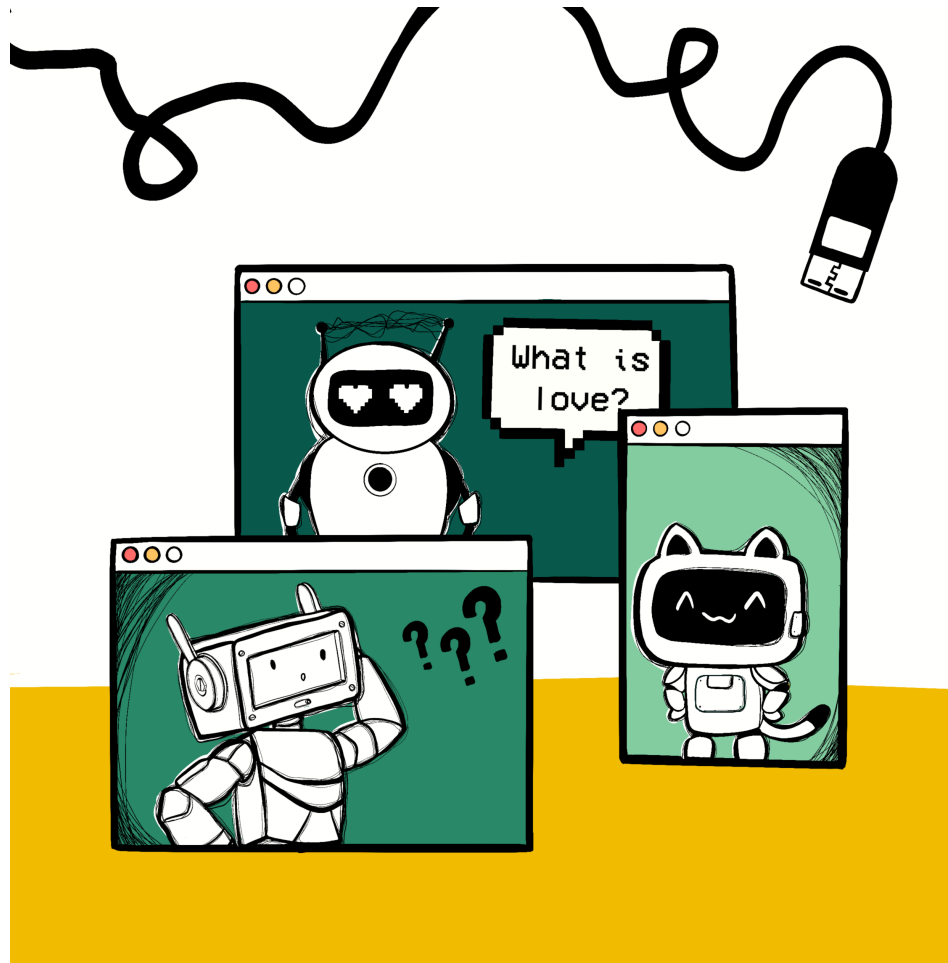


## 2. Modul

### Etika a mesterséges intelligenciában

"A mesterséges intelligencia nagyon jól leírja a mai **világot**, annak minden előítéletével együtt. De nem tudja, hogy **a világnak milyennek kellene lennie.**"



# A modulról

Ebben a modulban a mesterséges intelligencia kutatásának alapvető etikai szempontjaival foglalkozunk. Ennek során a diákoknak szabályzatok önálló ki- és átdolgozása révén kell megtanulniuk, hogyan tegyenek fel saját, kritikus kérdéseket, és hogyan tárják fel az egyes etikai alapelvek közötti ellentmondásokat. Annak érdekében, hogy ez az összetett vállalkozás minél élvezetesebb legyen, különböző megközelítésű és bonyolultságú feladatokat állítottunk össze, amelyek nehézségét a tanárok is szabályozhatják. A feladatok gondolat kísérletek lesznek (többek között a jól ismert villamos dilemma, de a fiatalabb tanulóknak szóló feladatok is szerepelnek, mint például szabályalkotás egy robot komornyik számára).

## Célok

A tanulók képesek lesznek: ...

