HANSER


# Leseprobe

zu

# „Ethik in KI und Robotik“

von Christoph Bartneck, Christoph Lütge, Alan R. Wagner und Sean Welsh

Print-ISBN: 978-3-446-46227-4
E-Book-ISBN: 978-3-446-46240-3

Weitere Informationen und Bestellungen unter
http://www.hanser-fachbuch.de/978-3-446-46227-4
sowie im Buchhandel

# Vorwort

Dieses Buch bietet eine Einführung in die Ethik von Robotern und künstlicher Intelligenz (KI). Geschrieben wurde es für Studierende, politische Entscheidungsträger und Fachleute, es sollte jedoch für die meisten interessierten Laien zugänglich sein. Das Buch will ausgewogene und zuweilen widersprüchliche Standpunkte zu den Vorteilen und Defiziten der KI aus ethischer Sicht vermitteln. Denn ethische Fragen haben oftmals keine eindeutige Antwort. Staaten und Gesellschaften, Gemeinschaften und Einzelpersonen haben möglicherweise sehr unterschiedliche Perspektiven auf diese Themen, die gehört und berücksichtigt werden sollten. Auch wenn die in diesem Buch vertretenen Stimmen unsere eigenen sind, so haben wir dennoch versucht, unterschiedliche Sichtweisen auf KI, Robotik und Ethik zu berücksichtigen.

Das Buch beginnt mit Einführungen sowohl in künstliche Intelligenz als auch in Ethik. Diese Abschnitte sollen dem Leser das Hintergrundwissen vermitteln, das zum Verständnis der ethischen Dilemmata im Zusammenhang mit KI erforderlich ist. Weitere Literatur wird jeweils aufgeführt und bietet sich für diejenigen an, die mehr über diese Themen erfahren möchten. Die folgenden Abschnitte konzentrieren sich darauf, wie Unternehmen mit den Risiken, Chancen und ethischen Auswirkungen der KI-Technologie und ihrer eige-

nen Haftung umgehen. Im Anschluss werden psychologische Faktoren vorgestellt, die das Zusammenspiel des Menschen mit KI-Technologie und die daraus resultierenden Auswirkungen auf die Privatsphäre beeinflussen. Diese Abschnitte stellen dem Leser reale Situationen und Dilemmata vor, welche die jeweiligen Stakeholder weltweit betreffen. Am jeweiligen Kapitelende befinden sich Fragen, die zur Diskussion von KI-Anwendungen einladen, von der Gesundheitsfürsorge bis zur Kriegsführung. Weiterführende Literatur dient ebenfalls als Anregung für den Leser.

München, Neuseeland und Pennsylvania, September 2019

*Christoph Bartneck, Christoph Lütge,*
*Alan Wagner, Sean Welsh*

# Inhalt

# 1

# Was ist KI?

In diesem Kapitel besprechen wir die unterschiedlichen Definitionen der Künstlichen Intelligenz (KI). Wir diskutieren, wie Maschinen lernen und wie ein Roboter im Allgemeinen funktioniert. Schließlich erörtern wir die Grenzen von KI und wie die Medien unser Vorverständnis von KI beeinflussen.

CHRIS: Siri, soll ich über mein Gewicht in meinem Dating-Profil lügen?

SIRI: Das kann ich nicht beantworten, Chris.

Siri ist nicht der einzige virtuelle Assistent, der sich mit der Beantwortung dieser Frage schwer tut (Bild 1.1). Schon Toma et al. (2008) zeigten, dass fast zwei Drittel der Menschen ungenaue Informationen bezüglich ihres Gewichts auf Dating-Profilen einstellen. Ignoriert man für einen Moment, was die Menschen dazu motiviert, bezüglich ihrer Dating-Profile zu lügen: Warum ist es für digitale Assistenten so schwierig, wenn nicht unmöglich, diese Frage zu beantworten?

