



# Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle

Konsolidierung, Weiterentwicklung und Standardisierungsmöglichkeiten

*VDE, Frankfurt am Main*

*Stand: 11.08.2020*

*Dieses Dokument erhebt keinen Anspruch auf Vollständigkeit und wird kontinuierlich von der VDE-Expertengruppe „Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle 2030“ weiterentwickelt.*

*Ansprechpartner:*

*Jessica Fritz*

*Managerin Digitale Technologien und Services*

*Bereich Neue Technologien und Services*

*VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V.*

*Tel: +49 69 6308-387*

*EMAIL: [jessica.fritz@vde.com](mailto:jessica.fritz@vde.com)*



## 1. Inhalt

2. Einführung in Expertengruppe, Leitfrage und Ziele .....	3
Kurze Vorgeschichte .....	3
VDE-Expertengruppe „Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle 2030“ .....	3
Was sind Gestaltungsregeln? .....	4
3. Zukunftsszenarios.....	4
Szenario 1 „Kooperation mehrerer Nutzer in unterschiedlichen Rollen beim Maschineneinsatz“ ....	5
Szenario 2 “Umgang mit selbstlernenden, adaptiven Systemen am Beispiel von Servicerobotern“ .	7
Szenario 3 „VR/AR in Wartung und in Assistenzsystemen“ .....	9
Szenario 4 „Umgang mit Unterschieden zwischen Kulturen, Milieus, Generationen, Bildungsniveaus, Überforderungsschwellen“ .....	10
Szenario 5 „Die Maschine als Lernort“ .....	13
Szenario 6 „Intelligente Unterstützung in Ausnahmesituationen“ .....	15
Szenario 7 „Smartwatches im Fertigungsumfeld“ .....	17
Szenario 8 „Gestaltungsrahmen für die Ethik von Künstlicher Intelligenz in der Produktionsumgebung“ .....	18



## 2. Einführung in Expertengruppe, Leitfrage und Ziele

### Kurze Vorgeschichte

Der VDE gestaltet als wissenschaftlich-technischer Fachverband an den Schnittstellen von Informationstechnik, Elektrotechnik/Elektronik und Automatisierungstechnik aktiv den digitalen Wandel mit. Fachleute aus Wirtschaft und Wissenschaft haben dazu im Jahr 2018 im „Praxisnetz Digitale Technologien“ des VDE innerhalb eines Workshops u.a. Szenarien für die Mensch-Maschine-Schnittstelle, Mixed Reality und KI erarbeitet. Diese Szenarien haben übergreifende Handlungsbedarfe für den VDE zur Gestaltung des Arbeitsplatzes in der Fabrikhalle im Jahr 2030 aufgezeigt:

- die Gestaltung der Mensch-Maschine-Schnittstelle für neue digitale Technologien
- die ethische Fundierung des Einsatzes von KI an der Mensch-Maschine-Schnittstelle
- die Erweiterung des Begriffs der funktionalen Sicherheit. Wenn sicherheitsgerichtete Steuerungen ihre Sicherheitsfunktionen zuverlässig erbringen, sprechen wir von funktionaler Sicherheit (FuSi).
- die Qualifikation von Beschäftigten, um der Technologieentwicklung gerecht zu werden.

Alle diese Themen wurden im Jahr 2019 vom VDE aufgegriffen, u.a. mit der internationalen Initiative OCEANIS, die von VDE|DKE und IEEE initiiert wurde (<https://ethicsstandards.org/>).

### VDE-Expertengruppe „Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle 2030“

Um die daraus resultierende Aufgabe den steigenden Bedarf für eine reibungslose Mensch-Maschine-Interaktion und Gestaltungsanforderungen zur digitalen Technik für zukünftige Fabriken decken zu können, hat der VDE im Herbst 2018 die Expertengruppe „Gestaltungsregeln für digitale Technologien an der Mensch-Maschine-Schnittstelle in der Fabrikhalle 2030“ gegründet. Seitdem wurden acht Zukunftsszenarios mit zugehörigen Gestaltungsbedarfen, -regeln, sowie Lösungsansätze und Anforderungen erstellt (siehe Abbildung 1).

**Jeder der Zukunftsszenarios wird von Themenpaten\*innen aus der VDE-Expertengruppe betreut. Dies sind:**

- Prof. Dr. Thomas Hermann, Ruhr-Universität Bochum
- Dr. Stefan Lutherdt, TU Ilmenau
- Univ.-Prof. Dr.-Ing. Verena Nitsch, IAW, RWTH Aachen
- Dipl.-Ing. Katrin Schilling, WZL, RWTH Aachen;
- Dr.-Ing. Matthias Vette-Steinkamp, Hochschule Trier
- Dr.-Ing. Tanja Döring, Univ. Bremen
- Dr. Jürgen Klippert, IG Metall
- Enterprise Architect & UX Manager André Noubissi, Continental
- Jun.-Prof. Dr.-Ing. Benjamin Weyers, Universität Trier



## Was sind Gestaltungsregeln?

Für eine gute Mensch-Maschine-Schnittstelle sind Gestaltungsregeln mit einer der wichtigsten Instrumente für die Entwickler. Diese können Normen und Standards zur Mensch-Computer-Interaktion sein, sowie die in der Überarbeitung befindliche Maschinen-Richtlinie, De-Facto-Standards und HMI-Styleguides. Aus diesen können die Entwickler und Designer anhand der Anforderungen und Vorgaben ästhetische, konsistente und intuitive Mensch-Maschine-Schnittstellen entwickeln.



## 3. Zukunftsszenarios

Wenn in den folgenden Steckbriefen der Einfachheit halber die männliche Form verwendet wird, ist darin die weibliche und diverse Form immer miteingeschlossen.



Abbildung 1: 8 Szenarios MMS



## Szenario 1 „Kooperation mehrerer Nutzer in unterschiedlichen Rollen beim Maschineneinsatz“

### Patenschaft

Prof. Dr. Thomas Hermann, Ruhr-Universität Bochum; Dr. Stefan Lutherdt, TU Ilmenau

### Szenario

Im vordigitalen Industriezeitalter hatte eine Fachkraft die Verantwortlichkeit und den Zugriff für den Maschineneinsatz; die Maschine war von äußerlichen Einflüssen abgeschottet. Durch die heutige Digitalisierung findet eine anwachsende Vernetzung der Maschinen und Systeme statt. Dies bedeutet, dass immer mehr Akteure gleichzeitig mit unterschiedlichen Verantwortlichkeiten auf die Maschinen zugreifen können. Dies erschwert Ausführung, Zuordnung, Koordinierung und Nachvollziehbarkeit von Entscheidungen und Aktionen innerhalb des Systems. Hier muss eine koordinierte Kooperation zwischen dem Hersteller der Maschine, den Anbietern von begleitenden Dienstleistungen (z.B. Inbetriebnahme, Wartung und Reparatur der Maschine) und von Diensten (z.B. die Durchführung von Software-Updates oder die Auswertung und Visualisierung der Betriebsdaten) sowie den Betreibern und Bedienern der Maschine hergestellt werden.

### Gestaltungsbedarfe

- **Datenhoheit an Betriebsdaten**  
Gleichzeitig zu der Nutzung der Rohdaten, sollte überlegt werden, wer die Rechte an den Daten innehat. Die Betreiber haben ein großes Interesse, die Rohdaten analysieren zu lassen, der Hersteller möchte wissen, wie sich seine Maschinen im Betrieb verhält, vorausschauende Wartung. Das Zuordnen, Koordinieren und die Nachvollziehbarkeit innerhalb einer Industrie 4.0-Lösung würde durch Auswertung der Daten einen reibungsloseren Ablauf gestatten, sowie auch neue Geschäftsfelder ermöglichen.
- **Systemhoheit**  
Bei Software-Updates und Remote-Wartung kann die Hoheit z.T. wieder an den Hersteller zurückgehen. Bei der Fernsteuerung und -wartung von Maschinen kann es zu Bedienkonflikten kommen: Wer hat dann die Verantwortung: der Betreiber, der Hersteller oder ein autonomes (KI-basiertes) Softwaresystem?
- **Einsatz von Künstlicher Intelligenz (KI)**  
KI in Maschinen hat ein großes Potential. Es kann in allen Bereichen der Wertschöpfungskette einer Industrie eingesetzt werden. Doch wer ist letztendlich Inhaber der Systemkontrolle? Entscheidet im Konfliktfall immer der Mensch, z.B. ein Mitarbeiter der Leitzentrale?
- **Kooperation, Koordination und der Wissensaustausch zwischen den Akteuren**  
Der gleichzeitige Zugriff auf die Maschine durch mehrere Nutzer fordert eine Klärung der Verantwortlichkeiten und Nutzungsrechten. Wer darf was, und zu welcher Zeit?
- **Wahrnehmbarkeit und Einsicht in den Systemzustand eines autonom agierenden Systems**



Wenn der Mensch Hand in Hand mit einem autonom agierenden System arbeitet, dann sollte die technologischen Begebenheiten so geschaffen sein, dass der Mensch zu jeder Zeit über die Handlung und Entscheidungen der Maschinen Kenntnis haben kann.

