KMU Forschung Austria
Austrian Institute for SME Research

Potenziale vertrauenswürdiger KI für die Digitalisierung in Österreich

Mit Fokus auf die Bereiche Produktion, Mobilität und Gesundheit

Wien, Januar 2022
www.kmuforschung.ac.at

Diese Studie wurde im Auftrag der aws - Austria Wirtschaftsservice Gesellschaft mbH erarbeitet.

KMU Forschung Austria
Austrian Institute for SME Research

Verfasser*innen der Studie

Joachim Kaufmann
Karin Petzlberger

Internes Review

Mario Steyer
Peter Kaufmann

Für Rückfragen zur Studie

Peter Kaufmann
Tel.: +43 1 505 97 61 - 31
p.kaufmann@kmuforschung.ac.at
www.kmuforschung.ac.at

Mitglied bei:

European Network for Social and Economic Research

acr austrian cooperative research

Inhalt

Zentrale Ergebnisse

Vielfältige Anwendungsmöglichkeiten von KI-Systemen im Verkehrs-, Industrie- und Medizinsektor

- Digitalisierung ist zum einen Voraussetzung für den Einsatz von KI, zum anderen kann der zunehmende Einsatz von KI die Digitalisierung weiter vorantreiben.
- Nicht nur die KI-Technologie selbst, sondern auch der Anwendungskontext ist für die Klassifikation von KI von großer Bedeutung, denn ein und dieselbe KI-Technologie kann je nach Verwendung unter Umständen für unterschiedliche Zwecke eingesetzt werden.
- Politik, Wirtschaft und Zivilgesellschaft sind gefordert, Rahmenbedingungen und KI-Anwendungen zu schaffen, die das Vertrauen der Nutzer*innen erhöhen.
- Die Anwendung von KI kann neue ethische Fragen aufwerfen, was Implikationen für die Schaffung vertrauenswürdiger KI-Systeme haben kann.

Klare Definitionen und Begriffsabgrenzungen erforderlich

- Aufgrund der Komplexität des Themas ist der Prozess der Begriffsdefinition sowohl von KI und vor allem von vertrauenswürdiger KI noch nicht abgeschlossen. Klare Definitionen schaffen Rechtssicherheit und erleichtern den Umgang mit KI-Systemen.

Schaffung von Rechtssicherheit und Standards notwendig

- Ethische Fragen bleiben bei der Entwicklung von KI-Anwendungen meist unberücksichtigt.
- Mit Hilfe von Standards können ethische Vorgaben in technische Kriterien übersetzt werden. Standards und Normen können eine Grundlage für weitere Innovationsaktivitäten bilden.
- Regulierungen können verbindliche Kriterien für KI-Systeme festlegen, die das Vertrauen erhöhen, aber unter Umständen Innovationsaktivitäten hemmen ("Überregulierung").
- Geeignete Zertifizierungsverfahren können Vertrauen in KI erhöhen, dabei ist aber die Dynamik selbstlernender Algorithmen zu berücksichtigen.

Internationale Kooperation und Zusammenarbeit oder globaler KI-Wettlauf?

- Einhaltung ethischer Standards bleibt beim globalen Rennen um Technologieführerschaft bislang auf der Strecke.
- Der Austausch zwischen KI-Akteuren und Einbindung von Stakeholdern ist von großer Bedeutung, um die Entwicklung vertrauenswürdiger KI zu fördern.
- Die Einbindung potenzieller Nutzer*innen und Expert*innen unterschiedlicher Disziplinen bei der KI-Entwicklung ist anzuraten, um unterschiedliche Blinkwinkel zu analysieren und potenzielle Schwierigkeiten rechtzeitig zu erkennen.

1 | Einleitung

Die Nutzung und Weiterentwicklung von Künstlicher Intelligenz (KI) birgt gemäß Literatur hohes Potenzial: KI kann demnach unter anderem zur Prozessoptimierung, zur schnelleren Entscheidungsfindung sowie zur Lösung akuter gesellschaftlicher Probleme wie dem Klimawandel, Ressourcenübernutzung, oder zu Verbesserungen im Gesundheitsbereich beitragen. Gleichzeitig gehen mit dem Einsatz von KI auch weitreichende Risiken einher, etwa hinsichtlich der Datensicherheit, eines möglichen Kontrollverlustes oder der Verantwortlichkeit im Schadensfall. Die umfassenden Einsatzmöglichkeiten von KI werden nur dann Akzeptanz erfahren, wenn die Risiken minimiert sowie Unsicherheiten und ethische Bedenken ausgeräumt werden. Daher ist es essentiell, den Fokus auf vertrauenswürdige KI zu richten, um einen ethisch verantwortlichen Umgang mit KI zu garantieren und für sämtliche Beteiligte die Vorteile nutzbar zu machen.

In dieser Überblicksstudie werden zunächst die Begrifflichkeiten geklärt, und mögliche Potenziale und Herausforderungen aufgezeigt. Der Fokus richtet sich auf die notwendige Digitalisierung der Bereiche Produktion, Mobilität und Medizin. Anschließend werden mögliche Anwendungen in der Industrie, dem Mobilitätssektor und Gesundheitswesen aufgezeigt und einschlägige politische Maßnahmen und Initiativen - darunter auch gute Förderpraktiken - präsentiert. Zudem werden zentrale Handlungsfelder aufgezeigt: dazu zählen die Wichtigkeit internationaler Kooperationen und der sektorübergreifende Austausch, die Schaffung von Standards um Rechtssicherheit gewährleisten zu können, oder die Einbindung sämtlicher Stakeholder bereits in der Entwicklungsphase von KI-Algorithmen um eine holistische Betrachtung zu gewährleisten. Abschließend werden noch offene Forschungsfragen thematisiert, etwa inwieweit die Berücksichtigung von Ethik-Richtlinien bei der Entwicklung und Implementierung von KI-Anwendungen in eine Zunahme von Bürokratisierung resultiert und sich dadurch womöglich als hinderlich für den Einsatz innovativer, vertrauenswürdiger KI erweist.

2 | Was ist (vertrauenswürdige) Künstliche Intelligenz?

Es existieren zahlreiche unterschiedliche Definitionen für den Begriff der Künstlichen Intelligenz und je nach Sichtweise wird die KI in Industrie, Forschung und Politik entweder über die zu erzielenden Anwendungen oder den Blick auf die wissenschaftlichen Grundlagen definiert. Bis heute gibt es allerdings keine allgemeingültige Definition von KI. Der Begriff bezieht sich meist nicht auf eine einzige Technologie, sondern umfasst eine Reihe von unterschiedlichen Ansätzen, Methoden und Technologieanwendungen. Die österreichischen KI-Strategie definiert KI als Computersysteme, “die intelligentes Verhalten zeigen, d. h. die in der Lage sind, Aufgaben auszuführen, die in der Vergangenheit menschliche Kognition und menschliche Entscheidungsfähigkeiten erfordert haben.“ (BMK & BMDW, 2021, S. 16) Im Vorschlag für eine Verordnung zur Festlegung harmonisierter Vorschriften für KI, veröffentlicht von der Europäischen Kommission 2021, stehen „Systeme künstlicher Intelligenz“ im Mittelpunkt. Was

genau darunter verstanden wird, ist im Dokument selbst allerdings nicht definiert. Stattdessen findest sich eine Forderung, den Begriff KI-Systeme klar zu definieren[1], um Rechtssicherheit gewährleisten zu können, was wiederum die Komplexität des Themas allein schon auf der Ebene der Begriffsdefinition verdeutlicht.

Kognitive Entscheidungen des Menschen können mit Hilfe von KI-Systemen nachgeahmt werden. Christen et al. (2020) bezeichnen KI beispielsweise als „den Versuch, Verstehen und Lernen mittels eines Artefakts nachzubilden, wobei in erster Linie auf Denken bzw. Handeln fokussiert sowie ein rationales Ideal oder eine Nachbildung menschlicher Fähigkeiten angestrebt wird." KI übernimmt dabei Aufgaben wie Wahrnehmung (akustisch, visuell, textuell, taktil, …), Entscheidungsfindung, Vorhersage, Wissenserschließung und Mustererkennung aus Daten, interaktive Kommunikation und logisches Schlussfolgern. Eine Möglichkeit, KI-Lösungen zu untergliedern, besteht darin, zwischen symbolischen und statistischen Ansätzen zu unterscheiden (OECD, 2020). Während KI in ersteren vorgegebenen Regeln folgt um zu Schlussfolgerungen zu kommen, wird KI bei statistischen Ansätzen zur Mustererkennung und Modellentwicklung eingesetzt. Auch die heute zunehmend häufiger eingesetzten Formen des Machine Learnings (ML) sowie des Deep Learnings (DL) zählen zu den statistischen Ansätzen.

Mit der steigenden Komplexität der Systeme und der Sorge vor einem „menschlichen Kontrollverlust" geht eine intensivere Auseinandersetzung mit ethischen, sozialen und rechtlichen Herausforderung beim Einsatz von Künstlicher Intelligenz einher. Forderungen nach einem verantwortungsvollen und nachvollziehbaren Umgang mit KI werden lauter. Bedenken bezüglich des Einsatzes von KI gibt es in allen europäischen Ländern, wie eine Eurobarometerumfrage aus dem Jahr 2019 zeigt (EK, 2019). Viele Europäer*innen sorgen sich beispielsweise, dass der Einsatz von KI zu Situationen führen könnte, in denen die Verantwortlichkeit nicht klar ist (etwa bei Unfällen mit selbstfahrenden Autos, 43 % der Befragten), oder dass der Einsatz von KI zur Diskriminierung (z. B. aufgrund des Alters, Geschlechts, etc.) führen könnte (36 % der Befragten). Die Hälfte der Befragten ist der Ansicht, dass es ein Eingreifen der Politik bedarf, damit Anwendungen künstlicher Intelligenz in ethischer Art und Weise entwickelt werden.

Auf politischer Ebene möchte insbesondere die EU die Entwicklung einer nachhaltigen und vertrauenswürdigen KI vorantreiben. Was genau künstliche Intelligenz vertrauenswürdig macht ist jedoch nach wie vor umstritten und viel diskutiert. Was den Term „vertrauenswürdig" betrifft, muss festgehalten werden, dass es in der Literatur auch diesbezüglich keine allgemein akzeptierte Definition von „Vertrauen" (bzw. Englisch „trust") gibt. Vertrauen als multidimensionales Konstrukt umfasst im Rahmen von Informationssystemen zumindest zwei Aspekte, nämlich Vertrauen in die Technologie selbst, sowie Vertrauen in die spezifische Anwendung bzw. den Anbieter einer KI-Lösung (Thiebes et al., 2020). Umso wichtiger, aber auch heikler, ist daher der Versuch, ein anwendbares Konzept einer vertrauenswürdigen KI zu entwickeln.