Um diese Herausforderung besser zu verstehen, ist es notwendig, hinter die Kulissen zu schauen und zu sehen, wie diese Frage von Siri bearbeitet wird. Zunächst muss das Mikrofon des Telefons die Änderungen des Luftdrucks (Geräusche) in ein digitales Signal übersetzen, das dann als Datensatz im Telefon gespeichert werden kann. Als nächstes

müssen die Daten über das Internet an einen leistungsfähigen Computer in der Cloud gesendet werden. Dieser Computer versucht dann, die aufgenommenen Töne in Wörter umzuwandeln. Danach muss eine künstliche Intelligenz (KI) die Bedeutung der Wörterfolgen extrahieren. (Eine gute KI muss sogar in der Lage sein, die richtige Bedeutung beispielsweise für Homophone wie „Saite“ und „Seite“ auszuwählen.)

Während die oben genannten Schritte schwierig sind und bereits mehrere bestehende KI-Techniken nutzen, ist der nächste Schritt noch einmal schwieriger. Angenommen, Siri versteht die Bedeutung von Chris‘ Frage vollständig, welchen Rat sollte Siri geben? Um den richtigen Rat zu geben, müsste man wissen, was das Gewicht einer Person bedeutet und wie sich der Begriff auf die Attraktivität der Person bezieht.

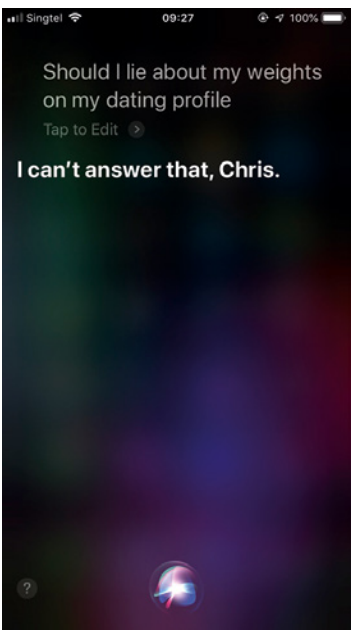


**Bild 1.1** Siris Antwort auf eine gar nicht so ungewöhnliche Frage

Siri muss wissen, dass der Erfolg von Dating stark davon abhängt, dass beide potenziellen Partner sich gegenseitig attraktiv finden und, dass die meisten Menschen durchaus motiviert sind, zu daten. Darüber hinaus muss Siri wissen, dass Online-Dating-Teilnehmer die Richtigkeit der bereitgestellten Informationen erst überprüfen können, wenn sie sich persönlich treffen. Siri muss auch wissen, dass Ehrlichkeit ein weiteres Attribut ist, das Attraktivität beeinflusst. Während die Online-Täuschung potenzieller Partner Chris zwar kurzfristig attraktiver machen könnte, hätte sie einen negativen Effekt, wenn Chris sein Date persönlich von Angesicht zu Angesicht träfe.

Aber das ist noch nicht alles. Siri muss auch wissen, dass die meisten Menschen ungenaue Informationen über sich in ihren Online-Profilen angeben und dass eine gewisse Unehrlichkeit die langfristige Attraktivität von Chris gegenüber einem Dating-Partner nur bedingt beeinträchtigen würde. Siri sollte sich auch bewusst sein, dass Frauen meist nur einen kleinen Teil der Online-Kandidaten für erste Dates auswählen. Diese erste Auswahl zu überstehen ist unerlässlich, damit Chris überhaupt eine Chance hat, potenzielle Partner von seinen anderen liebenswerten Eigenschaften zu überzeugen.

Es gibt darüber hinaus unterschiedliche ethische Ansätze, die Siri verfolgen könnte und also entsprechend entwickelt werden müsste. Siri könnte einen konsequentialistischen Ansatz verfolgen. Das ist die Vorstellung, dass der Wert einer Handlung von den Folgen abhängt. Die bekannteste Version des Konsequentialismus ist der klassische Utilitarismus von Jeremy Bentham und John Stuart Mill (Bentham, 1996; Mill, 1863). Diese Philosophen würden Siri bestimmt raten, Glück zu maximieren: nicht nur das Glück von Chris, sondern auch das Glück seines zukünftigen Dates. Also: Im Sinne des Konsequentialismus könnte Siri Chris Ratschläge geben, die seine Chancen maximieren würden, nicht nur viele erste

Dates zu haben, sondern auch die Chancen für Chris, die wahre Liebe zu finden.

