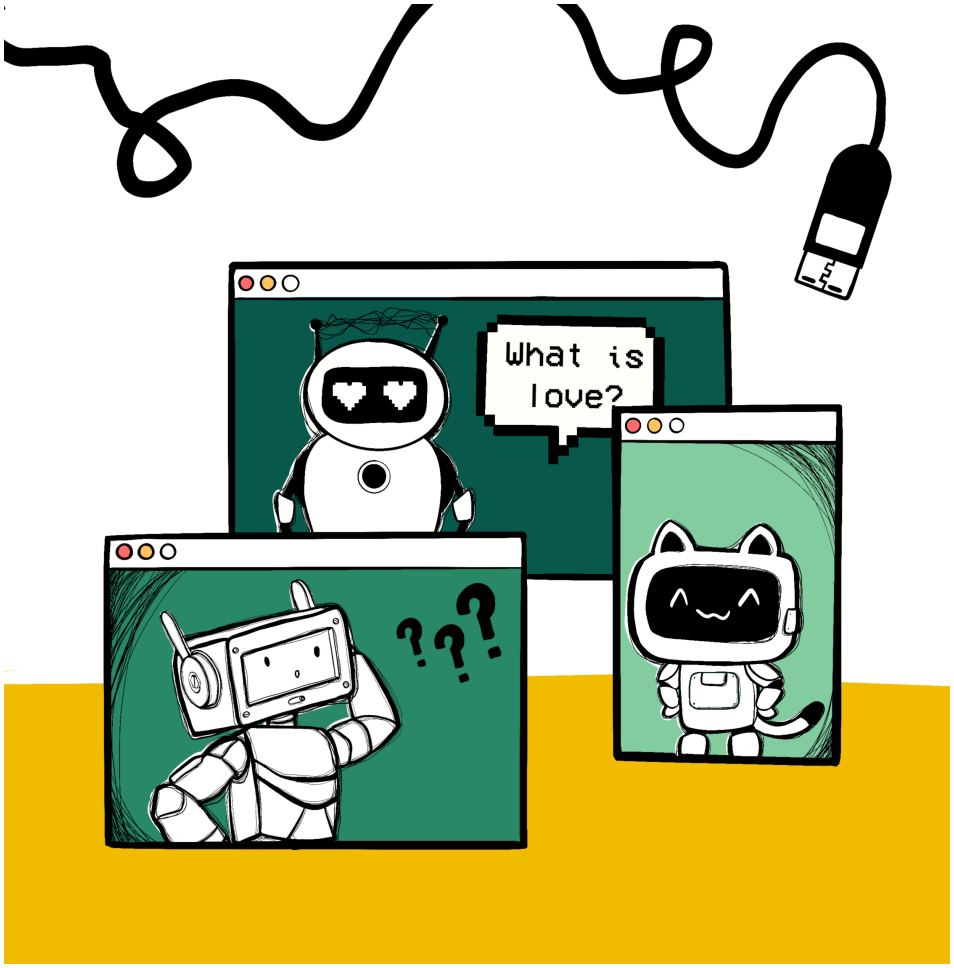
Modul 2

Ethik im Bereich der künstlichen Intelligenz

"***KI*** ist sehr gut darin, die **Welt zu beschreiben** , so wie sie heute ist, mit all ihren Vorurteilen. Aber ***KI*** weiß nicht, **wie die Welt sein sollte.**"



Über das Modul

In diesem Modul sollen grundlegende ethische Aspekte im Bereich der ***KI***-Forschung behandelt werden. Dabei sollen die Schüler*innen, indem sie selbstständig Regelkataloge er- und überarbeiten, lernen, selbstständig kritische Fragen zu stellen und Spannungsverhältnisse zwischen einzelnen ethischen Grundprinzipien aufdecken. Um dieses komplexe Unterfangen möglichst ansprechend zu gestalten, wurden Aufgaben mit unterschiedlichen Zugängen und Komplexitätsgraden zusammengestellt, wodurch der Schwierigkeitsgrad der Lerngruppe entsprechend angepasst werden kann. Bei den Aufgaben handelt es sich um Gedankenexperimente (sowohl altbekannte wie das Trolley-Dilemma, aber auch Übungen für jüngere Lernende wie dem Erstellen von Regeln für einen Roboter-Butler).

Lernziele

Die Schüler*innen können ...

