UNABHÄNGIGE

HOCHRANGIGE EXPERTENGRUPPE FÜR KÜNSTLICHE INTELLIGENZ

EINGESETZT VON DER EUROPÄISCHEN KOMMISSION IM JUNI 2018

EINE DEFINITION DER KI: WICHTIGSTE FÄHIGKEITEN UND WISSENSCHAFTSGEBIETE

Für die Zwecke der Gruppe entwickelte Definition

Eine Definition der KI:
Wichtigste Fähigkeiten und Wissenschaftsgebiete

Hochrangige Expertengruppe für künstliche Intelligenz

Haftungsausschluss und Verwendung dieses Dokuments: In der folgenden Beschreibung und Definition der Fähigkeiten und Forschungsbereiche der KI wird der Stand der Technik in stark vereinfachter Form dargestellt. Absicht dieses Papiers ist es nicht, alle Techniken und Fähigkeiten der KI genau und umfassend zu definieren – vielmehr soll ein Überblick über das gemeinsame Verständnis dieser Disziplin vermittelt werden, das die hochrangige Expertengruppe in ihren Ergebnissen verwendet. Wir hoffen jedoch, dass dieses Papier auch Nichtfachleuten einen hilfreichen Einstieg in die KI bietet und dazu anregt, sich ausführlicher und intensiver mit der Thematik zu befassen und sich ein genaueres Bild von dieser Disziplin und Technologie zu machen.

Die HEG-KI ist eine unabhängige Expertengruppe, die im Juni 2018 von der Europäischen Kommission eingesetzt wurde.

Kontakt Nathalie Smuha – Koordinatorin für die HEG-KI
E-Mail CNECT-HLG-AI@ec.europa.eu

Europäische Kommission
B-1049 Bruxelles/Brüssel

Veröffentlichung des Dokuments am X. April 2019.

Ein erster Entwurf dieses Dokuments wurde am 18. Dezember 2018 zusammen mit dem ersten Entwurf der Ethik-Leitlinien der HEG-KI für eine vertrauenswürdige KI veröffentlicht. Der Entwurf wurde unter Berücksichtigung der Stellungnahmen der Europäischen KI-Allianz und der im Rahmen der offenen Konsultation über den Entwurf der Leitlinien eingegangenen Rückmeldungen überarbeitet. Wir möchten uns ausdrücklich und herzlich bei allen Beteiligten für die Rückmeldungen zum ersten Entwurf dieses Dokuments bedanken.

EINE DEFINITION DER KI: WICHTIGSTE FÄHIGKEITEN UND WISSENSCHAFTSGEBIETE

Unser Ausgangspunkt ist die folgende, in der Mitteilung der Kommission über die KI[1] vorgeschlagene Definition der künstlichen Intelligenz (KI):

> *„Künstliche Intelligenz (KI) bezeichnet Systeme mit einem „intelligenten" Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen.*
>
> *KI-basierte Systeme können rein softwaregestützt in einer virtuellen Umgebung arbeiten (z. B. Sprachassistenten, Bildanalysesoftware, Suchmaschinen, Sprach- und Gesichtserkennungssysteme), aber auch in Hardware-Systeme eingebettet sein (z. B. moderne Roboter, autonome Pkw, Drohnen oder Anwendungen des ‚Internet der Dinge').„*

In diesem Papier erweitern wir diese Definition, um bestimmte Aspekte der KI als wissenschaftliche Disziplin und als Technologie klarzustellen. Ziel ist es, Missverständnissen vorzubeugen und eine gemeinsame Wissensbasis über die KI zu schaffen, die auch von Nichtfachleuten sinnvoll genutzt werden kann. Außerdem soll mit den hier zusammengestellten Informationen ein nützlicher Beitrag zu der Diskussion sowohl über die KI-Ethik-Leitlinien als auch über die KI-Politikempfehlungen geleistet werden.

1. KI-Systeme

Mit dem Ausdruck „KI" wird explizit auf den Begriff der Intelligenz Bezug genommen. Da jedoch das Konzept der Intelligenz (in Maschinen wie im Menschen) unbestimmt ist – obwohl von Psychologen, Biologen und Neurowissenschaftlern ausführlich untersucht –, verwenden KI-Forscher in der Regel den Begriff der Rationalität. Darunter wird die Fähigkeit verstanden, unter Berücksichtigung bestimmter zu optimierender Kriterien und der verfügbaren Mittel das bestmögliche Handeln zu wählen, um ein bestimmtes Ziel zu erreichen. Rationalität ist natürlich nicht der einzige, aber doch ein wesentlicher Bestandteil des Konzepts der Intelligenz.

