

Abgabe-Links

PDF-Präsentation (Beamer):

Präsentation als PDF

Link:

<https://drive.google.com/file/d/1Q7HzbsbAzQ0AW3bjNnzEqAKraJr2kG97/view?usp=sharing>

Gesamtes Abgabeverzeichnis:

Gesamtes Abgabeverzeichnis ansehen

Link:

<https://drive.google.com/drive/folders/1ePAF0ywCyLAlS6XrbKqRFH9p0aN9U19R?usp=sharing>

Ethische und moralische Probleme in der Künstlichen Intelligenz / EU AI-Act

Manizheh Hosseini, Zoha Elahi Awan

24.01.2025

Contents

Contents

1	Definition von KI	1
2	Die Rolle der Künstlichen Intelligenz	1
3	Ethik und Künstliche Intelligenz	1
3.1	Definition von Ethik	2
3.2	Künstliche Intelligenz (KI)	2
3.3	Die Veränderung der menschlichen Praxis durch KI	2
3.4	Die Rolle der Ethik in der KI-Entwicklung	2
3.5	Vorteile von KI	2
3.6	Risiken von KI	2
3.7	Ethische Fragestellungen im Zusammenhang mit KI	3
4	Der zentrale Unterschied zwischen Mensch und Maschine	3
4.0.1	Fazit	3
5	Diskriminierung	3
5.1	Ursachen und Ausprägungen der Diskriminierung durch Künstliche Intelligenz	4
5.2	Herausforderungen bei der Vermeidung von Diskriminierung durch KI . . .	5
5.3	Lösungsansätze zur Vermeidung von Diskriminierung durch Künstliche Intelligenz	5
6	KI in der Rekrutierung	6
6.1	Die vier Schritte	6
6.2	CRISP-Modell	6
7	Das Trolley-Problem als ethische Herausforderung für KI	7
8	KI in der Bildung	8
8.1	Handlungsempfehlungen	9
9	Der EU AI Act: Regulierung von KI-Systemen	9
9.1	Hochrisiko-KI-Systeme	9
9.2	KI mit begrenztem Risiko	10
9.3	KI mit minimalem Risiko	10
9.3.1	Fazit	10

1 Definition von KI

Künstliche Intelligenz (KI) ist ein Bereich der Informatik, der Systeme entwickelt, die Aufgaben ausführen können, die menschliche Intelligenz erfordern, wie logisches Denken, Problemlösung, Lernen, Sprachverarbeitung und Entscheidungsfindung. Diese Systeme analysieren Umweltdaten, treffen basierend darauf autonome Entscheidungen und verfolgen ein spezifisches Ziel.

KI nutzt Sensoren wie Kameras zur Datenerfassung, kombiniert mit Algorithmen zur Datenverarbeitung. Durch Fortschritte in Rechenleistung und Datenverfügbarkeit hat sich KI zu einem zentralen Motor der Digitalisierung entwickelt. Sie findet Anwendung in Bereichen wie Gesundheitswesen, Justiz und Verkehr, wirft jedoch auch ethische Fragen auf, die einen verantwortungsvollen Umgang erfordern.

2 Die Rolle der Künstlichen Intelligenz

Künstliche Intelligenz (KI) spielt eine zunehmend zentrale Rolle in der Optimierung von Prozessen, der Unterstützung von Entscheidungsfindungen und der Förderung von Innovationen in diversen Anwendungsbereichen wie Medizin, Mobilität und Industrie. Sie ermöglicht eine personalisierte Bereitstellung von Inhalten, die Automatisierung von Routineaufgaben sowie eine präzise Datenanalyse, wodurch sowohl wirtschaftliche als auch wissenschaftliche Fortschritte erzielt werden.

Prominente Beispiele für die Vielseitigkeit und den Einfluss von KI sind ChatGPT, das natürliche Sprachverarbeitung revolutioniert, DALL·E, das kreative Prozesse durch Bildgenerierung unterstützt, und AlphaFold, das die Proteinstrukturanalyse in der biomedizinischen Forschung maßgeblich vorangetrieben hat.

