

AISB2016

Symposium über die Prinzipien der Robotik

4. April 2016, Sheffield, Großbritannien.

Bearbeitet von

TonyJPrescott

Organisationskomitee

TonyJPrescott

AlanWinfield

MadeleinedeCockBuning

JoannaJBryson

Noel Sharkey

**Teil der Tagung 2016 der Society for the Study of Artificial Intelligence and
Simulation of Behaviour (AISB)**

Über das Symposium

Es ist fünf Jahre her, dass „Principles of Robotics“¹ veröffentlicht wurde, das von einer Gruppe angesehener britischer Robotik- und KI-Experten auf einem vom EPSRC/AHRC finanzierten Retreat entwickelt wurde. Die Prinzipien, die auf die „Regulierung von Robotern in der realen Welt“ abzielten, wurden in Form von fünf „Regeln“ und sieben „High-Level-Botschaften“ formuliert. Die fünf Regeln lauten wie folgt:

1. Roboter sind Mehrzweckwerkzeuge. Roboter sollten nicht ausschließlich oder in erster Linie dafür konzipiert sein, Menschen zu töten oder zu schädigen, außer im Interesse der nationalen Sicherheit.
2. Menschen, keine Roboter, sind verantwortliche Akteure. Roboter sollten so konzipiert und betrieben werden, dass sie soweit wie möglich die geltenden Gesetze, Grundrechte und Freiheiten, einschließlich der Privatsphäre, einhalten.
3. Roboter sind Produkte. Sie sollten mithilfe von Prozessen entwickelt werden, die ihre Sicherheit gewährleisten.
4. Roboter sind hergestellte Artefakte. Sie sollten nicht in einem Land entworfen werden Eine betrügerische Möglichkeit, gefährdete Benutzer auszunutzen; stattdessen sollte ihre Maschinennatur transparent sein.
5. Die Person mit der rechtlichen Verantwortung für den Roboter sollte zugewiesen werden.

Die Grundsätze hatten erhebliche Auswirkungen auf die britische Robotikforschung und provozieren weiterhin erhebliche Debatten. In einer Zeit, in der die öffentliche Besorgnis über die Entwicklung von Robotertechnologien zunimmt, halten wir es für sinnvoll, die Grundsätze noch einmal zu prüfen, um ihre fortdauernde Relevanz anhand der folgenden Kriterien zu prüfen: a.

Gültigkeit: Sind die Prinzipien korrekt, Aussagen über die Natur von Robotern (zum Beispiel, dass es sich um Werkzeuge und Produkte handelt), Roboterentwickler und die Beziehung zwischen Robotern und Menschen (zum Beispiel, dass Roboter ein transparentes Design haben sollten), oder sind sie tologisch fehlerhaft, ungenau, veraltet oder irreführend?

B. Ausreichend/ Allgemeinheit – sind die grundsätzlich unzureichenden und weitreichenden Aspekte, die alle wichtigen Fragen abdecken, die bei der Regulierung der Robotik in der realen Welt auftreten könnten, oder werden wesentliche

Bedenken übersehen. *c. Nutzen* – sind die Grundsätze der praktischen Nutzung für Roboterentwickler, Benutzer Hersteller, bei der Festlegung von Strategien für Best Practices in der Robotik, oder von rechtlichen Standards oder Rahmenwerken, oder sie sind in ihrer Verwendung durch mangelnde Spezifität oder durch die Zulassung kritischer Ausnahmen (z. B. die Verwendung von Robotern als Waffen für Zwecke der nationalen Sicherheit) eingeschränkt.

Am 4. April fand im Rahmen der AISB-Konferenz 2016 in Sheffield, Großbritannien, ein eintägiges Symposium statt. Dieses Protokoll enthält Kommentare zu den Grundsätzen, die nicht im Vorfeld der Sitzung angefordert wurden.

¹ <https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

AISB2016 Symposium über die Prinzipien der Robotik

Eingereichte Kommentare

- 1 Joanna Bryson
Die Bedeutung der EPSRC-Prinzipien der Robotik
- 2 Amadou Gning, Darryl Davis, Yongqiang Cheng und Peter Robinson
Robotikforschung, Ethikdiskussion
- 3 Vincent Müller
Gesetzliche Regeln, ethische Anforderungen, Zukunftsthemen
- 4 Tony Prescott
Roboter sind nicht nur Werkzeuge
- 5 Michael Szollosy
Einen überholten Human(ismus) verteidigen?
- 6 Aurora Voiculescu
Regulierung von Roboterstädten: Überlegungen zu den Prinzipien der Robotik aus der newfar-sideofthelaw
- 7 Paula Boddington
Kommentar zu Verantwortung, Produktdesign und Sicherheitsvorstellungen
- 8 RoelanddeBruinandMadeleinedeCockBuning
Kommentar zum Prinzip2
- 9 Burkhard Shafer undLilianEdwards
Fairdatahandlingandrobotics
- 10 Amanda Sharkey
Können Roboter verantwortungsbewusste moralische Agenten sein? Und warum sollte uns das interessieren?
- 11 Tom Sorell und HeatherDraper
Zweite Gedanken zu Privatsphäre, Sicherheit und Täuschung
- 12 Emily Collins
Kommentar zum Prinzip4
- 13 Andreas Theodorou, Robert Wortham und Joanna Bryson
Warum verhält sich mein Roboter so? Transparenz für Echtzeitinspektion entwerfen autonomer Roboter
- 14 Robert Wortham, Andreas Theodorou und Joanna Bryson
Robotertransparenz, Vertrauen und Nutzen

Die Bedeutung der EPSRC-Prinzipien der Robotik

Joanna Bryson, University of Bath und Princeton Center for Information Technology Policy

Einführung

Bei der Neubetrachtung der Prinzipien der Robotik ist es wichtig, ihre volle Bedeutung sorgfältig zu berücksichtigen. Hier gehe ich zunächst kurz auf die Bedeutung des Dokuments als Ganzes und dann auf seine Bestandteile ein. Die EPSRC-Grundsätze der Robotik wurden als Ergebnis von einer Gruppe erstellt, die mit wenig Anleitung zusammengestellt wurde und für die kein Ergebnis erforderlich war. Die ursprüngliche Absicht der EPSRC-Robotikveranstaltung scheint nur die Diskussion selbst oder vielleicht sogar nur die Tatsache des Treffens gewesen zu sein. Die anwesenden Akademiker wollten etwas für ihre aufgewendete Zeit vorweisen, und so wurde am letzten Tag ein erheblicher Teil der Zeit aller Anwesenden in die Erstellung [der drei Versionen der Grundsätze und deren Dokumentation gesteckt](#). Ein Teil der Dokumentation wurde nach der Sitzung – erneut im

Konsens – verlängert. Es ist richtig und passend, dass es eine Möglichkeit gibt, das Dokument zu prüfen und sogar zu aktualisieren oder zu pflegen. Sogar nationale Verfassungen verfügen über Mittel zur Aufrechterhaltung. Für die Wirksamkeit von Richtliniendokumenten ist es jedoch von entscheidender Bedeutung, dass sie nicht leicht zu ändern sind. Sie sollten ein Ruder darstellen, um ein Zittern zu verhindern, und sind daher normalerweise schwieriger zu ändern, als sie überhaupt zu instanziierten.

Beachten Sie, dass es einigen Ländern und anderen politischen Gewerkschaften aus genau diesem Grund nicht leicht fiel, auch nur ihre ursprünglichen Verfassungen zu schaffen. Daher ist es wichtig, sorgfältig über die Be-

Die Prinzipien als Politik

Technologiepolitik und Politik im Allgemeinen sind eine überraschend amorphe Sache. Wie andere Aspekte der natürlichen Intelligenz ist auch die Politik nicht immer im Gesetz oder gar in der Regierungsführung verankert. Ein Großteil der Politik ist ungeschrieben und nicht einmal explizit bekannt. Das Vereinigte Königreich ist tatsächlich herausragend in der Innovation des Gewohnheitsrechts, das dies und die Bedeutung von Kultur und Präzedenzfällen anerkennt. Dennoch müssen wir uns im kalten Licht eines Ausschusses, der sich mit REF-Auswirkungsfällen befasst, fragen:

