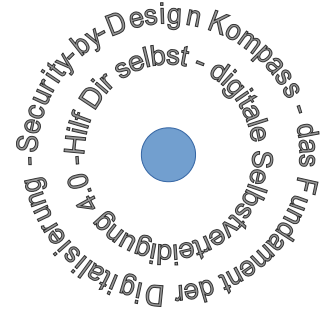


Ethik in Entwicklung und Benutzung künstlicher Intelligenz

Die digitalisierte Welt braucht fundamentale Grundsätze, denen man folgt.

Im Umgang mit künstlicher Intelligenz sind ethische Grundsätze zwangsläufig notwendig.

Grundlegende Prinzipien für künstliche Intelligenz werden dargelegt ebenso wie die Kernaussagen der Hambacher Erklärung zur künstlichen Intelligenz und offizielle Richtlinien zur Entwicklung und praktischen Verwendung von künstlicher Intelligenz der EU.



Quellen:

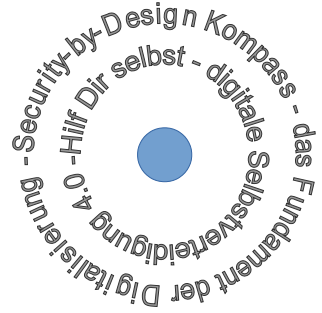
[1] From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, Jessica Morley, Luciano Floridi, Libby Kinsey, Anat Elhalal

[2] Hambacher Erklärung zur Künstlichen Intelligenz: Sieben datenschutzrechtliche Anforderungen (3.April 2019)
https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf

[3] [EU]-Ethik-Leitlinien für eine vertrauenswürdige KI (April 2019):
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60425

1. Wohltätigkeit [1]

- Stakeholderpartizipation
- Schutz fundamentaler Rechte
- Nachhaltige und umweltfreundliche KI
- Klarer Zweck für Bau des Systems sowie klarer Nutzen (kein Bauen zum Zwecke des Bauens)



Bekannt?

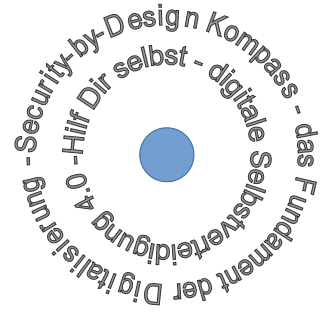
Gewissensbits
Fallbeispiele zu Informatik und Ethik

siehe unter
<https://gewissensbits.gi.de/>

„Die „Gewissensbits“ stellen Fallbeispiele zu verschiedenen Themen der Informatik dar. Diese beschreiben realitätsnahe Situationen, die verschiedene moralische und ethische Fragen aufwerfen. Entwickelt (...) von Mitgliedern der Fachgruppe Informatik und Ethik (fg-ie.gi.de). Zusammen mit einer Liste von Fragen laden diese Fallbeispiele zum Nachdenken und Diskutieren ein: (...) Blog der Fachgruppe (gi.de). (...) Wir freuen uns auf Ihre Beiträge und kommen gerne mit Ihnen ins Gespräch.“ - <https://gi.de/themen/beitrag/default-29872848c3> 9.9.2021

2. Unschädlichkeit [1]

- Sicherheit und Resilienz gegenüber Angriffen
- Notfallplan und generelle Sicherheit
- Genauigkeit (Dokumentation, korrekte Ergebnisse)
- Privatsphäre und Datenschutz
- Verlässlichkeit und Wiederholbarkeit
- Qualität und Integrität der Datenschutz
- Sozialer Einfluss



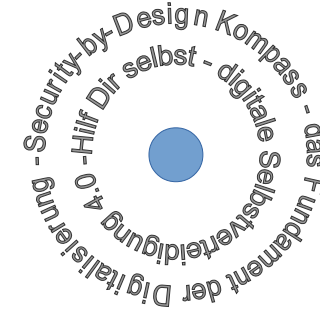
Verletzungen und das Nicht-Gegeben-Sein dieses Prinzips lassen sich als Beispiele in unzähligen Filmen und TV-Serien, die das Thema künstlicher Intelligenz zur Grundlage nehmen finden.

Oft fehlen Notfallpläne bzw. die generelle Sicherheit, also ein Plan B, meist wurden die ethischen Konsequenzen der Entwicklung der KI nicht berücksichtigt und Risiken unzureichend eingeschätzt.

Eine Verletzung der Unschädlichkeit hat dann meist Verletzungen weiterer Prinzipien zur Folge. Beispiele hierfür wären dann jene Filme, in denen sich die nicht „unschädlichen“ künstlichen Intelligenzen über die Menschen erheben, also z.B. deren Autonomie verletzen usw. (längst kein Science-Fiction mehr!)

3. Autonomie [1]

- Menschliche Informiertheit (Fähigkeit autonomer Entscheidungen)
- Menschliche Aufsicht
- **verhältnismäßiger** Einsatz von KI



Bedenklich ist die Verletzung dieses Prinzips z.B. bei der **Identifizierung und Ortung von Personen mithilfe von KI**. Auch z.B. Techniken zur Lügendetektion z.B. durch Auswertung sog. Mikroexpressionen. Biometrische Daten sind personenbezogene Daten, die es zu schützen gilt. Hier muss eine Definition eines gerechten Maßes erfolgen, wann ein Einsatz von KI als gerechtfertigt gilt.

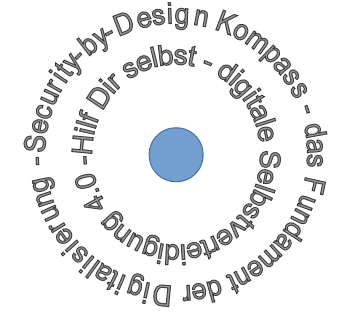
KI-gestützte Bewertungen von Bürgerinnen und Bürgern:

Sog. Scoring muss für seinen Einsatz eindeutig gerechtfertigt sein, seine Maßnahmen verhältnismäßig.

Keine allgemeine Bewertung der „moralischen Persönlichkeit“ oder der „ethischen Integrität“. Sie verletzt **alle** Aspekte dieser Werte! [3]

4. Gerechtigkeit [1]

- Vermeidung unfairer Verzerrungen
- Zugänglichkeit und universelles Design
- Gesellschaft und Demokratie
- Konsultierbarkeit
- Minimierung und Bericht negativer Einflüsse
- Kompromisse
- Rechtshilfe/Abhilfe/Regress



Konsequenzen und Probleme:

(Selbst, 2017) zeigen, dass die Anwendung von sog. „Predictive Policing“ zur vermehrten Verhaftung von farbigen Menschen führt/e. Selbst, Andrew D., Disparate Impact in Big Data Policing (February 25, 2017). 52 Georgia Law Review 109 (2017)

Hier zeigt sich insbesondere die **Abhängigkeit künstlicher Intelligenz von seinen Trainingsdaten**.

Jede Verzerrung in diesen wird sich ggf. als Diskriminierung von Teilen der Bevölkerung im Einsatz der KI übertragen und verfestigen.