[1] „Die Begriffsbestimmung sollte auf den wesentlichen funktionalen Merkmalen der Software beruhen, insbesondere darauf, dass sie im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren, sei es physisch oder digital." (EK, 2021, S. 21f)

Damit KI-Systeme von den User*innen als vertrauenswürdig wahrgenommen werden, sollte KI diverse ethische Prinzipien verfolgen, wie beispielswiese das Gemeinwohl, Schadensvermeidung, Autonomie, Gerechtigkeit und Fairness sowie Erklärbarkeit. Laut den ethischen Leitlinien der Europäischen Union für vertrauenswürdige KI gelten KI-Systeme als vertrauenswürdig, wenn folgende Anforderungen erfüllt sind: (1) Vorrang menschlichen Handelns und menschlicher Aufsicht, (2) technische Robustheit und Sicherheit, (3) Privatsphäre und Datenqualitätsmanagement, (4) Transparenz, (5) Vielfalt, Nichtdiskriminierung und Fairness, (6) Gesellschaftliches und ökologisches Wohlergehen und (7) Rechenschaftspflicht.

Die fortschreitende Digitalisierung bildet zum einen eine Voraussetzung für den Einsatz von KI, auf der anderen Seite können vorhandene und ausgefeilte KI-Lösungen selbst einen Anreiz darstellen, Digitalisierungsvorhaben voranzutreiben. Definitionen von Digitalisierung variieren je nach Kontext und es findet sich keine (einheitliche) Definition in entsprechenden Strategiepapieren der österreichischen Ministerien, der Europäischen Kommission oder der OECD. Im Rahmen dieser Studie meint Digitalisierung den Prozess der zunehmenden Verwendung digitaler Technologien in allen Bereichen der Gesellschaft, wobei der Fokus auf den Wirtschaftssektoren der Produktion bzw. Industrie, der Mobilität und der Gesundheit liegt. Es ergeben sich laufend neue Anwendungsfelder für KI-Lösungen durch Weiterentwicklungen in den Sensor- und Aktortechnologien, der Verbreitung von mobilen IT-Endgeräten, einer zunehmenden Vernetzung von Geräten im Internet of Things, und damit einhergehend durch das Erstellen, Speichern und Verarbeiten von digitalen Daten.

Die zentrale Frage für dieses Papier lautet somit: Welche Anwendungsmöglichkeiten und Potenziale ergeben sich für vertrauensvolle KI im Rahmen der zunehmenden Digitalisierung? Im Hinblick auf die wachsenden Fähigkeiten von KI und der Vielzahl an möglichen Anwendungsbereichen führt dies zu einer Reihe von Folgefragen im Umgang mit KI-Lösungen, die auch die Sicherheit und das Wohl von Menschen unmittelbar betreffen können.

3 | Potenziale und Herausforderungen

Der Fortschritt in der Entwicklung von KI geht neben technischen Fortschritten (Sensoren, Rechen- und Speicherkapazität, etc.) und wissenschaftlicher Forschung zum Teil auf die zunehmende Digitalisierung selbst zurück: das vermehrte Generieren von digitalen Daten, die Vernetzung von Geräten untereinander, die Entwicklung von Cloud Computing, etc. bereiten den Weg für immer neue Einsatzmöglichkeiten von KI. Daher ist die Digitalisierung zugleich ein Treiber für die KI-Entwicklung als auch eine Voraussetzung für den breiten Einsatz von KI-Anwendungen (Seifert et al., 2018, S. 29).

Die große Herausforderung ist bei der Entwicklung einer KI, die ethischen Richtlinien folgt und daher als vertrauenswürdig erscheint, dass bisher unbekannte ethische Probleme aufgeworfen werden und hierfür erst neue Richtlinien erarbeitet werden müssen. Dies führt dazu, dass neben der technischen Entwicklung einer KI verstärkt auch die soziale Welt berücksichtigt werden muss, sodass häufig von einer menschenzentrierten KI gesprochen wird (Cremens et al., 2019) Die

Vision[2] der Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE), an der sich auch österreichische Forschungseinrichtungen beteiligen, basiert ebenso auf dem Konzept einer menschenzentrierte KI wie die Strategie der Bundesregierung für Künstliche Intelligenz (AIM AT 2030)[3].

Für die Wirtschaft birgt KI ein großes Potenzial. Die Zahl der KI-Anwendungen wird in den kommenden Jahren voraussichtlich exponentiell zunehmen, was wiederum zum Wirtschaftswachstum beitragen kann (Cremens et al., 2019; Seifert et al., 2018). Vertrauen in die KI-Technologie ist hierbei von entscheidender Bedeutung und neben den technischen Schutzmechanismen (z. B. zum Schutz vor Cyberangriffen) muss auch sichergestellt werden, dass der Einsatz von KI ethisch vertretbar ist und bleibt. (Über-)Regulierungen neigen dazu, sich hemmend auf das Innovationspotenzial auszuwirken. Gleichzeitig schaffen gemeinsame Standards auch Sicherheit und Vertrauen unter den Innovationsakteuren, und bilden eine Basis für weitere Entwicklungen (Cihon, 2019). Mit dem Vorschlag eines ausgewogenen Regulierungsansatzes, der potenzielle Risiken minimiert ohne dabei technologische Entwicklungen zu behindern, versucht auch die Europäische Kommission (EK, 2021) hier die richtige Balance zu finden.

Das wirtschaftliche Potenzial zeigt sich auch daran, dass Investitionen in Startups im KI-Bereich in den letzten Jahren stark zugenommen haben (OECD, 2020). Im internationalen Vergleich zeigt sich, dass in den USA und China größere Volumina von Eigenkapitalinvestitionen getätigt werden, was auch dem schwach ausgeprägten private-equity Markt in Europa geschuldet sein dürfte (OECD, 2020, S. 43). Der Wettlauf zwischen den USA, China und Europa um die KI-Welt(markt)führerschaft verringert die Wahrscheinlichkeit der Etablierung technischer Vorsichtsmaßnahmen sowie der Entwicklung vertrauenswürdiger KI-Systeme, die Zusammenarbeit und den Dialog zwischen Forschungsgruppen und Unternehmen. Damit steht der KI-Wettlauf in krassem Gegensatz zur Idee der Entwicklung einer "AI4people" (Floridi et al., 2018). Der gewinnbringende Einsatz von maschinellen Lernsystemen ist häufig nicht primär von einer werte- oder prinzipienbasierten Ethik geprägt, sondern folgt vielmehr einer wirtschaftlichen Logik. Entwickler*innen sollen in Organisationen in erster Linie technische Lösungen erarbeiten, werden im Rahmen ihrer Ausbildung meist weder für ethische Fragen sensibilisiert noch finden sie in Organisationen Strukturen vor, die es ihnen ermöglichen würde ethische Bedenken vorzubringen. Im Geschäftsleben und Innovationsbereich ist Schnelligkeit maßgebend, wodurch ethische Überlegungen meist unberücksichtigt bleiben. So hat in der Praxis die Entwicklung, Implementierung und Nutzung von KI-Anwendungen sehr oft wenig mit ethischen Werten oder Prinzipien zu tun (Hagendorff, 2020).

Die OECD (2020) listet eine Reihe von Herausforderungen im Zusammenhang mit KI-Anwendungen auf. Statistische KI-Modelle können auf Basis ihrer Trainingsdaten systematische Fehler entwickeln und „Verzerrungen“ (z.B. einen Gender-Bias) aus der realen Welt in ihr digitales Modell übernehmen. Auch bewusste Manipulation der Trainingsdaten (etwa im Rahmen von Hackerangriffen) kann weitreichende Folgen für den Einsatz von KI haben. KI-Modelle werden

[2] https://claire-ai.org/wp-content/uploads/2019/10/CLAIRE-vision.pdf , 03.12.2021

[3] https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/aimat.html , 03.12.2021

zunehmend komplexer, ihre Funktionsweise ist auch für Personen mit Fachkenntnissen nicht mehr nachvollziehbar. Erklärbarkeit wird also zu einem wichtigen Gütekriterium. Dies stellt derzeit ein ernsthaftes Problem dar, da fast alle modernen ML-Methoden eine „Blackbox" darstellen (Eiling & Huber, 2021, S. 52).[4]

Durch die Verarbeitung großer Datenmengen durch KI-Systeme ist ebenfalls die Gewährleistung von Datenschutz im Fokus, vor allem sobald personenbezogene Daten von KI-Systemen verarbeitet werden (vgl. Cremens et al., 2019). Eine weitere Herausforderung im Zusammenhang mit KI-Anwendungen liegt in der Übernahme von Aufgaben von Beschäftigten. Mit der zunehmenden Automatisierung, die durch KI noch begünstigt werden könnte, wird ein Jobverlust befürchtet, etwa in der Produktion (automatisierte Fabrik), in der Mobilität (autonome Fahrzeuge ersetzen Fahrer) und im Gesundheitswesen (Roboter ersetzen Aufgaben von Pflegekräften). Derzeit ist nicht absehbar, welche Auswirkungen KI auf den Arbeitsmarkt haben wird. Es kann von ähnlichen Effekten ausgegangen werden, wie sie mit der Digitalisierung allgemein in Verbindung gebracht werden: Eine Aufwertung der Arbeit durch KI-Assistenzsysteme und den Wegfall monotoner Aufgaben, sowie eine Unterstützung bei komplexen Aufgaben. Aber auch ein kompletter Wegfall von Jobs und die Abwertung menschlicher Tätigkeiten in bestimmten Berufsfeldern scheinen im Bereich des Möglichen zu liegen. Zudem wird eine De-Qualifikation befürchtet, wenn zukünftig gewisse Tätigkeiten (etwa die Interpretation von Röntgenbildern) ausschließlich von der KI übernommen werden (WHO, 2021).

Nichtsdestotrotz sind die Chancen vielfältig, die sich durch KI ergeben: Eine Verbesserung der Gesundheitsleistungen, z.B. durch frühzeitiges Erkennen von Krankheiten (Ding et al., 2019), der betrieblichen Effizienz (z.B. KI-gesteuerte Produktion), des Ressourcenverbrauchs (z.B. KI-gesteuerte Routenführung im Verkehr), der Sicherheit (z.B. Vermeidung von Unfällen durch KI-gesteuerte Fahrzeuge), sowie eine Linderung des Fachkräftemangels (etwa den Mangel an Gesundheits- und Pflegepersonal (WHO, 2021) sind nur einige der Vorteile, die man sich von KI verspricht.

4 | Anwendungskontexte vertrauenswürdiger Künstlicher Intelligenz

4.1 | Anwendungen in der Industrie

Die digitale Transformation der Industrie fokussierte bisher vor allem darauf Abläufe effizienter zu machen, Kosten zu senken, die Qualität der Produkte zu erhöhen und den Betrieb insgesamt produktiver zu gestalten. In Zukunft werden sich in der Produktion vermehrt auch

[4] Hinsichtlich der Erklärbarkeit von KI-Modellen wird zwischen Blackbox-Modellen, Whitebox-Modellen und Greybox-Modellen unterschieden. Bei Blackbox-Modellen sind Entscheidungen nur schwer nachvollziehbar, in Whitebox-Modellen ist das Verhalten der KI vollständig nachvollziehbar und in Greybox-Modellen werden Ersatzmodelle für die Interpretation des Verhaltens der KI erstellt (Bauer et al., 2021, S. 19).

Geschäftsprozesse und -modelle durch den Einsatz digitaler Technologien verändern. KI kann diese Transformationsprozesse zusätzlich unterstützen (Gürtler, 2019).