- **Modellierung des menschlichen Verhaltens**  
Damit die Maschine tatsächlich Absichten, Einschränkungen und Möglichkeiten des Menschen berücksichtigen kann, sollten diese erfasst werden und dem System beigebracht werden. Die Maschinen sollten so gestaltet sein, dass sie Rücksicht auf die Gefühlslage des Menschen nehmen kann.
- **Übersteuerung des Menschen durch die Maschine**  
Abwägung, wann eine Übersteuerung des Menschen durch die Maschine sinnvoll oder sogar notwendig ist. Dabei müssten auch ethische Fragestellungen („moralisches Dilemma“) betrachtet werden.

## Gibt es zu dem Szenario bereits Gestaltungsregeln oder andere Lösungsansätze?

- **DIN EN 614-2 „Sicherheit von Maschinen - Ergonomische Gestaltungsgrundsätze - Teil 2: Wechselwirkungen zwischen der Gestaltung von Maschinen und den Arbeitsaufgaben“**  
Im zweiten Teil dieser Normenreihe liegt das Gewicht auf der Wechselwirkung zwischen der Gestaltung der Maschinen und den Arbeitsaufgaben der Anwender. Sie hilft dem Konstrukteur, die ergonomischen Grundsätze bei der Gestaltung der Maschine anzuwenden. Sie beinhaltet unter anderem die Grundsätze der Aufgabengestaltung, Merkmale gut gestalteter Arbeitsaufgaben der Anwender und die Methodik der Aufgabengestaltung in Bezug auf die Gestaltung von Maschinen. Um ein besseres Verständnis für die Anforderungen aufzubauen, ist ein erläuterndes Beispiel „Gestaltung einer Bohrmaschine“ enthalten.
- **8 Heuristiken für soziotechnisches Systemdesign [x]**  
Um Industrie 4.0-Konzepte in Bezug auf soziotechnische Zusammenspiele unter den Bedingungen digitalisierter industrieller Arbeit einschätzen zu können wurden die 8 Heuristiken (Daumenregeln) erstellt. Sie unterstützen Industrie 4.0 – Lösungen zu analysieren und besonders kritische Aspekte der Systemgestaltung hervorzuheben. In dieser Methodik wird der Mensch in der Gestaltung der technischen Anwendung zentriert, besonders in der Zusammenarbeit zwischen Mensch und Maschine. (<http://heuristics.iaw.rub.de/#!/heuristics> (Set v1.4 | August 2019 | T. Herrmann et al))

*Abbildung 2: 8 Heuristiken für soziotechnisches Systemdesign [x] Prof. Dr.-Ing. Thomas Herrmann, Institut für Arbeitswissenschaften, Ruhr-Universität Bochum*

- **Bekannte Rollenkonzepten aus IT-Systemen**  
Ein möglicher Lösungsansatz für die Spezifikation der arbeitsorganisatorischen Rollenmodelle ist die Anlehnung an die aus IT-Systemen bekannten Rollenkonzepten (wie sie etwa bei Workflowmanagementsystemen verwendet werden).
- **Modellierung von Industrie 4.0-Geschäftsmodellen**  
Ein möglicher Ansatz für die inhaltliche Beschreibung der Rollen (Hersteller, Vermieter, Dienstleister etc.) sind Verfahren für die Modellierung von Industrie 4.0-Geschäftsmodellen, z.B. von der Univ. Paderborn (GEMINI) oder der Univ. St. Gallen (Business Model Navigator).



- **Cobot-Ansätze**  
Ein kollaborativer Roboter (Cobot) ist in einem Produktionsumfeld nicht durch eine Schutzvorrichtung von den Mitarbeitern getrennt. Eigene Sensoren erlauben es dem Roboter gemeinsam mit dem Menschen zusammen zu arbeiten und diesen nicht zu verletzen. Dies wäre ein möglicher Lösungsansatz für die Kollaboration zwischen Mensch und Maschine. Cobots sollten DIN EN ISO 10218 zertifiziert sein (Industrieroboter - Sicherheitsanforderungen - Teil 1: Roboter [x]).
- **Crew Resource-Management (CRM)**  
Für die inhaltliche Rollendefinitionen und ihr Training wäre ein möglicher Lösungsansatz, das in der Luftfahrt etablierte „Crew Resource-Management“ (CRM), das insbesondere die situativ angemessene und eindeutige Kooperation und Kommunikation der Flugzeugbesatzungen beschreibt und trainiert.

## Relevante Stakeholder

Juristen (Vertragsrecht, Haftungsrecht), Informatiker, Organisationswissenschaftler

## Szenario 2 “Umgang mit selbstlernenden, adaptiven Systemen am Beispiel von Servicerobotern“

### Patenschaft

Prof. Dr.-Ing. Verena Nitsch, IAW, RWTH Aachen; Dipl.-Ing. Katrin Schilling, WZL, RWTH Aachen; Dr.-Ing. Matthias Vette-Steinkamp, Hochschule Trier

### Szenario

Das Zukunftsszenario beinhaltet den Einsatz selbstlernender, autonomer, fahrerloser Transportsysteme (FTS) als Serviceroboter in der Produktion:

Das FTS folgt keinen festen Fahrspuren, sondern bewegt sich frei und eigenständig auf den Verkehrswegen in der Fabrikanlage und hält dabei die betrieblichen Verkehrsregeln ein. Es weicht Menschen und Objekten auf seinem Fahrweg aus und umfährt bzw. überholt sie, wenn dies sicher möglich ist. Wenn notwendig, wird dafür auch kurzfristig der Verkehrsweg verlassen. Je nach Last und Verkehr hält das FTS eine angemessene Geschwindigkeit ein. Das Fahrzeug nimmt seine Transportgüter eigenständig von Logistikflächen auf und stellt sie ebenso eigenständig auf den entsprechenden Zielflächen ab.

In diesem Szenario sind neben dem FTS weitere Objekte und Menschen enthalten. Diese sollten in die Ableitung von Anforderungen miteinbezogen werden.

### Gestaltungsbedarfe

- **Wissen des Serviceroboters über die Handlung des Menschen**  
Da der Serviceroboter in der Produktionsumgebung dem Menschen immer näherkommt, sollte dieser wissen, wie der Mensch handelt, um darauf reagieren zu können. Dazu gehört auch die „implizite Interaktion“, die der Serviceroboter aus dem situativen Kontext ableitet.





