



INPUT-PAPIER

Ethik und KI

Ein kurzer Input für den 1. Workshop der Roundtable-Reihe ethische KI-Entwicklung (RTeKI)

21. Juni 2022, Linda Schwarz, Julia Meisner (Gesellschaft für Informatik e.V.)

Ziel der Reihe RTeKI ist es, die vielfältigen Probleme in der Umsetzung von KI-Ethik zu identifizieren und Best Practices für die wirksame Durchsetzung von Ethik-Richtlinien zu entwickeln.

Dabei steht in dem ersten Roundtable das gegenseitige Verstehen der unternehmerischen Positionen im Vordergrund. Dieses Papier soll helfen, die jeweiligen Perspektiven zu sortieren und einzuordnen. Dabei widmet es sich ① dem Begriff Ethik, gibt ② einen Überblick über die Ansätze der teilnehmenden Unternehmen und ③ einen kleinen Einblick in bestehende Forschungsliteratur zum Thema.

① ANGEWANDTE ETHIK UND SOZIALE VERANTWORTUNG

In der Philosophie gibt es verschiedene Arten von Ethik. Die insbesondere im KI-Zusammenhang relevante **Normative Ethik** beschäftigt sich vor allem mit der Frage, wie wir aufgestellte Normen begründen und inwiefern wir diese verallgemeinern können.¹ Dafür gibt es drei Ansätze, die als Impuls für die Beschäftigung mit KI und Ethik dienen können:

- Der **deontologische Ansatz** sieht es als zentral an, dass Menschen sich bei ihrem Handeln an Regeln orientieren. Diese Regeln stehen unabhängig von äußeren Umständen, Zwecken und Folgen fest, auch wenn sie kontraintuitiv erscheinen oder negative Folgen mit sich ziehen. Ein bekannter Vertreter ist Immanuel Kant.
- Der **teleologische Ansatz** nimmt die Ziele und Konsequenzen, die sich aus Handlungen ergeben, in den Blick. Ein Mensch, der danach handelt, möchte größtmöglichen Nutzen erzeugen. Eine bekannte Strömung ist der Utilitarismus von John Stuart Mill.
- Schließlich können im Sinne der **Tugendethik** auch Werte und Ideale die Handlungen von Menschen anleiten. Bei dieser (von Aristoteles vertretenen) Perspektive sind Regeln oder eine Reflexion über Konsequenzen nicht relevant. Es stellt sich die Frage: Was für ein Mensch möchte ich sein?

Die **Angewandte Ethik** ist ein Teilbereich der Normativen Ethik – und kann dementsprechend oben genannte Perspektiven einnehmen. Sie widmet sich bestimmten Bereichen, z.B. Medizin-Ethik, Roboter-Ethik oder KI-Ethik. Alle Bereichs-Ethiken entwickeln sich aufgrund konkreter Bedarfe. Im Fall der KI-Ethik ist es die Entwicklung „intelligenter“ oder „smarter“ Technologien: Das Besondere der KI-Ethik im Gegensatz zu anderen Bereichs-Ethiken ist, dass sie sich mit Technologien befasst, die nicht nur ethische Fragen für die Mensch-Maschine-Interaktion aufwerfen. Denn KI-Systeme handeln bis zu einem gewissen Grad selbst autonom, sodass ihre Handlungen sich nach ethischen Werten richten sollten.^{2,3}



② UNTERNEHMERISCHE ETHISCHE RICHTLINIEN

Seit etwa 2017 gibt es vermehrt ethische Richtlinien zum Einsatz Künstlicher Intelligenz in Unternehmen: So listet die gemeinnützige Forschungs- und Advocacy-Organisation AlgorithmWatch in einer Datenbank derzeit 167 Ethik-Richtlinien auf.⁴ Auch für die Teilnehmenden der Reihe RTeKI ist das Thema bereits sehr präsent. Mehr als die Hälfte der teilnehmenden Unternehmen haben bereits ethische Richtlinien zu KI aufgestellt.

Die meistgenannten Aspekte innerhalb dieser Richtlinien sind Sicherheit, Datenschutz, Fairness sowie das gesellschaftliche und ökologische Wohlergehen. Interessant ist, dass die Rolle des Menschen verschieden definiert wird: So heben einige Richtlinien gegenüber der KI den Vorrang des Menschen hervor (z. B. BMW Group, Lufthansa Group). Andere sehen Mensch und KI-System vorrangig als sich gegenseitig ergänzende bzw. kooperierende Akteure (z. B. Microsoft, Telekom).

Es gibt grob gesehen zwei Wege, auf denen ethische KI-Richtlinien in die Anwendung gebracht werden können:

i. Ethische Leitlinien werden **mittels Software** umgesetzt. Die KI-Anwendung soll ihre Entscheidungen etwa möglichst transparent machen oder Erfahrungswerte der Bediener*innen möglichst gut berücksichtigen. Beispiele hierfür sind:

- Die Continental Automotive AG nimmt in den Blick, wie Algorithmen die Mensch-Computer-Interaktion optimieren können. Dazu gehört beispielsweise die Frage, wie exakt die KI-Anwendung menschliche Eingaben auffasst.⁵
- Microsoft hat Software-Lösungen wie „FairLearn“ oder „Counterfit“ implementiert, die KI-Entwicklungsteams bei der Bewertung von Fairness und Robustheit unterstützen sollen.⁶
- Aleph Alpha sieht Open Source als eine Lösung, Transparenz herzustellen.⁷

ii. Die in die KI-Entwicklung involvierten Menschen stehen im Fokus. Beispiele **ethischer Überlegungen von Software-Entwickler*innen** sind: Welche (Menge an) Daten sind für das Training der KI ausreichend? Inwiefern berücksichtige ich dessen Energiebedarf? Wie viel Arbeitsaufwand stecke ich in die Sicherheit der KI? Ein Beispiel hierfür ist:

- Die BMW Group hat Checklisten für bestimmte Anwendungsfälle etabliert, z. B. für die Entwicklung der Notbremse beim autonomen Fahren. Damit sollen ethische Prinzipien konkretisiert werden.⁸
- Die Deutsche Telekom setzt u. a. darauf, Anwender*innen im Umgang mit KI zu bilden. Sie hebt hervor, dass klare Regeln zu Verantwortlichkeiten bzgl. der KI-Entwicklung etabliert werden sollen.⁹



3 HINDERNISSE BEI DER UMSETZUNG ETHISCHER RICHTLINIEN

Dr. Thilo Hagendorff, ein angewandter Philosoph, analysiert dass die meisten Unternehmen mit ihren ethischen Richtlinien auf technikgetriebene Lösungen setzen und ein deontologisches Verständnis von Ethik haben.¹⁰ Gleichzeitig sind die Richtlinien schwer umsetzbar: So stellten McNamara, Smith und Murphy-Hill fest, dass die bloße Kenntnis ethischer Richtlinien keinen Einfluss auf die Arbeit von Software-Entwickler*innen hat.¹¹ Hagendorff führt dazu beispielsweise folgende Ursachen auf.¹⁰:

- **Interpretationsspielräume:** Richtlinien sind wenig präzise. Sie enthalten abstrakte Begriffe. (Was genau bedeutet etwa „menschzentriert“?) und zu wenig technische Details.
- **Unterschiedliches Verständnis:** Einzelne Werte oder Begriffe, z. B. „Fairness“, werden kulturell sowie je nach Disziplin und Kontext unterschiedlich definiert und gewichtet.
- **Komplexe Team-Strukturen:** Viele verschiedene Menschen sind in die KI-Entwicklung involviert. Es gibt unterschiedliche Interpretationen und Erwartungen. Außerdem steht nicht genau fest, wer in der Entwicklung wofür verantwortlich ist.

Hagendorff¹⁰ und die Philosophin Aline Franke¹² regen daher an, Richtlinien konkreter zu formulieren und sie mit Beispielen sowie technischen Details zu unterfüttern. Verantwortlichkeiten sollten klar benannt werden. Sie schlagen außerdem vor, dass Unternehmen statt auf Regeln stärker auf Dialog und die Vermittlung von Werten setzen – damit werden Entwickler*innen als selbstverantwortliche Menschen angesprochen und ihr Handeln unmittelbarer beeinflusst.

4 DENKANSTÖSSE

- a) Wie versteht mein Unternehmen Ethik?
- b) Wie setzt mein Unternehmen Ethik um?
- c) Lassen sich manche Werte besser technologisch oder organisatorisch umsetzen? Wenn ja - welche?
- d) Welche Herausforderungen bei der Umsetzung ethischer Richtlinien kennen Sie?



LITERATURVERZEICHNIS

- ¹ „Ethik“. Spektrum.de - Nachrichten aus Wissenschaft und Forschung.
<https://www.spektrum.de/lexikon/philosophie/ethik/643> (Zugriff am 31. Mai 2022).
- ² M. Coeckelbergh, „Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability“, Science and Engineering Ethics, Bd. 26, Nr. 4, S. 2051–2068, Oktober 2019. <https://doi.org/10.1007/s11948-019-00146-8>
- ³ W. Schweidler, „Kleine Einführung in die Angewandte Ethik“. Wiesbaden: Springer Fachmedien Wiesbaden, 2018. <https://doi.org/10.1007/978-3-658-14032-8>.
- ⁴ „AI Ethics Guidelines Global Inventory by AlgorithmWatch“. AI Ethics Guidelines Global Inventory. <https://inventory.algorithmwatch.org/> (Zugriff am 31. Mai 2022).
- ⁵ C. Wirth, U. Schmid, and S. Voget, „Humanzentrierte Künstliche Intelligenz: Erklärendes interaktives maschinelles Lernen für Effizienzsteigerung von Parametrieraufgaben.“, Digitalisierung souverän gestalten II, S. 80–92, 2021, doi: 10.1007/978-3-662-64408-9_7.
- ⁶ „Ressourcen zu verantwortungsvoller KI – Microsoft AI“. Microsoft AI.
<https://www.microsoft.com/de-de/ai/responsible-ai-resources> (Zugriff am 31. Mai 2022).
- ⁷ <https://www.aleph-alpha.com/>
- ⁸ „KI-Kodex von BMW – Was steckt dahinter?“ corporate-digital-responsibility.de.
<https://corporate-digital-responsibility.de/article/ki-kodex-von-bmw/> (Zugriff am 31. Mai 2022).
- ⁹ <https://www.telekom.com/de/konzern/digitale-verantwortung/details/ki-leitlinien-der-telekom-523904>
- ¹⁰ T. Hagendorff, „The Ethics of AI Ethics: An Evaluation of Guidelines,“ Minds and Machines, Bd. 30, S. 99–120, 2020, doi: 10.1007/s11023-020-09517-8.
- ¹¹ A. McNamara, J. Smith, and E. Murphy-Hill, „Does ACM’s code of ethics change ethical decision making in software development?,“ Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018, 2018, doi: 10.1145/3236024.3264833.
- ¹² A. S. Franzke, „An exploratory qualitative analysis of AI ethics guidelines,“ Journal of Information, Communication and Ethics in Society, 2022, doi: 10.1108/jices-12-2020-0125.