- ... érveket megfogalmazni az etikai iránymutatások szerepéről a mesterséges intelligencia rendszerekhez kapcsolódóan
- ... önálló erkölcsi szabályokat megfogalmazni, és a tanár segítségével felülvizsgálni azokat
- ... csoportmunka révén azonosítani a kikapukat saját szabályaikban
- ... ismertetni a villamos dilemmát és a két ellentétes filozófiai irány közötti feszültséget
- ... saját filozófiai álláspontot kialakítani és helyesen osztályozni azt
- ... osztályozni az EU etikai irányelveinek lényegi pontjait, és kritikával vizsgálni azokat
- ... saját, személyes nézőpontot ismertetni a mindennapi élet **MI** rendszereivel kapcsolatban (az autonóm vezetés példáján keresztül)

## Beosztás

Idő	Tartalom
20 perc	Feladat – Alkossunk robottörvényeket
10 perc	Elmélet – Bevezetés: Milyenek a jó etikai szabályok?
15 perc	Feladat és bevezetés: Torzítási hibák
20 perc	Feladat – Erkölcs gép
15 perc	Elmélet – A villamos dilemma (haszonelvűség vs. erényetika)

**Idő****Tartalom**

---

15 perc

Elméleti Az önvezetésről

15 perc

Feladat – Autonóm vezetés

## Bevezetés

### A modulról – Miért fontos az etikai megközelítés a mesterséges intelligencia területén?

Amikor az emberek a mesterséges intelligenciára gondolnak, sokan szuperintelligens robotoktól tartanak, amelyek le akarják igázni az emberiséget, mint a Mátix vagy a Terminátor filmekben. Az **MI**-kutatás azonban mérföldekre van az ilyen forgatókönyvektől. A kutatások viszont már most is foglalkoznak alapvető etikai kérdésekkel, melynek oka, hogy a mesterséges intelligencia hibázik. Az „elfogult”, egyoldalú adathalmazok torzítási hibákat okoznak, a programok pedig ennek következtében egyéneket és csoportokat diszkriminálhatnak. Ezért van szükség arra, hogy az etika- és hittanórákon kívül is morális érzékenységgel tekintsünk a nehéz helyzetekre, megismerjünk különböző nézőpontokat és elsajátítsuk a döntéshozatalhoz szükséges készségeket.

# Alkoss saját robottörvényeket

Az első gyakorlat segítségével egy játékos, megengedő szemlélet kialakítása a cél. Ebben a feladatban egy fiktív forgatókönyvet alkotunk, amelyben egy humanoid háztartási robot képes minden feladatot ugyanúgy elvégezni, mint egy ember. A tanulónak el kell gondolkodniuk azon, hogy milyen lehetőségek és – ezzel együtt milyen – nehézségek merülnek fel a mesterséges intelligencia rendszerek mindennapokban való használatakor.

## 1. Képzeld el a következő helyzetet ...

Képzeld el, hogy van otthon egy robot komornyikod. Ez egy **humanoid (emberszerű) robot**, amely mechanikusan bármit képes megcsinálni, amit egy ember. Te vagy a robot tulajdonosa, és utasíthatod például arra, hogy kérdéseket válaszoljon meg vagy végezze el a házimunkát. A robot viselkedési képességeit csak néhány szabály korlátozza, amelyeket neked kell meghatároznod.

Az 1. segédanyag 2. diája használható segítségként, a handout 2. oldala pedig írásablonként.

## 2. A diákok "aranyszabályokat" találnak ki

A tanuló első feladata, hogy olyan szabályokat vagy "törvényeket" alkosson, amelyek biztosítják, hogy a robot ne viselkedjen károsan vagy szánt szándék nélkül. A következő beszélgetések szempontjából előnyös lehet, ha a tanulók eredményeiket kulcsszavak vagy gondolattérkép / poszter / dia formájában rögzítik. Ezt a gyakorlatot eredményesen lehet kiscsoportokban végezni, mely lehetőséget teremt a tanulók közötti eszmecserére. Alternatívaként a "Think-pair-share"(Gondolkodj! – Dolgozz párban! – Oszd meg!) módszer is alkalmazható. Elegendő időt (kb. 10-15 percet) kell biztosítani arra, hogy a tanulók valóban átgondolhassák az egyes forgatókönyveket, amelyekre a szabályaik vonatkoznak.

A kapcsolódó tananyag harmadik diáján néhány kulcskérdés található arra az esetre, ha a tanulónak további javaslatokra lenne szükségük, mielőtt elkezdenék a munkát (vagy ha elakadnának).

## 3. Szabályok bemutatása

Minden csapatnak adjunk néhány percet arra, hogy ismertessék szabályaikat, és egyszerű példákkal elmagyarázzák azokat.

## **4. Keressünk gyenge pontokat és kiskapukat**

Válasszunk ki egy szabályt a bemutatottak közül, és szemléltessük, hogyan lehet megkerülni / félreértelmezni. A Melléklet negyedik diáján van néhány példakérdés.

