

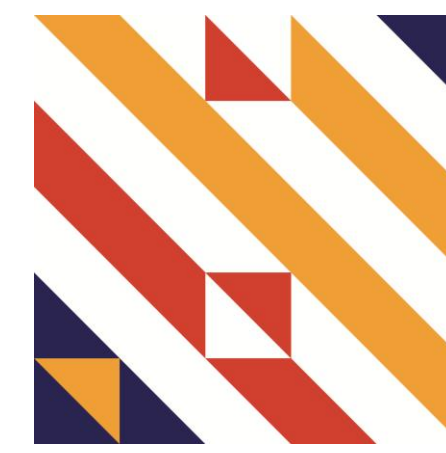
KI-Ethik in Innovationen verankern

Ergebnisse des Projekts ANKER (BMFTR, FKZ 16SV9223)

Dr. Mone Spindler, Dr. Simon Hirsbrunner,
Lisa Koeritz, Sonja Pfisterer

Juni 2026

INTERNATIONALES ZENTRUM
FÜR ETHIK IN DEN
WISSENSCHAFTEN (IZEW)



CLUSTER
INTEGRIERTE
FORSCHUNG



1. Verstehen: Wie müssen KI-Ethik Tools beschaffen sein, damit sie die Verankerung besser leisten?

Künstliche Intelligenz (KI) als komplexe soziotechnische Herausforderung

- ✓ **Integration von Ethik in KI-Entwicklung** als komplexer, unsicherer Prozess mit epistemischen, kommunikativen und strukturellen Herausforderungen
- ✓ **Den Einbezug von Ethik** in Entwicklung von und mit KI durch Tools **verstehen**, **unterstützen** & **vermitteln**
- ✓ **Ein kritisch-konstruktiver Beitrag** zu den Diskussionen über:
 - **Principle to practice gap** (Morley et al. 2020)
 - „**Toolification**“ (Völker et al. 2023)

Typologie basierend auf 29 digitalen Responsible AI Tools + Untersuchung der Verankerungsdimensionen



Interview-Studie: Interviews mit 5 Expert:innen zu Bedingungen für Wirkung & Nachhaltigkeit von Tools

- **Unterschiedliche Einsatzziele, vielfältige Zielgruppen**
- **Spannungsfeld Einfachheit-Komplexität:** Balance zwischen praxisnaher Gestaltung & ethischer Konzepte
- **Einfache Bedienbarkeit:** Betonung auf notwendiger Eingänglichkeit und intuitiver Nutzung
- **Einbettung in Workflows:** Kompatibilität mit Abläufen, Anpassung an verschiedene Kontexte
- **Erwartungsmanagement:** Rolle klarer Erwartungen an Funktionen und Grenzen der Tools
- **Institutioneller Support:** Langfristige Nutzung erfordert strukturelle Unterstützung und Ressourcen



2. Schaffen: Digitale KI-Ethik Tools in co-laborativen Designprozessen entwickeln.

AI Guidelines Guru

Dieses Tool wurde im Rahmen des Projekts ANKER am Internationalen Zentrum für Ethik in den Wissenschaften an der Universität Tübingen entwickelt und steht zur Zeit als Prototyp zur Verfügung.

Funktionen

- Durchsucht den Inhalt von über 200 Ethikrichtlinien und -frameworks.
- Bietet eine automatisch generierte Volltextantwort auf gestellte Fragen.
- Stellt die relevantesten Quellen mit Metadaten bereit.