Trotz der vielfältigen Potenziale gehen mit der Nutzung von KI auch erhebliche Herausforderungen und Risiken einher. Insbesondere werfen KI-Systeme zentrale ethische Fragestellungen auf, etwa im Hinblick auf Datenschutz, Transparenz und Diskriminierung. Zudem könnte die zunehmende Automatisierung dazu führen, dass bestimmte Arbeitsplätze ersetzt werden, was weitreichende gesellschaftliche und wirtschaftliche Folgen nach sich ziehen könnte. Gleichzeitig besteht die Gefahr, dass eine verstärkte Abhängigkeit von KI-Systemen zu einer Beeinträchtigung kreativer und kognitiver Fähigkeiten führt. Ein weiteres Risiko ergibt sich aus der ungleichen Verfügbarkeit von KI-Technologien, wodurch bestehende soziale Ungleichheiten potenziell verstärkt werden könnten.

Die Rolle der Künstlichen Intelligenz erfordert daher eine differenzierte Betrachtung, bei der sowohl die technologischen Fortschritte als auch die damit verbundenen gesellschaftlichen Auswirkungen kritisch reflektiert werden. [10]

3 Ethik und Künstliche Intelligenz

Die Fragestellung, die diesem Beitrag zugrunde liegt, lautet: Was unterscheidet den Menschen von der Maschine? Diese Frage ist insbesondere im Kontext der Künstlichen Intelligenz (KI) und der Ethik von Relevanz. Um diese Thematik zu analysieren, ist es notwendig, zunächst grundlegende Begriffe zu klären: Was versteht man unter Ethik? Was ist Künstliche Intelligenz? Und warum ist Ethik für KI von Bedeutung?

3.1 Definition von Ethik

Ethik ist die Wissenschaft, die sich mit den Kriterien des richtigen, gerechten, guten oder wünschenswerten Handelns befasst. Ihr Fokus liegt nicht auf Wissen oder Denken, sondern auf der menschlichen Praxis und dem Handeln. Gegenstand der Ethik ist demnach die individuelle und kollektive Praxis des Menschen.

3.2 Künstliche Intelligenz (KI)

KI bezeichnet die Nachbildung kognitiver Fähigkeiten durch Maschinen, insbesondere durch algorithmische Prozesse. KI-Systeme können Daten analysieren, Muster erkennen und Vorhersagen treffen. Dabei handelt es sich jedoch nicht um bewusstes oder intentionelles Handeln, sondern um algorithmisch gesteuerte Prozesse.

3.3 Die Veränderung der menschlichen Praxis durch KI

KI beeinflusst zahlreiche Lebensbereiche und wirkt sich direkt auf die menschliche Praxis aus. Drei zentrale Anwendungsfelder verdeutlichen dies:

- **Algorithmische Vorschlagssysteme:**
Streamingdienste und E-Commerce-Plattformen nutzen KI-gestützte Empfehlungssysteme, die Nutzerverhalten analysieren und präferenzbasierte Vorschläge generieren.
- **Entscheidungssysteme:**
Im Bereich Human Resources werden Bewerbungen zunehmend durch KI-gestützte Systeme vorsortiert, wodurch maschinelle Entscheidungsprozesse Einfluss auf Karrierechancen nehmen.
- **Robotik:**
Fortschritte in der Robotik führen zur Entwicklung KI-gestützter Assistenzsysteme, die beispielsweise in der Pflege oder im Haushalt Anwendung finden [7].

3.4 Die Rolle der Ethik in der KI-Entwicklung

Die Ethik reflektiert diese technologischen Entwicklungen kritisch und untersucht, ob sie wünschenswert, gerecht oder moralisch vertretbar sind. Dabei ergeben sich sowohl Vorteile als auch Risiken.

3.5 Vorteile von KI

- **Effizienz:** KI kann schneller und leistungstärker arbeiten als Menschen.
- **Neutralität:** KI kann potenziell objektivere Entscheidungen treffen, da sie keine subjektiven oder unbewussten Vorurteile besitzt.

3.6 Risiken von KI

- **Fehleranfälligkeit:** Algorithmen können intransparent oder diskriminierend sein.
- **Ethische Herausforderungen:** Entscheidungen können moralisch fragwürdig sein, insbesondere wenn sie tiefgreifende soziale Auswirkungen haben.