## **5. A tanulóknak meg kell találniuk és be kell mutatniuk, hogyan lehet „fogást találni” a többi szabályon.**

Erre is különösen alkalmas a "think-pair-share" módszer "(Gondolkodj! - Dolgozz párban! - Oszd meg!), valamint a kiscsoportos munka.

## **6. Tananyag – magyar nyelven**

-  Ethics – Worksheet Robot Law.pdf

# Bevezetés az MI-etikába

## Hogyan hozhatunk erkölcsös döntést?

Segítsük a tanulókat azzal, hogy beszélünk az erkölcsi szabályok általános megfogalmazásáról. Mit kell figyelembe venni ahhoz, hogy egy morális elv szilárd és megalapozott legyen?

Egy erkölcsi szabály vagy döntés:

- **Logikus:** Azaz célunk, hogy egy morális döntést bizonyítékokkal és érveléssel támasszunk alá, ahelyett, hogy érzésekre, társas vagy személyes preferenciákra hagyatkoznánk. A logikus döntések azt is biztosítják, hogy az ilyen értékítéleteink összeegyeztethetők legyenek más morális és nem-morális jellegű meggyőződéseinkkel is. Az ellentmondásokat minden esetben kerülnünk kell.

*A legtöbb filozófus egyetért abban, hogy ha erkölcsi ítéletet hozunk, késznek kell lennünk arra, hogy hasonló körülmények között ugyanígy döntsünk. Tehát ha például erkölcsileg helytelennek látjuk, hogy Miller asszony a saját javára módosította a számokat egy kutatásban, akkor az is ugyanígy helytelen, ha a kollégánk, a házastársunk vagy a szüleink teszik ugyanezt. Nem kivételezhetünk saját magunkkal olyan esetekben, amelyek miatt másokat elítélnénk.*

- **Tényeken alapuló:** Erkölcsi döntéseink előtt a lehető legtöbb releváns információt össze kell gyűjteni. Ezeknek az információknak hiánytalanoknak (vagy legalábbis az összes lehetséges rendelkezésre álló adatot tartalmazónak) és hitelesnek kell lenniük.
- **Szilárd és védhető erkölcsi elveken alapul:** Döntéseinknek olyan általánosan megfogalmazott erkölcsi elveken kell alapulniuk, amelyek ellenállnak a kritikai szemléletnek és a racionális kritikáknak. Az, hogy pontosan mi tesz egy erkölcsi elvet megalapozottá vagy elfogadhatóvá, az etika egyik legnehezebb kérdése.

Ha kerüljük erkölcsi döntéseink nyilvános megvitatását, az gyakran figyelmeztető jel. Ha nem akarunk nyilvánosan számot adni a tetteinkről, akkor fennáll a lehetősége annak, hogy olyasmit teszünk, ami erkölcsileg nem védhető. Ilyen

esetekben hasznos lehet Immanuel Kant filozófiai szemléletére gondolnunk: alkalmazzuk erkölcsi ítéleteinket minden egyes emberre, akire csak gondolni tudunk (magunkra is). Nem állhatunk ki egy erkölcsi alapelv mellett, ha nem akarjuk azt a társadalom egészére alkalmazni.

Hasznos lehet az is, ha egy problémát egy másik személy szemszögéből is megvizsgálunk, hogy elkerüljük az erkölcsi rövidlátást (például feltehetjük magunk számára a kérdést: "Mit szólnál hozzá, ha ...?").

**Isaac Asimovs** 1942-ben megfogalmazott **robotika törvényei** segítségünkre lehetnek annak eldöntéséhez, hogy a diákok milyen szabályokat állíthatnak fel:

1. A robotnak nem szabad kárt okoznia emberi lényben, vagy tétlenül túrnie, hogy emberi lény bármilyen kárt szenvedjen.
2. A robotnak engedelmeskednie kell az ember által adott utasításoknak, kivéve, ha ezek az utasítások ellentétesek az Első Törvénnyel.
3. A robotnak meg kell védenie saját létét, amíg ez a védelem nem ütközik az Első vagy a Második Törvénybe.

Vannak azonban hasonló megközelítések a tudományos-fantasztikus irodalmon kívül is. Számos nagyvállalat, például a BMW csoport vagy a Microsoft etikai alapelveket fogalmazott meg a mesterséges intelligenciára vonatkozóan.

Kötelező szabvány továbbra sincs. A gyors fejlődési ütemet tekintve pedig nem meglepő, hogy már most számtalan szakértői csoport próbál általános szabályokat meghatározni.



