

# Leitfaden für Lehrende

CHANGE Kurs: KI und Diskriminierung



## Inhalte:

1. Ziele und Anliegen des Kurses
2. Aufbau und Teile des Kurses
3. Impressum

## 1. Ziele und Anliegen des Kurses

Vorurteile sind für Menschen der Gegenwart ebenso überlebenswichtig wie zu früheren Zeiten. Sie ermöglichen es, sich in einer hochkomplexen und gefährlichen Umwelt zurechtzufinden. Allerdings führen sie zwangsläufig zu ungerechtfertigten Benachteiligungen, denn sie verhindern das Treffen faktenbasierter Entscheidungen. Computer erscheinen demgegenüber als objektiv und neutral, dabei sind insbesondere KI-basierte Systeme oft alles andere als fair. Das liegt selten (aber auch) an den Algorithmen, also den dahinter liegenden mathematischen Modellen und viel öfter an falschen oder unausgewogenen Trainingsdaten. Diese Daten werden für die Methoden des Maschinellen Lernens benötigt, welche wiederum die Grundlage der meisten modernen KI-Anwendungen bilden.

Daten basieren immer auf Entscheidungen, die Menschen zu einem früheren Zeitpunkt getroffen haben und Menschen haben Vorurteile. Künstliche Intelligenz tut damit das, was WIR ihr beibringen und wenn Programmierer\*innen einem KI-Algorithmus schlechte Trainingsdaten zur Verfügung stellen, wird er unzufriedenstellende Ergebnisse erzielen. Ein Computer übernimmt also die menschlichen Fehler und es gibt zahlreiche Beispiele, wie Maschinelles Lernen bestehende Strukturen reproduziert und auf diese Weise diskriminierte Gruppen weiterhin systematisch benachteiligt werden. So musste etwa der südkoreanische Chatbot "Lee Luda" 2020 nach wenigen Tagen vom Netz genommen werden, weil er rassistisches Verhalten erlernte. Das Schicksal teilte er übrigens mit seinem Microsoft-Vorgänger "Tay", der 2016 aus demselben Grund bereits nach wenigen Stunden wieder abgeschaltet werden musste.

Das Ziel dieses Kurses ist es, Schüler\*innen die Schwachstellen von Künstlicher Intelligenz aufzuzeigen und sie dafür zu sensibilisieren, dass auch Computer Fehler machen. Die Schülerinnen und Schüler lernen, wie und warum KI-basierte Systeme Menschen diskriminieren, was 'biased data' sind und welche Methoden angewendet werden können, um maschinelle Diskriminierungen zu verhindern oder zumindest zu verringern. Ein Hauptaugenmerk liegt dabei auf verschiedenen Gesichtserkennungstechnologien (GET) und ihren Diskriminierungspotenzialen.

## Zielgruppen

Der Kurs richtet sich an Schüler\*innen der Sekundarstufe 1. Anknüpfend an den Lehrplanentwurf für den Pflichtgegenstand Informatik, der im Schuljahr 2022/23 eingeführt wird, bietet sich die 4. Klasse (8. Schulstufe) für die Einbindung von Themen rund um Künstliche Intelligenz und Maschinelles Lernen besonders an:

- Kompetenzbereich Orientierung: Grenzen und Möglichkeiten von KI
- Kompetenzbereich Information und Daten: Gefahren der Erhebung, Auswertung und Verknüpfung von Nutzerdaten

Als Vorbereitung auf diesen Vertiefungskurs wird empfohlen, neben dem Grundlagenkurs "Künstliche Intelligenz" auch Teile der Kurse "Maschinelles Lernen" und "Big Data" zu absolvieren.

## 2. Aufbau und Teile des Kurses

Der Kurs besteht aus zwei voneinander unabhängigen Modulen: Wie KI diskriminiert und Digitale Gesichtserkennung.

### Modul 1: Wie KI diskriminiert

Das Basismodul *Wie KI diskriminiert* besteht aus den Teilbereichen:

- Fallanalyse: Diskriminierende KI
- Wissen: Interaktives Erklärvideo
- Reflexion: Mögliche Lösungsansätze
- Vertiefung: Film "Coded Bias" und Ethik-Leitlinien der EU-Kommission

Diese Teile bauen aufeinander auf (Ausnahme Vertiefung). Im ersten Teil lernen die Schüler\*innen verschiedene Diskriminierungsfälle von KI-Systemen kennen und erörtern in ihren Gruppen die jeweiligen Gründe für die Diskriminierungen. Im interaktiven Erklärvideo von Teil 2 vertiefen sie ihr Wissen über das Wie und Warum maschineller Diskriminierungen und können mit ihrem Wissen im dritten Teil miteinander über mögliche Lösungsansätze diskutieren.

### Modul 2: Digitale Gesichtserkennung

Modul 2 kann unabhängig von Modul 1 absolviert werden, um das Wissen von Schülerinnen und Schülern über Funktionsweise, Einsatzgebiete und Gefahren von verschiedenen Gesichtserkennungstechnologien (GET) zu vertiefen.

## 3. Impressum

CHANGE – EDUMINDSET FOR FUTURE

<https://edumakerspace.fll.wien/change/>

Die Lernmaterialien stehen zur Förderung von Open Educational Practices (OEP) unter einer „Creative Commons Lizenz“ vom Typ Namensnennung zur Verfügung; ausgenommen sind wenige Materialien, deren davon abweichende Nutzungslizenzen bzw. Quellen entsprechend gekennzeichnet sind.



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>