

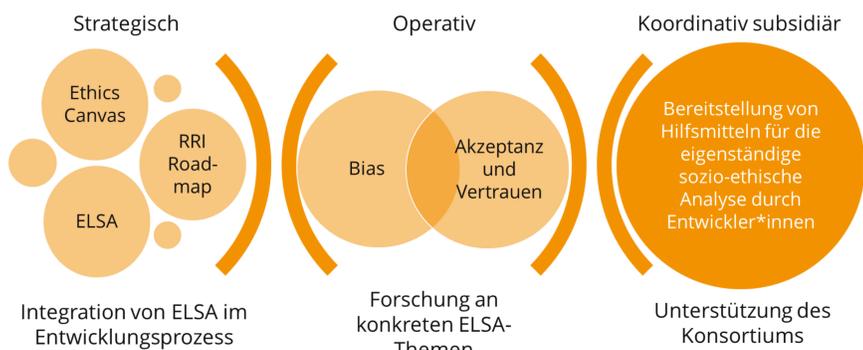
Responsible Innovation Plattform

Verantwortungsvolles Innovieren für vertrauenswürdige KI

Sabrina Blank, M.A. & Dr.-Ing. Christian Herzog, M.A. (Projektleitung)

Forschungsfelder der Responsible Innovation Plattform

Die Responsible Innovation Plattform steht für eine *frühzeitige*, *partizipative* und *iterative* Analyse ethischer, rechtlicher und sozialer Aspekte (ELSA) während der Entwicklung *vertrauenswürdiger* KI. Dafür wird synergetisch auf drei Ebenen geforscht: (i) Strategische Integration von ELSA in den Technikentwicklungsprozess, (ii) konkrete, operative Adressierung von ELSA im Projekt, (iii) Bereitstellung von Hilfsmitteln für die eigenständige sozio-ethische Analyse durch Entwickler*innen.



Quelle: Synergieeffekte der Forschung der Responsible Innovation Plattform (eigene Darstellung)



Quelle: Beispielhafte Skizzierung der Logik der RRI-Roadmap an einem ethischen Handlungsfeld in AP390: assistierendes Bewegungstraining; Farbgebung angelehnt an das RRI-Roadmap Template (eigene Darstellung)

Forschung an der operativen Adressierung von ELSA

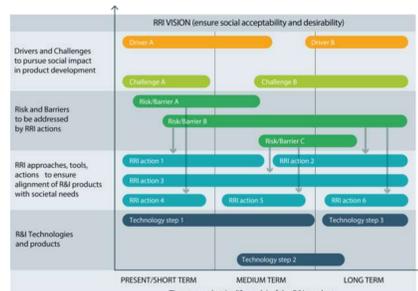
Die Implementierung von ELSA auf operativer Ebene bedarf einer projekt-spezifischen Ausgestaltung, da jedes Technik-Entwicklungsprojekt spezifische sozio-ethische Implikationen und Entwicklungskontexte aufweisen kann. Am Beispiel des Themas „Bias“ werden in **AP360 (VIKOOB)** potenzielle Bias-Eintrittspunkte, deren Dokumentation und Evaluation erforscht. Diese Erkenntnisse sollen das Risiko mindern, dass bspw. der KI-Algorithmus vorurteilbehafteten Output generiert. Damit kann diese Forschung zur Vertrauenswürdigkeit beitragen, da eine bessere Nachvollziehbarkeit des Outputs ermöglicht wird. Zudem wird ein Analyseprozess als möglicher Use Case für KI-basierte Technikentwicklungen skizziert.

Forschung an der strategischen Integration von ELSA

Die strategische Integration von ELSA in Technik-Entwicklungsprozesse bedarf Methoden, die bei der Antizipation sowie Reflexion unterstützen. Das Ethics Canvas ermöglicht eine themengeleitete und kollaborative Erarbeitung relevanter ethischer Fragestellungen und Stakeholder-Perspektiven, die als erste sozio-ethische Analyse verwendet werden kann.

Ethics Canvas		Project Title:	Date:	Prepared by:
1	2	3	4	5
6	7	8	9	10

Quelle: <https://ethicscanvas.org/>



Quelle: <https://www.rri-prisma.eu/road-map-rri-for-companies/>

Die Verwendung dieser Ergebnisse für die Erarbeitung einer sogenannten RRI-Roadmap hat das Ziel, die Risiken für den Entwicklungserfolg, sowie entsprechende Handlungsoptionen frühzeitig zu identifizieren.

In der Kollaboration mit dem Projekt **AP390 zum assistierenden Bewegungstraining** werden diese beiden Methoden erprobt. An der beispielhaften Darstellung einer der identifizierten potenziellen ethischen Herausforderungen ist zu erkennen, dass eine frühzeitige Identifikation von ethischen Risiken im Entwicklungsprozess eine Berücksichtigung bspw. im Produktdesign ermöglicht.

Befähigung zur eigenständigen sozio-ethischen Analyse

Methoden für eine eigenständige initiale sozio-ethische Analyse durch die Entwicklungsteams müssen an die Bedürfnisse und Anforderungen des KI-Ökosystems angepasst sein. In einem Workshop wurden diese Anforderungen partizipativ erarbeitet.



Quelle: Ziele und Fragestellungen zur Bearbeitung im Workshop mit dem Konsortium (eigene Darstellung)

Die Ergebnisse des Workshops zeigen, dass ein Bedarf an einem geleiteten Verfahren besteht, welches eng mit dem Entwicklungsprozess verknüpft ist und eine schnelle Identifikation von ethischen Problemfeldern bei KI-basierter Medizintechnik im Projekt ermöglicht.