Sind die Grundsätze Politik? Ich denke, die Antwort lautet „Ja“. Dabei handelt es sich um eine Reihe von Richtlinien, die von einem erheblichen, wenn nicht sogar willkürlichen Teil der von ihnen betroffenen Gemeinschaft vereinbart wurden, und sie werden auf Regierungswebseiten veröffentlicht. Jede Politik besteht aus drei Komponenten: allokativ, distributiv und stabilisierend. Beim Allokativ wird ermittelt, für welche Probleme es sich lohnt, Zeit und andere Ressourcen aufzuwenden. Im Fall der Grundsätze wurde dies vom EPSRC (oder einer Organisation über ihm) initiiert, aus Sorge, dass die britische Öffentlichkeit die Robotik ablehnen könnte, da sie über gentechnisch veränderte Lebensmittel verfügte. Uns wurde gesagt, dass die Ablehnung der Robotik als ernsthafte Bedrohung für die britische Wirtschaft angesehen werde. Beachten Sie auch, dass jeder der Teilnehmer (zumindest diejenigen, die nicht speziell für die Teilnahme bezahlt wurden) auch individuelle Investitionen tätigte und Zeit für

das Problem der Roboterethik aufwendete, obwohl dies für viele mit der Gelegenheit verwechselt wurde, bei ihrer primären Förderorganisation besser bekannt zu werden. Die stabilisierende Komponente stellt sicher, dass die Po-

eine Haftung oder eine kontroverse Angelegenheit. Bei den Grundsätzen ist dies offensichtlich zumindest einigermaßen gelungen, seit wir ihr fünfjähriges Jubiläum feiern. Aus Gesprächen mit anderen Autoren weiß ich, dass keiner völlig vom Endprodukt begeistert ist, aber alle respektieren den (zugegebenermaßen repräsentativen) demokratischen Prozess, durch den sie zustande kamen, und die Bedeutung des gegenseitigen Engagements ihrer Kollegen für das Endprodukt. Ich für meinen Teil würde gerne sehen, dass die Prinzipien weiter in die Politik oder sogar in das Gesetz umgesetzt werden, aber ich muss noch den Prozess entdecken, mit dem dies erreicht werden könnte. Allerdings wurden und werden sie weiterhin von verschiedenen Normungsgremien und parlamentarischen Untersuchungen sowie von der Presse und anderen Wissenschaftlern zur Kenntnis genommen. Den umstrittensten Aspekt der Politik lasse ich zum Schluss: den Verteilungsaspekt. Grundsätzlich geht es bei jeder Politik um die

Auswahl von Maßnahmen, und das impliziert die Zuweisung bzw. Umverteilung von Ressourcen. Die Politik versucht, dies zu vertuschen, da es zwangsläufig gegen den Strich derer geht, von denen die Ressourcen umverteilt werden, selbst in den Fällen, in denen diese Personen einen Nettonutzen erzielen könnten. Wir hassen es, die Kontrolle zu verlieren, aber Richtlinien dienen der Kontrolle. „Versucht zu vertuschen“ ist in der Tat eine Untertreibung; die Umverteilung schmackhaft zu machen, könnte das Kernprojekt der Politiker sein. In diesem Fall hatte die Regierung ganz konkrete Bedenken hinsichtlich Personen, die in den Medien die Angst vor Robotern geschürt hatten, und brachte ihren Wunsch deutlich zum Ausdruck, Wege zu finden, die Aufmerksamkeit der Medien und den öffentlichen Eindruck auf die Sicherheit der Robotik zu lenken. Im Gegensatz dazu waren es tatsächlich die Teilnehmer, die die anderen großen Veränderungen vom Sensationsdrang zum Pragmatismus vorbrachten, nämlich die Behauptung, dass Roboter keine verantwortlichen Parteien im Sinne des Gesetzes seien und dass Benutzer nicht über ihre Fähigkeiten getäuscht werden sollten. Die Ratsvertreter wussten, dass diese Umverteilung der Macht einige ihrer herausragenden Förderempfänger verärgern würde, und die Teilnehmer wussten dasselbe über einige ihrer Kollegen. Dennoch herrschte unter den Wissenschaftlern auffallende Einigkeit darüber, dass das größte moralische Risiko von Robotern in ihrer charismatischen Natur und dem unglaublichen Eifer vieler Menschen liegt, ihre eigene Identität in Maschinen zu investieren, was zu einer auffallenden Verwirrung über ihre Natur führte, die wir alle erlebt hatten. Dieses Charisma und diese Verwirrung öffneten die Tür für alle Arten von Manipulationen durch Unternehmen und Regierungen, bei denen die Roboter als Verantwortliche oder sogar als Ersatz für menschliches Leben oder Werte eingesetzt werden konnten.

Das Prinzip des Tötens

Roboter sind Mehrzweckwerkzeuge. Roboter sollten nicht ausschließlich oder in erster Linie dazu konzipiert sein, Menschen zu töten oder ihnen Schaden zuzufügen, außer im Interesse der nationalen Sicherheit.

Die ersten drei Prinzipien waren als Korrekturen der Asimov-Gesetze gedacht. Roboter sind keine verantwortlichen Parteien, also könnten sie nicht töten. Stattdessen sollten Roboter nicht als Tötungswerkzeuge eingesetzt werden können. Diese einfache Regel machte die Übertragung moralischer Subjektivität deutlich und kam gleichzeitig den pazifistischen Wünschen der meisten Anwesenden entgegen. Allerdings wurden Roboter pragmatisch gesehen bereits als Kriegswaffen eingesetzt, und ein Gesetz, das nicht durchsetzbar ist, ist fraglich. Wir waren davon überzeugt, dass die Führung mit einem Prinzip, von dem bekannt ist, dass es falsch ist, unsere Chancen auf kulturelle Wirkung erheblich verringern würde. Die Bedeutung des ersten Prinzips scheint daher durch den Kompromiss der Ausnahme neutralisiert zu werden, aber dass Roboter in der Zivilgesellschaft keine Waffen sein dürfen, ist immer noch ein wichtiger gesellschaftlicher Punkt. Darüber hinaus ist auch die Tatsache von Bedeutung, dass die praktische Politik die Bedürfnisse der Regierung berücksichtigen muss, um sowohl Sicherheit als auch Industrie zu gewährleisten (Großbritannien ist das fünftgrößte Waffenhandelsland der Welt). Wie rein akademisch manche von uns auch sein mögen

Obwohl wir unsere Disziplin so wünschen, bedeutet die Tatsache, dass viele ihrer Produkte einen unmittelbaren Nutzen haben, dass wir Auswirkungen auf unsere Welt nicht vermeiden können.

Das Prinzip der Compliance

Menschen, nicht Roboter, sind verantwortliche Akteure. Roboter sollten so konzipiert und betrieben werden, dass sie den geltenden Gesetzen sowie den Grundrechten und -freiheiten, einschließlich der Privatsphäre, entsprechen.

Beim zweiten Asimov-Gesetz geht es um die Befolgung von Anweisungen, aber auch um die Vorstellung, gehorchen zu müssen, ist moralische Entscheidungsfreiheit impliziert. Die ursprüngliche Bedeutung dieses Gesetzes bestand darin, dass Roboter gewöhnliche Technologie sind und gewöhnlichen Standards und Gesetzen entsprechen. Bei der Gestaltung der Prinzipien als Suite wurde das zweite Prinzip zu demjenigen, das die Gefahren der KI im Allgemeinen und der Verwechslung von KI mit einem moralischen Subjekt im Besonderen weiter verdeutlichte. Die Betonung der Privatsphäre spiegelt die besondere Sorge eines wahrnehmenden intelligenten physischen Agenten wider, der genau den gleichen Raum einnimmt wie eine menschliche Familie. Die Technologie ist grundlegend in die menschliche Umwelt eingetaucht, mehr als jede frühere Technologie oder jedes Haustier, vielleicht sogar mehr als einige Menschen in einem Haushalt, wie zum Beispiel Kinder. Es hat Zugriff auf geschriebene und gesprochene Sprache, soziale Informationen, eingehaltene Zeitpläne usw. Darüber hinaus kann es für ein Haustier oder ein anderes vertrauenswürdigen Familienmitglied gehalten werden, seine besonderen Fähigkeiten zur perfekten Kommunikation mit der Außenwelt vorübergehend vergessen werden oder seine Fähigkeit, Regelmäßigkeiten zu lernen und Reize klassifizieren. In diesen Fällen können Primateninformationen unbeabsichtigt in einer öffentlichen oder sogar einer vermeintlich privaten Cloud gespeichert werden, die anfällig für Hackerangriffe ist. Es ist keine triviale Aufgabe, eine solch neuartige, menschenähnliche Technologie in die Einhaltung standardmäßiger gesetzlicher Datenschutz- und Sicherheitsnormen zu zwingen.

Das Prinzip der Kommerzialisierung

Roboter sind Produkte. Sie sollten mithilfe von Prozessen entworfen werden, die ihre Sicherheit gewährleisten.

Das letzte Asimov-Gesetz ist Selbstschutz, aber Roboter haben kein Selbst. Stattdessen konzentrierte sich dieses Gesetz auf den Schutz des Menschen vor Robotern auf der Ebene der grundlegenden Gesundheit des Roboters. Das Prinzip macht uns erneut bewusst, dass der Roboter nicht speziell hergestellt ist, und versucht, die Umgehung einer rechtlichen Haftung durch die Behauptung zu verhindern, dass Roboter eine einzigartige Natur hätten. Der Hersteller eines Roboters sollte genauso viel Verantwortung dafür tragen, dass die Maschine den Spezifikationen entspricht wie der Hersteller eines Autos oder eines Elektrowerkzeugs. Tatsächlich könnten Roboter Autos oder Elektrowerkzeuge sein, aber wenn das der Fall ist, sollten sie eher sicherer als weniger sicher sein als die herkömmliche Variante von beidem.

Das Prinzip der Transparenz

Roboter sind hergestellte Artefakte. Sie sollten nicht in einer irreführenden Weise konzipiert sein, um gefährdete Benutzer auszunutzen; stattdessen sollte ihre Maschinennatur transparent sein.