3.7 Ethische Fragestellungen im Zusammenhang mit KI

Um KI ethisch zu bewerten, sind folgende Fragen relevant:

- **Fördert KI eine gerechtere Gesellschaft?** Wenn KI soziale Gerechtigkeit unterstützt, kann sie als positiv erachtet werden. Andernfalls bedarf sie kritischer Überprüfung oder regulatorischer Einschränkungen.
- **Unterstützt KI die menschliche Autonomie?** Beispielsweise können KI-gestützte Pflegeroboter dazu beitragen, dass ältere Menschen länger eigenständig leben können.
- **Trägt KI zu einem guten und wünschenswerten Leben bei?** Wenn KI monotone Aufgaben übernimmt, kann dies die Lebensqualität steigern [7, 5].

4 Der zentrale Unterschied zwischen Mensch und Maschine

Ein grundlegender Unterschied zwischen Mensch und Maschine lässt sich anhand der Konzepte der Humanität und der Algorithmizität verdeutlichen:

- **Humanität:** Menschen verfügen über Bewusstsein, mentale Zustände, Intentionalität und kognitive Fähigkeiten.
- **Algorithmizität:** KI ist eine Software, die mathematische Funktionen berechnet und durch festgelegte Algorithmen gesteuert wird.

4.0.1 Fazit

KI besitzt weder Bewusstsein noch Entscheidungsfreiheit, sondern operiert strikt algorithmisch. Diese grundlegenden Unterschiede helfen Ethikern, Verantwortlichkeiten zu klären und Risiken zu bewerten, beispielsweise im Zusammenhang mit diskriminierenden Algorithmen oder moralisch problematischen Entscheidungen. Die ethische Verantwortung liegt primär bei der Aufbereitung und Bereitstellung der Daten sowie bei den Entwicklern und Institutionen, die KI einsetzen. Eine kritische Reflexion ist notwendig, um sicherzustellen, dass KI-Technologien im Sinne der menschlichen Werte und Normen entwickelt und genutzt werden [7].

5 Diskriminierung

Ein zentrales ethisches Problem im Zusammenhang mit Künstlicher Intelligenz (KI) ist die Gefahr der Diskriminierung, die sowohl durch Verzerrungen in den zugrunde liegenden Daten als auch durch technische Designentscheidungen entstehen kann.

Der Einsatz von KI erstreckt sich mittlerweile auf hochsensible Bereiche wie das Justizwesen, das Gesundheitswesen, Überwachungssysteme und den Arbeitsmarkt. Obwohl KI-Systeme oft als objektiv wahrgenommen werden, reproduzieren sie in vielen Fällen bestehende gesellschaftliche Strukturen und Vorurteile.

Diese Problematik tritt insbesondere dann auf, wenn die verwendeten Trainingsdaten bereits diskriminierende Muster enthalten. Infolgedessen können KI-generierte Entscheidungen systematische Verzerrungen aufweisen, die zu Benachteiligungen bestimmter sozialer oder ethnischer Gruppen führen und bestehende Ungleichheiten weiter verstärken.

Die Gefahr algorithmischer Diskriminierung unterstreicht die Notwendigkeit eines verantwortungsvollen und transparenten Umgangs mit KI-Technologien. Um eine faire und gerechte Entscheidungsfindung zu gewährleisten, sind strenge ethische Richtlinien, eine kritische Überprüfung der verwendeten Datensätze sowie eine kontinuierliche Evaluierung der KI-Modelle unerlässlich. Nur durch eine konsequente Einhaltung dieser Prinzipien kann das Vertrauen in KI-Systeme gestärkt und ihr Einsatz gesellschaftlich legitimiert werden. [1]

5.1 Ursachen und Ausprägungen der Diskriminierung durch Künstliche Intelligenz

Die Diskriminierung durch Künstliche Intelligenz (KI) kann auf verschiedene Ursachen zurückgeführt werden. Ein zentraler Faktor ist der Bias in den Daten, also Verzerrungen innerhalb der Trainingsdaten, die sowohl bewusst als auch unbewusst entstehen können. Beispielsweise übernehmen algorithmische Entscheidungssysteme für Bewerbungsverfahren häufig historische Vorurteile aus traditionell männlich dominierten Belegschaften, was zu einer strukturellen Benachteiligung bestimmter Gruppen führen kann.

Zudem können technische Verzerrungen auftreten, die aus fehlerhaften technischen Implementierungen resultieren. Ein Beispiel hierfür sind Sensortechnologien, die bestimmte Hauttöne nicht adäquat erfassen und dadurch ungleiche Ergebnisse für verschiedene Personengruppen erzeugen.