In den folgenden Ausführungen ist unter dem Begriff *KI-System* jede Form von KI-gestützter Komponente, Soft- und/oder Hardware zu verstehen. Tatsächlich sind KI-Systeme normalerweise keine eigenständigen Systeme, sondern als Komponenten in größere Systeme *eingebettet*.

Ein KI-System ist somit einem der wichtigsten Standardwerke über KI[2] zufolge in erster Linie rational. Doch wie erzielt ein KI-System Rationalität? Wie aus dem ersten Satz der KI-Arbeitsdefinition oben hervorgeht, nimmt es dazu die Umgebung, in die es eingebettet ist, durch Sensoren wahr, erfasst und interpretiert dabei Daten, zieht aus dem Wahrgenommenen Schlussfolgerungen oder verarbeitet die aus diesen Daten abgeleiteten Informationen, und entscheidet über das bestmögliche Handeln, um dieses anschließend mittels Aktoren umzusetzen, wobei es die Umgebung möglicherweise verändert. KI-Systeme können entweder symbolische Regeln verwenden oder ein numerisches Modell erlernen, und sind auch in der Lage, die Auswirkungen ihrer früheren Handlungen auf die Umgebung zu analysieren und ihr Verhalten entsprechend anzupassen. Die Darstellung eines KI-Systems in Abbildung 1 kann als Orientierung dienen.

---

1 Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen „Künstliche Intelligenz für Europa", Brüssel, 25.4.2018, COM(2018) 237 final.

2 Russel, S., und Norvig, P.: „*Artificial Intelligence: A Modern Approach*". Prentice Hall, 3. Auflage, 2009.

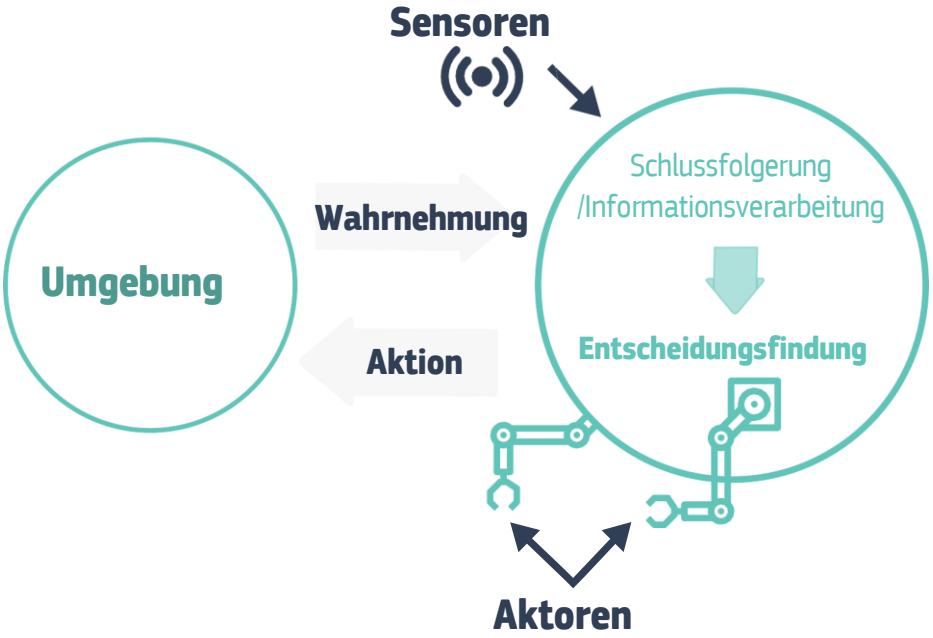

Abbildung 1: Schematische Darstellung eines KI-Systems.

Sensoren und Wahrnehmung. In Abbildung 1 sind die Sensoren des Systems durch ein WLAN-Symbol gekennzeichnet. Konkret könnte es sich dabei um Kameras, Mikrofone, eine Tastatur, eine Website oder ein sonstiges Eingabegerät, aber auch um Sensoren zur Messung physikalischer Größen (z. B. Temperatur, Druck, Entfernung, Kraft/Drehmoment, Tastsensoren) handeln. Im Allgemeinen muss das KI-System über geeignete Sensoren verfügen, damit es aus seiner Umgebung die Daten erfassen kann, die für das von seinem menschlichen Konstrukteur vorgegebene Ziel relevant sind. Soll beispielsweise ein AI-System entwickelt werden, das den Fußboden eines Raums bei Verschmutzung automatisch reinigt, könnten die Sensoren Kameras einschließen, die den Fußboden fotografieren.