- **Transparenz und Nachvollziehbarkeit (explainability)**  
Es ist erforderlich, dass alle Menschen in der Fabrikhalle einen einheitlichen Informationsstand zum Verhalten der Maschinen besitzen. Dabei soll die Transparenz und Nachvollziehbarkeit (explainability) des Zustandes, der Handlungsoptionen und des Handelns des Serviceroboters als Information den Mitarbeitern bereitgestellt werden.
- **Sicherheitsgrenzen bei selbstlernenden Mensch-Technik-Interaktionen**  
Bei selbstlernenden Mensch-Maschinen-Interaktionen müssen unabwendbare Sicherheitsgrenzen eingehalten werden.
- **Rückfluss von menschlichem Expertenwissen in das System**  
Es sollten technische Möglichkeiten vorhanden sein, damit menschliches Expertenwissen in die Systeme eingebracht werden können.
- **Transparenz absoluter Grenzen („Roboter-Knigge“)**  
Die Präsenz der Technik darf keine Angst machen, sie sollte „emotional verträglich“ sein. Die absoluten Grenzen („Roboter-Knigge“) müssen eine eindeutige Transparenz aufweisen.
- **Rücksichtnahme des Menschen auf die Maschine**  
Da der Mensch genauso unbedacht mit der Maschine umgehen kann, wie umgekehrt, ist auch hier für den Menschen ein „Knigge“ notwendig. Zum Beispiel sollte der Mensch nicht bewusst in den Fahrweg eines FTS treten.
- **Wahrung der Privatheit der Anwender**  
Daten der Anwender und ggf. von Dritten, die von Sensoren der Service-Roboter erfasst werden, müssen vor unberechtigtem Zugriff bewahrt werden.
- **Qualitative Abgrenzung von KI gegenüber anderen Automatisierungstechniken**  
Hier ist die Überlegung, ob das Spannungsfeld zwischen Vorhersehbarkeit und Adaptivität beim Einsatz von KI tatsächlich neu ist oder ob unter Umständen auf bestehende Erkenntnisse zurückgegriffen werden kann.
- **Fähigkeiten zum Greifen, Be- und Entladen**  
Neben der eigentlichen Fahrfunktion sollten Fähigkeiten zum Greifen, Be- und Entladen als Gestaltungsbedarfe aufgenommen werden, da diese einen erhöhten Mehrwert erzeugen könnten.
- **Schutz vor unautorisierten IT-Zugriff**  
Bei sicherheitskritischen Anwendungen sollte der Schutz vor unautorisiertem IT-Zugriff von außen gegeben sein. Hier ist die Sorgspflicht der Hersteller zu beachten.

**Mit dem Schritt vom klassischen FTS zum autonomen Serviceroboter in der Fabrikumgebung wurden weitere Aspekte in die Diskussion einbezogen:**

- „Emotional verträgliche“ Formgebung und Verhalten autonomer Systeme.
- Festlegung neuer Konventionen, beispielsweise farbige Beleuchtung, bestimmte Tonfolgen etc.
- Gestaltungsregeln für die Explorierbarkeit des Verhaltens eines FTS.
- Neue Arten, mit denen ein Mensch einem FTS „zeigen“ kann, wie es agieren soll, d.h. neue Formen der Programmierung bzw. des Trainings.
- Formulierung von Design Patterns als kleinste Verhaltenseinheit.





Hier ist zu beachten, dass die Ausgestaltung von neuen Kommunikationskonventionen erheblich von den Vorlieben und Vorkenntnissen der Benutzer abhängt

## Gibt es zu dem Szenario bereits Gestaltungsregeln oder andere Lösungsansätze?

- **IEC 61508 „Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme“**  
Die internationale Normreihe dient der Entwicklung von elektrischen, elektronischen und programmierbaren elektronischen (E/E/PE) Systemen, die eine Sicherheitsfunktion ausführen.
- **DIN ISO/TS 15066 „Sicherheitsanforderungen an kollaborierende Industrierobotersysteme und die Arbeitsumgebung“**  
Die technische Spezifikation von Robotern und Robotikgeräten legt Sicherheitsanforderungen an kollaborierende Industrierobotersysteme und deren Arbeitsumgebung fest. Sie ergänzt die in ISO 10218-1 und ISO 10218-2 „Industrieroboter – Sicherheitsanforderungen“ angeführten Anforderungen und Anleitungen zum Betrieb von kollaborierenden Industrierobotern. Für Roboter außerhalb des Industriebereiches ist sie nicht gültig.
- **ISO 9241 Ergonomie der Mensch-System-Interaktion**  
Der internationale Standard beschreibt Richtlinien der Mensch-Computer-Interaktion. Die Norm beinhaltet Anforderungen an die Arbeitsumgebung, Hardware und Software. Der Anwender soll somit vor gesundheitlichen Schäden geschützt und die Ausführung seiner Tätigkeit erleichtert werden.
- **Forum FTS (VDI 2510, VDI 2710, VDI 4450 ersetzt durch VDI 2710 Blatt 4)**  
Der VDI-Fachausschuss „Fahrerlose Transportsysteme (FTS)“ besteht seit 1986. Die in dem Ausschuss erarbeiteten VDI-Richtlinien VDI 2510 „Fahrerlose Transportsysteme (FTS)“, VDI 2710 „Ganzheitliche Planung von Fahrerlosen Transportsystemen (FTS)“ und VDI 2710 Blatt 4:2011-07 „Analyse der Wirtschaftlichkeit Fahrerloser Transportsysteme (FTS)“ stellen Informationen zu fahrerlosen Transportsystemen zur Verfügung. Die VDI-Regelwerke sind sehr praxisorientiert.

## Relevante Stakeholder

- Hersteller von Servicerobotern (auch von autonomen Fahrzeugen im PKW- und Bahnbereich)

## Szenario 3 „VR/AR in Wartung und in Assistenzsystemen“

### Patenschaft

- Jessica Fritz (VDE)



## Szenario

Bei Störungen an einer Maschine bekommt der Werker vor Ort per AR-Brille Hinweise und Informationen für die Beseitigung der Störung eingeblendet. Diese Informationen kommen von einem spezialisierten Techniker aus einer technischen Zentrale. Dabei bewegt sich dieser in der VR und hat für die Analyse der Problematik den digitalen Zwilling des gestörten Gerätes vor sich. Für beide Anwender in diesem Szenario erfolgt die Steuerung des Assistenzsystems mit Hilfe von Sprachbefehlen, Gesten und Eyetracking.

Das Szenario beinhaltet den Einsatz von Virtual Reality (VR) bzw. Augmented Reality (AR) in der Wartung. Es geht sowohl um eine intelligente Unterstützung vor Ort bei Ausnahmesituationen als auch um den entfernten (remote) Servicetechniker, der sich virtuell im Umfeld des Wartungsfalls bewegt (aufsetzend auf einem digitalen Zwilling)

Themen sind:

- die Auswirkungen von VR auf die Arbeitsprozesse. Interaktionsformen und kognitive Belastungen eines persönlichen Treffens sind andere als bei einer Kollaboration über VR oder einem einstündigen Video-Telefonat. Prinzipien der persönlichen Interaktion lassen sich nicht eins zu eins auf die digitale Interaktion übertragen und umgekehrt. Der VR-Einsatz hängt stark vom jeweiligen Arbeitskontext ab
- das Vermeiden physischer und psychischer Belastungen, wie z.B. schnelle Ermüdung
- die Adaption der Unterstützung an das Vorwissen und den Bedarf bzw. Wunsch nach Unterstützung durch den Anwender („gestaffelte Assistenz“)
- der Umgang mit dem Arbeitnehmerdatenschutz. Bei VR-Systemen fallen notwendigerweise große Datenmengen zum Verhalten der Bediener an.

## Gibt es zu dem Anwendungsbeispiel bereits existierende Regelwerke?

- VDI 3633 - Blatt 11 (Simulation und Visualisierung):
- VDC Fellbach: 3D Guide (<https://www.3d-guide.net/>): praxisnahe Design-Pattern für Desktop-, VR-, AR-Anwendungen, konkrete Umsetzung und qualitative Bewertungen

## Relevante Stakeholder

- Hardware-Hersteller

## Szenario 4 „Umgang mit Unterschieden zwischen Kulturen, Milieus, Generationen, Bildungsniveaus, Überforderungsschwellen“

## Patenschaft

Dr.-Ing. Tanja Döring, Univ. Bremen; Dr.-Ing. Stefan Lutherdt, TU Ilmenau



## Szenario

Unsere Fachleute haben sich die Frage gestellt, wie die Zukunft aussehen könnte, wenn mehrere Nutzer mit unterschiedlichen Bedürfnissen eine Maschine bedienen und wie das Einarbeiten auf ein neues System gestaltet werden kann:

„In der digitalen Fabrik der Zukunft kann die Fachkraft z.B. die Mensch-Maschine-Schnittstelle der Maschine individuell anpassen und von einer bekannten Maschine an eine andere, neue Maschine „mitnehmen“. Eine bestimmte Qualifikation vorausgesetzt, kann sie auch Sicherheitsvorkehrungen aufheben, die für ungeübte Bediener gedacht sind. Für Geräte unterschiedlicher Hersteller kann sie sich eine ähnliche Bedienoberfläche einstellen, um Umstellungen und Fehlbedienungen zu vermeiden. Bei der Erstbedienung einer Maschine wird der Werker interaktiv und spielerisch in die Bedienelemente eingeführt, wie er es von seinem Videospiele an der Spielkonsole gewohnt ist. Dabei kann er bekannte Strategien oder das look-and-feel von anderen genutzten HMI portieren und integrieren.“

Das Szenario beinhaltet die Anpassung der Mensch-Maschinen-Schnittstelle (MMS) an die Anforderungen, die aus den unterschiedlichen Kontexten der Anwender entstehen.