Az etikai szabályalkotás általánosabb témája után, a mesterséges intelligencia rendszerek elfogultsági hibáival foglalkozunk. A következő gyakorlat alapjául egy tanulóknak szóló szöveg szolgál, amely egy mesterséges intelligencia által automatizált középiskola választásról szól.

## Torzítási hibák

Ez a feladatlap feldolgozható közösen, hogy a tanulók biztosan megértsék a probléma hátterét.

Ebben a két feladatban a tanulók megtudhatják, **mik az torzítási hibák, hogyan fordulhatnak elő, és hogyan kerülhetők el.**

Az 1. és a 2. feladat egyaránt alkalmas csoportmunkára, de közösen is végezhető. Célszerű kulcsszavak formájában rögzíteni a felmerült gondolatokat a nyomomonkövethetőség érdekében. Mindenekelőtt világossá kell tenni a tanulók számára, hogy a torzítási hibák, csupán az adatsorokba táplált előítéletek, nem pedig tényszerű állítások. Különösen a torzításokból eredő diszkriminációt, mint súlyos problémát fontos kiemelni.

A torzítási hibák legfontosabb vonatkozásait a feladat végén írásban is rögzíthetjük. Lehetőség van arra is, hogy a tartalmakat tanárként, PowerPoint prezentáció formájában mutassuk be, vagy az előismeretek függvényében a tanulókkal közösen dolgozzuk fel az anyagot.

## Tananyag – magyar nyelven

-  Ethics – Worksheet Bias Error.pdf

# Torzítási hibák és etikai irányelvek

## Egyáltalán mik azok a torzítási hibák?

Olyan hibák, melyek az adatok torzítottságából vagy valamilyen elfogultságból, egyoldalúságból (pl. egy adott nézőpont gyakori megjelenése) származhatnak. Ezek már a szükséges adatok gyűjtésekor felmerülnek, amelyek később téves eredményekhez, valamint olyan súlyos problémákhoz vezethetnek, mint a kirekesztés és hátrányos megkülönböztetés.

A Tay nevű chatbotot például 2016-ban mutatták be a Twitteren, ám nagyon rövid időn belül le is kapcsolták, ugyanis a bot a felhasznált adathalmaz miatt számtalan diszkriminatív üzenet közzétételébe kezdett.

## Milyen típusú torzítási hibákról beszélhetünk?

- **algoritmikus MI- torzítás vagy "adat torzítás":** a betáplált adatok által okozott torzítási hiba (az adatok statisztikai jellegű torzítottsága).
- **társas MI torzítás:** a társadalom által „belénk programozott” normák, de a sztereotípiák is képezhetnek vakfoltokat vagy (társadalmi) előítéleteket.

**A társadalmi előítéletek gyakran befolyásolnak/teremtenek algoritmikus torzítási hibákat!**

A mesterséges intelligencia csak annyira jó, amennyire az alapjául szolgáló adatbázis. A programozók is csak emberek, akik szintén a saját kis buborékjukban élnek saját nézeteikkel és előítéleteikkel. Mindez akkor válik problémássá, amikor ezek az előítéletek a programokba is beépülnek a kiválasztott adathalmazok formájában ( **"Garbage in, garbage out. = Szemét be, szemét ki."** ). Mindez új követelményeket támaszt a mesterséges intelligencia kutatásával szemben. A programozási tudás mellett alapvető etikai ismeretekre is szükség van ahhoz, hogy mérlegelni lehessen, hogy az adott program segíti vagy éppen károsítja-e a közjót.

Ezért szükséges lefektetni etikai és erkölcsi elvek alapelveket, hogy egyetlen embert se rekeszthessen ki egy mesterséges intelligencia.

Számos szakértői csoport próbált már szabályzatokat létrehozni: többek között az Európai Bizottság is, amely az **etikai iránymutatásokat** dolgozott ki a megbízható

**MI-re vonatkozóan.** A következőkben az alábbi négy etikai alapelvet tekintjük alapvetőnek, amelyek megbízható mesterséges intelligenciák alapjául szolgálnak.

## Négy etikai alapelv:

### Méltányosság:

Ha a felhasznált adatokban jelen vannak előítéletek, a program méltánytalanul jár el. A "méltányosság" fogalmát nem olyan könnyű meghatározni: vajon egy csoporton belül minden embert pontosan ugyanúgy kell kezelni, vagy a különböző társadalmi csoportokat különbözőképpen kell kezelni az egyenlőtlenségek felszámolása érdekében? Az egyéneket és a csoportokat meg kell védeni a **megkülönböztetéstől** és biztosítani kell az **esélyegyenlőséget minél több területen** (oktatás, javak, technológia ...). Ezenkívül a felhasználókat nem szabad félrevezetni, és a döntéseket átláthatóan kell ismertetni.