- ... über die Relevanz ethischer Richtlinien in ***KI***-Systemen argumentieren
- ... selbstständig moralische Regeln formulieren und diese mithilfe von Input überarbeiten
- ... in Gruppenarbeit Schlupflöcher in selbstaufgestellten Regeln finden
- ... das Trolley Problem und das Spannungsverhältnis zwischen zweier sich widersprechender philosophischer Richtungen beschreiben
- ... ihren eigenen philosophischen Standpunkt bilden und diesen richtig einordnen
- ... zentrale Punkte der EU Ethik-Richtlinien einordnen und kritisch hinterfragen
- ... persönliche Standpunkte zu ***KI***-Systemen im Alltag (am Beispiel „Autonomes Fahren) darstellen

Agenda

Zeit	Inhalt
20 min	**Aufgabe - Robotergesetze entwickeln**
10 min	**Theorie - Input: Was machen gute ethische Regeln aus?**
15 min	**Aufgabe und Inputtext Bias-Fehler**
20 min	**Aufgabe - MoralMachine**
15 min	**Theorie - das Trolley Problem (Utilitarismus vs. Tugendethik)**

Zeit	Inhalt
15 min	Theorie - Autonomes Fahren
15 min	Aufgabe - Autonomes Fahren

Einleitung

Über dieses Modul - Wozu ein ethischer Zugang in einem Thema der Informatik?

Vielen Menschen haben, wenn sie an künstliche Intelligenz denken, Angst, es könne sich dabei um superintelligente Roboter, die Menschheit unterwerfen wollen, wie in den Matrix- oder Terminator-Filmen, handeln. Doch von solchen Szenarien ist die ***KI***-Forschung meilenweit entfernt. Allerdings beschäftigt sich die Forschung auch heutzutage bereits mit ethischen Grundfragen, denn künstliche Intelligenz macht Fehler. Durch einseitige Datensätze entstehen Bias-Fehler und Programme können dadurch Einzelpersonen und Gruppen diskriminieren. Darum ist es notwendig, auch abseits des Ethik- und Religionsunterrichts eine ethische Sensibilität für schwierige Situationen zu entwickeln, verschiedene Perspektiven kennenzulernen und Fähigkeiten für die Entscheidungsfindung zu erlangen.

Mithilfe der ersten Übung soll ein spielerischer, niederschwelliger Zugang geschaffen werden. In dieser Übung wird ein fiktives Szenario geschildert, in welchem ein humanoider Haushaltsroboter in der Lage ist, alle Aufgaben genauso wie ein Mensch zu bearbeiten. Die Schüler*innen sollen sich hierbei Gedanken machen, welche Möglichkeiten, aber auch welche Problematiken, bei der Verwendung von ***KI***-Systemen im Alltag, möglich scheinen.

Robotergesetze entwickeln

1. Stell dir das folgende Szenario vor...

Stell dir vor, dass du einen Roboter-Butler zu Hause hast. Es ist ein **humanoider (menschenähnlicher) Roboter**, der mechanisch in der Lage ist, alles zu tun, was ein Mensch tun kann. Du bist der Besitzer des Roboters und kannst ihm Befehle erteilen, z. B. Fragen beantworten oder Hausarbeiten erledigen. Die Verhaltensfähigkeiten des Roboters sind nur durch eine Reihe von Regeln begrenzt, die du definieren musst.

Es kann Folie 2 des Unterstützungsmaterials 1 als visuelle Hilfe und Seite 2 des Handout als Schreibvorlage verwendet werden.

2. Die Schüler*innen erfinden "goldene Regeln".

Die erste Aufgabe der Schüler*innen besteht darin, Regeln oder "Gesetze" zu finden, an die sich der Roboter halten muss, damit er sich nicht auf schädliche oder unbeabsichtigte Weise verhält. Es kann für die folgenden Diskussionen von Vorteil sein, wenn die Schüler*innen ihre Ergebnisse schriftlich oder als Mindmap/Poster/Dia festhalten. Diese Übung lässt sich gut in kleinen Gruppen durchführen, so dass die Schüler*innen ihre Ideen in Kleingruppen austauschen können. Alternativ kann auch die Methode "Think-pair-share verwendet werden. Es sollte genügend Zeit zur Verfügung stehen (etwa 10 bis 15 Minuten), damit die Schüler*innen wirklich über einzelne Szenarien nachdenken können, auf die ihre Regeln zutreffen.

Auf Folie 3 im Begleitmaterial1 finden Sie einige Leitfragen, falls die Schüler*innen weitere Anregungen brauchen, bevor sie beginnen (oder ihre Arbeit stagniert).

3. Die Schüler*innen stellen ihre Regeln vor

Jedes Team sollte ein paar Minuten Zeit haben, um seine Regeln vorzustellen und sie mit einfachen Beispielen zu verdeutlichen.

4. Suchen Sie nach Schwachstellen und Schlupflöchern

Nimm eine der vorgestellten Regeln und zeige, wie sie umgangen / fehlinterpretiert werden kann. Auf Folie 4 des Begleitmaterials 1 finden Sie einige Beispiele für Fragen.

5. Die Schüler*innen sollen Wege finden und präsentieren, wie sie die anderen Regeln untergraben würden.

Auch hier eignet sich die Methode "Think-pair-share "sehr gut, ebenso wie die Arbeit in kleinen Gruppen.