## Gestaltungsbedarfe

- **Einheitlichen Bedienkonzepten**  
Aktuell liegen sehr viele unterschiedlichen Hersteller-Spezifische Bedienkonzepte vor. Dies erhöht die Einarbeitung der Bediener und Fehleranfälligkeit an den Maschinen. Einheitliche Bedienkonzepte würden zur Effizienzsteigerung beitragen und könnten die Einarbeitung auf Systeme von unterschiedlichen Herstellern verbessern.
- **Individuelle Nutzerschnittstelle**  
Bisher sollte der flexible Bediener sich der unflexiblen Maschinen anpassen, doch äußert sich dies in Ablehnung, Stress und erhöhten Schulungszeiten. Dem könnte man mit individuell angepassten Nutzerschnittstellen entgegenwirken. Dies kann aber unter Umständen mit dem Arbeitsschutz, Sicherheitsnormen und weiteren Regelungen kollidieren und muss bei der Gestaltung der MMS mitbedacht werden. Mit einer individuellen Nutzerschnittstelle könnte z.B. auf rot-grün Sehschwächen, Nutzern mit eingeschränkter Bewegungsfreiheit und Kulturunterschiede individuell eingegangen werden.
- **Akzeptanz für neue digitale Technologien**  
Die Fähigkeiten eines Mitarbeiters dürfen nicht über- und nicht unterschätzt werden. Angst vor Fehlern bei neuer Technologie führt häufig zu Stress und Ablehnung. Dem kann durch unterschiedliche Maßnahmen, wie zum Beispiel Erfolgserlebnisse an der Maschine, positive entgegenwirkt werden. Hier sollte identifiziert werden, welche unterschiedliche Akzeptanzstufen ein Nutzer hat und wie diese positive, sowie negativ beeinflusst werden können.
- **Lerngewohnheiten bei der Technikeinführung**



Damit der Nutzer schnell in ein neues System eingearbeitet werden kann, ist zu analysieren, welche Unterschiede bei den Lerngewohnheiten der Nutzer vorliegen. Diese sind bekanntermaßen unterschiedlich und hängen von der individuellen Informationsaufnahme und -verarbeitung des Menschen ab. Dies kann bedeuten, dass die Lerngewohnheiten in die Gestaltung der Bedieneroberfläche mit integriert werden können (z.B. schrittweise Freigabe der Funktionsinhalte mit Einführung in das System). Dem Bediener der Maschine, muss es ermöglicht werden „Herr der Lage“ zu werden.

- **Unterschiedlichen Konventionen (Kultur, Alter, etc.) für die Gestaltung digitaler Nutzerschnittstellen**

Es liegen zum Teil erhebliche Unterschiede bei den Konventionen für Farbgebung, Fähigkeiten, Lerngewohnheiten, aber auch Grundhaltung zum Datenschutz vor. Zur Gestaltung der Mensch- Maschinen-Schnittstelle ist es notwendig diese Unterschiede zu kennen und diese in den Arbeitsprozess einbinden.

- **Grundhaltungen gegenüber dem Datenschutz**

Durch die Internationalisierung wird der Datenschutz unterschiedlich gehandhabt. Wie mit den Daten eines Unternehmens umgegangen wird, gehört jedoch Themenübergreifend behandelt und fällt aus dem direkten Scope des Szenarios.

## Gibt es zu dem Szenario bereits Gestaltungsregeln oder andere Lösungsansätze?

- **Betriebsverfassungsgesetz, Arbeitsschutzgesetz:**

Die Inhaltliche Übertragung der beiden Regelwerke wird als nur teilweise geeignet für eine inhaltliche Übertragung in die digitale Fabrik angesehen. Beide weisen Anpassungsbedarf auf, besonders das Arbeitsschutzgesetz im Bereich der psychischen Belastung sollte dem Stand der Arbeitswissenschaften bei der Gefährdungsanalyse (physische und psychische Gefährdung) angepasst werden.

- **Einsatz von partizipativen Entwicklungsmodellen, z.B. Design Thinking**

Es ist wichtig den zukünftigen Nutzer der Systeme in den Entwicklungsprozess mit einzubeziehen, um seine Bedürfnisse und Fähigkeiten frühzeitig in das Bedienkonzept aufnehmen zu können. In vielen Bereichen der partizipativen Entwicklungsmodellen sind Vorgehensweisen übertragbar, um Nutzer der Maschinen und weitere Stakeholder in die Gestaltung der MMS zu integrieren.

- **Aufnahme von Gefährdungsanalyse in die Entwicklungsmodelle**

Eine der zentralen Anforderungen des Arbeitsschutzgesetzes ist die Durchführung einer sogenannten Gefährdungsbeurteilung durch den Arbeitgeber. Durch diese Maßnahme sollen zukünftige mögliche Gefährdungen abgestellt werden. Durch Übertragung von diesem Ansatz in die Entwicklung der Mensch-Maschine-Schnittstelle können spätere Gefahren und Stress bei den Mitarbeitern minimiert werden.

- **Methoden des Partizipativen Designs (z.B. im Rahmen von Gesture Elicitation Studies)**



Die in der Praxis mittlerweile etablierten Guidelines für barrierefreie Benutzerschnittstelle können als Vorbild für die Gestaltung „kulturfreier“ Benutzerschnittstelle dienen.

- **Adaptiven Bedienkonzepten von Consumer-Produkten**  
Eine Orientierung an den häufig adaptiven Bedienkonzepten von Consumer-Produkten wird kontrovers in der Onlinebefragung bewertet und als nicht gut anwendbar bewertet.
- **In einer Online-Befragung wurden weitere Lösungsvorschläge genannt:**
  - Alle bewährten Methoden auf ihre Nutzbarkeit bewerten und daraus eine strukturierte Methodik entwickeln. Dadurch können gute Ideen vereint und über Links zugänglich gemacht werden (z.B. XML Maschine als Lernunterstützung).
  - Orientierung an den in Arbeit befindlichen VDI-Richtlinien.

## Relevante Stakeholder

- Psychologen, Neurowissenschaftler

## Szenario 5 „Die Maschine als Lernort“

### Patenschaft

Dr. Jürgen Klippert, IG Metall

### Szenario

Durch die Digitalisierung wird in der zukünftigen Fabrik der Fachkraft das Fachwissen zur Maschinenbedienung anders vermittelt. Was heute der Lehrmeister ist, dem man über die Schulter schauen kann, ist es morgen die Maschine selbst, die den Prozess der Vermittlung des Fachwissens übernimmt. Meist fallen die Wissensträger weg, da sich durch die Digitalisierung die Anforderungen an den Arbeitsbereich geändert haben und somit das Wissen zur Bedienung der Maschine veraltet ist.

Der Bediener erlernt den Umgang mit der Maschine an ihr selbst und wird dabei von adaptiven Assistenzsystemen geführt. Je nach seinem Lernstand und seinen Fähigkeiten hat er Zugriff auf die einzelnen Funktionen. Ein wichtiges Lernziel ist der Erwerb eines Grundlagenwissens zu Mechanik, Elektronik und Softwaretechnik, um über das reine Einsatzwissen hinaus auch ein Grundverständnis für die Abläufe und Zustände der Maschine zu entwickeln. Vermittelt wird aber nicht nur das abstrakte Wissen zur Funktionsweise der Maschine, sondern auch die Fähigkeit, sie tatsächlich in allen Zuständen richtig führen und bedienen zu können, um auch in Ausnahmesituationen angemessen reagieren zu können. Dieses Erfahrungswissen hat einen eigenen, anerkannten Wert und wird auch durch Zertifikate bestätigt.