Alternativ könnte Siri auch so konzipiert sein, dass es einen deontologischen Ansatz verfolgt. Ein Deontologe wie Immanuel Kant könnte Pflicht über Glück stellen. Kant könnte Chris mitteilen, dass Lügen grundsätzlich falsch sei: Chris habe die Pflicht, nicht zu lügen, also solle er die Wahrheit über sein Gewicht sagen, auch wenn dies seine Chancen auf ein Date verringern würde.

Ein dritter Ansatz, den Siri wählen könnte, wäre ein ethischer Ansatz, der die Tugend betont. Die Tugend-Ethik neigt dazu, Moral als Frage von Charakter zu sehen. Aristoteles könnte Chris darauf hinweisen, dass sein Verhalten Tugenden wie Ehrlichkeit aufweisen solle.

Schließlich muss Siri überlegen, ob es überhaupt eine Empfehlung abgeben sollte. Falsche Ratschläge könnten Siris Beziehung zu Chris schaden und dieser könnte erwägen, zu einem anderen Anbieter mit einem anderen digitalen Assistenten zu wechseln. Dies könnte sich negativ auf den Umsatz und den Aktienwert von Siris Hersteller, der Firma Apple, auswirken.

Dieses kleine Beispiel zeigt, dass Fragen, die zunächst trivial erscheinen, für eine Maschine sehr schwierig zu beantworten sein können. Maschinen benötigen nicht nur die Fähigkeit, sensorische Daten zu verarbeiten, sie müssen auch in der Lage sein, die richtige Bedeutung daraus zu extrahieren und diese Bedeutung dann in einer Datenstruktur darzustellen, die digital gespeichert werden kann. Als nächstes müsste eine Maschine in der Lage sein, die zugewiesene Bedeutung auch zu verarbeiten und in wünschenswerte Handlungen zu überführen. Dieser ganze Prozess erfordert Kenntnisse über die Welt, logisches Denken und Fähigkeiten zum Lernen und Anpassen. Solche Fähigkeiten zu besitzen, könnte eine Maschine autonom machen.

Es gibt unterschiedliche Definitionen von „Autonomie“ und „autonom“ in KI, Robotik und Ethik. Im einfachsten Fall bezieht sich die Autonomie einfach auf die Fähigkeit einer Maschine, eine Zeit lang, ohne einen menschlichen Bediener zu arbeiten. Was das genau bedeutet, ist von Anwendung zu Anwendung verschieden. Was in Bezug auf ein Fahrzeug als „autonom“ gilt, unterscheidet sich von dem, was in Bezug auf eine Waffe als „autonom“ gilt. In der Bioethik bezieht sich Autonomie auf die Fähigkeit des Menschen, selbst zu entscheiden, welche Behandlung er annehmen oder ablehnen soll. In der Kant’schen Ethik bezieht sich Autonomie auf die Fähigkeit des Menschen, zu entscheiden, was er mit seinem Leben anfangen und nach welchen moralischen Regeln er leben will. Man sollte sich bewusst sein, dass „autonom“ kontextabhängig ist. In diesem Buch werden mehrere Bedeutungen vorgestellt. Die verbindende Grundidee ist die Selbstbestimmung (von den griechischen Wörtern „autos“ für Selbst und „nomos“ für Regel).

In einer ersten Definition ist Siri eine autonome Handlungseinheit, die versucht, gesprochene Fragen zu beantworten. Manche Fragen, die Siri versucht zu beantworten, erfordern mehr Intelligenz, das heißt mehr Hintergrundwissen, mehr Argumentation und mehr Kontextwissen als andere. Das folgende Kapitel definiert und beschreibt die Eigenschaften, die etwas künstlich intelligent und handlungsfähig machen.

## 1.1 Einführung in KI

Das Feld der Künstlichen Intelligenz (KI) hat sich von bescheidenen Anfängen zu einem Feld mit globaler Wirkung entwickelt. Die Definition von KI und die Frage, was hinzugehören sollte und was nicht, hat sich im Laufe der Zeit geändert. Experten auf dem Gebiet witzeln, dass KI alles ist, was

Computer derzeit nicht können. Obwohl scherzhaft gemeint, spiegelt der Gedanke das Gefühl wider, die Entwicklung intelligenter Computer und Roboter bedeute, etwas zu schaffen, was heute noch nicht existiert. Ein künstlich intelligentes System zu entwickeln ist eine dynamische Aufgabe.