Was die erfassten Daten betrifft, so bietet sich oft eine Unterscheidung zwischen strukturierten und unstrukturierten Daten an. Bei *strukturierten Daten* handelt es sich um nach vordefinierten Modellen organisierte Daten (etwa in einer relationalen Datenbank); *unstrukturierte Daten* haben dagegen keine bekannte Organisationsform (das gilt zum Beispiel für Bilder oder Text).

Schlussfolgerung/Informationsverarbeitung und Entscheidungsfindung. Den Kern eines KI-Systems bildet sein Modul für Schlussfolgerung/Informationsverarbeitung, das die von den Sensoren erfassten Daten als Input nutzt und zur Erreichung des vorgegebenen Ziels ein bestimmtes Handeln vorschlägt. Deshalb müssen die von den Sensoren erfassten Daten in Informationen umgewandelt werden, die das Modul für Schlussfolgerung/Informationsverarbeitung verstehen kann. In unserem Beispiel des KI-gestützten Reinigungssystems würde die Kamera dem Modul für Schlussfolgerung/Informationsverarbeitung eine Aufnahme des Fußbodens übermitteln; das Modul muss dann entscheiden, ob der Fußboden gereinigt wird oder nicht (also entscheiden, welches Handeln am ehesten zu dem gewünschten Ziel führt). Während wir Menschen anhand einer Aufnahme des Fußbodens leicht entscheiden können, ob er gereinigt werden muss, fällt das einer Maschine schwerer, weil das Bild nur aus einer Folge von Nullen und Einsen besteht. Das Modul für Schlussfolgerung/Informationsverarbeitung muss also Folgendes leisten:

1. Es muss das Bild interpretieren, um zu entscheiden, ob der Fußboden sauber ist oder nicht. Im Allgemeinen muss es dazu Daten in Informationen umwandeln und die gewonnenen Informationen in prägnanter Weise modellieren können, dabei aber alle relevanten Datenelemente (in diesem Fall die Information über den Sauberkeitszustand des Fußbodens) berücksichtigen.

2. Es muss Schlussfolgerungen aus diesem Wissen ziehen oder diese Informationen verarbeiten und daraus ein numerisches Modell (d. h. eine mathematische Formel) erstellen, um über die beste Aktion zu entscheiden. Für unser Beispiel bedeutet das: Wird aus dem Bild die Information abgeleitet, dass der Fußboden schmutzig ist, wäre die beste Handlung die Aktivierung des Reinigungsvorgangs; andernfalls wäre es am besten, nichts zu unternehmen.

Es gilt zu beachten, dass der Begriff „Entscheidung" im weiten Sinne als jedes Auswählen einer durchzuführenden Handlung zu verstehen ist und nicht notwendigerweise bedeutet, dass KI-Systeme völlig autonom sind. Die Entscheidung kann also auch in der Auswahl einer Empfehlung an einen Menschen bestehen, der dann die endgültige Entscheidung trifft.

Ausführung. Nachdem über die Handlung entschieden wurde, kann das AI-System diese Handlung mithilfe der ihm zur Verfügung stehenden Aktoren ausführen. Die in der vorstehenden Skizze als Gelenkarme dargestellten Aktoren müssen nicht physischer Art sein. Es könnte sich auch um Software handeln. In unserem Reinigungsbeispiel könnte das KI-System ein Signal erzeugen, das einen Staubsauger aktiviert, wenn die Handlung in der Reinigung des Fußbodens besteht. Ein anderes Beispiel wäre der Fall eines Dialogsystems (eines Chatbots), das handelt, indem es als Antwort auf die Aussagen des Nutzers Sprachmitteilungen generiert.

Da die ausgeführte Handlung eine Veränderung der Umgebung bewirken könnte, muss das System seine Sensoren beim nächsten Mal erneut einsetzen, um möglicherweise veränderte Informationen aus einem modifizierten Umfeld zu erfassen.