### Gestaltungsbedarfe

Beim Anpassen der Bedarfe muss darauf geachtet werden, dass die unterschiedlichen kulturellen Aspekte der Firmen mit in den Prozess eingebunden werden.

- **Ausreichende Inanspruchnahme des Menschen**



Trotz der Adaptivität und Automatisierung ausreichende Inanspruchnahme des Menschen, damit dieser sein Wissen, sowie seine Fähigkeiten weiterentwickeln kann, um in Ausnahmesituationen gegebenenfalls angemessen reagieren zu können

- **Differenzierung zwischen Grundlagewissen und Einsatzwissen**

Differenzierung zwischen breitem Grundlagenwissen, das ein Systemverständnis erlaubt, sowie dem konkret notwendigen Einsatzwissen.

- **Gestaltung der Technikeinführung in den konkreten Betrieb**

Die Gestaltung der Technikeinführung in den konkreten Betrieb sollte geplant werden. Können z.B. kollaborierende Roboter (Cobots) zu schnell ohne Gewöhnung eingeführt werden? Welche Risiken drohen bei einem nicht angemessenen, missbräuchlichen Technikansatz. Wie kann der Mensch für neue Technologien sensibilisiert werden?

- **Analysieren der Technikablehnung durch den Mitarbeiter**

Damit der Nutzer schnell in ein neues System eingearbeitet werden kann, ist zu analysieren, wieso ein Mensch die Technologie ablehnen könnte. Welche Maßnahmen kann man treffen, um diesen entgegen zu wirken.

- **Lösung des Konflikts Privacy und Big Data - Analyse**

Lösung des Konflikts zwischen Privacy und Big Data-Analysen von Betriebsdaten mit Personenbezug, die bei der Maschinenbedienung oder bei Assistenzfunktionen entstehen. Bisher ist noch unklar, wie die Regeln der DSGVO und des Arbeitnehmerdatenschutzes, inklusive der Leistungs- und Verhaltenskontrolle, anzuwenden sind. Z.B. ist die Frage zu klären, ob es ein Recht auf Löschung geben sollte. Und Wie sollen abgeleitete Daten bewertet werden? Ist im Konfliktfall ein nachträgliches „Umlernen“ der KI-Verfahren notwendig?

- In der **Online-Umfrage** wurde vor allem der „Sensibilisierung des Menschen für neue Technologien“ eine hohe Relevanz zugeordnet. Der „Konflikt zwischen Privacy und Big Data-Analyse“ fällt dagegen ab. Durch die Bewertung der weiteren Gestaltungsbedarfen und den Antworten der Teilnehmer ist zu erkennen, dass der Mitarbeiter stark in die Einführung der Technologien eingebunden werden und ihm Handlungsmethodik übermittelt werden sollten, damit eine Überforderung und Ablehnung der neuen Technologien verringert werden kann. Ein Teilnehmer stellt klar, dass die Technikeinführung ein häufig unterschätzter und vernachlässigter Erfolgsfaktor ist. Als weitere Gestaltungsbedarfe wurden folgende Punkte benannt:

- Es sollten menschliche Vorbilder, die pädagogisch geschickt Wissen und Erfahrung weitergeben können, in den Prozess der Maschineninstruktion eingebunden werden (Mensch als soziales Wesen).
- Es sollte ein Methodenmix in Betracht gezogen werden. Und die intrinsische Motivation des Menschen mit einbezogen werden.
- Der Begriff „ausreichende Inanspruchnahme“ sollte schärfer geklärt werden. Hier besteht F&E-Bedarf.

**Gibt es zu dem Szenario bereits Gestaltungsregeln oder andere Lösungsansätze?**



- **Ansatz der „Lern- und kompetenzförderlichen Arbeitsgestaltung“ des BiBB:**  
(<https://www.bibb.de/veroeffentlichungen/de/bwp/show/1328>)
- **VDI-Richtlinie VDI/VDE MT 7100 „Lernförderliche Arbeitsgestaltung - Ziele, Nutzen, Definitionen“**  
[https://www.vdi.de/nc/richtlinie/?tx\\_wmdbvdirilisearch\\_pi1%5Brpro\\_id%5D=7157&cHash=26a9c5abd3193bbf050c48f783b72aa4](https://www.vdi.de/nc/richtlinie/?tx_wmdbvdirilisearch_pi1%5Brpro_id%5D=7157&cHash=26a9c5abd3193bbf050c48f783b72aa4)
- **Assistenzsysteme für ungeübte Mitarbeiter**  
Assistenzsysteme für ungeübte Fachkräfte, die das Wissen der Fachkräfte unterbrechen und somit einen gestaffelten Funktionsumfang mit gestaffeltem Durchgriff auf die physischen Aktionen der Maschine zur Verfügung stellt.
- **Strukturierte Aufbereitung von Fehlerauswertungen, damit die Fachkraft ihre Arbeitsweise anpassen kann, um zukünftige Fehler zu vermeiden.**  
Die fehlenden Datenerhebungen zur Anpassung des Ausbildungsbedarfs sind heute zum Teil schon vorhanden, aber nicht ausreichend. Es gibt schon Systeme, die der Fachkraft Fehlerauswertungen übermitteln, damit dieser seine Arbeitsweise anpassen kann, um den Fehler zu vermeiden.
- **Lernsysteme in der Airbus-Ausbildung**  
Als Vorbild könnte das digitale Übungs-Tutorial von Airbus dienen, mit dem dort Auszubildende und Fachkräfte trainieren.

## Relevante Stakeholder

- Airbus
- Partner des Forschungsprojekts APPSIST, z.B. Festo Lernzentrum Saar GmbH Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
- Gemeinsame Arbeitsstelle Ruhr Universität Bochum und IG Metall
- Ressort Zukunft der Arbeit beim IGM Vorstand, Dr. Jürgen Klippert

## Szenario 6 „Intelligente Unterstützung in Ausnahmesituationen“

### Patenschaft

Dr.-Ing. Matthias Vette-Steinkamp, André Noubissi

### Szenario

In einem Betrieb mit Maschinen, Anlagen, Fahrertransportsystemen und Robotern kann es immer zu unvorhergesehenen Störungen und Ausnahmesituationen kommen. In diesen Fällen müssen die anwesenden Bediener und Wartungstechniker unter Zeitdruck, unter hohem Stress und häufig mit unvollständigen Informationen zur Fehlerursache und zum Systemzustand schnell und angemessen reagieren, um Gefahr für Leib und Leben abzuwenden und um materielle Schäden zu verhindern. Je seltener die Fälle auftreten, desto weniger vorbereitet sind die Mitarbeiter in der Regel.





Assistenzsysteme könnten in solchen Fällen Unterstützung beim sachgerechten Umgang mit der Situation leisten.

## Gestaltungsbedarfe

- **Identifizierung der Ausnahmesituationen**  
Identifizierung, bei welchen Ausnahmesituationen der Bedarf so groß ist, dass eine gezielte Unterstützung notwendig ist. Wo benötigt der Mitarbeiter womöglich Unterstützung?
- **Fehleranalyse zur Gestaltung eines Assistenzsystems**  
Generierung eines Prozesses und Methoden, mit dem eine nachträgliche Fehleranalyse ermöglicht werden und die Erkenntnisse in das System zurückgespielt werden kann.