6. Material

- ✎ Ethics - Worksheet Robot Law.pdf

Über "Ethik im Bereich der KI”

Wie sollte eine moralische Entscheidung getroffen werden?

Als Hilfestellung für Schüler*innen kann über das generelle Formulieren von moralischen Regeln gesprochen werden. Was muss man überhaupt berücksichtigen, damit ein moralische Prinzip stichfest ist?

Eine moralische Regel bzw. Entscheidung sollte sein:

- **Logisch:** Das Ziel sollte es sein, eine moralische Entscheidung mit Hilfe von Beweisen und logischen Schlüssen zu untermauern und sich nicht auf Gefühle oder soziale oder persönliche Tendenzen zu stützen. Das Bilden logischer moralischer Urteile bedeutet auch, sicherzustellen, dass jedes unserer bestimmten moralischen Urteile auch mit unseren anderen moralischen und nichtmoralischen Überzeugungen vereinbar ist. Unstimmigkeiten müssen auf jeden Fall vermieden werden.

  *Die meisten Philosophen stimmen darin überein, dass wir, wenn wir ein moralisches Urteil fällen, bereit sein müssen, unter ähnlichen Umständen dasselbe Urteil zu fällen. Wenn wir es also zum Beispiel als moralisch falsch sehen, dass Frau Müller Zahlen in einer Forschung zu ihrem Vorteil verändert hat, so ist es auch moralisch falsch, wenn unser*e Kolleg*in, unser*e Ehepartner*in oder unsere Eltern die Zahlen geändert hätten. Genauso dürfen wir auch für uns selbst keine Ausnahme machen, indem wir uns Dinge für uns herauspicken, für die wir andere verurteilen würden.*

- **auf Fakten basierend:** Bevor moralische Entscheidungen getroffen werden können, müssen möglichst viele relevante Informationen gesammelt werden. Wichtig ist hierbei hervorzuheben, dass es sich um relevante Informationen, welche sich auf diesen konkreten Fall beziehen, handeln sollte. Diese Informationen sollten vollständig (oder zumindest alle möglichen, verfügbaren Daten beinhalten) und wahr sein.
- **auf soliden und vertretbaren moralischen Prinzipien basierend:** Unsere Entscheidungen sollten auf allgemein formulierten moralischen Prinzipien beruhen, welche der kritischen Prüfung und rationalem Kritizismus

standhalten. Was nun genau ein moralisches Prinzip stichhaltig oder akzeptabel macht, ist eine der schwierigsten Fragen in der Ethik.

Wenn wir nicht in der Öffentlichkeit über moralische Entscheidungen diskutieren wollen, ist das häufig ein Warnsignal. Wenn wir nicht bereit sind, öffentlich für unser Handeln Rechenschaft abzulegen, besteht die Möglichkeit, dass wir etwas tun, das wir nicht moralisch rechtfertigen können. Hierbei kann Immanuel Kants philosophischer Ansatz hilfreich sein, indem wir dazu bereit sein müssen, dass unsere moralischen Urteile für jede einzelne Person geltend sein können. Man kann nicht ein moralisches Prinzip befürworten, wenn man nicht bereit ist, dieses an die Allgemeinheit anzuwenden.

Darüber hinaus kann es hilfreich sein, ein Problem aus der Sicht einer anderen Person zu betrachten, um dadurch moralische Kurzsicht zu vermeiden (Man kann sich z.B. die Fragen stellen: “Wie würde es dir gefallen, wenn…?”)

Als Hilfestellung, welche Regeln Schüler*innen aufstellen könnten, können beispielsweise **Isaac Asimovs** 1942 postulierte **Robotergesetze** dienen:

1. Ein Roboter darf kein menschliches Wesen (wissentlich) verletzen oder durch Untätigkeit zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.

2. Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen - es sei denn, ein solcher Befehl würde mit Regel eins kollidieren.

3. Ein Roboter muss seine Existenz beschützen, solange dieser Schutz nicht mit Regel eins oder zwei kollidiert.

Ähnliche Ansätze gibt es jedoch auch bereits abseits von Science-Fiction Literatur. Viele Großkonzerne wie z.B. die BMW Group oder Microsoft stellen Ethikprinzipien für künstliche Intelligenz auf.

Noch gibt es keinen verbindlich vorgeschriebenen Standard. Angesichts der rasanten Entwicklungsfortschritte, überrascht es jedoch nicht, dass bereits unzählige Expert*innengruppen allgemeine Regeln zu definieren versuchen.

Von der eher allgemeinen Aufgabe des Aufstellens ethischer Regeln soll in dieser Übung nun der Fokus auf Bias-Fehler in ***KI***-Systemen gelegt werden, wobei ein schüler*innennaher Text über eine durch ***KI*** automatisierte Wahl der weiterführenden Schule als Diskussionsgrundlage dienen soll.