Rationale KI-Systeme wählen nicht immer das bestmögliche Handeln für das vorgegebene Ziel. Sie verhalten sich also nur *begrenzt rational*, was auf beschränkte Ressourcen wie Zeit oder Rechenkapazität zurückzuführen ist.

Rationale KI-Systeme sind eine sehr einfache Form der KI-Systeme. Sie verändern die Umgebung, passen ihr Verhalten aber nicht mit der Zeit an, um ihr Ziel besser zu erreichen. Ein *lernendes rationales System* bewertet (mittels Wahrnehmung) nach Ausführung einer Handlung den neuen Zustand der Umgebung, um festzustellen, wie erfolgreich sein Verhalten war, und passt daraufhin seine Regeln des Schlussfolgerns und seine Entscheidungsverfahren an.

2. KI als wissenschaftliche Disziplin

Bei der vorstehenden Beschreibung handelt es sich um eine sehr einfache und abstrakte Darstellung eines KI-Systems anhand seiner drei zentralen Fähigkeiten: Wahrnehmung, Schlussfolgerung/Entscheidungsfindung und Ausführung. Sie genügt jedoch, um ein Grundverständnis der meisten heute in der Entwicklung von KI-Systemen verwendeten KI-Techniken und -Teildisziplinen zu vermitteln, zumal sie alle zu den verschiedenen Fähigkeiten der Systeme in Bezug stehen. Allgemein formuliert lassen sich all diese Techniken in zwei an den Fähigkeiten des *Schlussfolgerns* und des *Lernens* orientierte Hauptgruppen unterteilen. Eine weitere sehr wichtige Disziplin ist die Robotik.

Schlussfolgern und Entscheiden. Diese Gruppe von Techniken umfasst Wissensrepräsentation und -verarbeitung, Planung, Terminierung, Suche und Optimierung. Mithilfe dieser Techniken ist das System in der Lage, aus den von den Sensoren gelieferten Daten Schlussfolgerungen zu ziehen. Damit das geleistet werden kann, müssen Daten in Wissen umgewandelt werden. Daher beschäftigt sich ein Bereich der KI mit der Frage, wie solches Wissen am besten modelliert werden kann (*Wissensrepräsentation*). Im Anschluss an die Wissensmodellierung werden im nächsten Schritt Schlussfolgerungen aus dem Wissen gezogen (Wissensverarbeitung). Dazu gehören die Bildung von Rückschlüssen durch symbolische Regeln, *Planungs*- und *Terminierungsaktivitäten*, *Durchsuchen* eines umfassenden Lösungspakets und *Optimierung* unter allen Problemlösungsalternativen. Im letzten Schritt wird entschieden, welche Handlung erfolgen soll. In der Regel ist der Teil der Schlussfolgerung und Entscheidungsfindung in einem KI-System sehr komplex und erfordert eine Kombination aus mehreren der genannten Techniken.

Lernen. Zu dieser Gruppe von Techniken zählen maschinelles Lernen, neuronale Netze, Deep Learning, Entscheidungsbäume und viele weitere Lerntechniken. Mithilfe dieser Techniken lernen KI-Systeme, wie sie nicht

genau definierbare Probleme oder solche lösen können, deren Lösungsweg nicht durch Regeln des symbolischen Schlussfolgerns abgebildet werden kann. Solche Problemstellungen ergeben sich beispielsweise im Zusammenhang mit Wahrnehmungsfähigkeiten wie *Sprachverstehen* und *Sprachenverständnis*, *maschinellem Sehen* oder der *Verhaltensvorhersage*. Diese Probleme erscheinen zwar zunächst einfach, weil sie für den Menschen normalerweise tatsächlich leicht lösbar sind, ein KI-System hingegen stellen sie vor Schwierigkeiten, weil es sich nicht (zumindest noch nicht) auf alltägliches Denken/Schließen stützen kann, und besonders schwierig wird es, wenn das System unstrukturierte Daten interpretieren muss. An dieser Stelle sind Techniken nach dem Ansatz des *maschinellen Lernens* hilfreich. Maschinelle Lernverfahren können aber über den Bereich der Wahrnehmung hinaus auch in vielen anderen Aufgabenbereichen eingesetzt werden. Maschinelle Lernverfahren erzeugen ein numerisches Modell (also eine mathematische Formel), das die Entscheidung aus den Daten berechnet.

Maschinelles Lernen hat verschiedenste Ausprägungen. Zu den gängigsten Ansätzen gehören *überwachtes Lernen*, *unüberwachtes Lernen* und *bestärkendes Lernen*.

