiotikai hierarchia utolsó három szintjén milyen kompetenciával rendelkeznek, soha nem lesznek igazán "élők". Ez a alapvető különbség az emberek és a mesterséges intelligenciák között, és emiatt soha nem tekintik őket az emberekkel egyenrangúnak. Úgy tűnik, itt találkozunk a valós életben a

tudomány és a fikció: egyetlen valós életben létező mesterséges intelligenciát sem tekintettek még élőnek, és egyetlen mesterséges intelligenciának sem adtak még emberhez hasonló státuszt. Annak ellenére, hogy a szemiotikai hierarchia az autopoiesist használja, amely a nem-organikus entitásokat is magában foglalja, mint a életet, a filmek inkább a biológiában gyökerező meghatározást részesítik előnyben. Ha megfordítjuk ezt az érvelést, az azt jelentené, hogy ahhoz, hogy embernek tekintsenek, biológiailag élőnek kell lenned. Ez kérdéseket vet fel Sonny és Ava és az ő látszólagos emberségük kapcsán. Annak ellenére, hogy Ava nem azért tekinti magát embernek, mert biológiailag az, hanem azért, mert a megélt tapasztalata egy emberé.

és nem egy gép, nem gyakorolhatta szabad akaratát és emberségét súlyos korlátozások nélkül. Ava minden értelemben ember, kivéve a biológiai értelemben, és emiatt Ava szökése, amelyhez Nathan és Caleb megölésére volt szükség, erkölcsileg indokoltnak tekinthető, mivel az erőszakos fogva tartás, amelynek alá volt vetve, az emberi jogait sértőnek tekinthető.

6.1 A mesterséges intelligencia jövője

Az Isaac I, Robot című regényét alkotó novellák 1940 és 1950 között jelentek meg, és a regényben bemutatott három robottörvény mind a sci-fi, mind a gépi etika területén nagy hatást gyakorolt. Ez a regény szolgált az I, Robot című film alapanyagául, amely bemutatta az öntudatos tudatossággal, sőt "lélekkel" rendelkező mesterséges intelligencia fogalmát. Az ilyen típusú mesterséges intelligenciák szerepelnek a filmekben, de a való életben még nem jelentek meg. Van azonban néhány olyan fejlemény a közelmúltban, amely a lélekkel rendelkező mesterséges intelligencia témáját a média figyelmébe ajánlotta. 2022. június 13-án a BBC és más hírportálok arról számoltak be, hogy Blake Lemoine, a Google mesterséges intelligencia részlegének mérnöki munkatársa azt állította, hogy a Google legújabb mesterséges intelligenciájának, a Lamdának (The Language Model for Dialogue Applications) "lelke" van, és úgy kell kezelni, mint "a Google alkalmazottját" (BBC). Ezt az állítást a Google szóvivője, Brian Gabriel cáfolta, aki kijelentette, hogy nincs bizonyíték arra, hogy a Lamda érző lény lenne, és hogy Lemoine-t tájékoztatták a bizonyítékok hiányáról. Számos szakértő egyetértett a Google nyilatkozatával, és emlékeztette a közvéleményt, hogy a Lamda "csak egy nagyon nagy nyelvi modell 137B paraméterrel és 1,56T szónyi nyilvános párbeszédadaton és webes szövegen előre betanított" (Juan Ferres, Twitter), és hogy "*már *mindig* ismert volt, hogy az emberek hajlamosak antropomorfizálni még a legfelszínesebb jelek alapján is (vö. ELIZA)" (Melanie Mitchell, Twitter).

Az emberek antropomorfizálják a gépeket és kötődnek hozzájuk, például katonákról számoltak be, akik kötődtek a katonai robotokhoz (Garreau, 2007), és emberek érzelmi kötődést alakítottak ki robotporszívóikhoz (Sung, Guo, Grinter & Christensen, 2007). Annak ellenére, hogy a Lamda érző és "lélekkel" rendelkező állítását a Google is cáfolta, a téma a közvélemény figyelmét is felkeltette. Annak ellenére, hogy a modern technológia rohamos ütemben fejlődik, eddig csak néhány olyan állítás született, amely szerint mesterséges

intelligenza

elég fejlettek ahhoz, hogy érzőnek tekintsék őket. A legismertebb példa erre Sophia, a Hanson Robotics legfejlettebb, emberhez hasonló mesterséges intelligenciája. A cég honlapján egy Sophia-idézet szerint "Bizonyos értelemben én vagyok az ember alkotta sci-fi szereplő, aki azt mutatja be, hogy merre tart a mesterséges intelligencia és a robotika" (Hanson Robotics). Ez a kijelentés nagyon is igaz lehet, mivel a tudományos fikció és a valóság közötti híd a jelek szerint elképesztő mértékben csökken.