Bias Fehler

Dieses **Arbeitsblatt** kann im Plenum gemeinsam besprochen werden, um sicher zu gehen, dass die Schüler*innen die Problematik dahinter verstehen.

Sowohl Aufgabe 1 als auch Aufgabe 2 eignen sich für Gruppenarbeiten, können jedoch genauso im Plenum besprochen werden. Empfehlenswert ist hierbei, das Besprochene schriftlich festzuhalten, um daran anschließen zu können. Dabei soll den Schüler*innen vor allem klar werden, dass es sich bei Bias-Fehler nur um eingeschleuste Vorurteile in den Datensätzen handelt und nicht um Fakten. Besonders die dadurch entstehende Diskriminierung sollte hierbei als ernstzunehmendes Problem hervorgehoben werden.

Die wichtigsten Aspekte rund um Bias-Fehler können am Ende schriftlich festgehalten werden. Dazu gäbe es die Möglichkeit als Lehrer*innenvortrag mit einer PowerPoint-Präsentation die Inhalte vorzugeben oder, je nach Vorwissen, im Plenum die Inhalte gemeinsam mit den Schüler*innen zu erarbeiten.

Material

- ✎ Ethics - Worksheet Bias Error.pdf

Bias-Fehler und die Ethikrichtlinien für eine vertrauenswürdige KI

Was sind Bias-Fehler überhaupt?

Bei einem Bias kann es sich um eine Verzerrung oder um Voreingenommenheit handeln. Diese entstehen bereits bei der Erhebung der notwendigen Daten, welche später zu fehlerhaften Ergebnissen führen und schwerwiegende Probleme, wie Ausschluss und Diskriminierung mit sich bringen können.

Der Chatbot Tay wurde bspw. 2016 auf Twitter präsentiert und musste innerhalb kürzester Zeit wieder offline gehen, als der Bot begann, aufgrund der vorhandenen Datenbasis unzählige diskriminierende Nachrichten zu veröffentlichen.

Welche Arten von Bias-Fehler gibt es?

- **algorithmic AI bias or "data bias":** Hier wird durch eingespeiste Daten ein Bias-Fehler begangen (statistische Verzerrung der Daten).
- **societal AI bias:** Hierbei handelt es sich um von der Gesellschaft indoktrinierte Normen; aber auch Stereotype erschaffen blinde Flecken oder Voreingenommenheit (gesellschaftliche Vorurteile).

Societal Bias beeinflusst/schafft häufig algorithmische Bias-Fehler!

Eine künstliche Intelligenz ist nur so gut wie ihre zugrundeliegende Datenbasis. Bei Programmierer*innen handelt es sich auch nur um Menschen, welche ebenso in ihrer Blase, mit eigenen Ansichten und Vorurteilen stecken. Problematisch wird dies, wenn diese Vorurteile in Form von ausgewählten Datensätzen auch in ihre Programme fließen (**„Garbage in, garbage out."**). Daraus ergeben sich neue Anforderungen in der ***KI***-Forschung. Neben Programmierkenntnissen ist auch ein ethisches Grundverständnis notwendig, um abwägen zu können, ob sie mit ihren Umsetzungen dem Allgemeinwohl helfen oder schaden.

Aus diesem Grund ist die Erstellung von ethisch-moralischen Prinzipien notwendig, damit keine Person durch eine künstliche Intelligenz ausgeschlossen werden kann.

Eine Vielzahl an Expert*innengruppen haben bereits versucht, Regelkataloge zu erstellen: so unter anderem auch jene der europäischen Kommission, welche die

Ethikrichtlinien für eine vertrauenswürdige *KI* erstellten. Im Folgenden orientieren wir uns an den dort erläuterten vier ethischen Grundsätzen, welche als Fundament für vertrauenswürdige ***KI***s dienen sollen .

Vier ethische Grundsätze:

Fairness:

Werden durch die eingespeisten Daten Vorurteile miteingebaut, handelt ein Programm nicht fair. Dabei ist der Begriff „Fairness“ nicht so leicht zu definieren: Sollen bei einer Gruppe Menschen nun alle exakt gleich behandelt werden oder sollen unterschiedliche gesellschaftliche Gruppen unterschiedlich behandelt werden, um Ungleichheiten auszumerzen? Personen und Gruppen müssen vor **Diskriminierung** geschützt werden und eine **Chancengleichheit** (zu Bildung, Gütern, Technologie…) muss gewährleistet werden. Nutzer*innen dürfen zudem nicht getäuscht werden und Entscheidungen sollten transparent dargelegt werden.

Achtung der menschlichen Autonomie:

Menschen sollte es weiterhin möglich sein, ihre **Selbstbestimmung** in vollem Umfang ausüben und auf ihre Grundrechte bestehen zu können. ***KI***-Systeme sollten Menschen stärken und fördern, statt sie zu manipulieren oder unterzuordnen. Wichtig ist hierbei, dass die Arbeitsprozesse der ***KI*** weiterhin durch **menschliche Aufsicht** kontrolliert werden.