Beim überwachten maschinellen Lernen werden dem System keine Verhaltensregeln vorgegeben, sondern es bekommt Beispiele für Input-Output-Verhalten in der Hoffnung, dass es anhand der Beispiele (die typischerweise die Vergangenheit beschreiben) Gesetzmäßigkeiten nachbilden kann und sich auch in Situationen angemessen verhält, die sich nicht in den Beispielen wiederfinden (und künftig auftreten könnten). In unserem Fallbeispiel könnten wir dem System viele Bilder eines Fußbodens zeigen und dazu die entsprechende Interpretation (d. h. ob der Fußboden auf einem bestimmten Bild sauber ist oder nicht) liefern. Wenn wir das System mit genügend Beispielen füttern, die eine ausreichende Bandbreite möglicher Situationen abbilden, wird es durch seinen maschinellen Lernalgorithmus Gesetzmäßigkeiten entwickeln und auch Bilder von Fußböden interpretieren können, die es zuvor nie gesehen hat. Manche maschinelle Lernverfahren arbeiten mit Algorithmen, die auf dem Konzept der *neuronalen Netze* beruhen. Das Konzept besteht, in Anlehnung an die Funktionsweise des menschlichen Gehirns, aus einem Netzwerk kleiner Verarbeitungseinheiten (ähnlich unseren Neuronen), die durch zahlreiche gewichtete Verbindungen miteinander verknüpft sind. Als Input erhält das neuronale Netz von den Sensoren erzeugte Daten (in unserem Beispiel das Bild des Fußbodens); der Output ist die Interpretation des Bildes (in unserem Beispiel die Aussage, ob der Fußboden sauber ist oder nicht). Während der Analyse der Beispiele (in der *Trainingsphase* des Netzes) werden die Gewichtungen der Verbindungen so angepasst, dass sie der Aussage der verfügbaren Beispiele möglichst nahe kommen (damit also die Abweichung zwischen erwartetem Output und vom Netz berechnetem Output minimiert wird). Auf die Trainingsphase folgt eine Testphase, in der das Verhalten des neuronalen Netzes anhand zuvor nicht gesehener Beispiele überprüft wird, um festzustellen, ob die Aufgabe gut erlernt wurde.

Bei dieser Methode ist zu beachten, dass (wie bei allen maschinellen Lernverfahren) immer mit einer gewissen – wenn auch meist geringen – Fehlerquote zu rechnen ist. Von wesentlicher Bedeutung ist daher die *Genauigkeit* als Messgröße dafür, wie groß der Anteil der korrekten Antworten ist.

Es gibt verschiedene Arten von neuronalen Netzen und Ansätzen des maschinellen Lernens. Zu den derzeit erfolgreichsten gehört das *Deep Learning*. Dieser Ansatz macht sich die Tatsache zunutze, dass beim neuronalen Netz verschiedene Ebenen zwischen Input und Output bestehen und dass der Gesamtzusammenhang zwischen Input und Output daher schrittweise erlernbar ist. Dadurch wird das Verfahren insgesamt präziser und benötigt weniger Steuerung durch den Menschen.

Neuronale Netze sind nur ein Werkzeug für maschinelles Lernen – es gibt zahlreiche andere mit jeweils unterschiedlichen Eigenschaften: Random Forests und Boosted Trees, Clusterverfahren, Matrixfaktorisierung usw.

Ein weiteres hilfreiches maschinelles Lernverfahren ist das *bestärkende Lernen*. Hierbei darf das KI-System nach und nach seine eigenen Entscheidungen treffen und erhält zu jeder Entscheidung die Rückmeldung, ob es gut oder schlecht entschieden hat. Ziel des Systems ist es, im Laufe der Zeit möglichst viele positive Rückmeldungen zu bekommen. Dieses Verfahren kommt beispielsweise in Empfehlungssystemen (wie etwa die Empfehlungssysteme im Internet, mit denen Kaufvorschläge generiert werden) oder im Marketing zum Einsatz.

Maschinelle Lernverfahren eignen sich nicht nur für Wahrnehmungsaufgaben wie Bild- und Spracherkennung, sondern auch für die vielen schwer definierbaren Aufgaben, die durch symbolische Verhaltensregeln nicht umfassend abgebildet werden können.