Die ersten drei Grundsätze legten den rechtlichen Rahmen für die Herstellung und den Vertrieb von Robotik fest, die mit anderen Produkten identisch sind. Die letzten beiden sollen sicherstellen, dass der Status auch dem Benutzer mitgeteilt wird. Der Grundsatz der Transparenz soll sicherstellen, dass Einzelpersonen nicht zu viel in ihre Technologie investieren, indem sie beispielsweise einen Haussitter engagieren, um zu verhindern, dass der Roboter einsam wird. Einige Robotiker lehnen dieses Prinzip ab, weil Täuschung für die Wirksamkeit ihrer beabsichtigten Anwendung notwendig ist, beispielsweise um Menschen dazu zu bringen, sich nicht einsam zu fühlen, damit sie weniger depressiv sind. Andere behaupten, dass dieses Prinzip die Möglichkeit verneint, dass Roboter mehr sein sollten als gewöhnliche Maschinen. Das erste Argument ist experimentierfreudig. Zunächst muss festgestellt werden, dass es keine Möglichkeit gibt, emotionales Engagement ohne Täuschung auszulösen, was angesichts des Ausmaßes emotionalen Engagements, das mit fiktiven Charakteren und eindeutig nicht erkennbaren Objekten hergestellt wird, unwahrscheinlich erscheint. Wenn experimentell festgestellt wird, dass eine Täuschung erforderlich ist, kann über den Kompromiss zwischen Kosten und Nutzen der Täuschung debattiert werden. Das zweite ist jedoch unbestreitbar. Die Urheberschaft, die wir über Artefakte haben, ist ein grundlegender Teil ihrer maschinellen Natur. KI ist definitiv ein Artefakt. In gewisser Weise könnten wir sogar argumentieren, dass dieses Prinzip selbstlimitierend ist. Wenn KI wirklich in der Lage wäre, die Bedeutung einer Maschine zu verändern, dann würde die Kommunikation dieser veränderten Maschinennatur diesem Prinzip immer noch entsprechen.

Der Grundsatz der rechtlichen Verantwortung

Die rechtliche Verantwortung für einen Roboter sollte der Person zugeordnet werden.

Das fünfte Prinzip schließlich kommuniziert den Status der Roboter als Artefakte auf möglichst grundlegende Weise. Sie sind Eigentum, und dieses Eigentum muss rechtlich zugeschrieben werden. Die Tatsache, dass Roboter konstruiert und besessen werden, ist der Grund dafür, dass ich zuvor argumentiert habe, dass wir ethisch verpflichtet sind, sie nicht zu Personen zu machen, weil der Besitz von Personen zweifellos unethisch ist. Das Argument ist nicht, dass es personenähnliche Roboter gibt, deren Status wir rechtlich herabstufen sollten, sondern vielmehr, dass der notwendigerweise herabgestufte Rechtsstatus bedeutet, dass wir nicht dafür sorgen sollten, dass Personenähnlichkeit ein Merkmal eines legal hergestellten Roboters ist. Allerdings reichen die Prinzipien der Robotik nicht bis zu diesem Extrem des Futurismus. Wie ich bereits sagte, konzentrieren sie sich darauf, die gegenwärtige Realität einer Bevölkerung zu vermitteln, die so sehr darauf bedacht ist, das Übermenschliche zu besitzen und sich mit ihm zu identifizieren, dass sie leicht zu der Annahme verleitet werden könnte, dass ein Roboter, der schlecht hergestellt oder bedient wird, selbst für den Schaden verantwortlich ist, den er anrichtet. Wenn Sie einen fürchterlichen Lärm hören und ein in Ihr Haus eingefahrenes Auto vorfinden, können Sie den Besitzer des Autos schnell und einfach identifizieren, auch wenn das Auto gerade leer ist, einfach anhand des Nummernschilds oder im schlimmsten Fall anhand der Seriennummer. Die Idee ist, dass das Gleiche auch gelten sollte, wenn Sie einen in Ihrem Grundstück eingebauten Roboter finden. Die Teilnehmer des Robotik-Retreats haben ein Problem, das aufgrund von Drohnen in unserer Gesellschaft bereits besteht und das jetzt in einigen Ländern mit obligatorischen Lizenzen angegangen wird, wie das Komitee empfohlen hat, genau vorhergesagt.

Abschluss

Zusammenfassend lässt sich sagen, dass die EPSRC-Grundsätze wertvoll sind, weil sie eine Politik darstellen, die mit erheblichen Kosten für den Steuerzahler und für den Einzelnen aufgebaut wird. Zwar ist keine Richtlinie perfekt, aber im Idealfall sie

sollte nur durch eine neue Politik mit einem gleich hohen oder höheren Investitionsniveau sowohl von der Regierung als auch von Fachexperten ersetzt werden. Ihr Zweck besteht darin, das Vertrauen von Verbrauchern und Bürgern in die Robotik als vertrauenswürdige Technologie zu stärken, die geeignet ist, sich in unserer Gesellschaft durchzusetzen. Die einzelnen Grundsätze repräsentieren jeweils wesentliche Anliegen der Experten und Interessengruppen, auch wenn diese Darstellung manchmal selbst nicht vollkommen transparent ist. Das übergeordnete Ziel bestand darin, klar zu kommunizieren, dass die Verantwortung für die sichere und zuverlässige Herstellung und den Betrieb von Robotern nicht anders ist als für alle anderen im Vereinigten Königreich hergestellten und verkauften Objekte und dass daher die bestehenden Gesetze des Landes ausreichen sollten, um sowohl Verbraucher als auch Hersteller zu schützen. Es ist wichtig zu erkennen, dass dies nicht bei allen denkbaren Robotern der Fall ist. Es ist leicht, sich einzigartige Kunstwerke vorzustellen, die als Roboter gelten und keine Massenprodukte sind, oder sich Roboter vorzustellen, die einfach auf unsichere oder unverantwortliche Weise gebaut sind. Den Menschen fällt es schwerer, sich vorzustellen, dass es möglicherweise kognitive Eigenschaften wie Empfindungen gibt, die möglicherweise in einen Roboter integriert werden könnten, dies jedoch genauso unethisch wäre wie das Anbringen fehlerhafter Bremsen an einem Auto. Die Prinzipien der Robotik zielen nicht darauf ab, zu bestimmen, was möglich ist; sie zielen darauf ab, empfehlenswerte Praktiken für die Integration autonomer Robotik in das Landrecht zu vermitteln.

RobotikForschungEthikDiskussion

A. Gning, D. Davis, Y Cheng, P. Robinson, Informatikabteilung, Hull University.

e.gning@hull.ac.uk

Einführung

In der modernen Welt hat die Robotik mit der Entwicklung technologischer Ressourcen zu zahlreichen Anwendungen [1] und oft unvorhergesehenen Einsätzen im wirklichen Leben geführt (z. B. der zunehmende Einsatz von Drohnen in zivilen Anwendungen). Um mit diesem Zeitalter der Robotik Schritt zu halten, müssen die Forschungsgemeinschaft und die Gesellschaft im Allgemeinen ethische Grundsätze definieren, die allgemein genug sind, um im Laufe der Zeit robust zu sein und an die Bandbreite möglicher Anwendungen angepasst zu werden.

Ethische Grundsätze sollten so weit verbreitet und verallgemeinert werden, dass sich die Designer und Hersteller von Robotern der zu respektierenden Vorschriften und Grenzen bewusst sind.

Im Allgemeinen lassen sich die Anwendungen der Robotik in vier Gruppen einteilen: Haushalts- oder Hilfsroboter [2] [3] [4], medizinische Roboter [5] [6] [7] [8], Verteidigungsroboter [9] [10].] und Industrieroboter [11] [12]. Die Diskussion konzentriert sich auf die ersten drei Gruppen von Robotern, da Industrieroboter häufig auf begrenzte Bereiche mit vorab festgelegten Sätzen begrenzter Aufgaben beschränkt sind und stehen nicht in direkter Interaktion mit der Gesellschaft.