In der österreichischen Sachgütererzeugung kommt KI noch in vergleichsweise wenig Unternehmen zum Einsatz. Einer Studie des Austrian Institutes of Technology (Zahradnik et al., 2019: 10) zufolge nutzten im Jahr 2018 etwa 2-3 % der Betriebe KI in ihrer Produktion. Der Erhebung des IKT-Einsatzes in Unternehmen 2021 der Statistik Austria zufolge verwenden 9 % der österreichischen Unternehmen KI-Technologien, die Nutzung hängt dabei aber stark mit der Größe des Unternehmens zusammen. So nutzen Großunternehmen (32 %) KI-Technologien deutlich häufiger als Mittel- (15 %) und Kleinunternehmen (7 %) Im produzierenden Bereich werden, sofern KI eingesetzt wird, am häufigsten Technologien zur Texterkennung und Text Mining (53 %) und zur Datenanalyse (rd. 35 %) eingesetzt. Die häufigsten Einsatzbereiche sind Produktionsprozesse (rd. 45 %), die Organisation betriebswirtschaftlicher Prozesse (rd. 29 %) und im Management oder der Führung des Unternehmens (rd. 21 %)[5].

Das größte Potenzial für den Einsatz von KI wird in Hochtechnologiebranchen wie Elektronik und Maschinenbau gesehen. Auch der Digital Innovation Hub „Artificial Intelligence for Production“ (AI4P) gibt auf seiner Website an, dass sich AI-Anwendungen in der Produktion derzeit bezogen auf den Anteil der Unternehmen nur im einstelligen Bereich bewegen (https://www.ai4p.at/). Für einen höheren Grad an Autonomie in der industriellen Produktion, wovon man sich deutliche Effektivitäts- und Effizienzsteigerungen erhofft, wird KI jedoch als Schlüsseltechnologie gesehen (Ahlborn et al., 2019).

Auch wenn KI vor allem in kleinen Produktionsbetrieben noch relativ selten angewandt wird, die Forschung und Entwicklung von KI hat hingegen in den Unternehmen bereits Fuß gefasst. So entfallen im Bereich der F&E-Aktivitäten (auf Basis von öffentlich geförderten F&E Projekten in Österreich) etwa 27 % auf die Sachgüterzeugung. Ein ähnlich hoher Anteil entfällt in diesem Sektor auf Stellenausschreibungen mit KI-Bezug: 30 % entfielen bereits mit Stand Oktober 2018 auf die Sachgütererzeugung. Ein reges Interesse an KI-Entwicklungen seitens der Produktionsbetriebe scheint also durchaus gegeben (Prem & Ruhland, 2019).

Weitere Studien (Bauer et al., 2021; Seifert et al., 2018) verweisen auf folgende Bereiche, in denen KI in der Produktion bereits angewandt wird:

- Im Rahmen von Predictive Maintenance optimiert KI die Wartungszeitpunkte von Maschinen und Anlagen, um so Ressourcen zu sparen und unnötige Stillstände zu vermeiden.
- In der Logistik kann KI Mitarbeiter*innen bei der Lagerdisposition helfen, d.h. sowohl bei der laufenden Überprüfung der Entwicklung von Lagerbeständen und beispielsweise beim Warentransport innerhalb des Lagers in Form von KI-gesteuerten Transportrobotern.
- Im Rahmen der Produktionssteuerung kann KI die effiziente Planung der Produktion unterstützen (Durchlaufzeiten erhöhen, etwaige Störungen identifizieren und

[5] https://www.statistik.at/web_de/statistiken/energie_umwelt_innovation_mobilitaet/informationsgesellschaft/ikt-einsatz_in_unternehmen/index.html, 14.12.2021

entsprechende Maßnahmen ergreifen). Intelligente Assistenzsysteme erleichtern Mitarbeiter*innen die Aufgaben und können dabei helfen, Arbeitsabläufe zu optimieren.

- In besonders komplexen Produktionsketten könnte KI zudem helfen Zusammenhänge zwischen Prozessen und Abläufen zu erkennen, was wiederum zu neuen Einsichten und Produktivitätssteigerungen führen kann.
- Roboter werden mithilfe von KI immer besser darin, menschliche Bewegungsabläufe nachzuahmen, und können beispielsweise Bauteile selbstständig erkennen und greifen. Über intelligente Sensorik, bei der Daten mittels KI vorverarbeitet werden, kann die Umgebungswahrnehmung verbessert und Prozesse effizienter und sicherer gestaltet werden (z.B. im Zuge einer Kollisionsvermeidung).
- KI kann in der Qualitätskontrolle eingesetzt werden, Qualitätsschwankungen in den Bauteilen und Produkten erkennen und entsprechende Informationen an nachgelagerte Prozesse weitergeben oder eventuelle Qualitätsprobleme vorhersagen.
- Auch zur Optimierung des Ressourcen- und Wissensmanagements (z. B. Beschaffungsplanung, Management von unternehmensinternen Informationen und Prozessen, siehe Seifert et al, 2018, S. 14) können KI-Anwendungen beitragen.
- Einen weiteren Einsatzbereich von KI stellen Automatisierungslösungen für kleine Losgrößen (batch size one) dar, denn hierbei stoßen bestehende Lösungen an ihre Grenzen, die auf Massenproduktion ausgerichtet sind. Eine Automatisierung für kleine Losgrößen, wie sie geraden im Produktion 4.0 angestrebt wird, wäre derzeit zu unflexibel, mit zu viel Aufwand verbunden und würde ein hohes Maß an Expertenwissen erfordern (Eiling & Huber, 2021, S. 46).

In all diesen Anwendungsbereichen spielen auch ethische Grundsätze wie z.B. Erklärbarkeit und Transparenz eine Rolle, damit einer KI-Anwendung vertraut wird und es zu keinen Schäden an Mensch und Maschine kommt. Transparenz bei der Datengrundlage, auf die KI zurückgreift, ist von großer Bedeutung damit KI keine fehlerhaften Ergebnisse produziert. Damit eng verbunden ist auch der Schutz der Daten (beispielsweise sichere Datenübertragung zwischen Maschinen) und der Schutz etwaiger personenbezogener Daten. Erklärbarkeit wiederum ist ein wichtiges Kriterium, damit die Funktionsweise der KI nachvollziehbar bleibt und auch Fachkräfte verstehen, auf Grundlage welcher Parameter, Kategorien und Annahmen die KI Ergebnisse erstellt (Pentenrieder et al., 2021, S 26f).

Bauer et al. (2021) nennen eine Reihe von Kriterien die KI in sozio-technischen Systemen erfüllen soll, um als robust und vertrauenswürdig wahrgenommen zu werden. Dazu zählt etwa, dass KI-Systeme von Menschen verstanden werden, dass Menschen von KI-Systemen wahrgenommen sowie deren Verhalten interpretiert werden kann oder, dass KI den kulturellen und gesellschaftlichen Kontext berücksichtigt und die Anwendung den ethischen Grundsätzen entspricht.

4.2 | Anwendungen im Mobilitätssystem

Auch im Mobilitätssystem versprechen KI-Technologien vielfältige Chancen. Vor allem die Sicherheit im Verkehr soll sich mittels KI verbessern lassen. Dies kann zum einen über autonome

Fahrzeuge erfolgen, da die meisten Unfälle auf menschliches Versagen zurückzuführen sind (EK, 2020). Fahrzeughersteller, Technologieunternehmen und Forscher arbeiten daher an Einsatzmöglichkeiten für autonome Fahrzeuge. Diese basieren auf einer Vielzahl an unterschiedlichen Sensoren und Aktoren sowie Softwareprogrammen. Ähnlich wie in der industriellen Anwendung ist man bei höheren Autonomiegraden für die Fahrzeugsteuerung auf eine KI angewiesen (Niestadt et al., 2019). Des Weiteren erhofft man sich durch KI auch im Verkehr Effizienzsteigerungen, etwa durch die Optimierung des Verkehrsflusses (Heusser et al., 2021). Auch die Entlastung des Fahrers, die Vermeidung von Staus über intelligente Routenführung und die Reduzierung von Treibstoffverbrauch stellen Chancen für die Anwendung von KI dar. Vor allem für die Steuerung des Verkehrsflusses in Städten im Sinne von Smart Cities wird KI von großer Bedeutung sein, ebenso wie für neue Formen der Mobilität („Mobility as a Service“) und der intermodalen Mobilität. Neben dem Monitoring von Fahrzeugflotten kann KI aber auch zur Überwachung der Fahrer*innen eingesetzt werden, um diese rechtzeitig zu warnen, wenn das Verhalten auf Müdigkeit schließen lässt (Paiva et al., 2021) Das sich eine solche Überwachung auf die Fahrer*innen nicht nur positiv auswirkt, zeigen mehrere Medienberichte[6] zur Anwendung einer solchen Überwachungsmethode bei Amazon in den USA. Zwar bemühte sich das Unternehmen zu betonen, dass die Videoüberwachung von Paketzusteller*innen in erster Linie die Sicherheit erhöhen soll und Fahrer*innen auf gefährliches Fahrverhalten aufmerksam gemacht werden. Letztlich geht aus den Stellungnahmen der Zusteller*innen selbst jedoch hervor, dass dadurch ihre Fahrweise auch bewertet wird was gegebenenfalls auch negative Konsequenzen für die Fahrer*innen selbst nach sich ziehen kann. Dieses Beispiel zeigt auch deutlich die Ambivalenz bzw. ethische Vielschichtigkeit von durch KI-Lösungen verarbeiteten Daten, die in ein und demselben Anwendungsfall sowohl zur Unterstützung als auch zur Überwachung und Kontrolle eingesetzt werden können. Es zeigt auch die Notwendigkeit der Berücksichtigung ethischer Prinzipien bei KI-Systemen, die menschliches Verhalten analysieren.

KI kann auch bei der Instandhaltung der Straßeninfrastruktur unterstützen: So kann mithilfe von Sensordaten die Straßenbeschaffenheit erkannt, mittels Datenübertragung in eine Cloud anderen Verkehrsteilnehmern aber auch Infrastrukturbetreibern mitgeteilt werden. Auf diese Weise kann beispielsweise auf Schlaglöcher, Glatteis und Aquaplaning entsprechend reagiert werden (Gürtler, 2019, S. 103f). Weitere Anwendungsfelder von KI sind beispielsweise Truck Platooning (das enge Hintereinanderfahren von LKWs), Prognosen über mögliche Unfallsituationen und die Analyse von Verkehrsdaten zur Verkehrsflusssteuerung (etwa mittels Lichtsignalen) (Niestadt et al., 2019).