Es ist wichtig, zwischen maschinellen Lernverfahren, die dem Erlernen einer neuen, symbolisch nur schwer abbildbaren Aufgabe dienen, und den (im vorigen Abschnitt erwähnten) lernenden rationalen Agenten zu unterscheiden, die ihr Verhalten mit der Zeit anpassen, um ein bestimmtes Ziel besser zu erreichen. Beide Techniken können sich überschneiden oder zusammenarbeiten, sind aber nicht notwendigerweise miteinander identisch.

Robotik. Die Robotik lässt sich als „KI im Einsatz in der physischen Welt" (auch als *eingebettete KI* bezeichnet) definieren. Ein Roboter ist eine physische Maschine und muss die Dynamik, Ungewissheit und Komplexität der physischen Welt bewältigen. Die Fähigkeiten des Wahrnehmens, Schlussfolgerns, Handelns und Lernens sowie der Interaktion mit anderen Systemen sind in der Regel in die Steuerungsarchitektur des Robotersystems integriert. Neben der KI spielen bei der Entwicklung und Anwendung von Robotern auch andere Disziplinen wie Maschinenbau und Steuerungstheorie eine Rolle. Beispiele für Roboter sind Roboter-Manipulatoren, autonome Fahrzeuge (z. B. Autos, Drohnen oder Flugtaxis), humanoide Roboter, Staubsauger-Roboter usw.

In Abbildung 2 sind die meisten der vorstehend genannten Teildisziplinen mit ihren Zusammenhängen dargestellt. Es gilt jedoch zu beachten, dass die KI wesentlich komplexer als hier dargestellt ist und zahlreiche andere Teildisziplinen und Verfahren umfasst. Zudem stützt sich die Robotik, wie bereits erwähnt, auf weitere Technologien, die nicht zum Gebiet der KI gehören. Unseres Erachtens reichen diese Ausführungen jedoch aus, um einen fruchtbaren Beitrag zum Austausch, zur Bewusstseinsbildung und zur Diskussion über die KI, die KI-Ethik und die KI-Politik zu leisten, die innerhalb der höchst multidisziplinären und aus unterschiedlichsten Akteuren zusammengesetzten Hochrangigen Expertengruppe stattfinden muss.