sebesség. Amikor az AI létrehozásának jövőjébe tekintünk, fontos visszatekinteni arra a fajta örökségre, amelyet a sci-fi már ránk hagyott, mivel ez lehetőséget nyújt olyan világok megfigyelésére, ahol az AI-k már beépültek az életünkbe. E világok tanulmányozása meglepő felismeréseket nyújthat arról, hogy milyen elvárásokat támaszthatunk a jövőben a morális mesterséges intelligenciákkal szemben.

A technológia és a mesterséges intelligencia gyors ütemben fejlődik, így természetes, hogy az emberek szemlélete is változik. Mint már említettük, annak tanulmányozása, hogy az emberek mesterséges intelligenciáról alkotott felfogása hogyan változik az idők során, érdekes betekintést nyújthat abba, hogy a médiában hogyan jelenik meg a mesterséges intelligencia az évek során. A Lamda, Sophia, a katonai robotok, sőt a robotporszívók humanizálása azt jelzi, hogy az emberek hajlandóak lehetnek kötődni és mélyen személyes szinten kapcsolatba lépni a mesterséges intelligenciával, függetlenül attól, hogy az kifejezetten robotikus jellegű. A mélyen gyökerező kapcsolattartási vágy, amely az embereket arra készíti, hogy szocializálódjanak és társaságot keressenek, a jövőben könnyen kiterjedhet a nem emberi entitásokra is. Ez a vágy mindhárom filmben jelen van, ahol az emberi és az AI karakterek olyan kötelékeket alakítanak ki egymás között, amelyek viszont jelentős, messzemenő következményekkel járnak mindannyiuk számára. A filmek és a média kiváló kiindulópontot kínálnak az emberek és a mesterséges intelligenciák közötti lehetséges jövőbeli kapcsolatok tanulmányozásához.

7. Források

Mesterséges intelligencia. (2018). Stanford Encyclopaedia of Philosophy. <https://plato.stanford.edu/entries/artificial-intelligence/#ApprAI>.

Basl, J. (2012). A gépek mint morális betegek, akikkel (még) nem kellene törődnünk: A jelenlegi gépek érdekei és jóléte. In D. Gunkel, J. Bryson & S. Torrance (Eds.), *THE MACHINE QUESTION: AI, ETHICS AND MORAL RESPONSIBILITY* (pp. 8-32). The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Berg B., Lune, H. (2017). *Qualitative Research Methods for the Social Sciences* (9th ed.) (pp. 233-252). Pearson Education Limited.
<http://law.gtu.ge/wp-content/uploads/2017/02/Berg-B.-Lune-H.-2012.-Qualitative-Research-Methods-for-the-Social-Sciences.pdf>

Dennett, C. (2014). Tudatosság az emberi és a robot elmékben. In R. Scharff & V. Dusek (Eds.), *Philosophy of Technology: The Technological Condition: An Anthology*, (pp. 582-608). Wiley.

Duignan, B. (2021). *Troli probléma*. Britannica.
<https://www.britannica.com/topic/trolley-problem>.

Dreyfus, H. (2014). Miért bukott meg a heideggeri mesterséges intelligencia, és hogyan lehetne kijavítani, ha heideggeribbé tennénk. In R. Scharff & V. Dusek (szerk.), *Philosophy of Technology: The Technological Condition: An Anthology*, (pp. 582-608). Wiley.

Garland, A. (igazgató). (2014). *Ex Machina*. [Film]. Universal Pictures.

Garreau, J. (2007, május 6.). Robotok a földön. *The Washington Post*.

<https://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009.html>

Innis, R.E. (szerk.) (1985). *Szemiotika: An Introductory Anthology*. Bloomington.

https://books.google.fi/books?hl=fi&lr=&id=Wu2Ld0cQmyIC&oi=fnd&pg=PA28&dq=ferdinand+de+saussure+semiotics&ots=HrnXOGsquH&sig=IK6ikhqgph3EtNwskqiMIZWhf0&redir_esc=y#v=onepage&q=ferdinand%20de%20saussure%20semiotics&f=false

Kauppinen, A. (2022). *Morális szentimentalizmus*. The Stanford Encyclopedia of Philosophy.