Eine weitere Form der Diskriminierung stellt die emergente Diskriminierung dar, die durch die Interaktion der KI mit ihrer Umgebung entsteht. Ein typisches Beispiel ist ein Chatbot, der diskriminierende Aussagen reproduziert, weil er aus Nutzereingaben lernt und somit gesellschaftliche Vorurteile verstärken kann.

Des Weiteren können ungerechtfertigte Gleich- und Ungleichbehandlungen auftreten, wenn eine KI entweder berechnete Unterschiede ignoriert oder unzulässige Differenzierungen vornimmt. Dies kann etwa im Bereich der sozialen oder gesundheitlichen Bedürfnisse problematisch sein, wenn Algorithmen individuelle Bedarfe nicht adäquat berücksichtigen.

Ein weiteres Problem ergibt sich aus gruppenbasierten Entscheidungen, bei denen KI-Systeme auf statistische Merkmale anstelle individueller Eigenschaften zurückgreifen. Dadurch besteht die Gefahr, dass Einzelpersonen aufgrund von Gruppenzugehörigkeiten pauschal bewertet und behandelt werden.

Diese Aspekte verdeutlichen, dass Diskriminierung durch KI auf vielfältige Weise entstehen kann und eine tiefgehende Auseinandersetzung mit den zugrunde liegenden Mechanismen erforderlich ist, um faire und diskriminierungsfreie Systeme zu entwickeln.

5.2 Herausforderungen bei der Vermeidung von Diskriminierung durch KI

Die Vermeidung von Diskriminierung in KI-Systemen stellt eine komplexe Herausforderung dar, die sowohl technologische als auch rechtliche und ethische Aspekte umfasst. Ein zentrales Problem ist die mangelnde Transparenz vieler KI-Modelle. Da Algorithmen häufig als „Black Box“ fungieren, sind ihre Entscheidungswege schwer nachvollziehbar, was eine Überprüfung auf mögliche Diskriminierung erschwert. Ohne klare Einblicke in die Funktionsweise der Modelle bleibt es schwierig, Verzerrungen frühzeitig zu identifizieren und zu korrigieren.

Ein weiteres Risiko besteht im unkontrollierten Weiterlernen von KI-Systemen. Modelle, die kontinuierlich aus neuen Daten lernen, können unbeabsichtigt diskriminierende Muster verstärken oder neue Verzerrungen entwickeln, insbesondere wenn keine geeigneten Kontrollmechanismen implementiert sind.[12]

Darüber hinaus spielt die Qualität der verwendeten Daten eine entscheidende Rolle. Vorurteilsbehaftete oder unausgewogene Trainingsdaten können bestehende gesellschaftliche Ungleichheiten reproduzieren und sogar verstärken. Ohne gezielte Maßnahmen zur Identifikation und Korrektur solcher Verzerrungen bleibt die Gefahr bestehen, dass KI-Systeme bestimmte Gruppen systematisch benachteiligen.

Ein weiteres Problem stellt die ungeklärte Verantwortlichkeit dar. Es bleibt oft unklar, wer letztlich für diskriminierende Entscheidungen haftet – die Entwickler der KI, die Betreiber der Systeme oder die Endnutzer, die auf die algorithmischen Empfehlungen vertrauen. Diese fehlende rechtliche Klarheit erschwert die Durchsetzung von Schutzmaßnahmen und die Etablierung von verbindlichen Standards zur Diskriminierungsvermeidung.

Diese Herausforderungen verdeutlichen, dass die Entwicklung fairer und diskriminierungsfreier KI-Systeme nicht nur technische Optimierungen erfordert, sondern auch klare regulatorische Rahmenbedingungen und ethische Richtlinien, um den gerechten und verantwortungsvollen Einsatz von KI-Technologien sicherzustellen. [13]

5.3 Lösungsansätze zur Vermeidung von Diskriminierung durch Künstliche Intelligenz

Zur Reduzierung der Diskriminierung durch Künstliche Intelligenz (KI) existieren verschiedene Lösungsansätze, die sowohl technischer, rechtlicher als auch gesellschaftlicher Natur sind. Ein zentraler Aspekt ist die Erklärbarkeit und Überprüfung von KI-Systemen. Damit Entscheidungen nachvollziehbar und überprüfbar bleiben, sollten Mechanismen zur Transparenz implementiert werden. Zudem könnten unabhängige Institutionen eine Kontrollfunktion übernehmen, um potenziell diskriminierende Praktiken zu identifizieren und zu korrigieren.