Hinweise

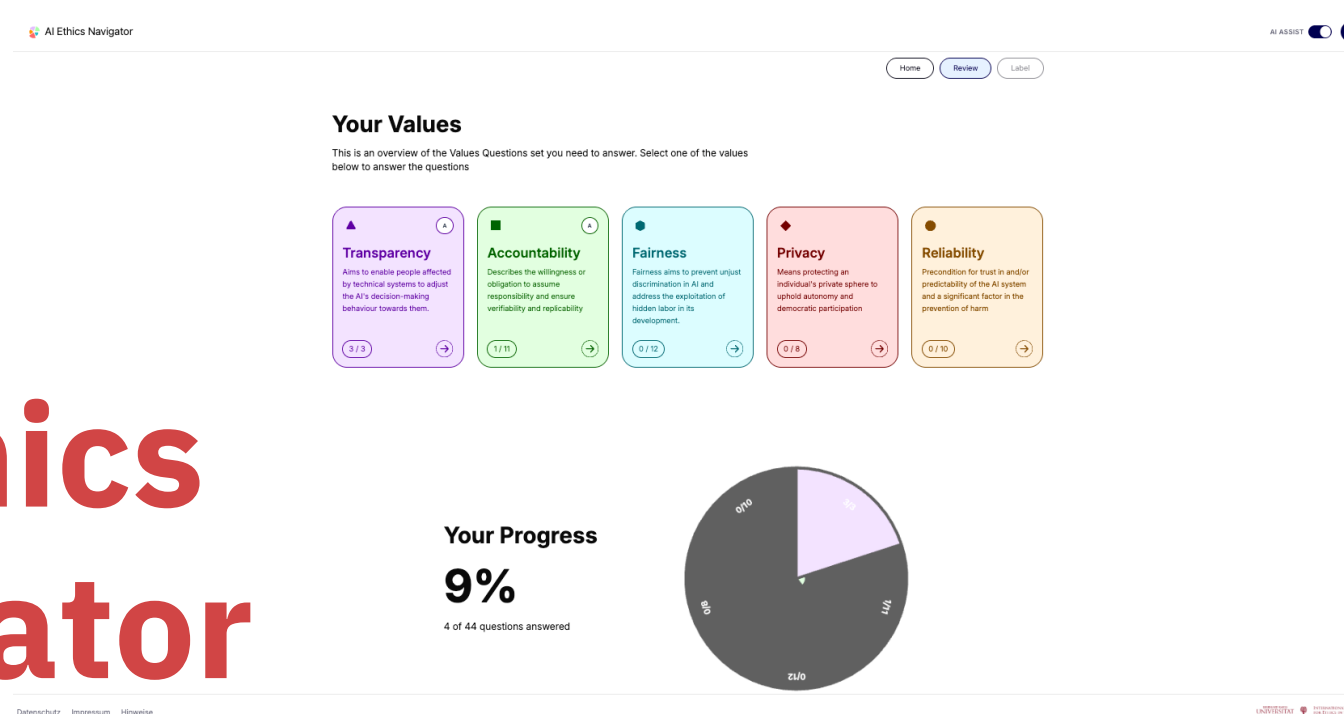
- Fragen müssen in der aktuellen Version in **englischer Sprache** formuliert werden.
- Klicken Sie auf **View Snippet** unter den Quellen, um die ausgewählten Textauschnitte einzusehen.
- Textauschnitte sind aktuell für den prototypischen Anwendungsfall absichtlich kurz gehalten.
- **Confidence Level:** Testierte enthält jede Antwort eine Einschätzung darüber, wie gut die gefundenen Textauschnitte die Antwort abdecken:
 - 1. High: Die Dokumente enthalten die Informationen explizit und umfassend.
 - 2. Medium: Informationen sind unvollständig oder erfordern logische Schlussfolgerungen.
 - 3. Low: Der Kontext ist unzureichend oder unpassend.

Enter your search query here

Problem: Wie findet man sich in den vielen KI-Ethik Richtlinien zurecht?

Lösung: Dem AI Guidelines Guru kann man Fragen zu KI-Ethik Richtlinien stellen. Das Frage-Antwort-Tool generiert aus über 200 KI-Ethik-Richtlinien eine Antwort und legt deren Quellen offen.