KI wird eine wichtige Rolle für die Entwicklung von Smart Mobility spielen, und stellt die Entwickler und Anwender zugleich vor große Herausforderungen. Zum einen wirft die Vielzahl an notwendigen Daten Fragen des Datenschutzes und der sicheren Datenübertragung auf. Vor allem bei autonomen Fahrzeugen muss gewährleistet werden, dass ausreichend und unverfälschte Daten aus der Umgebung erfasst und diese von der KI auch korrekt klassifiziert

[6] https://www.theverge.com/2021/3/24/22347945/amazon-delivery-drivers-ai-surveillance-cameras-vans-consent-form; https://threatpost.com/amazon-driver-surveillance-cameras/174843/; https://www.derstandard.at/story/2000129812783/amazons-ueberwachungs-ki-bestraft-fahrer-wenn-sie-ueberholt-werden; 03.12.2021

werden. Sollte es zu einem Unfall kommen, stellt sich die Frage, wer dafür haftet (OECD, 2020, S. 58f). Schließlich kann es auch zu Situationen kommen, in denen eine KI auf Basis der Unfallsituation letztendlich über Leben und Tod entscheidet. Wie eine KI in solchen Situationen entscheiden soll, ist derzeit aus ethischer Sicht nicht bestimmbar und daher noch ungeklärt. Daneben muss auch die Sicherheit der technischen Systeme gewährleistet werden, z. B. können sich Softwaresicherheitslücken unter Umständen fatal auswirken, wenn es dadurch Hackern ermöglicht wird, auf die Steuerung des Fahrzeugs zuzugreifen (EK, 2020). Neben technischen Herausforderungen besteht auch rechtlicher Handlungsbedarf, denn autonome Fahrzeuge sind im gegenwärtigen Rechtsrahmen nicht vorgesehen[7]. Auch Genehmigungsverfahren von Fahrzeugtypen werden vor diesem Hintergrund eine Überarbeitung benötigen.

Aktuelles Anwendungsbeispiel für KI im Bereich Mobilität: Projekt „SafeSign“

In diesem Projekt wird erforscht, wie KI bei der Erkennung von Verkehrszeichen so eingesetzt werden kann, dass Menschen ihr auch vertrauen. Die korrekte Erkennung von Verkehrszeichen unter widrigen Witterungsbedingungen oder bei Beschädigung und Verschmutzung ist für die Entwicklung von autonomen Fahrzeugen von großer Bedeutung. Dabei kommen Deep Learning Methoden zum Einsatz, die mit Daten basierend auf realen Verkehrszeichenbildern mit und ohne Störung sowie mit synthetisch erzeugten Störungsbildern trainiert werden. Das „Erkennen“ erfolgt dabei in zwei Schritten: Zuerst wird das Verkehrszeichen innerhalb eines Bildes identifiziert und anschließend in einem darauffolgenden Schritt klassifiziert. Wichtig ist, dass die KI robust ist, das heißt die Verkehrszeichen richtig erkennt, auch wenn es zu kleinen, für den Menschen nicht wahrnehmbaren, Änderungen in den Bildern kommt. Im Rahmen des Projekts beschäftigt man sich daher bewusst auch mit „Adversarial Attacks“, also mit der Möglichkeit, dass Bilder gezielt manipuliert werden, um die KI zu täuschen. Im Rahmen der KI-Entwicklung wurden zudem die Disziplinen Recht und Ethik einbezogen. Dabei wird auch grundsätzlichen Fragen zum Einsatzzweck der KI nachgegangen. Sollte KI etwa dabei helfen möglichst viele Unfälle zu vermeiden, oder sollte sie dahingehend Verwendung finden, möglichst schwere Unfälle zu verhindern? Eine weitere Frage ist, ob das Fahrzeug sich in erster Linie alleine steuern, oder doch der Mensch jederzeit das Steuer übernehmen können soll? Um diesen Fragen nachgehen zu können, wurden daher auch Teile der Bevölkerung und Stakeholder eingebunden. Durch diese Vorgehensweise erhoffen sich die am Projekt beteiligten Organisationen wichtige Aspekte in die Entwicklung einbeziehen zu können, die eine KI vertrauenswürdig machen. Die Projektergebnisse sowie eine Datenbank mit Störungsbildern sollen zudem öffentlich verfügbar gemacht werden, um die (Weiter-)Entwicklung auch durch andere Forschende in diesem Bereich zu unterstützen.[8]

[7] Vorreiter ist hier Deutschland, wo im Juni 2021 das Gesetz zum automatisierten Fahren in Kraft trat. Es stellt allerdings nur eine Übergangslösung dar, bis internationale Vorschriften erarbeitet werden. https://www.bmvi.de/SharedDocs/DE/Artikel/DG/gesetz-zum-autonomen-fahren.html , 03.12.2021

[8] https://www.ots.at/presseaussendung/OTS_20210907_OTS0135/projekt-safesign-fuer-den-strassenverkehr-vertrauenswuerdige-verkehrszeichenerkennung-der-zukunft , https://science.apa.at/power-search/5248855972967240455 , 27.09.2021

4.3 | Anwendungsbeispiele im Gesundheitswesen

Viele Bereiche des Gesundheitswesens können von der Nutzung der künstlichen Intelligenz profitieren. KI-Anwendungen wird bereits in unterschiedlichen Bereichen erfolgreich eingesetzt, u. a. zur Diagnostik, zur Bewertung des Risikos eines Krankheitsausbruchs (Ding et al., 2019), zur Abschätzung und Personalisierung des Therapieverlaufs, zur Optimierung der Abläufe im Krankenhaus (Antweiler et al., 2020), zur bedarfsgerechten Unterstützung der Patientinnen und Patienten während der aktiven Behandlung, in der Arzneimittelforschung zur schnelleren Herstellung von Medikamenten, im Bereich der Prävention, wo beispielsweise Menschen durch Gesundheitsassistenten zu einer gesünderen Lebensweise animiert werden, etc. KI-Anwendungen sollen vor allem Effizienzgewinne und damit eine Entlastung des medizinischen und pflegerischen Personals bewirken. Zudem können KI-Tools helfen in ressourcenarmen Ländern oder ländlichen Gemeinden bestehende Lücken beim Zugang zu Gesundheitsdiensten zu schließen, in denen Patient*innen oft nur eingeschränkten Zugang zu medizinischem Fachpersonal haben.

Trotz der vielfältigen und vielversprechenden Anwendungsmöglichkeiten und bereits erzielten Erfolge steckt die KI im Medizinbereich noch in den Kinderschuhen und wird in der Praxis noch recht zögerlich eingesetzt. Dies ist zum einen der mangelnden Transparenz von Entscheidungsprozessen geschuldet (Ainek et al., 2020). Hier besteht die Herausforderung darin, dass ärztliches Personal Verständnis hinsichtlich der eingesetzten Algorithmen benötigt und die eigesetzte KI keine „Blackbox"-Diagnose stellen darf. Zum anderen muss ausreichender Datenschutz sichergestellt werden, sodass Patientinnen und Patienten die Hoheit über ihre gesundheitsbezogenen Daten bewahren.

Ein konkreter Anwendungsfall ist ein KI-basiertes System, das den Blutzuckerspiegel von Intensivpatientinnen und -patienten automatisch stabilisiert. Ein intelligenter Algorithmus berücksichtigt dabei laufend, wie gut der*die Patient*in auf Nahrung und das verabreichte Insulin reagiert. Weiters werden die Krankengeschichte sowie das Körpergewicht des*der Patient*in mitberücksichtigt. Das KI-basierte System kann die Steuerung ständig verbessern und die in der Infusion enthaltene Insulinmenge individuell abstimmen, wodurch die Blutzuckerwerte stets im optimalen Bereich bleiben und die Pflegekräfte entlastet werden. Als erstes in Deutschland zugelassenes KI-Produkt im Bereich der apparativen Infusionstechnik zeigt dieses Beispiel die Anwendung von KI-Technologie in einem hochregulierten Markt.

Durch die COVID-19-Pandemie hat sich der Einsatz von KI-Anwendungen teilweise beschleunigt: Angesichts des Personalmangels und der immensen Zahlen von Patientinnen und Patienten greifen immer mehr medizinische Einrichtungen auf automatisierte Hilfsmittel zurück, um die Pandemie zu bewältigen. Ärztliches Personal setzt beispielsweise bereits KI zur Triage von COVID-19-Patientinnen und Patienten ein[9]. Haben solche Technologien, die von Entwickler*innen zunächst häufig als Testversion angeboten werden, erst einmal in den Klinikalltag Einzug gehalten, ist es wahrscheinlich, dass dies dann auf Dauer so bleibt. Auch in Deutschland will die Bundesregierung künftig Algorithmen einsetzen, die über die Dringlichkeit

[9] https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/ , 3.12.2021

eines Notfallpatienten entscheiden[10]. Kritiker*innen warnen jedoch eindringlich, dass die Gefahren größer wären als der Nutzen und beanstanden das Delegieren von Verantwortung an Systeme der Künstlichen Intelligenz.

Die Weltgesundheitsorganisation (WHO) hat mit der "dreifachen Milliarde" ein ambitioniertes Ziel definiert, wonach bis 2023 je eine Milliarde mehr Menschen von einer allgemeinen Gesundheitsversorgung profitieren sollen, wirksamer gegen gesundheitliche Notfälle geschützt sein sollen und eine Verbesserung ihrer Gesundheit bzw. ihres Wohlbefindens erfahren sollen. KI wird dabei als Instrument gesehen, um diese Ziele zu erreichen. Die Messung dieser Ziele wurde dabei an jene der Sustainable Development Goals (SDGs) angelehnt. Vinuesa et al. (2020) untersuchen in einer Studie den Einfluss von KI bei der Erreichung der SDGs und zeigen, dass Künstliche Intelligenz das Erreichen von 134 der 169 individuellen Zielvorgaben begünstigen kann, aber gleichzeitig bei 59 Zielvorgaben hinderlich sein kann.

Was die Kommerzialisierung von Gesundheitsdaten betrifft, so gibt es Bedenken in Bezug auf den Verlust der Autonomie des Einzelnen bzw. den Verlust der Kontrolle über die Daten, was bei hochsensiblen Gesundheitsdaten besonders problematisch ist (WHO, 2021).

Eine weitere Herausforderung ist die zunehmende Marktmacht, die einige Unternehmen über die Entwicklung, den Einsatz und die Nutzung von KI im Gesundheitsbereich (sowie in der Herstellung von Arzneimitteln) ausüben könnten. Monopolmacht kann die Entscheidungsfindung in den Händen einiger weniger Personen und Unternehmen konzentrieren, die dann als Gatekeeper für bestimmte Produkte und Dienstleistungen fungieren und den Wettbewerb einschränken, was sich letztendlich in höheren Preisen für Produkte und Dienstleistungen, weniger Verbraucherschutz oder geringerer Innovationsleistung niederschlagen kann (WHO, 2020). Die Zahl der Patentanmeldungen ist seit 2014 rasant gestiegen, insbesondere in China und den USA. Spitzenreiter ist Samsung Electronics (Südkorea), das viele sensorbasierte Geräte patentiert; Unternehmen wie Siemens und Philips sind die wesentlichen Player in Europa (De Nigris, 2020).

Anwendungsbeispiel Co-Design eines vertrauenswürdigen AI-Systems im Gesundheitswesen: Deep Learning-basiertes Klassifizierungssystem für Hautläsionen (Zicari et al., 2021)

Obwohl KI-Systeme im Hinblick auf die Diagnose von bösartigen Melanomen bereits menschliches Leistungsniveau erreichen, fehlt es beim Einsatz häufig an Akzeptanz unter den Usern. Maßgeblich dafür ist der meist verborgene, und damit nicht nachvollziehbare Entscheidungsfindungsprozess.