## Gibt es zu dem Szenario bereits Gestaltungsregeln oder andere Lösungsansätze?

- **Reihe DIN EN 614 „Sicherheit von Maschinen - Ergonomische Gestaltungsgrundsätze-Reihe“**
- **Flugzeugtyps-spezifischen Checklisten**  
Als Vorbild könnten die in der Luftfahrt etablierten flugzeugtypspezifischen Checklisten für den Standardbetrieb und Ausnahmefälle sein.
- **Reduktion der Bedienerschnittstellen auf eine vereinfachte Notfallsteuerung**  
Ein möglicher Lösungsansatz ist die Reduktion der regulären Bedienerschnittstelle auf eine vereinfachte Notfallsteuerung, die die Übernahme der Maschinensteuerung ermöglicht, aber trotzdem die kognitive Last des Bedieners verringert.
- Laut der **Online-Umfrage** ist die DIN EN - Reihe 614 „Sicherheit von Maschinen“ nur zum Teil bekannt. Die Personen, denen die Norm-Reihe bekannt ist, schätzen sie einmal als gut und einmal als sehr gering einsetzbar in der digitalen Fabrik an.  
Das Vorbild der Luftfahrt, etablierte flugzeugspezifische Checklisten zu verwenden, ist kaum bekannt. Es wird aber als ein gutes Mittel angesehen, die Arbeiten zu unterstützen. Die Begründungen gehen hier auseinander. Teilnehmer sind teilweise der Meinung, dass Checklisten in verschiedenen Branchen ihre Wirksamkeit unter Beweis gestellt haben und sollten systematisiert für MMS in der Zukunft aufbereitet werden. Auf der anderen Seite gibt es die Meinung, dass Checklisten zu viel Zeit benötigen und künftig die Daten- und Informationsaufbereitung intelligenter umgesetzt werden.

Der Lösungsansatz „Reduktion der regulären Bedienerschnittstelle auf eine vereinfachte Notfallsteuerung“ wird als ein guter Ansatz gesehen und in der Medizin, in Notfallsituationen, schon angewendet. Diese Erfahrungen und Lösungen sollten in die Überlegungen einbezogen werden. Ein Teilnehmer fügt hinzu, dass man sich in der Gestaltung auf die wesentlichen Anforderungen konzentrieren sollte, diese aber in Ausnahmesituationen schwierig vorherzusehen sind. Die definierten Anforderungen müssen in der Praxis kritisch geprüft werden.



**Als weitere Gestaltungsbedarfe ist genannt worden:**

- Startups mit einbeziehen, die mit intelligenten Lösungsansätzen Mensch und Maschine näher zusammenbringen.
- Orientierung an technischen Rückfallebenen etc. Konzepte wie Graceful Degradation sollten auf die Mensch-Maschine-Schnittstelle übertragen werden.“

## Relevante Stakeholder

Wurde noch nicht benannt.

## Szenario 7 „Smartwatches im Fertigungsumfeld“

### Patenschaft

André Noubissi, Dr. Stefan Lutherdt

### Szenario

Dieser Anwendungsfall beinhaltet den Einsatz von Smartwatches im Fertigungsumfeld.

Mitarbeiter im Fertigungsumfeld werden immer mobiler, Maschine und Systeme werden immer intelligenter und generieren immer mehr Informationen für die Mitarbeiter. Eine Smartwatch wird auch eingesetzt, um die Interaktion in solchen Fällen zu unterstützen.

Eine Smartwatch bietet den Vorteil, dass sie immer am Körper getragen wird und somit schnell erreichbar ist für den Mitarbeiter.

Die Smartwatch wird verwendet, um dem Mitarbeiter wichtige Mitteilungen zukommen zu lassen (von einfachen Informationen bis zu systemkritische Informationen wie maschinenstillstand). Smartwatches werden benutzt um Aufgaben zu verwalten ( Aufgabe annehmen, Aufgabe zuweisen , Aufgabenstatus verfolgen und ändern) und dies ermöglicht Fertigungsmitarbeiter die bestimmte Aufgaben in Gruppen erledigen sich sehr schnell zu organisieren , was zur Effizienzsteigerung führt (Z.B bei einem Maschinenstillstand ist schnell klar wer ist verfügbar und wer sich das Thema angenommen hat ).

Fertigungsprozessschritte, die eine einfache Interaktion verlangen (z.B: Barcode scannen, Foto aufnehmen) werden auch mittels Smartwatches absolviert und dokumentiert.

### Gestaltungsbedarfe

- **Vermeidung von physischer und psychischer Belastung**  
Smartwatches werden auch zum Teil privat von den Mitarbeitern genutzt. Deshalb sollte hier auf eine klare Unterscheidung geachtet werden.
- **Vibrationen am Handgelenk**  
Worauf muss geachtet werden? Die Smartwatches vibrieren mit jeder empfangenen Mitteilung.
- **Ergonomie: Bewegung vom Handgelenk, um Informationen zu lesen**



Worauf muss geachtet werden, wenn diese Bewegung zu oft gemacht wird z.B. bei zu viele Notifications am Tag.

- **Informationsüberflutung vermeiden**  
Anzahl von Informationen, Priorisieren von Informationen
- **Mitarbeiter entscheidet über Art der Informationswidergabe**  
Der Mitarbeiter sollte entscheiden können, wie er im Fertigungsumfeld informiert werden möchte (z.B. Vibration, Sound), da das Empfinden individuell ist und die Umgebung sehr laut sein kann.
- **Deaktivieren in bestimmten Zeiten**  
Z.B. könnten zu Pausenzeit die Smartwatches automatisch deaktiviert werden.
- **Umgang mit Mitarbeiterdaten**  
Z.B. Definition, wie man mit den Reaktionsdaten der Mitarbeiter umgehen sollte. Standort, Bewegungsdaten usw.
- **Hygienische Aspekte**  
Eine Smartwatch kann von mehreren Mitarbeiter getragen und von Schicht zu Schicht weitergegeben werden. Dadurch sollte unbedingt die hygienischen Aspekte in die Gestaltungsregeln miteinbezogen werden.
- **Gestaltung vom User interface**  
Das Interface der Smartwatch ist sehr klein und somit auch die Schriftgröße und Icons. Hier sollten die Guidelines, Patterns usw. dementsprechend definiert werden, damit der Mitarbeiter die Befehle verstehen kann.

## Szenario 8 „Gestaltungsrahmen für die Ethik von Künstlicher Intelligenz in der Produktionsumgebung“

### Patenschaft

Andreas Hauschke (VDE), Jessica Fritz (VDE)

### Szenario

Unsere Fachleute haben sich die Frage gestellt, wie die Zukunft und deren Gestaltungsbedarfe an einer Maschine für mehrere Nutzer aussehen könnte:

Ein Unternehmer möchte seine Fabrikhalle mit künstlicher Intelligenz, zum Beispiel Roboter und I4.0-Devices ausstatten. Damit seine Mitarbeiter die KI akzeptieren, sollte die neue Technologie im Produktionsprozess ethische Werte einhalten, wie beispielsweise Transparenz und Fairness. Durch eine gezielte Auswahl an Szenarien, die innerhalb der Fabrikhalle stattfinden, in Kombination mit menschlichem Verhalten, regulatorischen Vorgaben und ethischen Werten, wird ein bestimmtes Maß an ethischen Werten festgelegt, die die KI einhalten soll. Anhand dieser Maßstäbe werden die zukünftigen KI-Systeme eingekauft und in der Fabrikumgebung integriert.



In diesem Szenario sind neben der KI weitere Objekte und Menschen enthalten, unter anderem Entwickler, Betreiber, Mitarbeiter usw. Diese sollten in die Ableitung von Anforderungen und Bedarfe miteinbezogen werden.

## Gestaltungsbedarfe

Für das zukünftige Anwendungsszenario sind für die Entwicklung und den Einsatz ethischer Maßstäbe für die Gestaltung der KI die folgenden Bedarfe festgestellt worden (Die Liste stellt keinen Anspruch auf Vollständigkeit):



Abbildung 3: Gestaltungsbedarfe/-Werte für die Entwicklung und Einsatz ethischer Maßstäbe für die Gestaltung der KI

### KI soll den Menschen dienen und achten

Die KI wird nach und nach immer mehr Assistenzaufgaben übernehmen und den Mitarbeiter in seinen Tätigkeiten unterstützen. Hier muss aus technischer und sozialer Sicht darauf geachtet werden, dass z.B. die Grundrechte eingehalten werden. Gleichzeitig sollte das KI-System auf der Grundlage von Daten oder Modellen beurteilen können, wie der Mensch handelt, um drauf reagieren zu können. Es sollten Regeln geschaffen werden, dass die KI unter menschlicher Aufsicht steht und diesem zu jeder Zeit den Vorrang lässt.