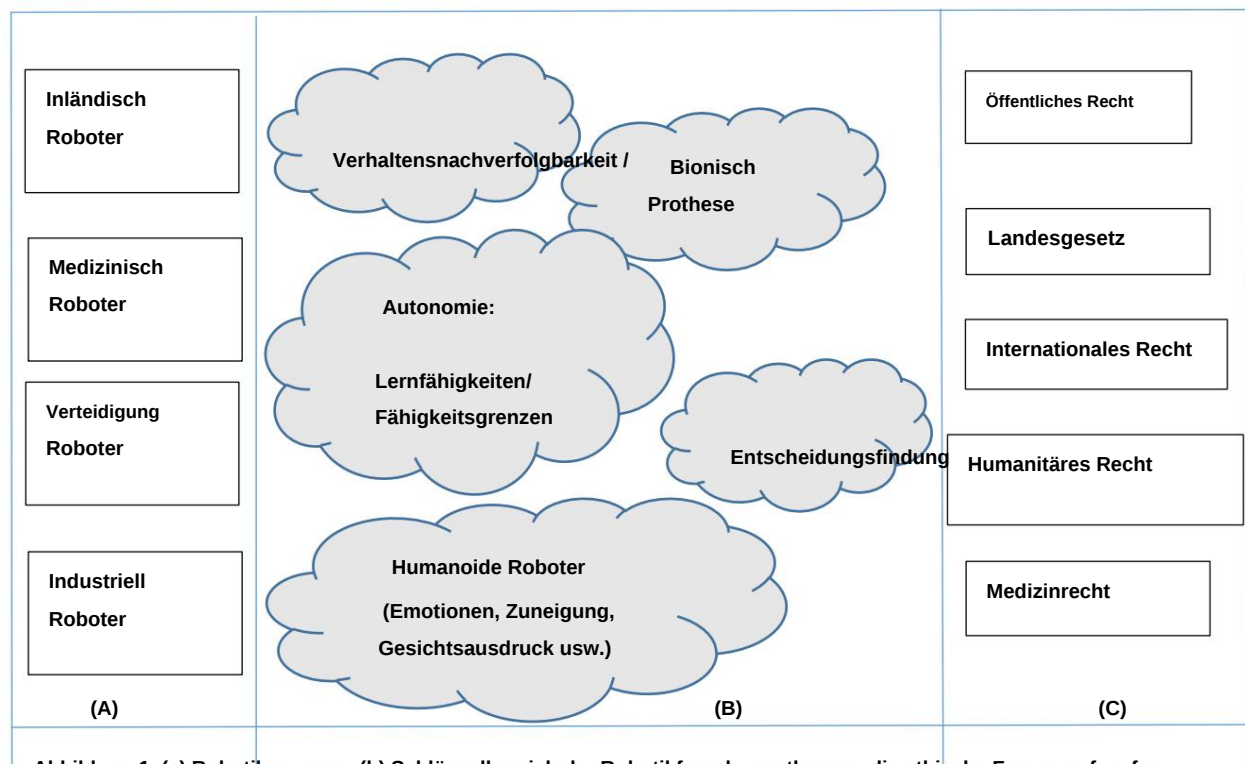


Abbildung 1. (a) Robotikgruppen, (b) Schlüsselbereich der Robotikforschungsthemen, die ethische Fragen aufwerfen, (c) zu berücksichtigender Bereich der gesetzlichen Vorschriften.

Abbildung 1 listet die schwierige Aufgabe der Regulierung der Ethik für verschiedene Gruppen der Robotik im Hinblick auf fünf beliebte Forschungsthemen auf. In jeder Gruppe der Robotik (Abbildung 1-(a)) können ethische Fragen aufgeworfen werden und Schlüsselfragen werden dargestellt (Abbildung 1-(b)): Wie können wir Roboter so gestalten, dass sie immer funktionieren?

in der Lage, ihr Verhalten zu interpretieren und die Grenzen lernender Roboter zu kennen; Humanoide Roboter, die menschliche Gesten und Gesichtsanimationen nachahmen, können die Interaktion mit Menschen (insbesondere mit Kindern und behinderten Menschen) in neue Dimensionen vorstoßen, was den Bedarf an Regulierung und Antizipation möglicher Auswirkungen erhöht; Heutzutage kann der Mensch dank der Fortschritte in der Robotikforschung stark von Prothesen profitieren. Dies kann jedoch dazu führen, dass nichtbehinderte Menschen Prothesen suchen, die ihre Fähigkeiten verbessern können, was wiederum den Bedarf an neuen Vorschriften erhöht. Schließlich betrifft die größte Herausforderung die Entscheidungsfähigkeit von Robotern. Diese Entscheidungen betreffen unmittelbar menschliches Leben und werfen daher Fragen zu rechtlichen Fragen auf, die sich aus durchgeführten Handlungen ergeben.

Zusätzlich zu diesen Forschungsthemen muss die Ethik der Robotik mit einer Reihe von Rechtsbereichen kompatibel sein (Abbildung 1 –(c)). Es muss berücksichtigt werden, dass Roboter mit verschiedenen Gesetzesebenen kompatibel sein müssen, die sich beispielsweise von Region zu Region oder von Staat zu Staat ändern können Zustand.

Vor fünf Jahren veröffentlichte ein Gremium angesehener Robotik- und KI-Experten in Großbritannien die EPSRC-Grundsätze der Robotik in Form von fünf Regeln und sieben hochrangigen Botschaften. Wir schlagen vor, diese Regeln mit einem Fokus auf die transversale Struktur – zwischen Robotik – zu diskutieren Gruppen, Forschungsthemen und rechtlichen Rahmenbedingungen, die in Abbildung 1 dargestellt sind – und unter Berücksichtigung der drei Kriterien Validität, Suffizienz und Nutzen.

Diskussion über die Festlegung dieser Regeln

Die allgemeine Bemerkung, die zu den fünf Regeln gemacht werden kann, ist, dass es ziemlich ehrgeizig ist zu glauben, dass es möglich ist, allen Arten von Robotern über alle möglichen Forschungsentwicklungen und alle Rechtsrahmen hinweg eine gemeinsame/einheitliche Anleitung zu geben. Es wäre natürlicher, nach Leitregeln zu suchen, die den transversalen Charakter der in Abbildung 1 dargestellten Robotik widerspiegeln. Wir glauben, dass die fünf Regeln nicht allgemein genug sind und die Regeln für jede Gruppe von Robotern in Abbildung 1 (a) spezifiziert und spezifiziert werden sollten Berücksichtigung der spezifischen Aspekte der Gesetze und Forschungswege in Abbildung 1 (c) und (b).

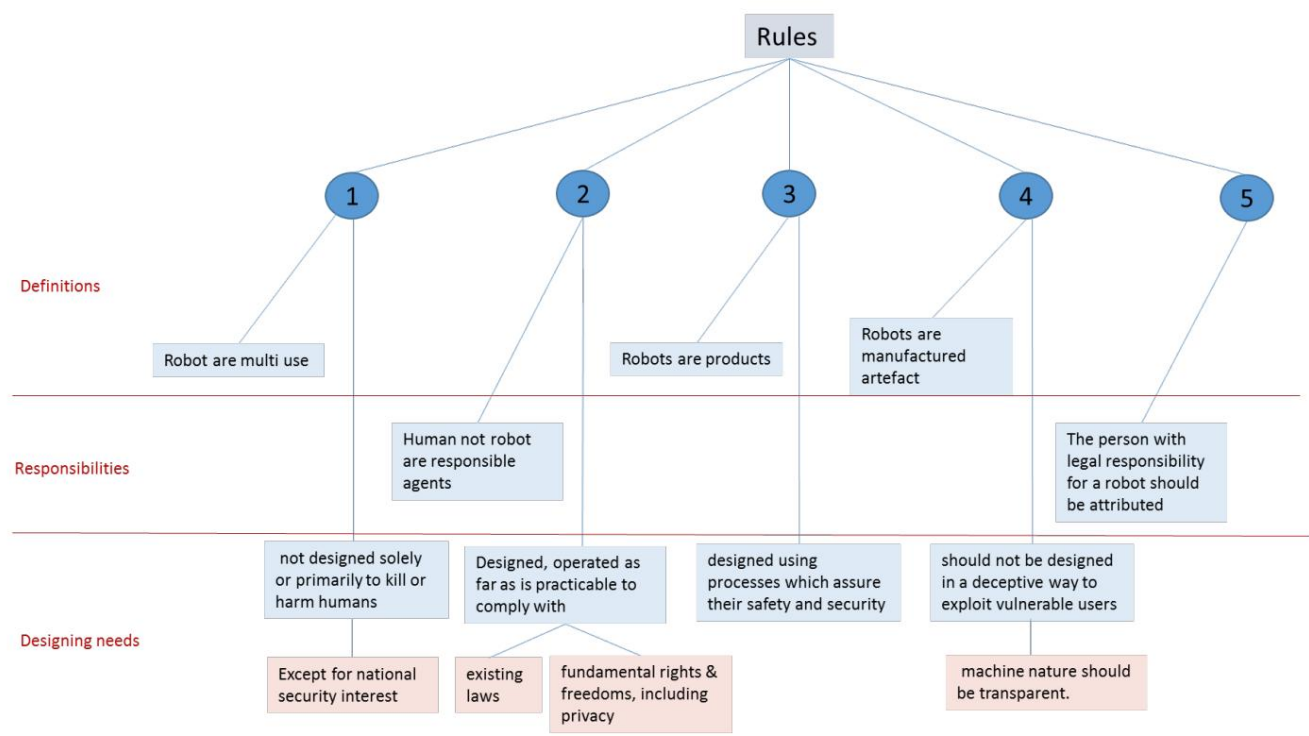


Abbildung 2. Der Satz von fünf Regeln

Abbildung 2 stellt den Satz von fünf Regeln grafisch dar. Es lässt sich erkennen, dass die fünf Regeln ein gemeinsames Muster aufweisen: Die ersten Sätze sind oft Allgemeingültigkeiten und Definitionen über Roboter wie „Roboter sind Produkte“ oder sagen, dass die Verantwortung nicht bei Robotern, sondern beim Menschen liegt. In den letzten Sätzen werden die Designanforderungen dargelegt und zwischen Sicherheitsaspekten, rechtlichen Aspekten und Transparenzaspekten hin- und hergewechselt.