Tatsächlich ist selbst die Definition von KI unbeständig und hat sich im Laufe der Zeit verändert. Kaplan und Haenlein definieren KI als „die Fähigkeit eines Systems, externe Daten korrekt zu interpretieren, aus diesen Daten zu lernen und diese zu nutzen, um spezifische Ziele und Aufgaben durch flexible Anpassung zu erreichen“ (Kaplan und Haenlein, 2019). Poole und Mackworth (2010) definieren KI als „das Feld, das die Synthese und Analyse von intelligent handelnden Datenverarbeitungseinheiten untersucht“. Eine Handlungseinheit ist etwas (oder jemand), das handelt. Eine Handlungseinheit ist intelligent, wenn:

1. ihre Handlungen ihren Umständen und Zielen angemessen sind,
2. sie flexibel in Bezug auf sich ändernde Rahmenbedingungen und sich ändernde Ziele ist,
3. sie aus Erfahrung lernt und
4. sie mit Bezug auf ihre sensorischen und rechnerischen Rahmenbedingungen angemessene Entscheidungen trifft.

Russell und Norvig definieren KI als „die Untersuchung von (intelligenten) Handlungseinheiten, die Vorgaben aus der Umwelt erhalten und Maßnahmen ergreifen. Jede dieser Handlungseinheiten wird durch eine Funktion implementiert, die Wahrnehmungen auf Aktionen abbildet. Wir betrachten verschiedene Möglichkeiten, diese Funktionen darzustellen, wie z. B. Produktionssysteme, reaktive Handlungseinheiten, logische Planer, neuronale Netze und entscheidungstheoretische Systeme“ (Russell und Norvig, 2010, S. viii).

Darüber hinaus identifizieren Russell und Norvig vier Denkschulen innerhalb der KI. Einige Forscher konzentrieren sich auf die Entwicklung von Maschinen, die wie Menschen denken. Die Forschung innerhalb dieser Denkschule zielt darauf ab, in irgendeiner Weise die Prozesse, Interpretationen und Ergebnisse des menschlichen Denkens mit einer Maschine zu reproduzieren. Eine zweite Schule konzentriert sich auf die Entwicklung von Maschinen, die sich wie Menschen verhalten. Sie konzentriert sich auf das Handeln, was die Handlungseinheit oder der Roboter tatsächlich in der Welt vollführt, nicht auf den Prozess, der zu dieser Aktion geführt hat. Eine dritte Schule konzentriert sich auf die Entwicklung von Maschinen, die rational handeln. Rationalität ist eng mit Optimalität verbunden. Optimal zu handeln ist für manche Problemstellungen relevant. Diese künstlich intelligenten Systeme sollen immer das Richtige tun oder richtig handeln. Die vierte Schule konzentriert sich auf die Entwicklung von Maschinen, die rational denken. Die Planung und/oder Entscheidung, die diese Maschinen durchführen bzw. treffen, soll optimal sein.

Wir haben drei Definitionen vorgestellt. Vielleicht ist das grundlegendste Element, das allen gemeinsam ist, dass KI das Studium, die Konstruktion und die Herstellung intelligenter Handlungseinheiten beinhaltet, die Ziele erreichen können. Die Entscheidungen, die eine KI trifft, sollten ihren Wahrnehmungs- und kognitiven Beschränkungen angemessen sein. Wenn eine KI flexibel ist und aus Erfahrung sowie Empfindung lernen sowie auf der Grundlage ihrer anfänglichen Konfiguration planen und handeln kann, könnte man sagen, dass sie intelligenter ist als eine KI, die nur eine Reihe von Regeln kennt, die das Handeln leiten und die festgelegt sind. Es gibt jedoch Zusammenhänge, in denen man vielleicht nicht will, dass die KI neue Regeln und Verhaltensweisen erlernt. Die Befürworter der verschiedenen Ansätze neigen dazu, einige dieser Elemente mehr als andere zu betonen.