Ein weiterer Ansatz ist die selektive Datennutzung, bei der diskriminierende Merkmale gezielt aus den Trainingsdaten entfernt werden. Dabei muss jedoch darauf geachtet werden, dass sogenannte Stellvertretermerkmale – also Merkmale, die indirekt auf sensible Eigenschaften hinweisen – ebenfalls keine ungewollten diskriminierenden Effekte hervorrufen.

Von besonderer Bedeutung sind zudem faire Algorithmen, die so entwickelt werden, dass Fairness als Zielgröße explizit berücksichtigt wird. Dies stellt jedoch eine erhebliche Herausforderung dar, da Fairness technisch schwer zu definieren und zu messen ist.

Verschiedene mathematische Definitionen von Fairness stehen teilweise in Widerspruch zueinander, sodass eine ausgewogene Abwägung erforderlich ist.

Neben technischen Lösungen spielt auch der rechtliche Schutz eine wesentliche Rolle. Betroffene sollten über ihre Rechte informiert sein und in der Lage sein, diskriminierende Entscheidungen anzufechten. Dazu bedarf es klar definierter gesetzlicher Regelungen, begleitender Aufklärungsinitiativen sowie institutioneller Unterstützung, um den Schutz vor algorithmischer Diskriminierung zu gewährleisten.

Schließlich ist ein gesellschaftlicher Diskurs essenziell, um zu definieren, welche Differenzierungen durch KI-Systeme legitim sind und welche als diskriminierend eingestuft werden sollten. Ein solcher Dialog kann dazu beitragen, gesellschaftliche Werte in die Entwicklung und Nutzung von KI zu integrieren und so die Akzeptanz und Fairness technologischer Systeme zu stärken.

Diese Lösungsansätze verdeutlichen, dass die Bekämpfung von Diskriminierung durch KI ein multidisziplinärer Ansatz erfordert, der technologische, rechtliche und gesellschaftliche Maßnahmen miteinander verbindet. [2]

6 KI in der Rekrutierung

6.1 Die vier Schritte

1. **Suche – Active Sourcing**

Active Sourcing bedeutet, dass Unternehmen proaktiv nach potenziellen Kandidaten suchen, anstatt nur auf Bewerbungen zu warten. Mithilfe von Schlagwörtern werden Kandidaten gefiltert, die den gesuchten Anforderungen entsprechen.

2. **Screening**

Der Screening-Prozess ist die automatische Vorauswahl von Bewerbungen. Bewerber, die nicht den Anforderungen entsprechen, werden automatisch aussortiert.

3. **Vorstellungsgespräch**

Unternehmen setzen standardisierte Video-Interviews ein, um Zeit zu sparen. Dabei versucht die KI, ein Persönlichkeitsprofil des Bewerbers zu erstellen.

4. **Auswahl**

Am Ende des Prozesses trifft die KI eine empfohlene Auswahl basierend auf den Daten aus den vorherigen Schritten.

6.2 CRISP-Modell

Das CRISP-DM-Referenzmodell hilft dabei, Schritte bei der Entwicklung von KI-Modellen klar zu strukturieren. CRISP-DM steht für „Cross Industry Standard Process for Data Mining“ und ist ein anerkanntes Modell für Datenanalyseprojekte.

Der Problemlösungsprozess besteht aus mehreren Phasen:

- **Geschäftsverständnis:** Zu Beginn wird das zu lösende Problem analysiert und die übergeordneten Ziele der Untersuchung werden definiert.
- **Datenverständnis:** Es werden relevante Daten identifiziert, gesammelt und hinsichtlich ihrer Qualität sowie Eignung für die Analyse überprüft.
- **Datenaufbereitung:** Die Daten werden bereinigt, transformiert und in ein geeignetes Format für die Modellierung überführt.
- **Modellierung:** Auf Basis der aufbereiteten Daten werden verschiedene Modelle entwickelt, getestet und hinsichtlich ihrer Leistungsfähigkeit optimiert.
- **Evaluierung:** Die erstellten Modelle werden hinsichtlich ihrer Anwendbarkeit in der Praxis validiert und auf ihre Effektivität geprüft.
- **Bereitstellung:** Die finalen Ergebnisse der Analyse werden dokumentiert, interpretiert und für die praktische Umsetzung bereitgestellt [8].