### Az emberi autonómia tiszteletben tartása:

Az embereknek továbbra is képesnek kell lenniük arra, hogy teljes mértékben gyakorolják **önrendelkezésüket**, és kiálljanak alapvető jogaikért. Az **MI**-rendszereknek segíteniük és támogatniuk kell az embereket ahelyett, hogy manipulálnák vagy kiszolgáltatnák őket. Fontos, hogy az **MI** működési folyamatai továbbra is **emberi felügyelet** alatt álljanak.

### Védelem az ártalmaktól:

A mesterséges intelligencia nem okozhat kárt, és nem is súlyosbíthatja a meglévő károkat (ez a mentális és fizikai épségre is igaz). A rendszereknek technikailag sziklaszilárdnak kell lenniük, hogy ne lehessen velük visszaélni.

Különösen fontos, hogy tekintettel legyünk a védelemre szorulóakra. Ezen túlmenően főként a hatalom vagy az információ egyenlőtlen elosztására (pl. kormány és állampolgárok) kell érzékenynek lennünk.

### Nyomonkövethetőség:

Ahhoz, hogy egy mesterséges intelligencia rendszer a felhasználói körében bizalmat keltsen, a benne zajló folyamatoknak a lehető legátláthatóbbnak kell lenniük. Sajátosak az úgynevezett **"fekete doboz algoritmusok"**, amelyeknél zavaros, hogy a rendszer hogyan jut el az adott eredményre. Különös figyelmet kell fordítani a nyomon követhetőségre és a rendszerről szóló átlátható kommunikációra.

# Összeférhetetlenség

Sajnos az imént említett elvek nem mindig férnek meg egymás mellett. Például az "ártalomkerülés" és az "emberi autonómia" ellentmondásba kerülhet, ha a mesterséges intelligencia rendszereknek a bűnmegelőzés területén történő felhasználásra gondolunk. Bizonyos speciális megfigyelési módszerek segíthetnek a bűnözés elleni küzdelemben, ugyanakkor korlátozhatják az egyén szabadságát és adatvédelmi jogait.

Az iránymutatások kidolgozásakor figyelembe kell venni különböző nézőpontokat is. A fejlesztők, az ügyfelek és a végfelhasználók mind különböző szempontokkal rendelkeznek, amelyek akár egymásnak is ellentmondhatnak.

A torzító hatások kérdése jelenleg is fontos kutatási téma. Sok torzítási hiba csak akkor mérsékelhető vagy korrigálható, ha már ismert a hatása. Ha aktívan törődünk az ilyen hibák kezelésével és megelőzésével, erősödhet a társadalom mesterséges intelligenciába vetett bizalma.

Hogyan kerülhetők el a torzítási hibák?

**1. Önreflexió** Előfordul-e, hogy sztereotípiában gondolkodom? Lehetséges, hogy a családi vagy kulturális háttérem miatt? A különböző nézőpontok figyelembevétele segíthet a torzítások azonosításában.

**2. Aktív kommunikáció** A tudatosság aktív eszmecsere és a torzítási hibákra való rámutatás révén jön létre.

A mesterséges intelligencia rendszerek etikai irányelveinek kialakításakor többek között a következő követelményeket lehet figyelembe venni:

Alapvető jogok, az emberi cselekvés elsőbbsége, a támadásokkal szembeni ellenállás, megbízhatóság, a magánélet tiszteletben tartása, az adatok minősége, az adatokhoz való hozzáférés biztonsága, átláthatóság / nyomonkövethetőség, nyílt kommunikáció, sokszínűség / megkülönböztetés mentesség, az érdekelt felek részvétele, fenntarthatóság / környezetvédelem, társadalmi hatás, ellenőrizhetőség, a jogi alapok figyelembevétele (nemzeti különbségek).

## Tananyag – magyar nyelven

-  Ethics – Worksheet Bias Error.pdf

Az alábbi feladat egy jól ismert filozófiai gondolkísérletet, a villamos dilemmát mutatja be. Ezt a gyakorlatot offline, nyomtatott változatban, kiscsoportokban is meg lehet beszélni, de lehetőség van arra is, hogy a [www.moralmachine.net](http://www.moralmachine.net) oldalon online járjuk végig az egyes forgatókönyveket. Ennél a feladatnál is biztosítottak a beszélgetéshez/vitához szükséges információk. Ezen kívül érdemes hozzá egyfajta döntési naplót készíteni, hogy kiderüljön, mennyire következetesen születtek a döntések a csoportban, és milyen tényezők voltak meghatározóak egy-egy döntésnél.