Absicht des Projektes war es daher anstelle statischer Checklisten einen ganzheitlichen Ansatz (mittels Co-Design) zu verfolgen, um Techniker*innen bei der Entwicklung und Implementierung eines vertrauenswürdigen KI-Systems bei diesem konkreten Anwendungsfall zu unterstützen. Bei der Co-Design-Methodik arbeitet ein interdisziplinäres Team von Expert*innen (unterschiedliche

[10] https://www.zeit.de/digital/2021-05/triage-software-notfallmedizin-algorithmus-kuenstliche-intelligenz-ethik, 14.12.2021

Stakeholdern aus dem medizinischen Anwendungsbereich, aber auch Jurist*innen, Ethiker*innen) mit KI-Entwickler*innen, Manager*innen und Patient*innen zusammen, um eine Identifizierung der unterschiedlichen Blickwinkel vorzunehmen und ethische, rechtliche und technische Fragestellungen, die sich aus dem künftigen Einsatz des KI-Systems ergeben können zu thematisieren. Dadurch können Risiken und Schäden unter verschiedenen Gesichtspunkten abgewogen und vorab berücksichtigt werden. So wurde beispielsweise Das Problem der Überdiagnose von Melanomen im ersten Prototyp vom Technikteam nicht bedacht. Der ganzheitliche Ansatz schafft Vorteile in Bezug auf die allgemeine Akzeptanz oder Bedenken inner- und außerhalb der Institution.

5 | Politikmaßnahmen und Initiativen

Das große Potenzial von KI-Anwendungen und der nach wie vor hohe Forschungs- und Entwicklungsbedarf ruft auch die Politik auf den Plan. In den EU haben mittlerweile fast alle Mitgliedsstaaten eine KI-Strategie erstellt bzw. befindet sich eine solche in Ausarbeitung[11].

International will die EU mit anderen Ländern, insbesondere den USA und China bei der Technologieentwicklung mithalten. Im Rahmen der Förderprogramme „Horizont Europa“ und „Digitales Europa“ soll daher bis 2030 jährlich € 1 Mrd. in KI-Vorhaben investiert werden.[12]

Daneben haben sich zahlreiche (länderübergreifende) Initiativen mit dem Ziel gebildet, einen ethischen Rahmen für die Entwicklung und den Einsatz von KI auszuarbeiten, damit nach Möglichkeit alle Menschen von dieser technologischen Revolution profitieren können. Zudem beabsichtigen diese Initiativen die Forcierung des internationalen Dialogs und Austausches, um weltweit eine gerechte und integrative KI-Entwicklung zu verwirklichen.

Einige ausgewählte Initiativen, die keinerlei Anspruch auf Vollständigkeit erheben, sind:

- Partnership on AI ist eine non-profit-Partnerschaft unterschiedlicher akademischer, ziviler, industrieller Akteure und Medienorganisationen, die sich vor allem der Öffentlichkeitsarbeit widmet, Empfehlungen und Best-Practices erarbeitet, die im Sinne ihrer Vision stehen, die Menschheit zu stärken und zu einer gerechteren, ausgewogeneren und wohlhabenderen Welt beizutragen.
- The Montréal Declaration for responsible AI development ist eine Initiative der Universität Montreal, die von jeder/jedem unterzeichnet werden kann und sich u.a. die Entwicklung eines ethischen Rahmens für AI zum Ziel setzt.
- Ethics and Governance of AI Initiative ist ein gemeinsames Projekt des MIT Media Labs und des Harvard Berkman-Klein Centers for Internet and Society. Es will sicherstellen, dass KI-Technologien in einer Weise erforscht, entwickelt und eingesetzt werden die den

[11] Für einen Überblick über nationale Strategien und KI-Initiativen siehe: https://oecd.ai/en/dashboards , 14.12.2021

[12] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_de (17.9.2021)

gesellschaftlichen Werten der Fairness, menschlicher Autonomie und Gerechtigkeit entspricht.

- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems setzt sich zum Ziel, "dass alle an der Konzeption und Entwicklung autonomer und intelligenter Systeme beteiligten Akteure so ausgebildet, geschult und befähigt werden, dass sie ethischen Erwägungen Vorrang einräumen, damit diese Technologien zum Wohle der Menschheit weiterentwickelt werden."
- Das CLAIRE Research Network ist eine breit aufgestellte europäische Initiative die von mehr als 1.000 AI Expert*innen unterstützt wird und durch Vernetzung eine höhere Sichtbarkeit der Europäischen AI Community anstrebt. Der Fokus der Initiative liegt auf „vertrauenswürdiger KI, die die menschliche Intelligenz ergänzt, anstatt sie zu ersetzen, und die somit den Menschen in Europa zugutekommt".
- European network of Human-Centered Artificial Intelligence ist ein Projekt der Knowledge 4 All Foundation, dass die Entwicklung von KI-Systeme erleichtern möchte, die menschliche Fähigkeiten verbessern, dadurch Einzelne wie auch die Gesellschaft insgesamt stärken und dabei menschliche Autonomie und Selbstbestimmung wahren.
- AI for Good ist eine Konferenzreihe der Vereinten Nationen in der KI-Lösungen vor allem in Hinblick auf die Social Developments Goals (SDGs) diskutiert werden. Die Website enthält zudem Informationsmaterial und bietet Vernetzungsmöglichkeiten.
- Das Global Partnership on Artificial Intelligence ist eine Initiative unterschiedlicher Stakeholder welche darauf abzielt, die Kluft zwischen Theorie und Praxis durch Unterstützung von Spitzenforschung und angewandten KI-Aktivitäten zu überwinden.

Während diese Initiativen stärker der Entwicklung von Grundlagen für eine vertrauenswürdige KI gewidmet sind, ist an dieser Stelle insbesondere von Interesse, ob es in einzelnen Ländern Initiativen gibt, die bereits gezielt die Umsetzung vertrauenswürdiger KI-Vorhaben (vor allem im betriebswirtschaftlichen Kontext) fördern.

5.1 | Internationale Beispiele guter Förderpraktiken

Dieses Kapitel präsentiert ausgewählte internationale Förderprogramme zu vertrauenswürdiger Künstlicher Intelligenz und Digitalisierung.

Deutschland – „Forschungsförderung des Bundes im Bereich Bioethik – Ethische, rechtliche und soziale Aspekte der Lebenswissenschaften. Fördermaßnahme: Digitalisierung"

Die Fördermaßnahme unterstützt eine frühzeitige Identifizierung und Reflexion der ethischen, rechtlichen und sozialen Fragen, die durch die Digitalisierung in der Gesundheitsforschung und -versorgung aufgeworfen werden.

Die geförderten Forschungsarbeiten sollen Chancen und Risiken der gewählten Themenbereiche in interdisziplinärer Zusammenarbeit systematisch analysieren, bewerten und Lösungskonzepte für die Grundsatz- und/oder Handlungsebene entwerfen. Die thematische Bandbreite umfasst

eHealth-Anwendungen bzw. Künstliche Intelligenz zur Diagnosefindung und Patientensteuerung, Assistenzsysteme in der Pflege, Datenzugang und -nutzung in der Forschung und Versorgung, die elektronische Patientenakte, die Veränderung von Werten, Konzepten und Praktiken in der Gesundheitsversorgung sowie die Bewertung von Mensch-Maschine-Schnittstellen in der Diagnostik.

Die Resultate der einzelnen Forschungsprojekte fließen in Leitlinien, Handbücher, Publikationen und Stellungnahmen ein und leisten damit essentielle Beiträge für wissenschaftlichen bzw. gesellschaftlichen Diskurs zu einem reflektierten Umgang der Digitalisierung in der Gesundheitsforschung und -versorgung. Der Förderzeitraum betrifft die Jahre 2019 bis 2023, die gesamte Fördersumme beträgt bis zu € 10 Mio. Damit werden insgesamt 10 Verbundprojekte und ein Einzelvorhaben (CoCoAI - Kooperative und kommunizierende KI-Methoden für die medizinische bildgeführte Diagnostik) gefördert. Das Verbundprojekt vALID will die Auswirkungen der neuen Technologien auf Wissenschaft und Gesellschaft untersuchen und auf einen gesellschaftlich akzeptierten und verantworteten Rahmen für ihren Einsatz hinzuwirken. Dabei wird eine umfassende Analyse der Frage durchgeführt, wie KI-gesteuerte klinische Entscheidungsunterstützungssysteme mit dem Ideal der Arzt- und Patientensouveränität in Einklang gebracht werden können.

Elektroniksysteme für vertrauenswürdige und energieeffiziente dezentrale Datenverarbeitung im Edge-Computing (OCTOPUS)[13]

Hierbei handelt es sich um eine Technologieförderprogramm, das Verbundprojekte mittels Zuschuss für eine Dauer von bis zu 3 Jahren fördert. Voraussetzung für die Förderung ist die Zusammenarbeit mehrerer unabhängiger Partner aus Wissenschaft und Wirtschaft zur Lösung von gemeinsam vereinbarten Forschungsaufgaben (Verbundvorhaben). Die Forschungsaufgaben und -ziele müssen den Stand der Technik deutlich übertreffen und durch ein hohes wissenschaftlich-technisches sowie wirtschaftliches Risiko gekennzeichnet sein. Unternehmen erhalten 50 % ihrer förderfähigen Kosten als Zuschuss, KMU können zudem unter bestimmten Voraussetzungen ein Bonus erhalten. Das Antragsverfahren ist zweistufig, in der 1. Stufe wird eine Projektskizze eingereicht, in der 2. Verfahrensstufe muss ein förmlicher Förderantrag eingebracht werden.

Die Entwicklung einer speziellen KI-Anwendung ist keine Voraussetzung für die Inanspruchnahme dieses Zuschusses, unterstützt werden explizit aber auch die Entwicklung intelligenter und ressourcensparsamer Elektronik, intelligenter Netzwerkssteuerung und künstliche Intelligenz für die Datenverarbeitung sowie die Verwaltung und Integration von Netzen. Zudem müssen neben den Technologiebereichen auch mindestens zwei Querschnittsthemen berücksichtigt werden, und KI-Methoden, Methoden des maschinellen Lernens und verteilten Rechnens sind eines von fünf dieser Querschnittsthemen.

[13] https://www.foerderdatenbank.de/FDB/Content/DE/Foerderprogramm/Bund/BMBF/octopus.html , 14.12.2021

Richtlinie zur Förderung von deutsch-französischen Projekten zum Thema Künstliche Intelligenz[14]

Im „Aachener Vertrag“ haben Deutschland und Frankreich am 22. Januar 2019 eine Kooperation auf dem Gebiet der Forschung und des digitalen Wandels beschlossen, insbesondere auf dem Gebiet der Künstlichen Intelligenz (KI). Die thematischen Schwerpunkte der Förderung sind an aktuellen Herausforderungen im Forschungs- und Anwendungsfeld von KI ausgerichtet. Die Projektkonsortien sollen vorrangig mindestens eine der im Folgenden genannten Fragestellungen bearbeiten:

- Verteilte KI, wie z.B. verteiltes Lernen oder Edge-Computing
- Grüne KI, für geringeren Ressourcenverbrauch, z.B. Algorithmen, die weniger Energie, weniger Speicher und -weniger Kommunikationsbandbreite benötigen
- Hybride KI, z.B. die Kombination von maschinellem Lernen und Wissen
- KI in anderen Wissenschaften, z.B. KI und numerische Simulationen, KI und Physik, KI und Chemie, etc.
- Vertrauenswürdige KI, z.B. zertifizierbare, erklärbare oder interpretierbare Modelle und Verarbeitungspipelines
- KI für Spitzentechnologien, z.B. Dialogsysteme für den Medienzugang

Die Forschungsarbeiten können ein breites Set an Anwendungsfeldern und Branchen adressieren, unter anderem Mobilität, Energie, Umwelt und Ressourceneinsatz, Intelligente Industrie und Produktionstechnologien, Gesellschaft sowie Smart Health.