7 Das Trolley-Problem als ethische Herausforderung für KI

Ein klassisches Gedankenexperiment zur moralischen Entscheidungsfindung ist das Trolley-Problem. Es veranschaulicht die ethischen Fragestellungen, die sich insbesondere im Zusammenhang mit KI-Systemen ergeben.

Das Trolley-Problem beschreibt ein Szenario, in dem eine außer Kontrolle geratene Straßenbahn auf fünf Personen zurast, die sich auf den Gleisen befinden. Ein Weichensteller hat die Möglichkeit, die Weiche umzulegen und die Bahn auf ein anderes Gleis zu lenken, auf dem sich jedoch eine einzelne Person befindet. Dadurch ergibt sich ein moralisches Dilemma:

- Soll der Weichensteller untätig bleiben und fünf Menschen sterben lassen?
- Oder soll er aktiv eingreifen und eine Person opfern, um die anderen zu retten?

Dieses ethische Problem wird häufig auf autonome Fahrzeuge übertragen. In einem Unfall-Szenario muss die KI unter Umständen eine Entscheidung darüber treffen, welche Personen priorisiert werden.

Die „Moral Machine“-Studie, ein großangelegtes Experiment zur moralischen Entscheidungsfindung in autonomen Fahrzeugen, zeigt, dass die Mehrheit der Testpersonen eine utilitaristische Perspektive einnimmt, d. h. eine Entscheidung bevorzugt, die die Opferzahl minimiert.

Die Studie umfasste verschiedene Szenarien, die bewusst kontrovers gestaltet waren und keine eindeutig „richtige“ Lösung besaßen. Eine Analyse der 40 Millionen gesammelten Entscheidungen ergab, dass die meisten Teilnehmenden eine Priorisierung nach bestimmten Kategorien vornehmen. So wurden in Analogie zum ursprünglichen Trolley-Problem tendenziell mehr Personen vor weniger Personen, Menschen vor Tieren, gesetzestreue Bürger vor Straftätern und Fußgänger vor Fahrzeuginsassen geschützt.



Figure 1: *Ein Dilemma aus dem „Moral Machine“ Projekt [14]*

Diese Ergebnisse zeigen, dass ethische Entscheidungsprozesse in der KI-gestützten Automatisierung nicht nur technische, sondern auch gesellschaftliche und kulturelle Faktoren berücksichtigen müssen [6].

8 KI in der Bildung

Der Einsatz von KI-Tutoren, datenbasierten Analysen und automatisierten Prüfungsverfahren bietet das Potenzial, Bildungssysteme zu modernisieren. Gleichzeitig gehen diese Technologien jedoch mit erheblichen ethischen und praktischen Herausforderungen einher, die einer kritischen Auseinandersetzung bedürfen.

Ein wesentliches Problem ist die Reproduktion und Verstärkung bestehender Vorurteile, da KI-Modelle Verzerrungen aus den zugrunde liegenden Trainingsdaten übernehmen können. Dies birgt insbesondere für Schüler aus sozial benachteiligten Gruppen ein Diskriminierungsrisiko, beispielsweise wenn Trainingsdaten überwiegend aus privilegierten Bildungseinrichtungen stammen und somit bestehende Ungleichheiten weiter manifestieren. Ein weiteres zentrales Risiko betrifft die potenzielle Beeinträchtigung der Eigenständigkeit von Lernenden. Eine übermäßige Abhängigkeit von KI-Tutoren könnte dazu führen, dass Schüler kritisches Denken sowie individuelle Lernstrategien vernachlässigen und sich stärker auf automatisierte Unterstützung verlassen, anstatt eigene Problemlösungsansätze zu entwickeln.

Darüber hinaus stellt der Datenschutz eine essenzielle Herausforderung dar. KI-Systeme sammeln und analysieren umfangreiche Daten zum Lernverhalten der Schüler, wodurch ohne klare und transparente Regelungen ein Missbrauch dieser Informationen möglich wäre – sei es durch gezielte Werbestrategien, unautorisierte Überwachung oder anderweitige kommerzielle Zwecke.