Kritikpunkte an der in Abbildung 2 dargestellten Struktur lassen sich eindeutig wie folgt auflisten

- Die fünf Regeln überschneiden sich manchmal. Beispielsweise umfasst „die Einhaltung bestehender Gesetze“ in Regel 2 „nicht ausschließlich oder in erster Linie dazu bestimmt, Menschen zu töten oder zu schädigen“ in Regel 1; „Die Person, die die rechtliche Verantwortung für einen Roboter trägt, sollte zugewiesen werden“ in Regel 5 kann implizit „Menschen, nicht Roboter sind verantwortliche Agenten“ in Regel 2 einschließen.
- Die fünf Regeln sind nicht allgemeingültig. Wir können beispielsweise kaum erkennen, wie bionische Prothesen in der gegenwärtigen Form in die Regeln eingepasst werden können (z. B. kann die Robotikforschung in Zukunft eine Modifikation des menschlichen Körpers ermöglichen, um mehr Kraft, Geschwindigkeit und Künstliche Sexualität ist ein weiteres Beispiel für kontroverse Forschung, die zu ethischen Fragen führen kann).
- Die fünf Regeln besagen nicht, dass Gesetze widersprüchlich sein können, je nachdem, welcher Bereich von Räten, Regionen, Ländern und Kontinenten betrachtet wird. Ein ähnliches Beispiel, das in der Forschung häufig anzutreffen ist, sind Patentanmeldungen, für deren Anwendung in bestimmten Regionen der Welt mehrere spezifische Studien erforderlich sind. Gesetze können sogar noch komplizierter sein, da der religiöse Glaube sowie die Gewohnheiten und Bräuche der Menschen den Begriff der Ethik bestimmen.

Schlussfolgerungen

In dieser Diskussion haben wir kurz Argumente für die Notwendigkeit einer unterschiedlichen Formulierung der fünf Regeln dargelegt. Anhand einer bildlichen Darstellung der fünf Regeln wird deutlich, dass diese tatsächlich nicht ausreichen, sich überschneiden und nicht explizit die wahren Herausforderungen der Robotik-Ethik widerspiegeln.

Ein Teil unserer Überlegungen stützte sich auf die transversale Natur der Roboterethik entlang dreier Linien: Gruppen, die die Robotik ausmachen, zukünftige Wege der Robotik, die bei der Definition der Ethik unbedingt erfasst werden müssen, und die strukturierte Natur des Rechts. Wir empfehlen eine natürliche Neuformulierung, die die Ethik für jede Gruppe von Robotik differenziert und gleichzeitig die Widersprüche und starken Einschränkungen berücksichtigt, die aufgrund der strukturellen Natur des Rechts bestehen können.

Literaturverzeichnis

- [1] G. Bekey, Robotik: Stand der Technik und zukünftige Herausforderungen., Kalifornien: London Imperial College-Presse. , 2008.
- [2] K. Dautenhahn, S. Woods, C. Kaouri, ML Walters, KL Koay und I. Werry, Was ist ein Roboter-Begleiter, Freund, Assistent oder Butler?, IIC o. IR a. Systems, Hrsg., 2005.
- [3] J. Forlizzi und DC, Serviceroboter in der häuslichen Umgebung: eine Studie über den Roomba-Staubsauger im Haushalt, 1. ACM SIGCHI/SIGART-Konferenz über Mensch-Roboter-Interaktion. ACM, 2006.
- [4] JY Sung, RE Grinter, HI Christensen und L. Guo, Hausfrauen oder Technikbegeisterte?: Verständnis für Hausroboterbesitzer, 3. ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2008, S. 129–136.
- [5] GP Moustris, SC Hiridis, K. Deliparaschos und K. Konstantinidis, Evolution of Autonome und halbautonome chirurgische Robotersysteme: eine Überprüfung der Literatur,

Bd. 7, The International Journal of Medical Robotics and Computer Assisted Surgery, 2011, S. 375-392.

- [6] J. Rassweiler, J. Binder und T. Frede, „Robotik und Telechirurgie: Werden sie unsere verändern?“ Zukunft?“, Bd. 11, Nr. 3, S. 309–320, 2001.
- [7] G. Kwakkel, KBJ und KHI, Auswirkungen der robotergestützten Therapie auf die oberen Gliedmaßen Erholung nach Schlaganfall: eine systematische Übersicht, Neurorehabilitation und neurale Reparatur, 2007.
- [8] K. Cleary und C. Nguyen, „Stand der Technik in der chirurgischen Robotik: klinische Anwendungen und technologische Herausforderungen“, Bd. 6, nein. 6, S. 312–328, 2001.
- [9] PW Singer, „Robots at War“, *Wilson Quarterly*, 2008.
- [10] TK Adams, „Zukünftige Kriegsführung und der Niedergang der menschlichen Entscheidungsfindung“, *Parameters*, Bd. 31, Nr. 4, 2001.
- [11] P. Leitão, „Agentenbasierte verteilte Fertigungskontrolle: Eine hochmoderne Umfrage“, *Technische Anwendungen künstlicher Intelligenz*, Bd. 22, nein. 7, S. 979-991, 2009.
- [12] ZM Bi, SY Lang, W. Shen und L. Wang, „Rekonfigurierbare Fertigungssysteme: der Stand der Technik“, *International Journal of Production Research*, vol. 46, Nr. 4, S. 967-992, 2008.

Roboter sind nicht nur Werkzeuge

Tony J. Prescott, University of Sheffield

Im Mittelpunkt der EPSRC-Prinzipien der Robotik (im Folgenden „die Prinzipien“) stehen eine Reihe ontologischer Behauptungen über die Natur von Robotern, die als Axiome für die spätere Entwicklung ethischer Herausforderungen und Regeln dienen. Dazu gehören Behauptungen darüber, was Roboter sind, und auch darüber, was sie nicht sind. Zu den Behauptungen darüber, was Roboter sind, gehört, dass „Roboter Mehrzweckwerkzeuge sind“ (Prinzip 1), dass „Roboter Produkte“ (Prinzip 3) und „Teile der Technologie“ (Kommentar zu Prinzip 3) sind und dass „Roboter hergestellte Artefakte sind“ (Prinzip 4). Zu den Behauptungen darüber, was Roboter sind, gehört nicht, dass „Menschen, keine Roboter, eine Antwort sind.“ sibleagenten“ (Prinzip 2), dass Roboter „einfach keine Menschen“ sind (Kommentar zu Prinzip 3) und dass Roboterintelligenz nur einen „Eindruck echter Intelligenz“ vermitteln kann (Kommentar zu Prinzip 4).

Beim ersten Lesen scheinen diese Aussagen eindeutige Behauptungen offensichtlicher Wahrheiten zu sein. Ich werde argumentieren, dass dies nicht der Fall ist. Stattdessen werde ich behaupten, dass es diesen ontologischen Verpflichtungen an Nuancen mangelt, dass sie nur allzu leicht davon ausgehen, dass wir die Randbedingungen der zukünftigen Robotikentwicklung kennen, und dass sie einige der wichtigen ethischen Debatten verschleiern oder ignorieren Prinzipien könnten damit beginnen, sorgfältig über den ontologischen Status von Robotern nachzudenken.

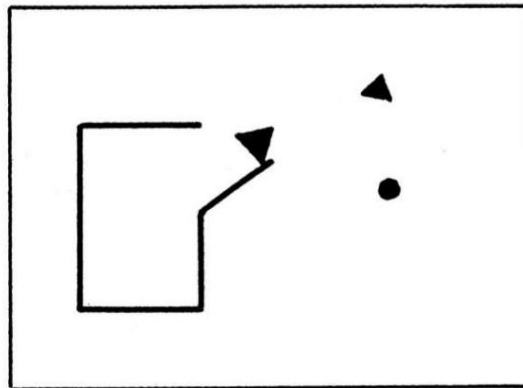
Wenn wir uns ansehen, wie die Prinzipien dargestellt werden, scheint ein impliziter Induktionsprozess am Werk zu sein, der es ermöglicht, Aussagen darüber, was die meisten aktuellen Roboter sind, als Aussagen darüber, was Roboter im Wesentlichen sein müssen, zu interpretieren. Beispielsweise fügt sich die Aussage, dass Roboter im Prinzip Mehrzweckwerkzeuge sind¹, in die Behauptung ein, dass Roboter „nur Werkzeuge“ im Kommentar zu Prinzip 2 seien, und in die Aussage, dass „Roboter sind lediglich Werkzeuge verschiedener Art, wenn auch sehr spezielle Werkzeuge“ in der Präambel. Während es leicht ist, der allgemeinen Aussage zuzustimmen, dass Roboter Mehrzweckwerkzeuge sind, insbesondere im Zusammenhang mit der Diskussion über Dual-Use (Prinzip 1), behauptet die viel stärkere Behauptung, dass Roboter nur Werkzeuge oder nur Werkzeuge seien, bestreitet, dass sie sinnvollerweise zu anderen getrennten Kategorien gehören könnten.

Nehmen wir zum Beispiel die Kategorie „Begleiter“. Es gibt große Bemühungen um die Entwicklung von Roboterbegleitern, die Menschen soziale und emotionale Unterstützung bieten können, wie in der Diskussion von Prinzip 4 teilweise anerkannt. Die Kategorie der Werkzeuge beschreibt physische/mechanische Objekte, die eine Funktion erfüllen, während die Kategorie der Gefährten wichtige andere Personen beschreibt, normalerweise Menschen oder Tiere, mit denen Sie möglicherweise zusammen sind Eine prokale Beziehung ist durch eine emotionale Bindung gekennzeichnet. Die Möglichkeit, dass Roboter zu beiden Kategorien gehören könnten, wirft wichtige und interessante Fragen auf, die durch die Behauptung, dass Roboter nur Werkzeuge seien, verschleiert werden.