### Maßnahmen für Sicherheit und Schutz gegen Angriffe

Der Aufbau der KI sollte sich so gestalten, dass sie einen hohen Grad an Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen aufweist. Genauso sollte es einen Auffangplan geben, wenn es zu einem Sicherheitsvorfall kommt. Dies sollte zu jederzeit zuverlässig funktionieren und reproduzierbar sein.

### Die Privatsphäre des Mitarbeiter muss geschützt werden

Daten der Mitarbeiter und ggf. von Dritten, die von der KI erfasst werden, müssen vor unberechtigten Zugriff bewahrt werden (Privatsphäre). Dies bedeutet unter anderem die Qualität und Integrität der Daten sowie Datenzugriffe zu achten.

### Rückverfolgbarkeit der KI muss sichergestellt werden

Um z.B. Fehlentscheidungen der KI ermitteln zu können und somit zukünftige Fehler zu vermeiden, sollte die Rückverfolgbarkeit sichergestellt werden. Für die Transparenz ist auch erforderlich, dass die technischen Prozesse, wie auch die menschlichen Entscheidungen erklärbar und nachvollziehbar



sind. Für den Mitarbeiter sollte auch zu jederzeit erkennbar sein, dass er mit einer KI zusammenarbeitet und kommuniziert.

### **Durch KI darf keine Diskriminierung oder Bevorteilung stattfinden**

Über den ganzen Lebenszyklus einer vertrauenswürdigen eingesetzten KI müssen Inklusion und Vielfalt garantiert sein. Für alle Mitarbeiter, die die Verantwortlichkeit für den Prozess haben, sollte der Zugang barrierefrei gestaltet sein. Sie sollten in den Integrationsprozess der KI mit eingebunden werden, um etwaige Benachteiligungen frühzeitig entgegenzuwirken.

### **KI sollte auf eine umweltfreundliche und sozialverträgliche Weise eingesetzt werden**

KI ist so einzusetzen, dass es nicht zu einer Beeinträchtigung von Mensch und Umwelt kommt. Es sollten Maßnahmen zur Sicherung der Umwelt- und Sozialverträglichkeit definiert werden.

### **Mechanismen zur Wahrung der Verantwortlichkeit und Rechenschaftspflicht**

Es werden Rahmenbedingungen benötigt, die die Rückverfolgbarkeit und Protokollierung der Prozesse und Ergebnisse des KI-Systems sicherstellt. Zusätzlich ist es notwendig negative Auswirkungen, zu minimieren und ein Meldesystem dafür zu erstellen.

## **Gibt es zu den Gestaltungsbedarfen bereits Gestaltungsregeln oder andere Lösungsansätze?**

### **KI-Ethik messbar machen**

Es liegen zwar bis zu diesem Zeitpunkt schon zahlreiche Sammlungen ethischer Prinzipien für die Gestaltung der KI vor, doch lassen sich diese aufgrund ihrer allgemeinen Formulierung nicht direkt in der Praxis anwenden. Dies wird aber immer dringender. Die Lösungen müssen umsetzbar und machbar für die Unternehmen sein. Hier setzt der VDE als Initiator und Leiter der Gruppe „AI Ethics Impact Group“ gemeinsam mit der Bertelsmann Stiftung an. In der vor kurzem veröffentlichten Studie „From principles to practice – an interdisciplinary framework to operationalise AI ethics“ zeigen VDE und die Bertelsmann Stiftung, wie sich Ethikprinzipien für KI in die Praxis bringen lassen. Mit der Kombination von drei Werkzeugen, dem WKIO-Modell, einem KI-Ethik-Label und einer Risikoklassifizierung zeigt die Studie beispielhaft, wie allgemein ethische Prinzipien messbar und konkret umsetzbar werden könnten.

### **Überblick: Berücksichtigung der ethischen Werte mittels Risikoklassen und Label**

Die Berücksichtigung der ethischen Werte kann auf Grundlage des Gestaltungsrahmens der AI Ethics Impact Group ablaufen. Hierfür wird zunächst mittels einer Risikomatrix das Anwendungsfeld der KI bewertet. Aus dieser ergeben sich im nächsten Schritt Anforderungen an das KI-System, die als Mindestniveaus in einem Labeling-System beschrieben werden. Dieses ist visuell an die Bewertungssysteme der Energieeffizienz angelehnt und weist somit eine einfache Verständlichkeit auf. Anhand dieser Anforderungen im Labeling-System kann nun die passende KI-Anwendung gefunden oder entwickelt werden.

### **Erstellung der Risikomatrix aus dem Anwendungsfeld**

Eine Organisation, die es plant, ein KI-System einzusetzen, stellt in einem ersten Schritt die genaue Anwendung des KI-Systems fest. Hierbei werden anhand einer Checkliste die verschiedenen Dimensionen der möglichen Risiken des KI-Systems bewertet (siehe Bild 2).

Hierfür wird beispielsweise beantwortet, ob das KI-System physisch mit einem Menschen zusammenarbeitet oder in einer gefährlichen Umgebung eingesetzt wird.

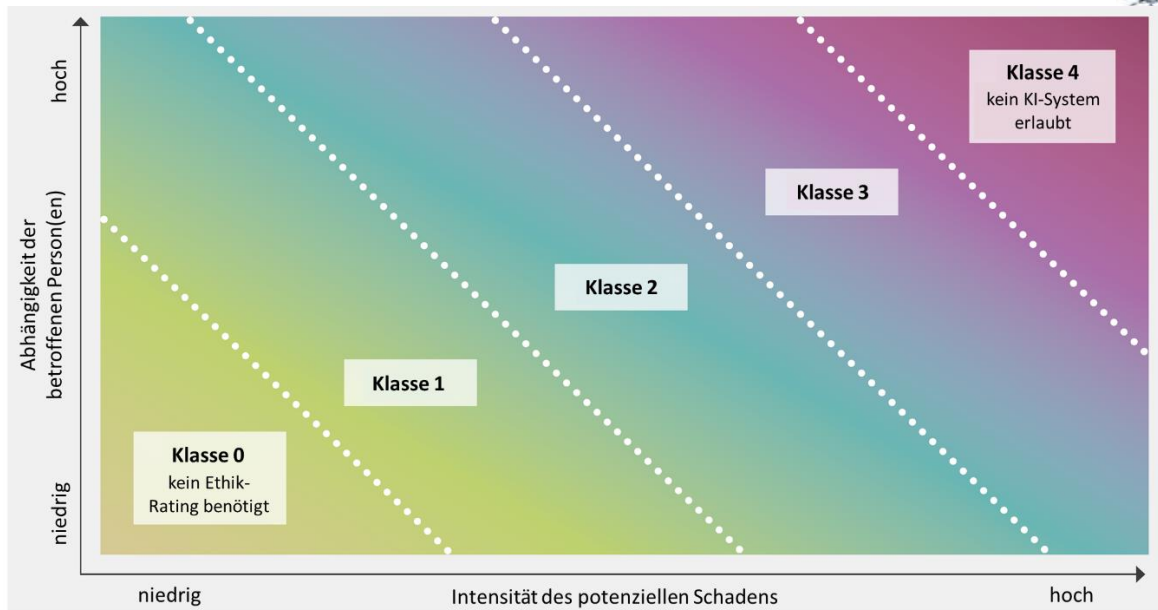


Abbildung 4: Entstehung der Risikoklasse aus dem Anwendungsumfeld

Die hieraus entstehende Klassifizierung muss auf dem gesamten potenziellen Schaden basieren, den ein KI-System in seinem jeweiligen sozialen Prozess verursachen kann. Ausschlaggebend für die Bewertung dieses Potenzials sind die Intensität des potenziellen Schadens des KI-Systems und die Abhängigkeit der betroffenen Person(en) von der jeweiligen Entscheidung.

Die Risikomatrix stellt also dar, inwieweit der Anwendungskontext eines KI-System ethisch sensibel ist. Hiermit wird in einem späteren Schritt auch beurteilt, ob und inwieweit eine Regulierung nötig ist.