Neben den ethischen und datenschutzrechtlichen Bedenken bleibt auch die Frage der Verantwortung ungeklärt: Wer trägt letztlich die Haftung für potenziell diskriminierende oder fehlerhafte Entscheidungen – die Entwickler der KI, die Lehrkräfte oder die Bildungseinrichtungen? Diese rechtliche Unsicherheit erfordert klare regulatorische Rahmenbedingungen.

Trotz der technologischen Fortschritte kann KI zudem die Rolle von Lehrkräften nicht ersetzen. Pädagogen spielen eine unverzichtbare Rolle, indem sie soziale Interaktion, Motivation und emotionale Unterstützung bieten, die für den Lernerfolg entscheidend sind.

Auch automatisierte Bewertungssysteme sind problematisch, da sie keine individuellen oder familiären Herausforderungen berücksichtigen und somit in bestimmten Fällen zu ungerechten Beurteilungen führen können.

Schließlich trägt der Einsatz von KI in der Bildung potenziell zur Verschärfung der digitalen Kluft bei, da nicht alle Schüler über den erforderlichen Zugang zu moderner Technologie oder stabiler Internetverbindung verfügen.

Diese Herausforderungen verdeutlichen, dass der Einsatz von KI in der Bildung nur unter Einhaltung klar definierter ethischer Standards und mit einer verantwortungsbewussten Implementierung erfolgen darf, um einen gerechten und inklusiven Bildungszugang für alle Lernenden zu gewährleisten. [4]

8.1 Handlungsempfehlungen

Nachdem die ethischen Herausforderungen von Künstlicher Intelligenz (KI) im Bildungsbereich analysiert wurden, stellt sich die zentrale Frage, welche Maßnahmen erforderlich sind, um diesen Risiken wirksam zu begegnen. Eine ganzheitliche ethische Reflexion muss bereits im Entwicklungsprozess von KI-Systemen erfolgen, um sicherzustellen, dass diese fair, inklusiv und frei von diskriminierenden Verzerrungen gestaltet werden.

Von entscheidender Bedeutung ist eine interdisziplinäre Zusammenarbeit zwischen Informatikern, Pädagogen und Ethikexperten, um Transparenz in den Entwicklungs- und Implementierungsprozessen zu gewährleisten und potenzielle Biases in den verwendeten Daten zu minimieren. Darüber hinaus sollte KI nicht als Ersatz für pädagogische Interaktionen fungieren, sondern die Autonomie der Lernenden fördern, indem sie individuelle Lernprozesse unterstützt und aktiv zur Partizipation am Bildungsgeschehen beiträgt. [11]

9 Der EU AI Act: Regulierung von KI-Systemen

Der EU AI Act hat das Ziel, Künstliche Intelligenz in der Europäischen Union zuverlässig, verständlich und nichtdiskriminierend zu gestalten. Ein zentraler Aspekt ist, dass Menschen die Systeme überwachen, um schädliche Ergebnisse zu vermeiden. Die Regelungen unterteilen KI-Anwendungen in drei Risikoklassen.

9.1 Hochrisiko-KI-Systeme

Hochrisiko-KI-Systeme kommen in sicherheitskritischen Bereichen zum Einsatz. Dazu zählen beispielsweise die Infrastruktur, in der die Sicherheit des Transportwesens gewährleistet werden muss, sowie der Bildungs- und Prüfungsbereich. Ebenso gelten KI-Systeme im Gesundheitswesen, insbesondere bei robotergestützter Chirurgie, sowie im Arbeitsmarkt, etwa bei automatisierter Bewerberauswahl, als hochriskant. Weitere Anwendungsfälle betreffen Bonitätsprüfungen und polizeiliche sowie justizielle Anwendungen, etwa bei Grenzkontrollen.

Aufgrund der potenziell schwerwiegenden Folgen von Fehlern oder diskriminierenden Entscheidungen unterliegen solche Systeme strengen Anforderungen. Dazu gehören:

- die Durchführung von Risikobewertungen,
- die Sicherstellung der Datenqualität,

- eine umfassende Dokumentation.

Zudem ist eine menschliche Aufsicht verpflichtend, um sicherzustellen, dass die Systeme robust und sicher operieren.