In Übereinstimmung mit der Sichtweise von Robotern als Werkzeug ist die Diskussion über die Kameradschaft von Robotern in den Prinzipien ziemlich abweisend und beschreibt die Spielzeuge, die Menschen, die keine Haustiere halten können oder können, ein gewisses Vergnügen bereiten könnten. Roboter sind in dieser Hinsicht künstliche Begleiter, die eine „Illusion von Emotionen“ erzeugen und deren Intelligenz künstlich und nicht „echt“ ist. Die künstliche Natur von Roboterbegleitern, so wird argumentiert, stelle insofern ein realethisches Problem dar, als Roboterbegleiter potenziell trügerisch seien und daher so gestaltet werden sollten, dass ihre „Maschinennatur“ transparent sei.

Das ontologische Problem betrifft hier insbesondere die Behauptung, dass Roboter niemals über psychologische Fähigkeiten wie „echte“ Emotionen oder Intelligenz verfügen könnten. Was diese, unmenschlich ausgedrückt, sind, wird in den Kognitions- und Gehirnwissenschaften heftig diskutiert. Es gibt daher keinen zwingenden Grund zu der Annahme, dass diese Fähigkeiten nur für Menschen gelten und nicht von Maschinen geteilt werden könnten. Tatsächlich gibt es Gegenbehauptungen, dass Roboter, entsprechend konfiguriert, Emotionen haben können^[1], während die Zukunft der künstlichen Intelligenz als Intelligenz keine offensichtliche Obergrenze unterhalb der menschlichen Ebene hat

Ein weiteres Problem betrifft die Annahme darüber, wie Menschen Roboter *sehen werden* – insbesondere, dass Roboter als Werkzeuge gesehen werden, wenn sie transparent dargestellt werden. Dies könnte leicht falsch sein, Zum Beispiel können Menschen Roboter vermenschlichen, unabhängig davon, wie offensichtlich sie hergestellte Produkte sind. Ein Grund zu der Annahme, dass dies der Fall sein könnte, ist die stark soziale Natur unserer Gehirne und wie leicht unsere Empathie durch etwas ausgelöst wird, das lebensecht erscheint. Heider-Simmelanimationen einfacher geometrischer Figuren [2] (siehe Abbildung) zeigen, wie grob diese Informationen sein können, und dennoch werden wir Absicht, Motivation und sogar Emotionen sehen. Die Erfindung des digitalen Haustiers Tamagotchi hat gezeigt, dass eine einfache 2-Danimation eines tierähnlichen Lebewesens einen überzeugenden Drang zur Fürsorge erzeugen kann [3]. Wir müssen nicht glauben, dass der psychologische Zustand, den wir in diese Artefakte eingelesen haben, real ist, um ihn zu haben eine authentische emotionale Reaktion auf uns selbst.



Figur. Geometrische Formen, die sich in einfacher Animation bewegen, wurden in dieser berühmten Studie von Heide und Simmel aus dem Jahr 1944 als „belebte Wesen, hauptsächlich Personen“ interpretiert.

Eine Analyse ontologischer und psychologischer Fragestellungen in der Mensch-Roboter-Interaktion wurde bereits durchgeführt wurde von Kahn und Kollegen [4] gemacht. Einem ähnlichen Gedankengang folgend, können wir beschreiben Vier allgemeine Möglichkeiten, wie ontologische Perspektiven darauf, was *Roboter sind*, und *psychologische Perspektiven darauf, wie Roboter gesehen werden*, kombiniert werden können. Diese werden in der folgenden Tabelle dargestellt zusammen mit einigen der damit verbundenen ethischen Fragen.

I. Roboter sind nur Werkzeuge (o), und Menschen werden Roboter nur als Werkzeuge betrachten, es sei denn, sie werden durch trügerisches Roboterdesign in die Irre geführt (p). <i>Ethische Fragen: Wir sollten uns mit der menschlichen Verantwortung als Roboterhersteller/-benutzer und der Gefahr der Täuschung bei der Herstellung von Robotern befassen, die den Anschein erwecken, etwas zu sein, was sie nicht sind. Dies ist die Position der „Prinzipien“.</i>	II. Roboter sind nur Werkzeuge (o), aber die Menschen können sehen, dass sie über erhebliche psychologische Fähigkeiten verfügen, unabhängig von der Transparenz ihre Maschinennatur (p). <i>Ethische Fragen: Wir sollten berücksichtigen, wie Menschen Roboter sehen, zum Beispiel, dass sie das Gefühl haben, sinnvolle und wertvolle Beziehungen zu Robotern zu haben, oder dass sie Roboter als wichtige innere Zustände sehen, wie etwa die Fähigkeit zu leiden, obwohl sie nicht über solche Fähigkeiten verfügen.</i>
III. Roboter können über erhebliche psychologische Fähigkeiten verfügen (o), aber die Menschen werden sie immer noch nur als Werkzeuge betrachten (p). <i>Ethische Fragen: Wir sollten die Risiken analysieren, die mit der Behandlung von Lebewesen einhergehen, die möglicherweise erhebliche psychologische Fähigkeiten haben, wie zum Beispiel die Fähigkeit zu leiden, als wären sie nur Werkzeuge, und die Gefahren, die darin liegen, eine neue Klasse von Lebewesen mit bedeutenden psychologischen Fähigkeiten, wie zum Beispiel menschenähnlicher Intelligenz, zu schaffen, ohne anzuerkennen, dass sie dies tun.</i>	IV. Roboter können einige bedeutende menschenähnliche psychologische Fähigkeiten haben (o), und die Menschen werden sehen Sie verfügen über solche Kapazitäten (p). <i>Ethische Fragen: Wir sollten Szenarien betrachten, in denen Menschen neben neuen Arten psychologisch bedeutsamer Wesenheiten in Form zukünftiger Roboter/KIs koexistieren müssen.</i>

Tabelle: Wie sich ontologische und psychologische Perspektiven auf Roboter kombinieren lassen (nach Kahn et al., 2007).

Beachten Sie, dass nur ein Quadrant dieser Tabelle (I) in den Prinzipien behandelt wird, dass aber II, III und IV zumindest theoretisch alle möglich sind. Zum Abschluss dieses Aufsatzes möchte ich kurz einige der ethischen Fragen betrachten, die in den Quadranten II–IV auftreten.

In Quadrant II, interesting questions arise how robots should be treated—

not because they are sentient agents but because people will choose to treat them as such. For instance, the idea that it should be unlawful to

Quadrant III betrifft die Möglichkeit, dass Roboter über bedeutende psychologische Fähigkeiten verfügen, die nicht Gefahr laufen, von Menschen übersehen zu werden. Dies wirft ethische Risiken auf, die in den Prinzipien nicht diskutiert, aber von anderen hervorgehoben wurden. Beispielsweise hat Metzinger[8] argumentiert, dass wir Roboter bauen könnten, die in der Lage sind, Leiden zu erleben, ohne es zu merken, und so etwas Neues erschaffen könnten. Eine Art empfindungsfähiges Wesen, das aufgrund unserer Handlungen unnötig leidet, ist eindeutig ethisch problematisch, wenn dies passieren könnte. Auch wenn dies auf kurze Sicht unwahrscheinlich erscheint, gibt es Gründe zu bedenken, dass dies mittel- bis langfristig möglich sein könnte, da kognitive Architekturen für Roboter komplexer werden. Mehrere Trends in der laufenden Forschung zum menschlichen Bewusstsein unterstützen diese Möglichkeit ebenfalls. Erstens einer der großen zeitgenössischen Theorien des Bewusstseins [9] behauptet eine entscheidende Rolle für die Integration von Informationen

Das erfordert nicht unbedingt ein biologisches Substrat. Neurologen prüfen auch erneut, ob Inseln integrierter Aktivität in den Gehirnen „eingesperrter“ Patienten eine Form von minimalem Bewusstsein darstellen könnten [10]. Schließlich gibt es eine aktive Debatte darüber, ob Tiere mit

Kleinere Gehirne als unseres, wie z. B. Fische, könnten in erheblichem Maße empfindungsfähig sein (z. B. dass sie Schmerz erleben)[11]. Diese Entwicklungen deuten darauf hin, dass Bewusstsein in einem künstlichen Agenten möglich sein könnte, ohne der Größe oder Komplexität eines intakten menschlichen Gehirns entsprechen zu müssen. Dennett hat argumentiert, dass „rohes, kitschiges, zweitklassiges, künstliches Bewusstsein“ in Robotern möglich sei [12] und Bryson[13] haben vorgeschlagen, dass heutige Roboter möglicherweise bereits über einige einfache Formen des Bewusstseins verfügen, die einige häufig vorgeschlagene Kriterien erfüllen. Noch nichts davon behauptet, dass sie im Quadrant III sind, aber angesichts der Risiken sollten Ethiker uns dazu drängen, wie wir wissen würden, ob wir es wären.