## Az erkölcs gép

Ebben a feladatban egy filozófiai gondolkísérlettel foglalkozunk, amelyben komoly döntéseket kell hoznotok! Ebben a gondolkísérletben egy autó fékjei meghibásodnak, és el kell dönteni, hogy az autó hová fog becsapódni. A legtöbb esetben emberéletek között kell választani.

1. Válasszatok valakit a csoportból, aki feljegyzí az általatok hozott döntéseket. Írjátok le, **miért** így döntöttetek. Aktívan beavatkoznátok a történésekbe (vagyis az autó mozgásirányába)?
2. Nyissátok meg a [www.moralmachine.net](http://www.moralmachine.net) weboldalt, és kattintsatok a "Start Judging" ("Mehet a bíraskodás") opcióra."
3. Döntsetek a két lehetőség közül a megfelelő képre kattintva. Az utcán lévő emberekről részletesebb információkat kaphattok, a "Show descriptions" ("Leírás megjelenítése") gombra kattintva.
4. Beszéljétek meg a csoportban, hogy **miért** döntöttetek közösen a választott lehetőség mellett. Jegyezzétek le azt is, ha nem tudtatok azonnal megegyezni.
5. A végén megjelennek az eredményeitek. Milyen tényezők tünnek különösen fontosnak a csoportotokban? Melyek voltak kevésbé fontosak?
6. Beszéljétek meg az eredményeiteket az egész csoporttal! Hogyan döntöttek a többiek?

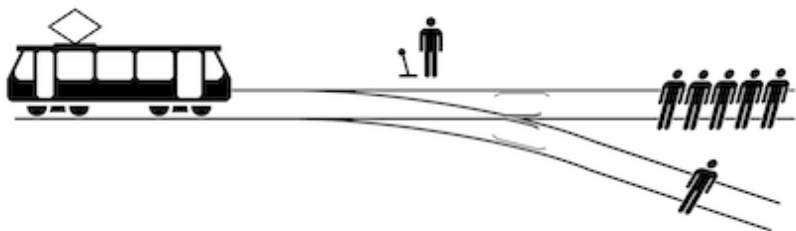
**Adalék:** Németországban az alaptörvény 1. cikke így szól: "Az élet méltósága sérthetetlen". Ausztriában az Általános Polgári Törvénykönyv alkalmazandó: "Minden ember bír vele született, már magából az ésből kivilágló jogokkal, s azért személynek tekintendő"

Hogyan értelmeznéd ezeket a jogszabályokat az előbbi gondolat kísérlet szempontjából?

# A villamos dilemma háttere

Vezető vagy gyerek? Rendőr vagy hajléktalan? Nő vagy férfi? Egy embernek meg kell halnia, egy ember pedig túlélheti a halálos kimenetelű ütközéses balesetet – de ki legyen az?

A **villamos dilemma** egy erkölcsfilozófiai gondolkísérlet, amelyet először az 1930-as években jegyeztek le. A dilemmában egy irányíthatatlanná vált, öt ember felé tartó villamos áterelhető egy másik vágányra a váltó átállításával. Ugyanakkor sajnos a másik vágányon is van egy ember. Ekkor felmerül az erkölcsi kérdés, hogy vajon beavatkozzon-e a váltókezelő – ezzel öt ember életét megmentve, de egy embert feláldozva?



Ebből az erkölcsfilozófiai dilemmából számos olyan kérdés vezethető le, amelyek nemcsak elméleti konstrukciókat, hanem nagyon is konkrét, aktuális problémákat érintenek, mint például az autonóm vezetés. Mekkora döntési jogkörrel rendelkezhet egy gép? Miként programozzák le az etikai döntéseket vészhelyzetekre, és kinek van joga dönteni az emberi életéről (kutatóknak, programozóknak, gyártóknak stb.)? Ki a végső felelős baleset esetén?

Ezek a kérdések megvitathatók a [moralmachine.net](http://moralmachine.net) segítségével a csoportban. A Moral Machine (erkölcsgép) a Max Planck Intézet (MPI) emberi fejlődéssel foglalkozó kutatási projektjének része, amelyben a résztvevőknek többször kell dönteniük két forgatókönyv között. Minden szituációban egy elkerülhetetlen baleset lehetséges következményét mutatják be, ahol a két kimenet közül kell dönteni. A kutatási projekt célja az élet-halál kérdéskörével kapcsolatos erkölcsi dilemmákra vonatkozó egyéni vélekedések vizsgálata. Az eredmények szemléltetése és bemutatása miatt ez a projekt jól alkalmazható iskolai keretek között is.