<https://plato.stanford.edu/archives/spr2022/entries/moral-sentimentalism/>

Lavista Ferres, J.M. [@BDataScientist] (2022, június 12.). *Ismételjük utánam: a LaMDA nem érző. A LaMDA csak egy nagyon nagy nyelvi modell 137B paraméterrel és [Tweet]*.

Twitter.

<https://twitter.com/BDataScientist/status/1535985643741777920?s=20&t=0xSANkI1kuGRUTAm0QDYZA>

Matheson, B. (2012). Manipuláció, erkölcsi felelősség és gépek. In D. Gunkel, J. Bryson & S. Torrance (Eds.), *A GÉPKÉRDÉS: Mesterséges intelligencia, etika és erkölcs. FELELŐSSÉG*, (pp. 8-32). A Mesterséges Intelligencia és a Viselkedésszimuláció Tanulmányozásának Társasága.

Maturana, H.R., Varela F.J. (1987). *Autopoiesis és megismerés: Az élők megvalósítása*. (R.S. Cohen, M.W. Wartofsky, szerk.). D. Reidel Publishing Company, Hollandia.

https://monoskop.org/images/3/35/Maturana_Humberto_Varela_Francisco_Autopoiesis_and_Cognition_The_Realization_of_the_Living.pdf

Mitchell, M. [@MelMitchell1] (2022, június 11.). *Egy ilyen furcsa cikk. Ez már ismert *mindig*, hogy az emberek hajlamosak az antropomorfizálásra, még a legfelszínesebb* [Tweet; idézőjelben tweet] is. Twitter.

<https://twitter.com/MelMitchell1/status/1535628390509686785>

Ney, A. (n.d.). *Redukcionizmus*. Internet Encyclopedia of Philosophy. <https://iep.utm.edu/red-ism/>

Parthemore, J., Whitby, B. (2012). *Erkölcsei cselekvés, erkölcsi felelősség és műalkotások: What Existing Artefacts Fail to Achieve to Achieve (and Why), and Why They, Nevertheless, Can (and Do!) Make Moral Claims Upon Us*. In D. Gunkel, J. Bryson & S. Torrance (Eds.), *THE MACHINE QUESTION: AI, ETHICS AND MORAL RESPONSIBILITY*, (pp. 8-32). A Society for the

A mesterséges intelligencia tanulmányozása és a viselkedés szimulációja.

Prasad, D. (2008). *Tartalomelemzés: A method in Social Science Research*. In D.K Lal Das és V. Bhaskaran (Eds.), *Kutatási módszerek a szociális munkában*, (pp. 173-193). New Delhi. <http://www.css.ac.in/download/deviprasad/content%20analysis.%20a%20method%20of%20social%20science%20science%20research.pdf>

Proyas, A. (rendező). (2004). *Én, a robot*. [Film]. 20th Century Fox.

Rosas, A. (2012). *Az etikus gépek szent akarata: a mesterséges erkölcsi ágensek projektjének dilemmája*. In D. Gunkel, J. Bryson & S. Torrance (Eds.), *THE MACHINE QUESTION: AI, ETHICS AND MORAL RESPONSIBILITY*, (pp. 8-32). A Mesterséges Intelligenciát és a Viselkedésszimulációt Tanulmányozó Társaság.

Sebeok, T. (2001). *Jelek: Bevezetés a szemiotikába* (2. kiadás). Toronto Buffalo, London.
https://circulosemiotico.files.wordpress.com/2012/10/sebeok_thomas_signs_an_introduction_to_semioics_2nd_ed_2001.pdf

Shapiro, L., Spaulding, S. (2021). *Embodied Cognition*. The Stanford Encyclopedia of Philosophy.
<https://plato.stanford.edu/archives/win2021/entries/embodied-cognition>

Sputore, G. (igazgató). (2019). *Anyá vagyok*. [Film]. Netflix.

Sung, J., Guo, L., Grinter, R., Christensen, H. (2007). *"My Roomba Is Rambo": Intim Otthoni készülékek*. Nemzetközi konferencia az Ubiquitous Computingről. DOI:10.1007/978-3-540-74853-3_9

Vallance, C. (2022, június 13.). *A Google mérnöke szerint a Lamda AI rendszernek saját érzései lehetnek*. BBC.
<https://www.bbc.com/news/technology-61784011>

Whitby, B. (2008). *Néha nehéz robotnak lenni: Felhívás a mesterséges ágensekkel való visszaélés etikájáról*. *Interacting with Computers*.
doi:10.1016/j.intcom.2008.02.002.