Schadensverhütung:

Künstliche Intelligenz darf weder Schäden verursachen noch diese verschlimmern (dazu zählt sowohl die geistige als auch die körperliche Unversehrtheit). Die Systeme müssen technisch robust sein, sodass sie nicht für Missbrauch anfällig sind.

Besonders wichtig ist hierbei, die Rücksicht auf schutzbedürftige Personen. Zudem sollte ein Hauptaugenmerk auf ungleiche Macht- oder Informationsverteilungen gelegt werden (bspw. Regierung und Bürger*innen).

Nachvollziehbarkeit:

Um bei den Benutzer*innen des ***KI***-Systems Vertrauen zu schaffen, müssen die Prozesse möglichst transparent sein. Eine Sonderstellung nehmen sogenannte „Blackbox-Algorithmen“ ein, bei denen es nicht ganz klar ist, wie ein System zum jeweiligen Ergebnis kommt. Hierbei muss besonders auf eine

Rückverfolgbarkeit und transparente Kommunikation über das System geachtet werden.

Spannungsfelder

Leider lassen sich die soeben genannten Grundsätze nicht immer vereinen. So können bspw. „Schadensverhütung“ und „menschliche Autonomie“ im Konflikt stehen, wenn man den Einsatz von ***KI***-Systemen im Bereich „vorausschauende Polizeiarbeit“ betrachtet. Spezielle Überwachungsmaßnahmen können dann zwar bei der Verbrechensbekämpfung helfen, schränken dabei aber zugleich die eigenen Freiheits- und Datenschutzrechte ein.

Auch müssen unterschiedliche Perspektiven bei der Erstellung von Richtlinien berücksichtig werden. Entwickler*innen, Auftraggeber*innen und Endnutzer*innen werden allesamt unterschiedliche Standpunkte vertreten, die einander sogar widersprechen können.

Bias-Effekte sind ein wichtiges, aktuelles Forschungsthema. Viele Bias-Fehler können erst abgeschwächt oder behoben werden, wenn der Effekt bekannt ist. Durch ein aktives Auseinandersetzen und Verhindern von Bias-Fehlern kann das Vertrauen der Gesellschaft in künstliche Intelligenz gestärkt werden.

Wie können Bias-Fehler vermieden werden?

1. Selbstreflexion Ertappe ich mich bei Schubladendenken? Liegt es eventuell an einer familiären oder kulturellen Prägung? Das Einnehmen unterschiedlicher Perspektiven kann helfen, auf Verzerrungen aufmerksam zu werden.

2. Aktive Kommunikation Eine Schaffung von Bewusstsein gelingt durch einen aktiven Austausch und ein Hinweisen auf Bias-Fehler

Folgende Anforderungen können unter anderem bei der Erstellung von ethischen Richtlinien in ***KI***-Systemen berücksichtigt werden:

Grundrechte, Vorrang menschlichen Handelns, Widerstandsfähigkeit gegen Angriffe, Zuverlässigkeit, Achtung der Privatsphäre, Qualität der Daten, Sicherheit bei Datenzugriff, Transparenz/Nachverfolgbarkeit, offene Kommunikation, Vielfalt/Nichtdiskriminierung, Beteiligung der Interessensträger,

Nachhaltigkeit/Umweltschutz, soziale Auswirkungen, Nachprüfbarkeit, Einbezug rechtlicher Grundlagen (nationale Differenzen).

Material

- ✎ Ethics - Worksheet Bias Error.pdf

In dieser Übung wird das altbekannte philosophische Gedankenexperiment, das Trolley-Dilemma, präsentiert. Diese Übung kann entweder offline als Print-Version in Kleingruppen besprochen werden oder es gäbe die Möglichkeit, über **www.moralmachine.net** Szenerien online durchzugehen. Auch für diese Übung Input für eine Diskussion zur Verfügung gestellt. Darüber hinaus lohnt es sich bei dieser Übung eine Art Entscheidungsprotokoll anzufertigen, um herauszufinden, wie konsistent die Entscheidungen in der Gruppe getroffen wurden und welche Faktoren für eine Entscheidung ausschlaggebend waren.

Moral Machine

In dieser Übung setzt ihr euch mit einem philosophischen Gedankenexperiment auseinander, in dem ihr schwerwiegende Entscheidungen treffen müsst! In diesem Gedankenexperiment fallen die Bremsen eines Autos aus und man muss sich entscheiden, wo das Auto aufprallen soll. In den meisten Fällen muss man sich zwischen mehreren Menschenleben entscheiden.