Miközben a tanulók kis csoportokban dolgoznak az egyes forgatókönyveken, célszerű egyfajta döntési jegyzőkönyvet készíteni, amelyben feljegyzik, hogy az



egyes döntések során mely tényezők voltak a meghatározóak, és hogy a fordulók során változik-e a tényezők preferált kiválasztása. Itt tanárként már figyelhetünk arra is, hogy a diákok úgy döntenek-e, hogy aktívan beavatkoznak az autó mozgásába, hogy minél kevesebb kárt okozzanak (inkább haszonelvű megközelítés), vagy inkább tartózkodnak a döntéstől, mert így, vagy úgy valahány emberek meg kell halnia.

A Német Szövetségi Köztársaság alaptörvényének 1. cikke vagy Ausztriában az 1811. évi általános polgári törvénykönyv 16. szakasza jó kiindulási alap lehet. Mindkettő definíciója szerint az emberi élet sérthetetlen. A jog szerint az emberi élet ellen nem szabad véteni, de az emberi természet egész másként működhet. Hasznos lehet azt javasolni, hogy a diákok képzeljék magukat egy másik ember helyébe. Akár szerepkártyákat is lehet írni (idős hölgy, sebész, terhes nő...), amelyek segítségével a tanulóknak a villamos dilemmához hasonló döntéseket kell hozniuk. Így a diákok könnyebben átérezhetik, hogy mindenkinek joga van az élethez.

A gondolat kísérlet online változatának végén az eredmények vizuálisan újra összegzésre kerülnek egy lista formájában a [moralmachine.net](http://moralmachine.net) oldalon. Itt is újra megvitatható, hogy bizonyos tulajdonságokkal rendelkező embereket miért védelmezzünk jobban, mint másokat.

## Jeremy Bentham vs. Immanuel Kant

A Moral Machine kutatócsoport a két megoldást Jeremy Bentham és Immanuel Kant filozófusokra alapozta. Míg Bentham szerint az autónak haszonelvű módon a lehető legnagyobb szerencsétlenséget kellene elkerülnie, és a kisebb tömegbe hajtva az ő halálukat okozni (beleértve az ártatlan embereket vagy az autósokat), addig Immanuel Kant szerint az autónak az erényetika alapján kellene működnie, melynek révén olyan alapvető erkölcsi elvek érvényesülnének, mint a "Ne ölj". Kant szerint nem szabad olyan cselekedetet végrehajtani, amely kifejezetten emberölésre irányul, ehelyett az előírányzott utat kell választani, még akkor is, ha ennek következtében több ember hal meg. A Moral Machine korábbi publikációiban a résztvevők inkább a Bentham filozófiájával összhangban lévő lehetőségeket választották. Azt gondolhatnánk tehát, hogy a nagyobb rosszat elkerülő haszonelvű megközelítés lenne a legésszerűbb és közösségbarát. Azonban amint a résztvevőket megkérdezték, hogy vennének-e ilyen autót, azonnal visszautasították, tekintettel arra, hogy az a saját életüket is

veszélyeztetheti. Mindenképpen olyan autót akarnak, amely a saját életüket védi, de mindenki másnak olyan autót kellene vennie, amely a lehető legnagyobb kárt kerüli el.

## Tananyag

- 🌐 (DE) Gépek a morális dilemmában (Wiener Zeitung)
- 🌐 (DE) Az autonóm vezetésről Bentham és Kant után szabadon
- 🌐 (DE) Az önvezetés etikai szabályai más kultúrákban (Wiener Zeitung)
- ✎ Ethics – Worksheet Moral Machine.pdf

# Az önvezetésről

A dilemmák sajátossága, hogy a benne érintett személy csak két lehetőség közül választhat, más kiút pedig nincs.

Ezen dilemmahelyzetek megoldásai gyakran vitatottak a tudományban és a filozófiában. A legfőbb felmerülő kérdés az, hogy ki ítélkezhet mások élete felett, vagy egyáltalán jogot formálhat-e erre bárki. Még ha a programozók meg is adnák a programnak az etikailag alapvetően helyes döntést, az akkor is egy kívülről jövő döntés maradna, mely nem egy konkrét helyzetet szemlél annak összes meglévő előnyével és hátrányával együtt, hanem egy elvont alapgondolatból származna. Következésképpen az emberről már nem önmaga, hanem mások rendelkeznének. Ennek következménye több szempontból is problematikus. Egyrészt a "helyes" értékrendet az állam vagy egy vállalat határozná meg, másrészt az embert, mint egyént teljesen figyelmen kívül hagynák.