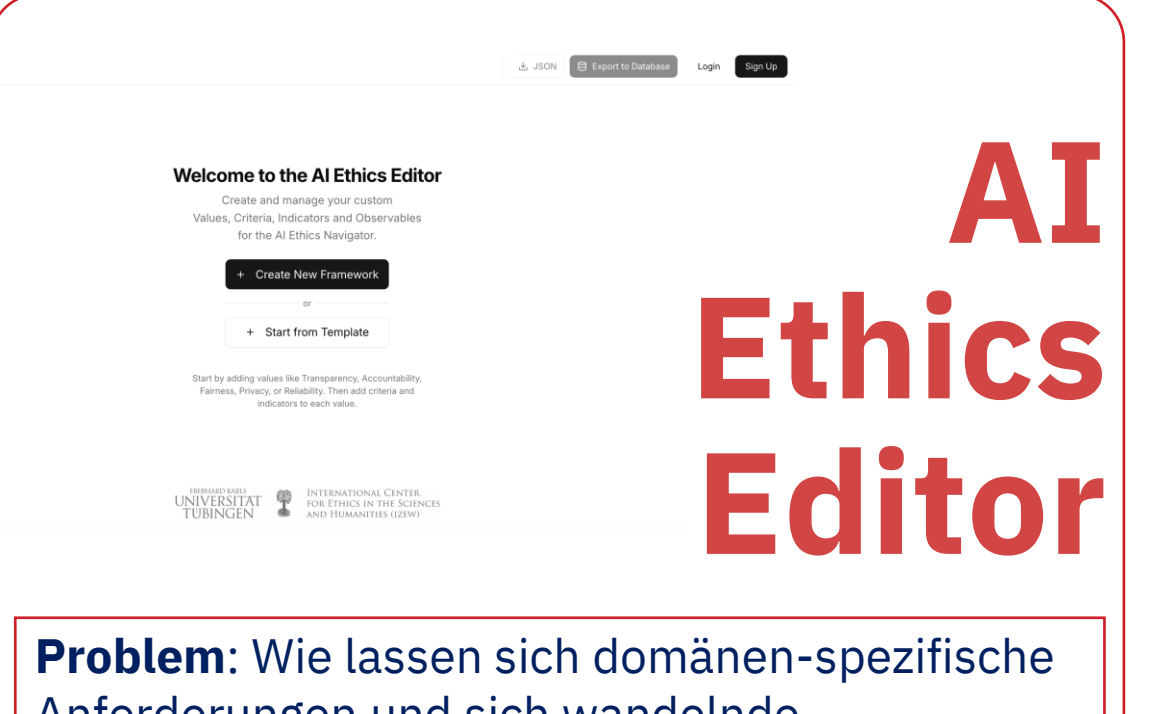
AI Ethics Navigator



Problem: Wie entwickelt man ein vertrauenswürdiges KI-System?

Lösung: Der AI Ethics Navigator unterstützt Entwicklungsteams, ethische Prinzipien gemeinsam für ihr spezielles KI-System zu konkretisieren, zu reflektieren und zu bewerten. Dafür haben wir die bestehende VDE-Spezifikation 90012 für vertrauenswürdige KI (VCIO-Ansatz) zu einer interaktiven, co-laborativen, flexibel anpassbaren Web-Anwendung ausgebaut.

AI Ethics Editor

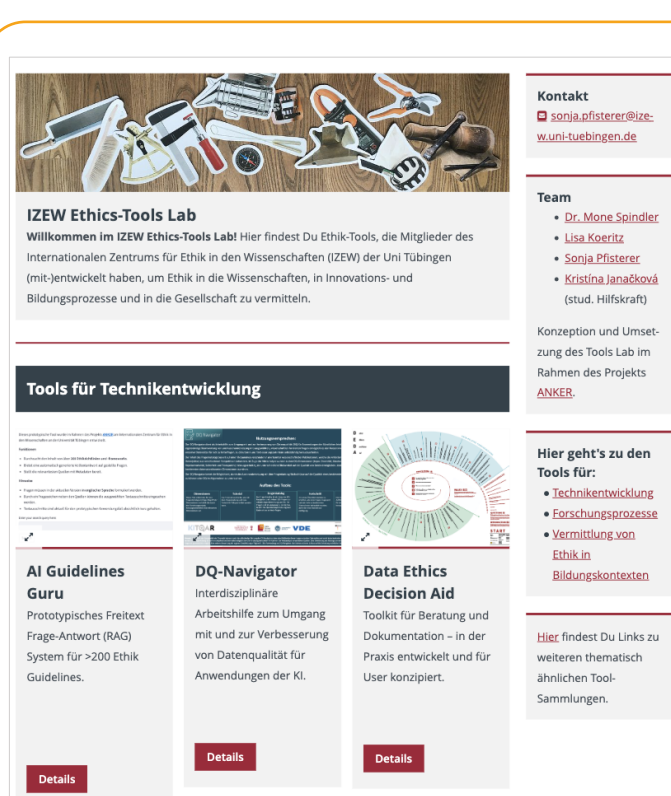


Problem: Wie lassen sich domänen-spezifische Anforderungen und sich wandelnde Regulierungen und Wertvorstellungen im AI Ethics Navigator abbilden?

Lösung: Der AI Ethics Editor ermöglicht das Bearbeiten bestehender sowie das Erstellen neuer Werte, Indikatoren und Kriterien für den AI Ethics Navigator.



3. Vermitteln: Tools & Ergebnisse nachhaltig nutzbar machen.



Ethics Tools Lab

➢ **Veröffentlichung von Tools**, die am IZEW zur Vermittlung von Ethik in Technikentwicklung, Forschungs- und Bildungsprozesse (mit)entwickelt wurden.

