7-Schritte-Leitfaden für ethische Entscheidungsfindung

Kompaktes Infoblatt • basierend auf dem Framework des Carleton College (SERC)

Ethische Rahmenwerke helfen sicherzustellen, dass Entscheidungen – insbesondere beim Einsatz von KI – keinen unbeabsichtigten Schaden verursachen. Dieser Leitfaden bietet eine strukturierte Methode in sieben Schritten.

Die 7 Schritte im Überblick

1	**Problem benennen**	Was genau bereitet Unbehagen? Wo liegt ein Interessenkonflikt? Das Problem möglichst klar formulieren.
2	**Fakten prüfen**	Relevante Daten und Studien recherchieren – ohne Voreingenommenheit, auch Gegenargumente betrachten.
3	**Faktoren identifizieren**	Alle relevanten Einflussfaktoren auflisten: interne (Werte, Vorurteile) und externe (Strukturen, Gesetze, Markt).
4	**Optionen entwickeln**	Konkrete Handlungsmöglichkeiten erarbeiten – technische, organisatorische und kommunikative.
5	**Optionen testen**	Jede Option anhand ethischer Tests prüfen (siehe Tabelle unten).
6	**Entscheidung treffen**	Auf Grundlage der Analyse eine Option wählen und die Wahl begründen.
7	**Überprüfung**	Rückblick: Wie lässt sich Wiederholung vermeiden? Welche Perspektiven fehlten?

Die 6 Ethik-Tests (Schritt 5)

Mit diesen Fragen lassen sich Handlungsoptionen systematisch prüfen:

Schadenstest	Ist diese Option weniger schädlich als die anderen?
Datenschutztest	Greife ich mit dieser Option in die Privatsphäre anderer ein?
Öffentlichkeitstest	Würde ich wollen, dass diese Entscheidung in der Zeitung steht?
Verteidigungstest	Könnte ich diese Entscheidung vor einer Gruppe verteidigen?
Reversibilitätstest	Wäre ich einverstanden, wenn ich selbst davon betroffen wäre?
Kollegentest	Was würden Kollegen sagen, wenn ich diese Option vorschlage?

Reale Beispiele für KI-Ethik-Probleme

Amazons KI-Bewerbungsfilter (2014–2017)

Amazon entwickelte ein KI-System zur automatischen Bewerbungsauswahl. Das System wurde mit historischen Einstellungsdaten trainiert – da in der Vergangenheit überwiegend Männer eingestellt wurden, bewertete die KI Bewerbungen von Frauen systematisch schlechter. Lebensläufe mit dem Wort „women‘s“ (z.B. „women‘s chess club“) wurden abgewertet. Amazon stellte das Projekt 2017 ein.

Quelle: Reuters/Dastin (2018); MIT Technology Review

Rassistischer Gesundheits-Algorithmus (Optum, 2019)

Ein in den USA weit verbreiteter Algorithmus zur Identifikation von Hochrisikopatienten betraf ca. 200 Millionen Menschen. Er nutzte Gesundheitsausgaben als Proxy für Krankheitsschwere. Da für schwarze Patienten historisch weniger ausgegeben wird, stufte der Algorithmus sie systematisch als gesünder ein – obwohl sie kränker waren. Nach der Aufdeckung konnte der Bias um 84% reduziert werden.

Quelle: Obermeyer et al. (2019), Science 366(6464); NBC News; Scientific American

Social-Media-Algorithmen und Jugendliche

Untersuchungen zeigen, dass Empfehlungsalgorithmen von Plattformen wie TikTok und Instagram gezielt Inhalte an Jugendliche ausspielen, die Suchtverhalten fördern können. Eine Harvard-Studie beziffert die jährlichen Werbeeinnahmen durch minderjhrige Nutzer auf ca. 11 Milliarden US-Dollar. Der US-Surgeon General sprach 2023 eine offizielle Warnung zur psychischen Gesundheit Jugendlicher aus.

Quelle: Costello et al. (2023), American Journal of Law & Medicine; Cureus (2025)

Rahmenwerk-Quelle: *Science Education Resource Center, Carleton College – „Ethical Decision-Making“*
https://serc.carleton.edu/geoethics/Decision-Making Adaptiert für KI-Ethik auf Basis von: Intel Digital Readiness – AI for Youth, Modul 16