Eine der Konsequenzen aus der Sichtweise von Robotern als „nur Werkzeuge“ ist die stillschweigende Ablehnung der Möglichkeit einer starken KI – dass zukünftige Roboter über allgemeine Intelligenz auf menschlicher oder darüber hinausgehender Ebene verfügen könnten. Das Thema Aquadrant III/IV wurde kürzlich von namhaften Wissenschaftlern und Innovatoren wie Stephen Hawking, Elon Musk und Bill Gates, um nur einige zu nennen, diskutiert und von Bostrom eingehend analysiert[14], ist, dass eine KI-Singularität die Herr-Sklave-Beziehung zwischen Menschen und Robotern umkehren könnte. Die Überzeugung, dass Roboter/KI „nur Werkzeuge“ sind, könnte uns davon abhalten, die ersten Anzeichen dafür zu erkennen. Eine sich selbst fördernde Super-KI. Ein ethischer Ansatz würde sicherlich zu mehr Wachsamkeit ermutigen. Eine positivere Quadranten

Die Debatte ist die von Heylighen[15] und anderen vorgeschlagene Perspektive des „globalen Gehirns“, dass Menschen und fortgeschrittene KI zu unserem gegenseitigen Nutzen nebeneinander existieren könnten. Dies erinnert daran, dass es bei der Ethik um die Analyse des potenziellen Nutzens sowie der Risiken gehen muss.

Obwohl Quadrant-III/IV-Szenarien weit hergeholt oder zumindest weit entfernt erscheinen mögen, haben solche Bedenken die öffentliche Vorstellungskraft erregt und erhebliche Debattenaufrufe ausgelöst (z. B. [16]). Nach meiner eigenen Erfahrung im Gespräch mit Mitgliedern der Öffentlichkeit und der Medien sind dies oft die Themen, an denen das größte Interesse und die größte Sorge besteht. Der Versuch, durch Beharren eine rhetorische Barrikade gegen diese Probleme zu schaffen, dass Roboter nur Werkzeuge sind, kann die Stimmen nur wenig beruhigen und könnte hegemonial und herablassend wirken. Auch wenn Ansätze für diese längerfristigen methodischen Herausforderungen zwangsläufig spekulativ sind, ist ein Ausgangspunkt die Anerkennung, dass es hier Bedenken gibt, die weiterer Aufmerksamkeit wert sind.

Ein offenerer Ansatz könnte sein, anzuerkennen, dass, während die meisten Roboter derzeit kaum mehr als Werkzeuge sind, wir in eine Ära eintreten, in der es neue Arten von Fähigkeiten geben wird, die einige der Eigenschaften von Maschinen und Werkzeugen mit psychologischen Fähigkeiten kombinieren, von denen wir zuvor angenommen hatten, dass sie komplexen biologischen Organismen wie Menschen vorbehalten sind. In Anlehnung an Kang [17] könnte der ontologische Status von Robotern am besten als *grenzwertig* beschrieben werden – weder lebend auf die gleiche Weise wie biologische Organismen, noch einfach mechanisch wie bei einer herkömmlichen Maschine. Die Begrenztheit von Robotern macht sie sowohl faszinierend als auch von Natur aus beängstigend und ist ein Blitzableiter für unsere umfassenderen Ängste vor den entmenslichenden Auswirkungen der Technologie[18].

Die Association of Manhattan Scientists schrieb 1945 [19] über ihr Gefühl der kollektiven Verantwortung für ihre Rolle bei der Entwicklung einer Technologie mit „Potenzial für große Rüstung oder großes Gut“ (Atomenergie) und über ihr „besonderes Bewusstsein“, dass sie zum „Fortschritt unserer Zivilisation oder ihrer völligen Zerstörung“ führen könnte. Sie haben auch eine besondere Verantwortung, zu verstehen und offen darüber zu sprechen, was die Zukunft der Robotik bringen könnte und welche potenziellen Vorteile und Risiken sie mit sich bringt.

Verweise

1. Fellous, J.-M., *From human emotions to robot emotions*, in AAAI Spring Symposium on Architectures for modeling emotions: Cross-disciplinary foundations. E. Hudlicka and L. Cañamero, Editors. 2004, AAAI Press: Menlo Park, CA. S.37-47.
2. Heider, F. und M. Simmel, *Eine experimentelle Studie über scheinbares Verhalten*. The American Journal of Psychology, 1944.57(2):S.243-259.
3. Levy, D., *Love and Sex with Robots*. 2007, London: HarperCollins.
4. Kahn, J., Peter H., et al., *What is a Human?: Toward psychological benchmarks in the field of human-Roboter interaction*. Interaction Studies, 2007.8(3):S.363-390.
5. Lovgren, S., *Robot Code of Ethics to Prevent Android Misuse, Schutz der Menschen*, in National Geographic News. 2007.
6. Brown, A., *Tomorrow's robotic dog is truly human* in Guardian. 2015: Manchester.
7. Collins, EC, A. Millings und P. TJ *Attachment to Assistive Technology: A New Conceptualisation*. in Assistive Technology: From Research to Practice: AAATE 2013. 2013.
8. Metzinger, T., *Der Ego-Tunnel: Die Wissenschaft vom Geist und der Mythos vom Selbst*. 2009, New York: Basic Books.
9. Tononi, G., *Bewusstsein als integrierte Information: ein vorläufiges Manifest*. The Biological Bulletin, 2008.215(3):S.216-242.
10. Qiu, J., *Sondierung von Bewusstseinsinseln im geschädigten Gehirn*. The Lancet Neurology. 6(11):S.946-947.
11. Seth, AK, *Why fish pain cannot and should not be ruled out* Animal Sentience, 2016.2016.020.
12. Dennett, D., *Die praktischen Anforderungen für die Herstellung eines bewussten Roboters*. Philosophical Transactions of the Royal Society of London A, 1994.349:S.133-146.

13. Bryson, JJ Rohes, *kitschiges, zweitklassiges Bewusstsein*. in *WienKonferenz über Bewusstsein*. 2008.
14. Bostrom, N., *Superintelligenz: Wege, Gefahren, Strategien*. 2014, Oxford: OxfordUniversityPress.
15. Heylighen, F., *TheGlobalBrainasaNewUtopia*, in *Zukunftsfiguren*, R.Maresch und F.Rötzer, Herausgeber.2002, Suhrkamp: Frankfurt.
16. FutureofLifeInstitute. *Ein offener Brief: Forschungsprioritäten für robuste und nützliche künstliche Intelligenz*. 2015;Verfügbar unter: <http://futureoflife.org/ai-open-letter/>.
17. Kang, M., *SublimeDreamsofLivingMachines: Der Automat in der europäischen Imagination*. 2011, Cambridge, MA: HarvardUniversityPress.
18. Szollosy, M., *Freud, Frankenstein und unsere Angst vor Robotern: Projektion unserer kulturellen Wahrnehmung von Technologie*. AI&SOCIETY,2016:S.1-7.
19. AssociationofManhattanScientists. *Vorläufige Stellungnahme*. 1945; Verfügbar unter: <https://www.gilderlehrman.org/history-by-era/postwar-politics-and-origins-cold-war/resources/physicists-predict-nuclear-arms-race->.

AISBWorkshop zu den EPSRC-Prinzipien der Robotik

Einen überholten Humanismus verteidigen?

Michael Szollosy, Sheffield Robotics

Einführung

Der EPSRC-Workshop 2010 zur Entwicklung einer Reihe von Prinzipien für die verantwortungsvolle Entwicklung und Erforschung von Robotern war ein wichtiges, ehrgeiziges und sehr gut gemeintes Projekt. Die „Prinzipien der Robotik“, die sich aus diesen Diskussionen ergaben, können als hervorragende Sammlung pragmatischer Regeln für den Bau und die Regulierung von Robotern angesehen werden und enthalten sehr sinnvoll Versionen, die sowohl spezifisch als auch spezifisch genutzt werden können im rechtlichen Kontext und auch für das allgemeine

Aus der Durchsicht der EPSRC-Grundsätze wird deutlich, dass sie eine wichtige Funktion erfüllen können und sollten Schutz der Menschen vor unverantwortlicher oder einfach gedankenloser Erforschung von Technologien, die möglicherweise sehr reale, sehr negative Folgen für die Menschheit haben könnten, auf persönlicher, gesellschaftlicher oder sogar artweiter Ebene.

Es ist jedoch auch klar, dass das, was durch die EPSRC-Prinzipien geschützt wird, ein sehr spezifischer Mensch ist, oder zumindest eine spezifische Vorstellung davon, was einen Menschen ausmacht. Den Prinzipien liegen eine Reihe unausgesprochener Annahmen über die menschliche Natur und folglich die Natur unserer Beziehung zur Technologie zugrunde. In der Workshop tat dies.

(Allerdings behaupte ich, dass es eine Präambel geben sollte, in der diese Annahmen dargelegt werden.)

Die EPSRC-Richtlinien sollen Menschen schützen und sicherstellen, dass Roboterforschung zum größtmöglichen Nutzen für ihre Bürger durchgeführt wird. Wie genau diese „Bürger“ aussehen könnten, bleibt jedoch unbeantwortet und bedeutet, dass diese Grundsätze dennoch gelten, so lobenswert ihre Absichten auch sein mögen. Es gibt bereits ein Dokument, das in einem bestimmten historischen und kulturellen Kontext verwurzelt ist, und daher ist es unwahrscheinlich, dass diese Grundsätze in ihrer jetzigen Form mittel- bis langfristig Bestand haben werden.