Dabei sind zwei Förderlinien vorgesehen, nämlich Forschungskooperationen (Linie A) sowie Verbünde aus Wissenschaft und Wirtschaft mit dem Ziel, risikoreiche industrielle Forschungs- und vorwettbewerbliche Entwicklungsvorhaben in bilateraler Zusammenarbeit durchzuführen (Linie B). Zuwendungsempfänger sind Hochschulen, Forschungseinrichtungen und Industriepartner (auch KMU). Die beantragte Förderung der deutschen und französischen Partner darf für Projekte der Förderlinie A insgesamt maximal € 400.000 und für Projekte der Förderlinie B insgesamt maximal € 800.000 betragen. Die Förderdauer für die Verbundvorhaben der Förderlinie A darf bis zu vier Jahre, die der Förderlinie B bis zu drei Jahre betragen. Die Zuwendungen werden im Wege der Projektförderung als nicht rückzahlbare Zuschüsse gewährt.

Bayern: Initiative „Künstliche Intelligenz – Big Data“

Mit der Initiative „Künstliche Intelligenz – Big Data“ fördert das Bayerische Staatsministerium für Wirtschaft, Landesentwicklung und Energie (StMWi) anwendungsoffene Innovationen im Bereich Datenanalyse, Data Science und Künstliche Intelligenz, welche die Digitalisierung in Bayern vorantreiben und die Bewältigung zukünftiger, gesellschaftlicher Herausforderungen unterstützen.

Das StMWi beabsichtigt im Rahmen der Strategie BAYERN DIGITAL und der Hightech Agenda Bayern innovative Forschungsprojekte zu fördern. Das StMWi gewährt die Zuwendung gemäß

[14] https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2020/10/3205_bekanntmachung.html , 14.12.2021

der Richtlinie zur Durchführung des Bayerischen Verbundforschungsprogrammes des StMWi in der Förderlinie Digitalisierung, Förderbereich Informations- und Kommunikationstechnik. Gegenstand der Förderung sind Forschungs- und Entwicklungsaufwendungen im Rahmen vorwettbewerblicher Verbundvorhaben. Es werden ausschließlich Vorhaben gefördert, die wesentliche Innovationen auf dem Gebiet Künstliche Intelligenz – Big Data beinhalten.

Im Rahmen dieses Aufrufes sollen Projekte aus den Gebieten Künstliche Intelligenz (KI) und Data Science unterschiedlichster Anwendungsdomänen, aber auch domänenübergreifend (cross-industry), gefördert werden, die insbesondere Forschungs- und Entwicklungsarbeiten in einem oder mehreren der folgenden Themenbereiche beinhalten:

- Vertrauenswürdige KI: Nachvollziehbarkeit und Transparenz von Methoden, Algorithmen und deren Entscheidungen (Explainable Artificial Intelligence – XAI), Berücksichtigung der Unschärfe bzw. Verzerrung (Bias) von Daten bzw. Algorithmen (Dateninsensitivität), Entwicklung resilienter Lernverfahren
- Entwicklung, Weiterentwicklung und Kombination unterschiedlicher KI-Methoden
- Digital Twin
- Dateneffizienz
- Domain Know-how
- KI-Werkzeuge
- Predictive und Prescriptive Analytics
- Datensynthese
- Automatisiertes maschinelles Lernen

Großbritannien - Ethics in artificial intelligence research and development

In den Forschungsprojekten soll untersucht werden, wie ethische Ansätze in die frühen Phasen der KI-Forschung und Entwicklung eingebettet werden können. Geldgeber des Programmes Ethics in artificial intelligence research and development ist das Arts and Humanities Research Council (AHRC). Die Förderung erfolgt in Form eines Zuschusses: Insgesamt stehen Fördermittel in der Höhe von £ 320.000 zur Verfügung, die maximale Zuwendung pro Vorhaben liegt bei £ 81.250. Die Projekte müssen in einer Forschungszusammenarbeit durchgeführt werden, es muss sich um frühe Forschung handeln und kann bis zu 12 Monate dauern.

Australien

Das Center for AI and Digital Ethics (CAIDE) fördert die fächerübergreifende Forschung, indem es eine Anschubfinanzierung für 2 oder 3 Projekte bereitstellt, die ihren Fokus auf pervasive Geräte (z. B. Überwachungsgeräte zur digital unterstützten Pflegebeobachtung) richten. Das Programm ist darauf ausgelegt, Forschungsideen zu entwickeln und neue Kooperationen innerhalb der Universität zu fördern, um Forschungskapazitäten im Bereich der digitalen Ethik aufzubauen. Dabei ist eine interdisziplinäre, fakultätsübergreifende Zusammensetzung der Forschungsteams erforderlich. Die Antragsteller haben Anspruch auf eine Förderung von bis zu AUD 20.000. Projekte, die im Rahmen des Seed-Programms gefördert werden, bewerben sich um einen größeren, wettbewerbsorientierten Zuschuss.

5.2 | KI-Register

Neben Förderprogrammen stellt auch die Einführung von KI-Registern einen wichtigen Schritt zur weiteren Durchsetzung von KI und vor allem der vertrauenswürdigen Anwendung dieser dar.

Im September 2020 haben die Städte Amsterdam und Helsinki ihre jeweiligen KI-Register lanciert, die einen Überblick über die von den Städten verwendeten KI-Systeme und Algorithmen bieten sollen. Die Intention dahinter war, allen Bürger*innen Zugang zu Informationen darüber zu verschaffen, wie Algorithmen ihr Leben beeinflussen und auf welcher Grundlage algorithmische Entscheidungen getroffen werden. Die KI-Register sind ein standardisiertes, abfragbares und archivierbares Verfahren, um jene Entscheidungen und Annahmen zu dokumentieren, die bei der Entwicklung, Implementierung, Verwaltung und schließlich bei der Abschaffung eines Algorithmus getroffen wurden. Damit soll Transparenz und Erklärbarkeit geschaffen und das Vertrauen in die Technologie gestärkt werden. Zudem ist es den Bürger*innen möglich Feedback zu geben und sich so an der Entwicklung von Algorithmen zu beteiligen.

Das KI-Register der Stadt Amsterdam (City of Amsterdam Algorithm Register) befindet sich erst im Aufbau, derzeit werden drei Anwendungsfälle aufgelistet:

- die automatisierte Parkkontrolle[15]
- Betrugserkennung bei Ferien-unterkünften
- Meldung von Problemen im öffentlichen Raum (Beschwerde-management).

In Helsinki informiert das City of Helsinki AI Register über jene bisher fünf Anwendungsfälle, in denen künstliche Intelligenz als Teil der städtischen Dienstleistungen eingesetzt wird. Dazu zählen u. a. ein „Parking chatbot“ sowie ein „Gesundheitszentrum chatbot“.

Automatisierte Parkkontrolle in Amsterdam

In Amsterdam prüft die Stadtverwaltung ob ein geparktes Auto auch tatsächlich dazu berechtigt ist. Die Überprüfung erfolgt mit Hilfe von Scan-Autos, die mit Kameras ausgestattet sind und den Prozess der Nummernschilderkennung und Hintergrundprüfung mit speziellen Scan-Geräten und KI-basierten Identifizierungsdiensten automatisieren. Der Dienst wird derzeit für mehr als 150.000 Straßenparkplätze in der Stadt Amsterdam genutzt.
Dabei kommt ein dreistufiges Verfahren zum Einsatz. Im ersten Schritt fahren die Scan-Autos durch die Stadt und verwenden eine Objekterkennungssoftware, um die Nummernschilder der umliegenden Autos zu scannen. Nach der Identifizierung wird das Kennzeichen mit dem Nationalen Parkregister abgeglichen, um festzustellen, ob das Auto eine Parkerlaubnis für den betreffenden Standort hat. Gibt es keine gültige Parkerlaubnis, wird der Fall zur weiteren Bearbeitung an eine*n menschliche*n Kontrolleur*in weitergeleitet. Im letzten Schritt bewertet der*die Parkinspektor*in die gescannten Bilder, um festzustellen, ob das Nummernschild richtig erkannt wurde und ob besondere Umstände wie Be- oder Entladen vorliegt bzw. prüft vor Ort. Wenn kein triftiger Grund für das unbezahlte Parken gefunden wird, wird ein Parkschein ausgestellt.

[15] https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf

5.3 | Standards, Normen und Zertifizierungen

Eine essentielle Voraussetzung für den Einsatz von KI auch in sensiblen Anwendungsbereichen ist das nachhaltige Vertrauen der Nutzer*innen in diese Technologie. Unternehmen und Entwickler*innen sind daher angehalten, die Vertrauenswürdigkeit ihrer KI-Systeme zu gewährleisten und darzulegen. Die Europäische Kommission hat im April 2021 den weltweit ersten Rechtsrahmen für KI[16] vorgelegt und beabsichtigt durch die Entwicklung allgemeingültiger Normen die Vertrauenswürdigkeit von KI sicherzustellen. Jene KI-Systeme, die die Sicherheit, die Lebensgrundlagen und die Rechte der Menschen bedrohen, werden hingegen verboten. Für KI-Systeme mit hohem Risiko müssen künftig strenge Vorgaben erfüllt sein, die auch durch Normen abgedeckt werden können.[17]

Standardisierungsbemühungen im Themenfeld KI sind bereits im Gange, wie ein Blick auf die Website der internationalen Organisation zeigt, wo derzeit an der Entwicklung zahlreicher Standards zu KI gearbeitet wird[18]. Auf europäischer Ebene veröffentlichten das Europäische Komitee für Normung (CEN) und das Europäische Komitee für elektrotechnische Normung (CENELEC) eine KI Road Map[19]. Zudem haben DIN, DKE und das BMWK in Deutschland ebenfalls eine Standardization Roadmap on AI erarbeitet. Im Draft des Annual Union Work Programmes for Standardization 2022[20] der Europäischen Kommission findet sich schon konkret ein Eintrag zu sicheren und vertrauenswürdigen KI-Systemen, für die neue europäische Standards entwickelt werden sollen.