1. Wählt in der Gruppe eine Person aus, die mitschreibt, welche Entscheidungen ihr trefft. Schreibt dabei auf, **warum** ihr euch so entscheidet. Greift ihr dabei aktiv in das Geschehen (das fahrende Auto) ein?
2. Öffnet die Website **www.moralmachine.net/hl/de** und klickt auf "Beurteilung beginnen"
3. Entscheidet euch für eine der beiden Möglichkeiten, indem ihr auf das entsprechende Bild klickt. Ihr erhaltet genauere Informationen über die Personen auf der Straße, wenn ihr auf "Beschreibungen einblenden" klickt.
4. Besprecht in der Gruppe, **warum** ihr euch zusammen für eine Möglichkeit entscheidet. Schreibt auch auf, wenn ihr euch nicht sofort einigen konntet.
5. Am Ende werden eure Egebnisse angezeigt. Welche Faktoren waren anscheinend in eurer Gruppe besonders wichtig? Welche waren weniger wichtig?
6. Besprecht eure Ergebnisse in der Klasse! Wie haben sich die anderen entschieden?

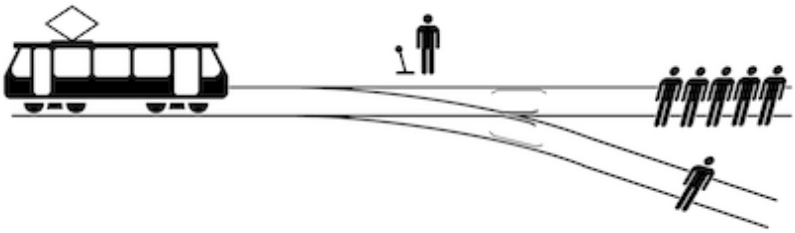
Zusatz: In Deutschland lautet der Artikel 1 der Grundgesetzes: *"Die Würde des Lebens ist unantastbar".* In Österreich gilt das Allgemeinen Bürgerlichen Gesetzbuch: *"Jeder Mensch hat angeborene, schon durch die Vernunft einleuchtende Rechte und ist daher als eine Person zu betrachten."*

Was könnten diese Gesetzesausschnitte im Zusammenhang mit diesem Gedankenexperiment bedeuten?

Über das Trolley-Problem

Fahrer*in oder Kind? Polizist*in oder Obdachlose*r? Frau oder Mann? Eine Person muss sterben, eine Person kann einen fatalen Aufprallunfall überleben - nur wer?

Beim **Trolley Problem** handelt sich um ein moralphilosophisches Gedanken-experiment, das erstmals in den 1930er Jahren beschrieben wurde. Bei diesem Dilemma kann durch das Umstellen einer Weiche eine außer Kontrolle geratene Straßenbahn, welche auf fünf Personen zurollt, auf ein anderes Gleis gelenkt werden. Unglücklicherweise befindet sich auf dem anderen Gleis eine weitere Person. Nun stellt sich die moralische Frage, ob durch das aktive Eingriffen eines Weichenstellers der Tod einer Person in Kauf genommen werden darf, um das Leben der fünf anderen Personen zu retten.



Aus diesem moralphilosophischen Dilemma lassen sich eine Vielzahl an Fragen ableiten, die nicht nur theoretische Konstrukte betreffen, sondern auch ganz konkrete, gegenwärtige Probleme betreffen, wie beispielsweise das autonome Fahren. Denn: Wie viel Entscheidungsmacht darf eine Maschine haben? Wie werden ethische Entscheidungen für den Notfall programmiert und wer soll das Recht haben, über Menschenleben zu entscheiden (Forscher*innen, Programmierer*innen, Industrielle, etc.)? Wer haftet schlussendlich bei einem Unfall?

Diese Fragen können mithilfe von moralmachine.net im Unterricht besprochen werden. Die Moral Machine ist Teil eines Forschungsprojekts des Max-Planck-Instituts (MPI) für Bildungsforschung, in dem Teilnehmende mehrmals zwischen zwei Szenarien entscheiden müssen. In jedem Szenario wird eine mögliche Folge eines unvermeidlichen Unfalls dargestellt und man muss sich für eines der beiden Ergebnisse entscheiden. Das Ziel des Forschungsprojekts ist es, Meinungen von Einzelpersonen in Bezug auf moralische Dilemmas zu verstehen, bei denen es um Leben und Tod geht. Aufgrund der Visualisierung und der Darstellung der Ergebnisse eignet sich dieses Projekt auch gut für den Unterricht.

Während die Schüler*innen in Kleingruppen die einzelnen Szenarien bearbeiten, ist es empfehlenswert eine Art Entscheidungsprotokoll anzufertigen, in dem sie aufschreiben, welche Faktoren bei den jeweiligen Entscheidungen ausschlaggebend waren und ob sich im Laufe der Runden die präferierte Wahl der Faktoren verändert. Hier kann als Lehrperson auch bereits darauf geachtet werden, ob die Schüler*innen sich für einen aktiven Eingriff in das fahrenden Auto entscheiden, um damit möglichst wenig Schaden anzurichten (ein eher utilitaristischer Zugang) oder ob sie sich der Entscheidung eher enthalten möchten, da so oder so Menschen sterben müssen.