No.	Tool	Last time updated	Type of Tool	Publisher	Year of publication
1	AI Ethics Navigator	2023-05-26	Self-Assessment	Research project ANKER	2023
2	AI Explainability 360 (AIX360)	2023-05-26	Technical	IBM / Linux Foundation	2019
3	Fairlearn	2023-05-26	Technical	Fairlearn Open Source Community	2019 (last updated: 2023)
4	InterpretML	2023-05-26	Technical	Microsoft / Microsoft	2019
5	Learning Responsibility Toolkit (LRT)	2023-05-26	Self-Assessment	Google	2020
6	NLP Checklist	2023-05-26	Self-Assessment	Google	2019
7	Responsible AI Framework	2023-05-26	Self-Assessment	Google	2019
8	AI Ethics Checklist	2023-05-26	Self-Assessment	Google	2019
9	Model Cards Toolkit	2023-05-26	Technical	Google	2019
10	Data Cards Toolkit	2023-05-26	Technical	Google	2019
11	High-Quality AI Model Toolkit	2023-05-26	Technical	Google	2019
12	ELI5-4747 Health and Care	2023-05-26	Technical	ELI5-4747 Health and Care	2022
13	Card for Humanity	2023-05-26	Technical	Card for Humanity	2022
14	The Digital Ethics Compass	2023-05-26	Technical	Digital Ethics Compass	2021
15	AI Ethics Emergency Kit	2023-05-26	Technical	Research project ANKER	2023
16	AI Ethics Checklist	2023-05-26	Self-Assessment	Google	2019
17	AI Ethics Quiz	2023-05-26	Self-Assessment	Research project ANKER	2023
18	The Bias Dynamics of AI Principle	2023-05-26	Technical	AI Ethics Lab	2023
19	NLP Toolkit	2023-05-26	Technical	Microsoft	2019
20	Open AI Ethics Checklist	2023-05-26	Self-Assessment	Open AI	2019
21	AI Maturity Tool	2023-05-26	Self-Assessment	IBM Research Center	2019
22	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
23	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
24	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
25	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
26	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
27	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
28	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
29	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019
30	AI Ethics Checklist	2023-05-26	Self-Assessment	IBM Research Center	2019

Typologie

Dimensionen, die Ethik verankern:

- **Ethische Deliberation:** kritische & offene Reflexion fördern, diverse Perspektiven einbeziehen, Handlungsalternativen erkunden
- **Im Praxisfeld:** in bestehende Arbeitsprozesse integrierbar, über Hierarchien hinweg wirksam, kollaborativ
- **Zeitliche Dimension:** flexibel einsetzbar, wiederholt nutzbar und auf nachhaltige Wirkung ausgelegt
- **Gesellschaftlicher Diskurs:** Zivilgesellschaft aktiv mitemdenken, öffentlich zugängliche, gesellschaftlich wirksame Outputs

Spannungsfelder von Responsible AI:

- Ethische Deliberation vs. Verankerung in den Arbeitsweisen des Anwendungsfeldes.
- Verzahnung von reflexiver Ethik und Standardisierungsansätzen.
- Pacing-Problem: KI-Ethik Tools können kaum Schritt halten mit dem sich rapide verändernden Technikfeld.

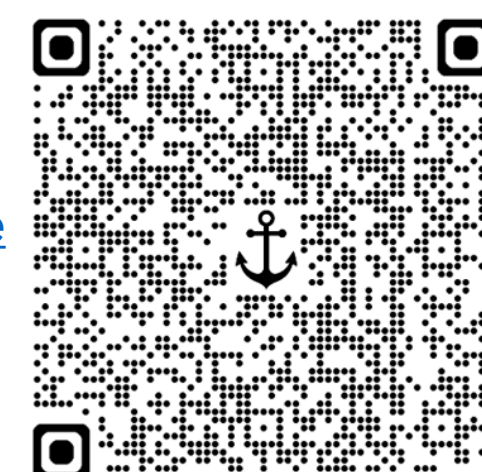
Weiterer Forschungsbedarf:

- Ethische Bewertung automatisierter Assessment-Verfahren für und mit GenAI.



Kontakt

Dr. Mone Spindler
mone.spindler@uni-tuebingen.de
Dr. Simon Hirsbrunner
simon.hirsbrunner@uni-tuebingen.de
Lisa Koeritz
lisa.koeritz@uni-tuebingen.de
Sonja Pfisterer
sonja.pfisterer@uni-tuebingen.de



<https://uni-tuebingen.de/de/257900>



Bundesministerium
für Forschung, Technologie
und Raumfahrt

References:

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 26(4), 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>

Völker, T., Mazzonetto, M., Slaattelid, R., & Strand, R. (2023). Translating tools and indicators in territorial RRI. *Frontiers in Research Metrics and Analytics*, 7, 1038970. <https://doi.org/10.3389/frma.2022.1038970>

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