Standards und Normen, die Kriterien und Spezifikationen vorgeben, können die Entwicklung von Produkten, Services und Prozessen erleichtern, beispielsweise indem sie Interoperabilität oder sichere Prozesse gewährleisten. Neben wirtschaftlich positiven Effekten können Standards speziell auch technische Kriterien festlegen, die die Implementierung von ethischen Werten etwa in KI-Systemen unterstützt (DIN & DKE, 2020). Standards werden in Komitees entwickelt, die sich aus Experten und Expertinnen aus Wirtschaft, Wissenschaft, Verwaltung und Nicht-Regierungsorganisationen zusammensetzen, und stehen prinzipiell allen offen.[21]

Ein Weg zu vertrauenswürdiger Künstlicher Intelligenz kann somit die Zertifizierung von KI-Systemen sein. Sie ist eine meist zeitlich begrenzte Bestätigung durch unabhängige Dritte, dass

[16] https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206&from=EN , 16.12.2021

[17] Die Plattform lernende Systeme plädiert im Hinblick auf den Vorschlag der Europäischen Kommission auf die Berücksichtigung des jeweiligen Anwendungskontextes bei der Regulierung von KI-Systemen. Denn ein und das selbe System kann in einem Anwendungskontext völlig unproblematisch und in einem anderen jedoch durchaus kritisch sein. Ein Staubsaugerroboter beispielsweise gilt zwar als relativ unbedenklich, sammelt er aber persönliche Daten, die er an seinen Hersteller übermittelt, dann kann die Bewertung kritischer ausfallen. (Heesen et al. 2021, bzw. siehe auch Beispiel zum Einsatz von KI zur Überwachung von Fahrer*innen in Kapitel 4.2)

[18] https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0 , 16.12.2021

[19] https://standict.eu/sites/default/files/2021-03/CEN-CLC_FGR_RoadMapAI.pdf , 16.12.2021

[20] https://ec.europa.eu/docsroom/documents/45727?locale=en , 16.12.2021

[21] Für die Teilnahme können unterschiedliche Voraussetzungen gelten, wie etwa Kompetenz im Themenfeld, Fremdsprachenkenntnisse, Interesse am Thema, etc. Siehe auch: https://www.austrian-standards.at/de/standardisierung/standards-mitgestalten/in-komitees-teilnehmen , 16.12.2021

vorgegebene ethische und technische Standards, Normen oder Richtlinien erfüllt werden. Dies setzt allerdings voraus, dass es bereits entsprechende Standards und Normen gibt.

Dass ein gewisser Bedarf an Standards im Zusammenhang mit KI-Anwendungen besteht, ist daran ersichtlich, dass immer häufiger verschiedene Institutionen Prüfkataloge und Bewertungskriterien erstellen, um Entwickler*innen anzuleiten, wie KI-Anwendungen vertrauenswürdig gestaltet werden können. Gleichzeitig sollen KI-Prüfer*innen mit derartigen Katalogen und Checklisten bei der systematischen Evaluierung und Qualitätssicherung unterstützt werden.

Ein Bespiel dafür ist der vom **Fraunhofer-Institut** für Intelligente Analyse- und Informationssysteme herausgegebene Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz.

In Österreich bietet beispielsweise die Initiative **Trust your AI** mit Beteiligung der TU Graz eine 360°-Zertifizierung für Künstliche Intelligenz. Unternehmen werden dabei unterstützt, KI-Anwendungen sicher und vertrauenswürdig zu gestalten, es werden etwa Trainingsdaten als auch Vorhersagen von KI-Systemen im Hinblick auf mögliche Diskriminierung, unfaire Ungleichgewichte und Verzerrungen analysiert und die Robustheit des KI-Systems getestet. Der TÜV AUSTRIA bestätigt mit dem Zertifikat **TRUSTED AI** Robustheit, Sicherheit und Eignung einer zertifizierten KI-Anwendung für definierte Verwendungszwecke und Einsatzgebiete. Dazu wird in einem drei- bis viermonatigen Zertifizierungsverfahren ein Anforderungskatalog geprüft (u. a. funktionale Anforderungen zu Ethik und Datenschutz), Audits und technische Inspektionen durchgeführt. Alle drei Jahre ist zudem eine neuerliche Zertifizierung vorgesehen. Das Bundesrechenzentrum BRZ - das Kompetenzzentrum für Digitalisierung des heimischen Public Sectors - hat ebenfalls einen Prüfkatalog erarbeitet, der ein gemeinsames Verständnis zum Thema vertrauenswürdiger KI vermitteln soll. Ziel ist es, die mit dem Einsatz von KI-Systemen verbunden Risiken aufzuzeigen und zu reduzieren. Die Prüfung vertrauenswürdiger KI erfolgt in den Prüfbereichen, Transparenz, Verantwortung, Datenschutz, Zuverlässigkeit und Gerechtigkeit, wobei diese Kriterien auf 22 Prüfkriterien, 70 Prüfpunkte und mehr als 250 Merkmale heruntergebrochen werden.

Eine besondere Herausforderung für Zertifizierungen besteht jedoch in der Eigenschaft von lernenden KI-System sich kontinuierlich und eigenständig (und unter Umständen auch unvorhersehbar) weiterzuentwickeln. Das KI-System wird zwar vor dem erstmaligen Praxiseinsatz geprüft, wird aber womöglich einige Zeit nach seiner Inbetriebnahme den Kriterien der Zertifizierung nicht mehr gerecht. Deshalb ist es notwendig KI-Systeme in regelmäßigen Abständen zu re-zertifizieren bzw. sollten Zertifikate für KI diese Dynamik idealerweise berücksichtigen (Heesen et al., 2020). Madaio et al. (2020) zeigen jedoch, dass Checklisten zur Überprüfung von Vertrauenswürdigkeit teilweise falsch eingesetzt werden, wenn sie nicht den praktischen Bedürfnissen der Anwender*innen entsprechen. Demnach ist es essentiell, dass derartige Checklisten auf bestehende Arbeitsabläufe der Teams abgestimmt sind und von der Unternehmenskultur entsprechend mitgetragen werden. Jobin et al. (2019) analysieren 84 Dokumente zu ethischer KI und finden dabei nicht nur abweichende ethische Prinzipien, sondern auch unterschiedliche und mitunter sogar widersprüchliche Interpretationen sowie inkonsistente

Handlungsempfehlungen. Auch die Europäische Kommission (EK, 2021) stellt fest, dass „der reibungslose unionsweite Waren- und Dienstleistungsverkehr im Zusammenhang mit KI-Systemen durch das Entstehen eines Flickenteppichs potenziell abweichender nationaler Vorschriften behindert [wird]… Einzelstaatliche Konzepte zur Lösung der Probleme werden nur zu mehr Rechtsunsicherheit und zu Hemmnissen sowie zu einer langsameren Markteinführung von KI führen." Denn potenziell widersprüchliche nationale Vorschriften würden den freien Waren- und Dienstleistungsverkehr von Produkten mit KI-Technologien verunmöglichen. Daher schlägt die Kommission gemeinsames Handeln auf Unionsebene vor, um ausreichend Schutz zu gewährleisten und gleichzeitig die Wettbewerbsfähigkeit Europas und die Industriebasis im KI-Bereich zu stärken.

6 | Schlussfolgerungen und Handlungsoptionen

Die Literatur zu vertrauenswürdiger KI für die Digitalisierung in den Bereichen Produktion, Mobilität und Gesundheit zeigt folgende zentrale Handlungsfelder auf, die Gegenstand von Förderprogrammen und -maßnahmen sein könnten:

- **Beitrag zur Schaffung von Rechtssicherheit und Standards:** Derzeit mangelt es an (verpflichtenden) Standards und. Rechtssicherheit bei der Anwendung Künstlicher Intelligenz. Es gilt daher, ein Haftungssystem einzurichten, das für sämtliche Stakeholder Rechtssicherheit und Vertrauen gewährleistet, und möglicherweise dadurch sogar einen Wettbewerbsvorteil schafft. Während die Gesetzgebung in den Händen politischer Akteure liegt, könnte im Rahmen von Förderprogrammen die Schaffung von KI-Standards in Anwendungsbereichen oder Branchen unterstützt werden. Dies erscheint insofern als bedeutsam, da KI-Anwendungen zwar häufig ähnliche Aufgaben übernehmen (z. B. Bilderkennung, Vorhersagen, etc.) aber häufig jeweils sehr spezifisch in einem bestimmten Anwendungsbereich trainiert wurden. Eine spezialisierte KI-Lösung kann daher nicht ohne Weiteres in einen anderen Bereich eingesetzt werden. Auch für die technische Implementierung ethischer Vorgaben können Standards und Normen genutzt werden.
  - ⇨ Gremien zur Entwicklung von **Normen und Standards** für vertrauenswürdige KI einsetzen, um einheitliche technische Kriterien vertrauenswürdiger KI zu definieren und die Entwicklung und Anwendung vertrauenswürdiger KI zu erleichtern.
  - ⇨ **Zertifizierungen** können das Vertrauen in KI-Anwendungen weiter erhöhen. Eine Chance besteht darin, marktfähige Zertifizierungen von KI-Systemen „Made in Europe" zu etablieren, die Vertrauen und Orientierung schaffen die und die Dynamik von KI-Systemen berücksichtigen.
- **Stakeholder-Einbindung und Wissensdiffusion** ermöglicht es ethische Fragen aufzuwerfen sowie Bedenken zu äußern und sensibilisiert Akteur*innen im Bereich KI für ethische Fragestellungen.
  - ⇨ **Einbindung potenzieller Nutzer*innen** und sämtlicher direkter und indirekter Stakeholder von Beginn an in die KI-Entwicklung.

- ⇨ Durch eine **Diversifizierung von Entwickler*innen-Teams** können Merkmale wie Transparenz, Fairness und Erklärbarkeit von KI-Systemen stärker berücksichtigt werden.
- ⇨ **Wissensvermittlung** hinsichtlich der Bedeutung des Themas vertrauenswürdiger KI im Rahmen von KI-Förderungen betreiben und entsprechende Kriterien in anderen KI-Förderprogrammen berücksichtigen.

- **Sichtbarkeit erhöhen,** indem erfolgreich umgesetzte Projekte als Best Practice dienen und weitere Akteure aus der Wirtschaft auf das Thema aufmerksam machen sowie konkrete Anwendungsfälle vermitteln.
  - ⇨ Erfolgreiche Projekte und Anwendungsfälle als **Best Practice Beispiele** öffentlichkeitswirksam darstellen.
  - ⇨ **Leuchtturmprojekte** können die Sichtbarkeit vertrauenswürdiger KI-Systeme für die Öffentlichkeit erhöhen und deren praktische Relevanz für Entwickler*innen, Anwender*innen und Stakeholder aus Gesellschaft, Wirtschaft und Politik verdeutlichen.
- **Internationale Kooperation**: Derzeit findet ein globaler KI-Wettlauf statt: Nationen sind bestrebt die Technologieführerschaft zu übernehmen und Unternehmen versuchen Wettbewerbsvorteile durch den Einsatz von KI zu lukrieren: Die Einhaltung ethischer Standards bleibt dabei auf der Strecke.
  - ⇨ Verstärkte **länder- und sektorübergreifende Zusammenarbeit** könnte Fortschritte im Hinblick auf die Umsetzung einheitlicher Kriterien von vertrauenswürdiger KI erzielen (vgl. OECD, 2021).
- **Qualifizierung**: Einerseits ist eine adäquate Schulung von Arbeitnehmer*innen unerlässlich, um KI ordnungsgemäß einsetzen zu können (OECD, 2021). Andererseits muss ein allgemeines Verständnis von Vertrauenswürdigkeit und Ethik gegeben sein, um ethische Leitlinien und das Konzept der Vertrauenswürdigkeit von KI-Anwendungen erfolgreich implementieren zu können.
  - ⇨ **Aufnahme von ethischen Aspekten in Ausbildungsinhalte** – speziell in der IT-/Technikausbildung.
  - ⇨ Förderprogramme könnten **Fortbildungen zu ethischen Themen** und Aspekten unterstützen oder durch die Förderung interdisziplinärer Projektteams eine Sensibilisierung anregen.