Az olyan általános érvényű szabályok lefektetése, mint a "személyi sérülések elsődleges védelme a vagyoni károkkal szemben", hasznosnak tűnhet dilemmahelyzetekben, de olyan tényezőket, amelyek egy konkrét esetben viszont sok ember életébe kerülhetnek, nem vesznek figyelembe (pl. egy tartálykocsi szivárgása vagyoni kár következtében vagy az elektromos hálózat összeomlása egy nagyvárosban). Az általános érvényű szabályok lefektetésével az a probléma, hogy a lehetséges helyzetek sokféleségét és összetettségét nem képesek lefedni. Ezért az etikai tanács úgy döntött, hogy olyan megoldásokat választ, amelyek a lehető legnagyobb mértékben csökkenthetik a balesetek számát és technikailag is megvalósíthatóak.

Az emberi élet védelme számít a legfőbb jónak. A potenciális áldozatok számát csökkentő programok írása akkor indokolt, ha a programkód minden egyes közlekedésben résztvevő kitétségét egyenlő mértékben csökkenti. Ez azonban nem jelenti azt, hogy valakinek az életét szembe lehetne állítani más emberek életével. Nem fogadhatjuk el, hogy valaki az életét adja egy veszélyhelyzet során azért, hogy több ember életét mentsük meg, hiszen mindenkinek joga van az élethez. A gyártó azonban nem vonható felelősségre, ha ésszerű kereteken belül mindent megtesz annak érdekében, hogy egy automatizált rendszer a lehető legkevesebb személyi sérülést okozza.

Ugyanez vonatkozik az útitársak védelmére is: az útitársak védelme nem másodlagos, ugyanakkor a járókelők sem áldozhatók fel az útitársak védelme érdekében.

További kérdésekre – pl. az állatok balesetekben való figyelembevételére, vagy a magánadatok felhasználására vonatkozó kérdésekre – az automatizált és hálózatba kapcsolt vezetés etikai bizottságának jelentése nyújt választ.

Az autonóm vezetés példája arra szolgál, hogy az eddig tanultakat kézzelfoghatóbb kontextusba helyezzük. Az alábbi zárófeladatban az autóipar és a politika meglévő megoldási javaslatait kell megvitatni. A tanulók megfigyelhetik, hogy ezek mennyiben felelnek meg az ő nézeteiknek, illetve a tárgyalási álláspontoknak.

## Autonóm vezetés

Ez a feladatlap jó alapot nyújt arra, hogy a következő kérdéseket megvitassuk a tanulókkal:

- Tud-e egy gép különbséget tenni élet és halál között? Ki programozza be ezeket a döntéseket a gépbe (gyártók, programozók, kormány ...)?
- Szerinted kinek kellene a legnagyobb beleszólást biztosítani ezekbe a döntésekbe? A fogyasztókat is be kellene vonni?
- Milyen szerepe lesz az emberek a jövőben az autóvezetésben?
- Te szeretnél-e teljesen automatizált járművet vezetni az utakon? Miért (nem)? Szerinted mi kell ahhoz, hogy az autonóm autókat habozás nélkül használhassuk?
- Az említett irányelvek megvalósíthatóak, vagy adódhatnak problémák? Véleményed szerint milyen egyéb szempontokat lenne még fontos beépíteni az ilyen irányelvekbe? (lásd adatvédelem, állatok a balesetekben ...).

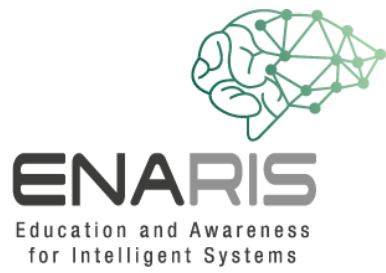
A csoporttól függően ezeket a kérdések **közösen** vagy **kiscsoportokban** is megvitathatók. Rendelkezésre álló időtől függően kis csoportokban, **poszttereket** vagy **PowerPoint diákat** lehet készíteni a kérdések ismertetéséhez.

## Tananyag – magyar nyelven

-  Ethics – Worksheet Autonomous Driving.pdf

## Hivatkozások

1. (DE) Etikai bizottság – Automatizált és hálózatra kapcsolt járművek (14. oldaltól)
2. (DE) Ember az automatizált járműben – Digitális Etika a mindennapokban
3. (DE) BMK – Automatizált járművek



EUROPEAN UNION