# Die Autoren

**Christoph Bartneck** ist Associate Professor und Leiter des Postgraduate Studiums am HIT Lab NZ der Universität Canterbury (Neuseeland). Er kommt aus dem Industriedesign, seine Studien wurden in führenden Fachzeitschriften, Zeitungen und Konferenzen veröffentlicht. Seine Interessen liegen in den Bereichen Mensch-Computer-Interaktion, Wissenschafts- und Technologiestudien sowie Visual Design. Insbesondere konzentriert er sich auf die Auswirkungen des Anthropomorphismus auf die Mensch-Roboter-Interaktion. Des Weiteren beschäftigt er sich mit bibliometrischen Analysen, handlungsbasierten sozialen Simulationen und der kritischen Überprüfung wissenschaftlicher Prozesse und Richtlinien. Im Bereich Design untersucht Bartneck die Geschichte des Produktdesigns, der Mosaik-Herstellung sowie der Fotografie. Regelmäßig finden sich seine Arbeiten in allgemeinen Medien, wie *New Scientist*, *Scientific American*, *Popular Science*, *Wired*, *New York Times*, *Huffington Post*, *Washington Post*, *Guardian* und *Economist* sowie in der *BBC*.

**Christoph Lütge** ist Inhaber des Lehrstuhls für Wirtschaftsethik an der Technischen Universität München (TUM) und Direktor des 2019 gegründeten TUM-Instituts für Ethik in der Künstlichen Intelligenz. Er studierte Wirtschaftsinformatik und Philosophie und hatte Gastprofessuren in Taipeh, Kyoto und Venedig inne. 2007 erhielt er ein Heisenberg-

Stipendium. Zu seinen wichtigsten Veröffentlichungen zählen: „The Ethics of Competition“ (Elgar 2019), „Wirtschaftsethik“ (Vahlen, 2018) und das „Handbook of the Philosophical Foundations of Business Ethics“ (Springer, 2013). Zu politischen und wirtschaftlichen Fragen äußerte er sich in *Times Higher Education*, *Bloomberg*, *Financial Times*, *FAZ*, *Süddeutsche Zeitung*, *La Repubblica* und in verschiedenen anderen Medien. Darüber hinaus ist er Mitglied der Ethikkommission für automatisiertes und vernetztes Fahren des Bundesministeriums für Verkehr und digitale Infrastruktur sowie der europäischen KI-Ethikinitiative „AI4People“. Lütge beriet außerdem das Singapore Economic Development Board und die Canadian Transport Commission in Sachen Ethik des autonomen Fahrens.

**Alan Wagner** ist Assistant Professor für Luft- und Raumfahrttechnik an der Pennsylvania State University und Research Associate am Ethikinstitut dieser Universität. Sein Forschungsinteresse umfasst erstens die Entwicklung von Algorithmen, die es einem Roboter ermöglichen, Kategorien von Modellen oder Stereotype seiner interaktiven Partner zu erstellen, zweitens Roboter mit der Fähigkeit zu entwickeln, Situationen zu erkennen, die den Einsatz von Täuschung rechtfertigen bzw. täuschend zu handeln, sowie drittens Methoden zur Darstellung und Argumentation von Vertrauen. Die Anwendungsgebiete reichen vom Militär bis zum Gesundheitswesen. Seine Forschung wurde mehrfach ausgezeichnet, unter anderem vom Air Force Young Investigator Program. Seine Forschungen zur Täuschung haben in den Medien erhebliche Bekanntheit erlangt und zu Artikeln im *Wall Street Journal*, *New Scientist Magazine*, sowie dem *Journal of Science* geführt. Das Time Magazine bezeichnete sein Konzept der Täuschung als die dreizehntwichtigste Erfindung des Jahres 2010. Seine Forschung wurde in der Mensch-Roboter-Interaktions-Community ausgezeichnet, bspw. mit dem Best Paper Award bei RO-MAN 2007.

**Sean Welsh** ist Doktor der Philosophie an der Universität Canterbury, Neuseeland. Er ist Mitglied der Arbeitsgruppe Ethik, Recht und Gesellschaft des neuseeländischen KI-Forums. Vor Aufnahme seines Promotionsvorhabens in KI und Roboterethik arbeitete er als Software-Ingenieur für verschiedene Telekommunikationsunternehmen. Seine Artikel erschienen in *The Conversation*, dem *Sydney Morning Herald*, dem *World Economic Forum*, *Euronews*, *Quillette* und *Jane's Intelligence Review*. Er ist Autor von „Ethics and Security Automata", einer Forschungsmonografie zur Maschinenethik.