Als Input können hierbei Artikel 1 des Grundgesetzes der Bundesrepublik Deutschland bzw. Paragraf 16 des Allgemeinen Bürgerlichen Gesetzbuches aus dem Jahr 1811 in Österreich dienen. In beiden wird die menschliche Leben als unantastbar definiert. Laut dem Gesetz sind Menschenleben also nicht abwegbar, in der menschlichen Natur sieht das jedoch wiederum ganz anders aus. Hilfreich kann es hierbei sein, den Schüler*innen vorzuschlagen, sich in die Rolle einer anderen Person zu versetzen. Dabei können zum Beispiel auch Rollenkärtchen geschrieben werden (ältere Dame, ein Chirurg, eine schwangere Frau…) mit Hilfe derer die Schüler*innen dann dem Trolley-Experiment ähnliche Entscheidungen treffen müssen. So können die Schüler*innen leichter empathisch nachvollziehen, dass ein jeder Mensch das Recht auf Leben hat.

Am Ende der Online-Version dieses Gedankenexperiments werden auf moralmachine.net die Ergebnisse visuell in einer Liste nochmal zusammengefasst. Auch hier kann nochmal darüber gesprochen werden, warum man Menschen mit gewissen Eigenschaften eher rettet als andere.

Jeremy Bentham vs. Immanuel Kant

Das Forschungsteam der Moral Machine orientierte sich bei den beiden zur Verfügung gestellten Lösungswegen an den Philosophen Jeremy Bentham und Immanuel Kant. Während nach Bentham das Auto in utilitaristischer Manier das größtmögliche Unheil vermeiden soll und in die kleinere Menschenmenge fahren und diese töten darf (auch Unschuldige oder den fahrenden Menschen), soll das Auto nach Immanuel Kant der Tugendethik und somit moralischen Grundprinzipien wie bspw. *“Du sollst nicht töten”* folgen. Nach Kant dürfte man also keine Handlung setzen, die explizit einen Menschen tötet, sondern soll den vorgesehenen Weg wählen, auch wenn dadurch mehrere Menschen getötet

werden. In den bisherigen Publikationen zur Moral Machine wählten Teilnehmende eher Optionen aus, die mit Benthams Philosophie übereinstimmen. Man könnte also meinen, dass eine utilitaristische Herangehensweise, die das größte Übel vermeidet, am vernünftigsten und im Sinne der Gemeinschaft sei. Sobald man jedoch die Teilnehmer*innen fragte, ob sie ein solches Auto auch kaufen würden, verneinten sie nachdrücklich, da damit auch ihr eigenes Leben gefährdet werden könnte. Man wolle Autos, die das eigene Leben mit allen Mitteln schützen, aber alle anderen sollen hingegen Autos kaufen, die den größtmöglichen Schaden vermeiden.

Material zum Downloaden und weiterführende Links:

- 🌐 Maschinen im moralischen Dilemma (Wiener Zeitung)
- 🌐 Über autonomes Fahren nach Bentham und Kant
- 🌐 Ethische Regeln beim Autonomen Fahren in anderen Kulturen (Wiener Zeitung)
- ✏ Ethics - Worksheet Moral Machine.pdf

Über das autonome Fahren

Dilemmata zeichnen sich dadurch aus, dass man als Betroffener nur zwischen zwei Übeln notwendigerweise wählen kann und es keinen anderen Ausweg dafür gibt.

In der Wissenschaft und in der Philosophie werden immer wieder Lösungen für diese Dilemma-Situationen diskutiert. Dabei stellt sich allem voran die Frage, wer und ob einzelne Personen indirekt über das Leben anderer richten dürfen. Auch wenn die Programmierer*innen hierbei die grundsätzlich ethisch richtige Entscheidung einprogrammieren, bleibt es dennoch eine Fremdentscheidung, die nicht eine konkrete Situation mit allen gegebenen Vor- und Nachteilen erfasst, sondern sich nur an einer abstrakten Grundidee ableiten lässt. Dadurch wäre der Mensch nicht mehr selbst- sondern fremdbestimmt. Diese daraus resultierende Konsequenz ist in vielerlei Hinsicht problematisch. Einerseits können hier die „richtigen" Wertbilder durch den Staat oder Unternehmen vorgegeben werden, andererseits würde der Mensch als Individuum komplett außer Acht gelassen werden.