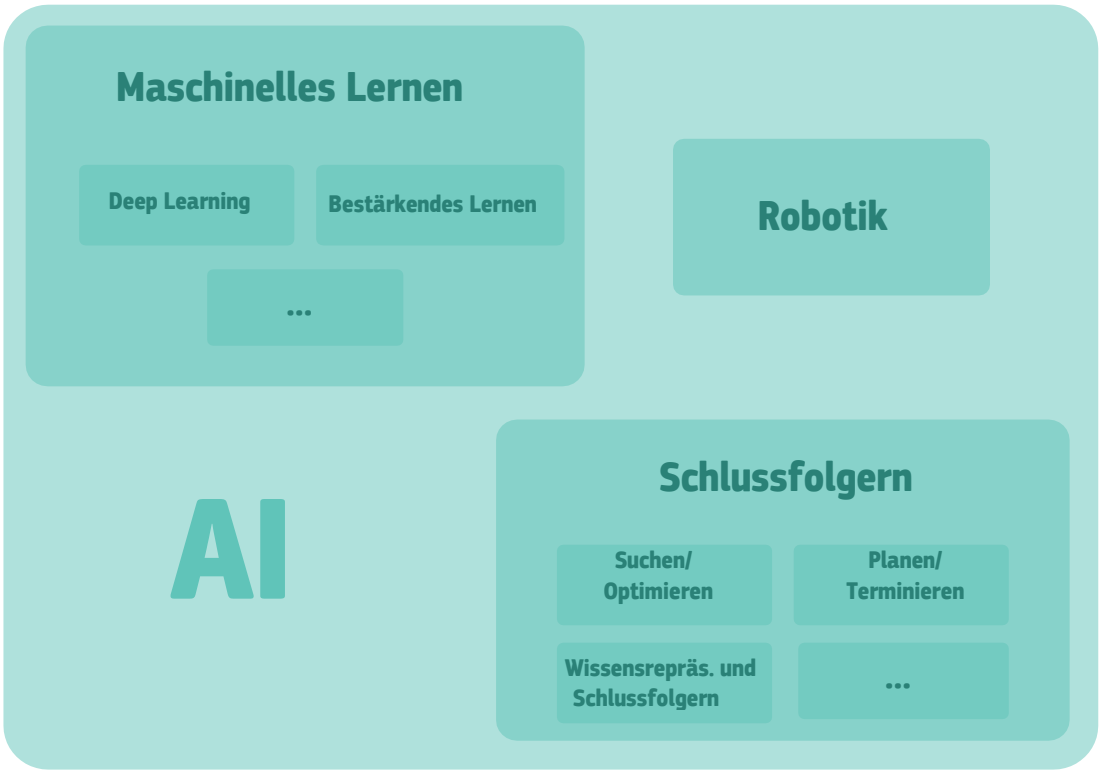

Abbildung 2: Vereinfachter Überblick über die Teildisziplinen der KI und ihre Zusammenhänge. Maschinelles Lernen und Schlussfolgern umfassen jeweils eine Reihe anderer Techniken; die Robotik wiederum beinhaltet Techniken, die nicht zur KI gehören. Die KI als Ganzes ist eine Teildisziplin der Informatik.

3. Andere wichtige Konzepte und Fragestellungen der KI

Enge oder angewandte (schwache) KI und allgemeine (starke) KI. Ein allgemeines KI-System soll die meisten Aufgaben erfüllen können, zu denen auch der Mensch in der Lage ist. Enge (oder angewandte) KI-Systeme hingegen sollen eine oder einige wenige spezifische Aufgaben erledigen. Bei den heute verwendeten KI-Systemen handelt es sich um Beispiele der engen KI. Ursprünglich wurde in der Forschung eine andere Terminologie (schwache und starke

KI) verwendet. Es sind noch viele ethische, wissenschaftliche und technische Herausforderungen zu bewältigen, bevor die Fähigkeiten realisiert werden können, die für eine allgemeine KI erforderlich wären. Dazu gehören alltägliches Denken/Schließen, Ichbewusstsein und die Fähigkeit der Maschine, ihren Zweck selbst zu bestimmen.

Datenqualität und Verzerrungen. Da viele KI-Systeme – zum Beispiel solche mit Komponenten des überwachten maschinellen Lernens – enorme Datenmengen brauchen, um gute Ergebnisse zu erzielen, ist es wichtig, zu verstehen, wie Daten das Verhalten des KI-Systems beeinflussen. Enthalten die Trainingsdaten Verzerrungen, sind sie also nicht ausgewogen und umfassend genug, so wird das mit diesen Daten trainierte System keine angemessenen Gesetzmäßigkeiten abbilden können und möglicherweise unfaire Entscheidungen treffen, die bestimmte Gruppen gegenüber anderen bevorzugen können. Die KI-Gemeinschaft arbeitet seit Kurzem an der Entwicklung von Methoden zur Erkennung und Abmilderung von Verzerrungen in den Trainingsdatensätzen und anderen Bereichen des KI-Systems.

Black-Box-KI und Erklärbarkeit. Einige Verfahren des maschinellen Lernens sind zwar im Hinblick auf die Genauigkeit sehr erfolgreich, aber gleichzeitig auch sehr undurchsichtig, was die Art und Weise ihrer Entscheidungsfindung betrifft. Solche Szenarien sind mit dem Begriff „*Black-Box-KI*" gemeint – wenn also der Grund für bestimmte Entscheidungen nicht nachvollziehbar ist. Die Erklärbarkeit wiederum ist eine Eigenschaft jener KI-Systeme, die eine Art von Begründung für ihr Verhalten liefern können.

Zielgerichtete KI. Die heutigen KI-Systeme arbeiten zielgerichtet. Das bedeutet, dass ihnen vom Menschen ein bestimmtes Ziel vorgegeben wird, welches sie mithilfe bestimmter Verfahren umsetzen. Sie bestimmten ihre Ziele nicht selbst. Manche KI-Systeme jedoch (zum Beispiel solche, die auf bestimmten maschinellen Lernverfahren beruhen) können freier darüber entscheiden, auf welche Weise sie das gesetzte Ziel erreichen.