9.2 KI mit begrenztem Risiko

Neben Hochrisiko-KI gibt es Anwendungen mit begrenztem Risiko. In dieser Kategorie liegt der Fokus auf Transparenz. Nutzer müssen darüber informiert werden, wenn sie mit einer KI interagieren, etwa bei Chatbots. Auch KI-generierte Inhalte müssen klar als solche erkennbar sein, um das Vertrauen in diese Technologien zu stärken.

9.3 KI mit minimalem Risiko

Schließlich gibt es KI-Anwendungen mit minimalem Risiko, wie etwa in Videospielen oder Spamfiltern. Da diese Systeme kaum Gefahrenpotenzial bergen, dürfen sie ohne besondere Einschränkungen genutzt werden.

9.3.1 Fazit

Zusammenfassend schafft der EU AI Act eine klare Struktur, um die Risiken von KI-Systemen zu minimieren. Gleichzeitig fördert er einen verantwortungsvollen Umgang mit KI, der Innovationen nicht hemmt, sondern gezielt lenkt [3, 9].

References

- [1] Mohamed Akrouf and Robert Steinbauer. “Machine Ethics: The Creation of a Virtuous Machine”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, pp. 135–142. URL: <https://arxiv.org/abs/2002.00213>.
- [2] S. Beck et al. *Künstliche Intelligenz und Diskriminierung – Whitepaper aus der Plattform Lernende Systeme*. 2019. URL: <https://www.plattform-lernende-systeme.de>.
- [3] European Commission. *Regulatory framework proposal on artificial intelligence*. 2025. URL: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
- [4] “Empirical Essays in Artificial Intelligence Ethics”. PhD thesis. Technische Universität Delft. URL: <https://research.tudelft.nl/files/124689843/dissertation.pdf>.
- [5] Alexander Filipović. “Ethische Aspekte der Künstlichen Intelligenz”. In: *Zeitschrift für Digitale Ethik* 1.2 (2019), pp. 45–60. URL: <https://www.econstor.eu/bitstream/10419/224465/1/173246250X.pdf>.
- [6] S. Frisch. *Künstliche Intelligenzen und automatisierte Entscheidungsfindung: Schwachstellen und rechtliche Rahmenbedingungen in kontroversen Anwendungsfällen*. Technische Universität Wien, 2024. URL: https://www.law.tuwien.ac.at/BA_Frisch.pdf.
- [7] T. Greger. *Ethische Herausforderungen durch Künstliche Intelligenz*. YouTube Video. Dec. 2021. URL: <https://www.youtube.com/watch?v=m8vPb96ZXrs>.

- [8] C. Mach. *Ethik in der KI: Vermeidung von Vorurteilen und Diskriminierung in den Bewerbungsprozessen internationaler Unternehmen*. HWZ Hochschule für Wirtschaft Zürich, 2024. URL: <https://barrueto.ch/wp-content/uploads/2024/08/Bachelorarbeit-2024-Ethik-in-der-KI-bei-Bewerbungsprozessen-Mach.pdf>.
- [9] Bundeskanzleramt Österreich. *AI Act. Digitales Österreich*. 2025. URL: <https://www.digitalaustria.gv.at/Themen/KI/AI-Act.html>.
- [10] Europäisches Parlament. *Was ist künstliche Intelligenz und wie wird sie genutzt?* Abgerufen am 23. Februar 2025. Aug. 2020. URL: <https://www.europarl.europa.eu/topics/de/article/20200827ST085804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt>.
- [11] K. Porayska-Pomsta, W. Holmes, and S. Nemorin. “The ethics of AI in education”. In: *Handbook of Artificial Intelligence in Education*. Ed. by B. du Boulay, A. Mitrovic, and K. Yacef. University College London, 2024. Chap. 26.
- [12] Stefan Ravizza et al. “Artificial Intelligence Transparency”. US11651276B2. Patent granted on May 16, 2023. May 2023. URL: <https://patents.google.com/patent/US11651276B2>.
- [13] Hamza Ait Si Selmi. “Ethische Probleme beim Einsatz von Künstlicher Intelligenz im Gesundheitswesen”. Masterarbeit. Ostfalia Hochschule für angewandte Wissenschaften, 2021. URL: https://opus.ostfalia.de/files/1279/Hamza_2021_KI_Gesundheitswesen.pdf.
- [14] Studyflix. *Trolley Problem*. Abgerufen am 26. Februar 2025. n.d. URL: <https://studyflix.de/allgemeinwissen/trolley-problem-7879>.