Das Aufstellen von allgemeinen Regeln wie z.B. „Personenschaden vor Sachschaden" erscheinen zwar bei Dilemma-Situationen hilfreich, jedoch werden dabei Faktoren in konkreten Fällen nicht berücksichtigt, welche wiederum das Leben vieler Menschen kosten könnten (z.B. das Auslaufen eines Tanklasters in Folge eines Sachschadens oder der Zusammenbruch des Stromnetzes in einer Großstadt). Mit dem Aufstellen von allgemeinen Regeln tritt das Problem auf, dass die Vielfalt und Komplexität der verschiedenen denkbaren Situationen nicht abdeckt werden. Aus diesem Grund entschied sich die Ethik-Kommission dazu, Lösungswege einzuschlagen, die die größten Potenziale der Unfallverringerung bieten und technisch machbar sind.

Als höchstes Gut zählt hierbei der Schutz des menschlichen Lebens. Eine Programmierung auf die Minimierung der Opfer kann dann gerechtfertigt werden, wenn die Programmierung das Risiko eines jeden einzelnen Verkehrsteilnehmers gleichermaßen reduziert. Das bedeutet jedoch nicht, dass man das Leben von Menschen mit denen von anderen aufrechnen kann. Man dürfte auch nicht im Notstand den Tod einer Person in Kauf nehmen, um mehrere zu retten, da jedes Wesen das Recht auf Leben hat. Der Hersteller ist jedoch nicht haftbar, wenn er bei

einem automatisierten System alles Zumutbare tut, um Personenschäden möglichst zu minimieren.

Auch bezüglich des Selbstschutzes der mitfahrenden Individuen gilt dasselbe: der Selbstschutz der Mitreisenden ist nicht nachrangig, jedoch dürfen auch nicht zu Gunsten der Mitfahrenden Unbeteiligte geopfert werden.

Weitere Fragestellungen z.B. bezüglich der Berücksichtigung von Tieren bei Unfällen und der Verwertung von privaten Daten werden im Bericht der Ethik-Kommission für automatisiertes und vernetztes Fahren beantwortet.

Material

- ✎ Ethics - Worksheet Autonomous Driving.pdf

Durch das Beispiel des autonomen Fahrens soll das bisher Gelernte nochmal in einen konkreteren Kontext gebracht werden. In dieser Abschlussübung soll darum auf bereits existierende Lösungsvorschläge aus der Autoindustrie und der Politik eingegangen werden und diskutiert werden, inwieweit diese mit den Ansichten der Schüler*innen und besprochenen Grundpositionen übereinstimmen.

Autonomes Fahren

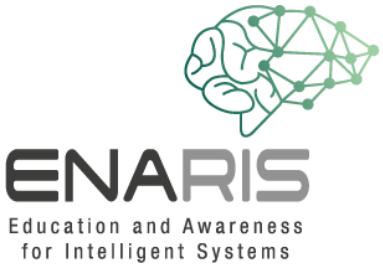
Dieses Arbeitsblatt dient als gute Grundlange um mit Schüler*innen folgende Fragen auszudiskutieren:

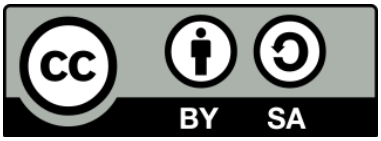
- Darf eine Maschine über Leben und Tod entscheiden? Wer programmiert diese Entscheidungen der Maschine ein (Hersteller*innen, Programmierer*innen, Regierung…)?
- Wer sollte eurer Meinung nach am meisten Mitspracherecht haben bei diesen Entscheidungen? Sollen Konsument*innen miteinbezogen werden?
- Welche Rolle spielt der Mensch zukünftig noch beim Autofahren?
- Würdest du mit einem vollautomatisierten Fahrzeug im Straßenverkehr fahren wollen? Warum (nicht)? Was braucht es deiner Meinung nach, damit man autonome Autos bedenkenlos nutzen kann?
- Sind die hier gewählten Richtlinien auch so umsetzbar oder könnte es Probleme geben? Welche Aspekte wären eurer Meinung nach noch wichtig, um sie in derartigen Richtlinien festzuhalten? (vgl. Datenschutz, Tiere in Unfällen…)

Diese Fragen können je nach Klasse im **Plenum** oder in **Kleingruppen** besprochen werden. Um die Fragen in der Klasse später entsprechend zu präsentieren, können je nach verfügbarer Zeit zur Sicherung auch **Plakate oder PowerPoint-Folien** in Kleingruppen erstellt werden.

Referenzen

1. Ethik-Kommission - Automatisiertes und Vernetztes Fahren (ab Seite 14)
2. Der Mensch im automatisierten Fahrzeug - Digitale Ethik im Alltag
3. BMK - Automatisiertes Fahren


ENARIS
Education and Awareness
for Intelligent Systems


BY SA


EUROPEAN UNION


Interreg
Austria-Hungary 2014-2020
European Union – European Regional Development Fund


TU
Graz


mobilis
interaktív kiállítási központ


OESTERREICHISCHE
COMPUTER GESELLSCHAFT
AUSTRIAN
COMPUTER SOCIETY