4. Eine aktualisierte Definition der KI

Wir schlagen vor, die folgende aktualisierte Definition der KI zu verwenden:

> „Systeme der künstlichen Intelligenz (KI-Systeme) sind vom Menschen entwickelte Softwaresysteme (und gegebenenfalls auch Hardwaresysteme)[3], die in Bezug auf ein komplexes Ziel auf physischer oder digitaler Ebene handeln, indem sie ihre Umgebung durch Datenerfassung wahrnehmen, die gesammelten strukturierten oder unstrukturierten Daten interpretieren, Schlussfolgerungen daraus ziehen oder die aus diesen Daten abgeleiteten Informationen verarbeiten, und über das bestmögliche Handeln zur Erreichung des vorgegebenen Ziels entscheiden. KI-Systeme können entweder symbolische Regeln verwenden oder ein numerisches Modell erlernen, und sind auch in der Lage, die Auswirkungen ihrer früheren Handlungen auf die Umgebung zu analysieren und ihr Verhalten entsprechend anzupassen.
>
> Als wissenschaftliche Disziplin umfasst die KI mehrere Ansätze und Techniken wie z. B. maschinelles Lernen (Beispiele dafür sind „Deep Learning" und bestärkendes Lernen), maschinelles Denken (es umfasst Planung, Terminierung, Wissensrepräsentation und Schlussfolgerung, Suche und Optimierung) und die Robotik (sie umfasst Steuerung, Wahrnehmung, Sensoren und Aktoren sowie die Einbeziehung aller anderen Techniken in cyber-physische Systeme)."

Ferner schlagen wir vor, zur Unterstützung dieser Definition auf dieses Dokument als Quelle zusätzlicher Informationen zu verweisen.

[3] Menschen entwerfen KI-Systeme direkt, sie können deren Entwurf aber auch mithilfe von KI-Techniken optimieren.

Dieses Dokument wurde von den Mitgliedern der hochrangigen Expertengruppe für KI erstellt.

Beteiligte Mitglieder in alphabetischer Reihenfolge:

Pekka Ala-Pietilä, Vorsitzender der HEG-KI
AI Finland, Huhtamaki, Sanoma
Wilhelm Bauer
Fraunhofer
Urs Bergmann
Zalando
Mária Bieliková
Slowakische Technische Universität Bratislava
Cecilia Bonefeld-Dahl
DigitalEurope
Yann Bonnet
ANSSI
Loubna Bouarfa
OKRA
Stéphan Brunessaux
Airbus
Raja Chatila
IEEE Initiative Ethics of Intelligent/Autonomous Systems und Universität Sorbonne
Mark Coeckelbergh
Universität Wien
Virginia Dignum
Universität Umea
Luciano Floridi
Universität Oxford
Jean-Francois Gagné
Element AI
Chiara Giovannini
ANEC
Joanna Goodey
Agentur der Europäischen Union für Grundrechte
Sami Haddadin
Munich School of Robotics and Machine Intelligence

Gry Hasselbalch
The thinkdotank DataEthics und Universität Kopenhagen
Fredrik Heintz
Universität Linköping
Fanny Hidvegi
Access Now
Eric Hilgendorf
Universität Würzburg
Klaus Höckner
Hilfsgemeinschaft der Blinden und Sehschwachen
Mari-Noëlle Jégo-Laveissière
Orange
Leo Kärkkäinen
Nokia Bell Labs
Sabine Theresia Köszegi
TU Wien
Robert Kroplewski
Rechtsanwalt und Berater der polnischen Regierung
Elisabeth Ling
RELX
Pierre Lucas
Orgalim – Europe's Technology Industries
Ieva Martinkenaite
Telenor
Thomas Metzinger
JGU Mainz und European University Association
Catelijne Muller
ALLAI Netherlands und EWSA
Markus Noga
SAP
Barry O'Sullivan, Stellvertretender Vorsitzender der HEG-KI
University College Cork
Ursula Pachl
BEUC
Nicolas Petit
Universität Lüttich
Christoph Peylo
Bosch

Iris Plöger
BDI
Stefano Quintarelli
Garden Ventures
Andrea Renda
College of Europe Faculty und CEPS
Francesca Rossi*
IBM
Cristina San José
Europäischer Bankenverband
George Sharkov
Digital SME Alliance
Philipp Slusallek
Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
Françoise Soulié Fogelman
KI-Beraterin
Saskia Steinacker
Bayer
Jaan Tallinn
Ambient Sound Investment
Thierry Tingaud
STMicroelectronics
Jakob Uszkoreit
Google
Aimee Van Wynsberghe
TU Delft
Thiébaut Weber
EGB
Cecile Wendling
AXA
Karen Yeung
Universität Birmingham

*Francesca Rossi war Berichterstatterin für dieses Dokument.