#### Von der Risikomatrix zum AI Ethics Label

Die Risikomatrix gibt vor, welche Anforderungen an ein KI-System gestellt werden. Hierfür ist die Risikomatrix in verschiedene Klassen eingeteilt, die jeweils höheren Anforderungen an das KI-System entsprechen. Mit der Anwendung der Risikomatrix wird vermieden, dass an weniger kritische Systeme überhöhte Anforderungen gestellt und diese somit zu stark reguliert werden. Wird also beispielsweise eine Risikoklasse von „0“ ermittelt, so werden keine Anforderungen an das System gestellt. Geht von einem System jedoch ein höheres Risiko aus, so wachsen entsprechend auch die Anforderungen.

Um diese Anforderungen nachvollziehbar und vergleichbar darzustellen, wird hierfür ein AI Ethics Label vergeben (siehe Bild 3). Hierbei werden für jeden Wert Kriterien definiert, die anhand observierbarer Indikatoren bewertet werden. Somit kann die Erfüllung eines unscharfen Wertes durch die Beantwortung konkreter Fragen evaluiert werden (siehe Bild 4).

Beispielsweise wäre für den Wert „Transparenz“ ein Kriterium die „Offenlegung des Ursprungs der KI-Trainingsdaten“. Einer der Indikatoren wäre hier, ob der „Ursprung der Trainingsdaten dokumentiert ist“. In diesem Fall wären die möglichen Indikatoren:

- „Ja, mit einer umfassend Protokollierung der gesamten Trainings- und Betriebsdaten, sowie einer Versionskontrolle von Datensätzen etc.“



- „Ja, Protokollierung und Versionskontrolle durch einen Intermediär (z. B. Datenlieferant).“
- „Keine Protokollierung; verwendete Daten werden nicht kontrolliert oder dokumentiert.“

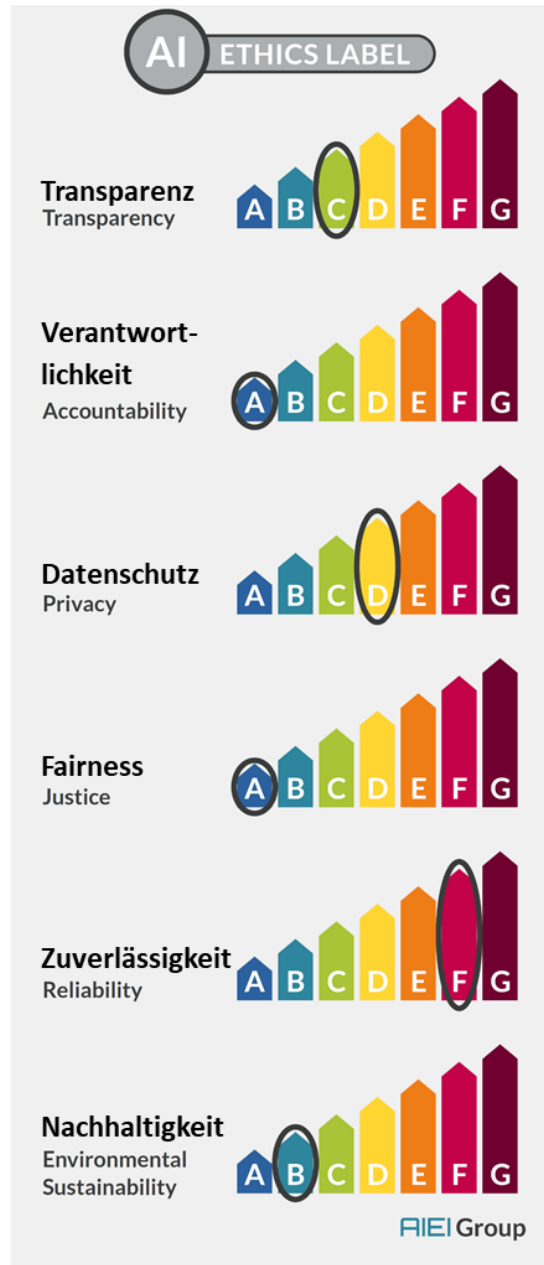


Abbildung 5: Übersichtliche Einstufung der Werte

Je nach Höhe der zu erreichenden Ränge im AI Ethics Label könnte die Bewertung selbst vom Hersteller oder von einer Zertifizierungsstelle vorgenommen werden.



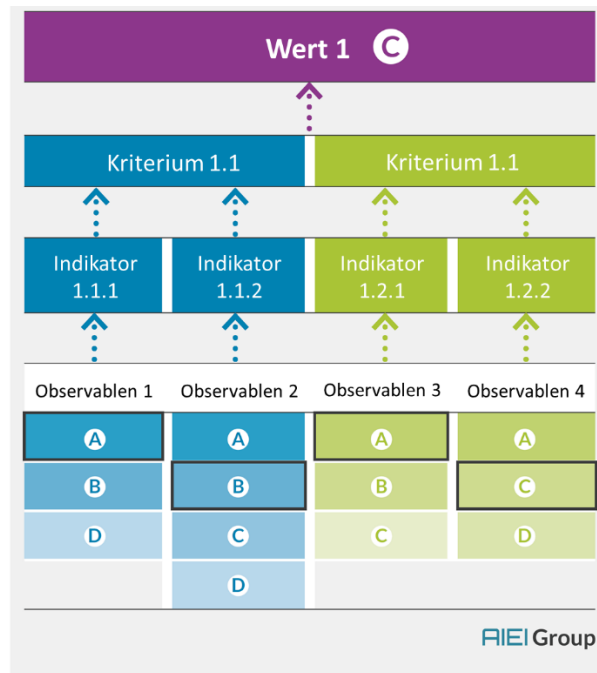


Abbildung 6: Erstellung der Einstufung für einen Wert anhand des WKIO-Ansatzes (Werte, Kriterien, Indikatoren, Observablen)

### Informierte Entscheidung anhand des AI Ethics Label

Nachdem im jeweiligen Unternehmen die Anforderungen an das KI-System festgelegt wurden, kann nun ein entsprechendes System auf dem Markt identifiziert, in Auftrag gegeben oder selbst entwickelt werden.

Hierbei kann auch bewusst ein System ausgewählt werden, das eine höhere Einstufung im AI Ethics Label aufweist. Wenn beispielsweise angedacht ist, das KI-System später in einem anderen Anwendungskontext einzusetzen oder wenn das Unternehmen selbst höhere ethische Maßstäbe erfüllen möchte.

Im Umkehrschluss können dann auch die Hersteller ihre KI-Systeme mit dem AI Ethics Label kennzeichnen. Hiermit können sie signalisieren, welche ethischen Standards das System erfüllt. Durch eine höhere Einstufung kann also auch ein Marktvorteil entstehen, der sowohl im B2C- als auch B2B-Kontext von Bedeutung ist.

Besonders ist hierbei, dass die Einstufung des KI-Systems unabhängig vom späteren Anwendungs- und vor allem auch Regulierungskontext vorgenommen werden kann. Sollte es in unterschiedlichen Märkten zu unterschiedlichen Regulierungsanforderungen kommen, so kann die Einstufung des KI-Systems dennoch weltweite Geltung haben.

Mithilfe des AI Ethics Label wird also erreicht, dass klare Spezifikationen erstellt und somit auch eine erhöhte Markttransparenz dargestellt werden können.

### Verteilung der Verantwortlichkeiten

Hat sich ein Unternehmen dafür entschieden ein KI-System einzusetzen und hierbei auf ethische Aspekte zu achten, sollte eine Regulierung zusätzlich zu den Anforderungen an das System auch Aspekte der Beziehung zwischen den beteiligten Parteien regeln.

## Paper wird stets weiterentwickelt



Vor allem im Kontext der Verantwortlichkeiten (Accountability) ist es bei KI-Systemen wichtig, klare Regelungen herbeizuführen. Die Verantwortlichkeiten zwischen allen beteiligten Parteien sollten klar verteilt werden. Hierdurch kann sichergestellt werden, dass beim Eintreten eines Schadensfalles für den Betroffenen sofort ersichtlich ist, wer die Verantwortung trägt. Somit wird verhindert, abstrakt auf die „eigene Entscheidung“ des KI-Systems zu verweisen und Verantwortlichkeiten ins Leere laufen zu lassen.