Der Literaturstand zeigt auch einige Forschungslücken bzw. wirft folgende **Forschungsfragen** auf:

- Wie kann künftig im Zuge von immer größeren verfügbaren digitalen Datenmengen und technisch ausgereifteren KI-Systemen der Datenschutz sichergestellt werden?
- Wie kann sichergestellt werden, dass bestehende Ethik-Richtlinien tatsächlich Berücksichtigung finden bei der Entwicklung von KI-Anwendungen? Und wie müssen diese Richtlinien formuliert sein, damit ihre Beachtung bei der Entwicklung und Implementierung von KI-Anwendungen nicht zu einer Zunahme der Bürokratisierung führt und sich dadurch als hinderlich erweist?

- Wie können insbesondere spezialisierte KMU bei der Entwicklung von vertrauenswürdigen KI-Anwendungen im Hinblick auf die notwendige Interdisziplinarität (Ethik, Recht, Technik) unterstützt werden?
- Auf welche Weise gelingt es Awareness und Akzeptanz bei allen Beteiligten zu schaffen und sie in der Entwicklung und Anwendung von KI-Lösungen zu sensibilisieren und zu schulen?
- Wie müssen geeignete Zertifizierungsverfahren aussehen, die zum einen international akzeptierte Standards und Normen und auch die Dynamiken von selbstlernenden KI-Systemen berücksichtigen?

7 | Literaturverzeichnis

Ahlborn et al (2019): Technologieszenario “Künstliche Intelligenz in der Industrie 4.0“. Working Paper. BMWi. Berlin.

Ainek, M., Kors, J. Rijnbeek, P. (2020) The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, Volume 113. https://doi.org/10.1016/j.jbi.2020.103655.

Antweiler, D., Beckh, K., Sander, J., Rüping. S. (2020) Künstliche Intelligenz im Krankenhaus. Potenziale und Herausforderungen – eine Fallstudie im Bereich der Notfallversorgung. Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS.

Bauer W., Riedel, O., Renner, T., Preissner, M. (2021): Menschenzentrierte KI-Anwendungen in der Produktion. Praxiserfahrungen und Leitfaden zu betrieblichen Einführungsstrategien. Fraunhofer IAO. Stuttgart.

BMK, BMDW (2021): Strategie der Bundesregierung für künstliche Intelligenz. Artificial Intelligence Mission Austria 2030 (AIM AT 2030). Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie. Wien

Cihon, P. (2019): Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development. Technical Report. Future of Humanity Institute. University of Oxford. https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf

Christen, M., Mader, C. Čas, J., Abou-Chadi, T., Bernstein, A., Braun Binder, N., Dell' Aglio, D., Fábián, L., George, D., Gohdes, A., Hilty, L., Kneer, M., Krieger-Lamina, J., Licht, H., Scherer, A., Som, C., Sutter, P., Thouvenin, F. (2020) Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, TA-SWISS Publikationsreihe (Hrsg.), TA 72/2020. Zürich.

Cremens, A., Englander, A., Gabriel, M., Hecker, D., Mock, M., Poretschkin, M., Rosenzweig, J., Rotalski, F., Sicking, J., Volmer, J., Voosholz, J., Voss, A., Wrobel, S. (2019) Vertrauenswürdiger Einsatz von künstlicher Intelligenz. Handlungsfelder aus philosophischer, ethischer, rechtlicher und technologischer Sicht als Grundlage für eine Zertifizierung von künstlicher Intelligenz. Fraunhofer IAIS. Sankt Augustin.

De Nigris S., Craglia M., Nepelski D., Hradec J., Gómez-González E, Gomez E , M.Vazquez-Prada Baillet, R.Righi, G.De Prato, M.López Cobo, S.Samoili, M.Cardona (2020) AI Watch: AI Uptake in Health and Healthcare 2020, EUR 30478 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-26936-6, doi:10.2760/948860, JRC122675.

DIN, DKE (2020) German Standardization Roadmap on Artificial Intelligence. Berlin / Frankfurt am Main. https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf

Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, Behr SC, Flavell RR, Huang SY, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Hernandez Pampaloni M, Hadley D, Franc BL. (2019) A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology. 2019 Feb;290(2):456-464. doi: 10.1148/radiol.2018180958.

Gürtler, O. (2019) Künstliche Intelligenz als Weg zur wahren digitalen Transformation. In: Buxmann, P., Schmidt, H. (Hrsg): Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Springer Gabler. Berlin. S.95-104

Eiling, F., Huber, M. (2021) Automatische Programmierung von Produktionsmaschinen. In: Hartmann, E.A. (Hrsg.): Digitalisierung souverän gestalten. Innovative Impulse im Maschinenbau. Springer Vieweg. Berlin.

EK (2018b) Communication from the Commission to the European Parliament, The European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brüssel.

EK (2019) Die Europäer und die künstliche Intelligenz. Standard-Eurobarometer 92. Herbst 2019. Europäische Union.

EK (2020) Bericht der Kommission an das Europäische Parlament, den Rat und den europäischen Wirtschafts- und Sozialausschuss. Bericht über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung. Brüssel.

EK (2021) Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung Harmonisierter Vorschriften für Künstliche Intelligenz (Gesetzt über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union). Brüssel.

Niestadt, M, Debyser, A., Scordamaglia, D., Pape, M. (2019) Artificial intelligence in transport. Current and future developments, opportunities and challenges. EPRS. Europäische Union.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & More Authors (2018) AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Hagendorff, T. (2020) The Ethics of AI Ethics: An Evaluation of Guidelines. Minds & Machines 30, 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8.

Hao, K. (2019) In 2020 let's stop AI ethics-washing and actually do something, MIT Technology Review. Available from: https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/.

Heesen J., Müller-Quade, J. & Wrobel, S. (2020) Zertifizierung von KI-Systemen – Impulspapier aus der Plattform Lernende Systeme. München.

Heesen J., Müller-Quade, J. & Wrobel, S. (2021) Kritikalität von KI-Systemen in ihren jeweiligen Anwendungskontexten – Ein notwendiger, aber nicht hinreichender Baustein für Vertrauenswürdigkeit. Whitepaper aus der Plattform Lernende Systeme, München. DOI: https://doi.org/10.48669/pls_2021-3.

Heusser D., Schmidt, A., Frederiksen, A., Ayaz, B., Lange, H., Kesic, M, Petri, R. (2021) Autonomes Fahren. VDE Faktencheck.

Jobin, A.; Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Leslie, D. (2019) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529.

Madaio, M. A., Stark, L., Wortman Vaughan, J., and Wallach, H. (2020) Co-Designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI. Proc. 2020 CHI Conf. Hum. Factors Comput. Syst., 1–14. doi:10.1145/3313831.3376445.

McNamara, A., Smith, J., Murphy-Hill, E. (2018) Does ACM's code of ethics change ethical decision making in software development?" In G. T. Leavens, A. Garcia, C. S. Păsăreanu (Eds.) Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering—ESEC/FSE 2018 (pp. 1–7). New York: ACM Press.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342.

OECD (2020a) Künstliche Intelligenz in der Gesellschaft, OECD Publishing, Paris, https://doi.org/10.1787/6b89dea3-de.

OECD (2020) Trustworthy AI in health. Background paper for the G20 AI Dialogue, Digital Econmy Task Force. Saudi Arabia, 1-2 April 2020.

OECD (2021) "An overview of national AI strategies and policies", *Going Digital Toolkit Note*, No. 14, *https://goingdigital.oecd.org/data/toolkitnotes/No14_ToolkitNote_AIStrategies.pdf.*

Paiva, S.; Ahad, M.A.; Tripathi, G.; Feroz, N.; Casalino, G. (2021) Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges. Sensors 2021, 21, 2143. https://doi.org/10.3390/s21062143

Pentenrieder, A., Bertini, A., Künzel, M. (2021) Digitale Souveränität als Trend? Der Werkzeugmaschinenbau als wegweisendes Modell für die deutsche Wirtschaft. In: Hartmann, E.A.: Digitalisierung souverän gestalten. Innovative Impulse im Maschinenbau. Springer Vieweg, Berlin.

Poretschkin, M., Schmitz, A., Akila, M., Adilova, L., Becker, D., Cremers, A., Hecker, D., Houben. D., Mock, M., Rosenzweig, J., Sicking, J., Schulz, E., Voss, A., Wrobel, S. (2021) Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. KI-Prüfkatalog. Fraunhofer IAIS.

Prem, E., Ruhland, S. (2019) AI in Österreich. Eine Annäherung auf Basis wirtschaftlicher Analysen. BMVIT. Wien.

Seifert, I., Bürger, M., Wangler, L., Christmann-Budian S., Rohde, M., Gabriel, P., Zinke., G. (2018) Potenziale Künstlicher Intelligenz im produzierenden Gewerbe in Deutschland. Studie im Auftrag des BMWi im Rahmen der Begleitforschung zum Technologieprogramm PAiCE – Platforms | Additive Manufacturing | Imaging |Communication | Engineering. Iit-Institut für Innovation und Technik in der VDI/VDE Innovation + Technik GmbH. Berlin.

Thiebes, S., Lins, S., Sunyaev, A. (2020): Trustworthy artificial intelligence. Electronic Markets. https://doi.org/10.1007/s12525-020-00441-4.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, SD., Tegmark, M., F., Nerini, F. (2020) The role of artificial intelligence in achieving the sustainable development goals. Nat Commun. 2020; 11(1): 233.

WHO (2021) Ethics and Governance of artificial Intelligence for health: WHO guidance.Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

Zahradnik, G., Dachs, B., Rhomberg, W., Leitner, K.-H. (2019): Trends und Entwicklungen in der österreichischen Produktion. Highlights aus dem European Manufacturing Survey 2018. Austrian Institute of Technology, Wien.

Zicari, R.V., Ahmed, S., Amann, J., Braun S.A., Brodersen, J., Bruneault, F., Brusseau, J., Campano, E., Coffee, M., Dengel, A., Düdder, B., Gallucci, A., Gilbert, T.K., Gottfrois, P., Goffi, E., Haase, C.B., Hagendorff, T., Hickman, E., Hildt. E., Holm, S., Kringen, P., Kühne, U., Lucieri, A., Madai, V.I., Moreno-Sánchez, P.A., Medlicott, O., Ozols, M., Schnebel, E., Spezzatti, A., Tithi, J.J., Umbrello, S., Vetter, D., Volland, H., Westerlund, M., Wurth, R. (2021) Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion. Classifier. *Front. Hum. Dyn* 3:688152. doi: 10.3389/fhumd.2021.688152