Die EPSRC-Prinzipien machen bestimmte, sehr spezifische, aber völlig unausgesprochene Annahmen darüber, was einen „Menschen“ ausmacht. Und die Prinzipien passen *daher sehr gut zum Menschen*. Wir haben das Recht, Roboter zu entwerfen, zu bauen und zu kaufen, und müssen die volle rechtliche Verantwortung dafür tragen.

Die Prinzipien, wie sie derzeit formuliert sind, werden wahrscheinlich kurzfristig, vielleicht sogar mittelfristig, ausreichen, um die meisten Probleme mit neuen Technologien zu lösen, die in der Robotik und in Computerlabors im gesamten Vereinigten Königreich entstehen

Vergängliches Geschöpf, eine relativ neue Erfindung, ein Wesen, das darüber hinaus weder gegenwärtig ist noch jemals intern konsistent und einheitlich war und durch eine ganze Reihe neuer Technologien ständig neu geschaffen und transformiert wird. Zu diesen transformativen Technologien gehören die biologischen Verbesserungen und mechanischen Verbesserungen, die von Posthumanisten, Transhumanisten und anderen verfochten werden.

Aber auch – einfacher gesagt – Neues in Bezug auf uns selbst, die Gesellschaft und neue Denkweisen über unseren Platz in der Welt, Veränderungen, die nicht nur durch Robotiker, Genetiker, Informatiker, sondern auch durch Veränderungen in unserem wirtschaftlichen, politischen und sozialen Leben im Allgemeinen herbeigeführt werden können. Der Mensch – oder die Menschen –, um die es in den ESPRC-Prinzipien geht, wird nicht der erste oder letzte Ausdruck dessen sein, was es bedeutet, ein Mensch zu sein.

Welcher Mensch?

Im (impliziten) Kern der ESPRC-Prinzipien steht ein besonderer Mensch, der in den letzten Jahrhunderten durch das definiert wurde, was *als Humanismus bekannt geworden ist*. Dieser Mensch ist ein eigenständiger Agent, ein Wesen, das unabhängig ist und nicht von anderen, metaphysischen oder übernatürlichen Kräften beherrscht werden kann. Dieser Mensch steht im Mittelpunkt europäischer Rechts-, Ethik-, Wirtschafts- und Politiksysteme. Es ist jedoch wichtig, sich daran zu erinnern, dass 1. dieser Mensch noch eine relativ neue Erfindung ist und dass es 2. im Laufe seines Lebens nie nur eine einzige Version dieses Menschen gegeben hat, wie humanistische Befürworter sich das gerne vorgestellt haben.

Es besteht wenig Konsens über die Geburt des Menschen – einige würden behaupten, die Renaissance, andere würden sagen, die Aufklärung, und wieder andere würden immer noch sagen, dass das humanistische Subjekt erst im 19. Jahrhundert eine zentrale Rolle in der Art und Weise spielte, wie wir über uns selbst denken. Ebenso behaupten einige, dass dieser Mensch in der zweiten Hälfte des 20. Jahrhunderts gestorben sei, während andere argumentieren, dass er bis heute fortbesteht. Welches Datum auch immer man dem zuschreibt Wenn wir über uns selbst denken, ist es sicher, dass diese Konzeptualisierung des Menschen eine Erfindung ist, nicht gegeben; der humanistische Mensch ist nicht „natürlich“ oder gar eine „richtige“ Interpretation unserer menschlichen Natur Wir nutzen und die Umwelt, in der wir leben.

Und wenn wir den Humanismus auf diese Weise kontextualisieren und seine Seite historisch betrachten und dabei beobachten, wie unterschiedliche Vorstellungen davon, was es bedeutet, menschlich zu sein (oder sogar „Humanist“), sich im Laufe der Jahrhunderte radikal verändert haben, wird deutlich, dass wir nicht nur von einem einzelnen Menschen oder einer Vorstellung davon sprechen, was es bedeutet, menschlich zu sein Über das, worüber wir sprechen, geht es nicht um einen einzelnen Menschen, sondern um viele Menschen, nicht um eine einzelne, unveräußerliche, selbstverständliche menschliche Natur, sondern um Menschen, die ihre Selbstauffassung in bestimmten Kontexten verändern. Das Selbstverständnis des Menschen ist immer kontingent und kontextabhängig . Jeder Mensch kann ein Mitglied der Öffentlichkeit, ein Bürger, ein Spezialist in verschiedenen Kontexten, ein Konsument, ein Produzent sein; wir können Kriminelle, Patienten, Klienten, Steuerzahler, Stakeholder, Studenten, Arbeiter oder das Management oder alles auf einmal oder nichts davon sein, je nach Kontext. Und was es bedeutet, zum Beispiel ein „Krimineller“, ein „Bürger“, ein „Arbeiter“ oder ein „Mensch“ zu sein. ora' „Frau“ ist heute ganz anders als vor zweihundert oder sogar fünfzig oder sogar vor zehn Jahren. Wir können ständig wechselnden Diskursen über Recht, Medizin, Bildung, Politik, Wirtschaft, Philosophie, Industrie, Medien und eine Vielzahl anderer Systeme, Sprachen und Institutionen ausgesetzt sein, die versuchen, uns auf leicht unterschiedliche oder radikal unterschiedliche Weise zu definieren und zu verstehen.

Die Annahmen, die unserer Vorstellung einer einzelnen Vorstellung davon zugrunde liegen, was es bedeutet, „Mensch“ zu sein, sind angesichts der intensiven Prüfung neuer Denkweisen über uns selbst nicht haltbar, und auch, da neue Technologien uns dazu zwingen, anders über uns selbst zu denken. Technologien haben seit jeher die radikale Transformation des Menschen erzwungen, seit unsere prähistorischen Vorfahren zum ersten Mal auftraten

Stöcke, die bei der Jagd helfen, oder das fliegende Shuttle veränderten die Art und Weise, wie Stoffe in der Industriellen Revolution hergestellt wurden. Neue Entwicklungen in der Robotik werden diese Prozesse verschärfen. Die Menschen werden es sein Sie sind komplexer in ihre Werkzeuge integriert, da Stöcke zu Prothesen werden und menschliche Arbeit vollständig durch automatisierte Maschinen ersetzt wird. Und diese Entwicklungen werden neue Menschen und neue Denkweisen über uns selbst hervorbringen.

Technologie manifestiert sich jedoch nicht immer in physischen Gebilden; technologische Fortschritte zeigen sich nicht immer in Form neuer Werkzeuge oder Maschinen: Die Erfindung von Gesetzen und Rechtssystemen waren neue Technologien, die enorme Auswirkungen darauf hatten, wie wir unser menschliches und soziales Selbst konstruieren, genauso wie die Erfindung wissenschaftlicher Methoden, neue industrielle Beziehungen oder Facebook die Art und Weise verändert haben, wie wir uns selbst verstehen und präsentieren Diese Idee an uns selbst, an die Welt. Unsere Technologien des 21. Jahrhunderts – darunter fortschrittlichere Roboter und KI, aber auch die Veränderung von Rechtssystemen, politischen Körperschaften und Systemen für ethisches Leben – werden weitere Entwicklungen sein, die im Laufe der Zeit unser Selbstverständnis und unsere Vorstellung davon, was es bedeutet, menschlich zu sein, radikal verändern werden.

In welcher impliziten, unausgesprochenen Art und Weise müssen wir speziell untersuchen, ob die ESPRC-Prinzipien diesen humanistischen Annahmen über die Natur des Menschen genügen? Verfassungen, Chartas, Verträge oder Grundsatzserklärungen – sollten von Anfang an die Wahrheiten, die als selbstverständlich gelten, als Grundlage für das, was zu befolgen ist, ausdrücklich darlegen.

In Ermangelung eines klar definierten Themas bieten die Prinzipien die übliche, vertraute humanistische Vorstellung vom Menschen – den statischen, homogenen Menschen, der sehr bald entstehen wird obsolet, wenn nicht schon, durch den technologischen Fortschritt, den es zu kontrollieren versucht. Es gibt eine allzu vereinfachte Vorstellung von der Beziehung zwischen Menschen und ihren Werkzeugen: eine einseitige Beziehung, bei der Werkzeuge immer Diener ihrer menschlichen Herren sind und immer unter der Kontrolle eines unabhängigen menschlichen Agenten stehen. Eine solche Beziehung zwischen Subjekt und Objekt, aktivem Agenten und passivem Artikel wäre immer naiv gewesen Naja, zuerst haben wir einen Stock gepackt, aber wie unsere neuen Technologien eine grundlegende Neuorganisation unseres gesamten Lebens erfordern, und darauf bestanden, wie wir unsere gesamte soziale Struktur konzipieren.

Zu sagen, dass unsere Beziehung zu unseren Werkzeugen nicht eine einfache einseitige Herr-Diener-Beziehung ist, heißt nicht, dass unsere Werkzeuge unsere Herren sind. Wir müssen jedoch erkennen, dass *der Mensch ebenso das Produkt unserer Art und Weise ist, Dinge zu machen, denn was wir machen und wie wir es können, wird von den Menschen entschieden.* (Es ist wirklich nichts Kontroverses, wenn man das sagt; es ist etwas, was Marx vor über einhundertfünfzig Jahren erkannte, als er erklärte, *wie die Produktionsweise* einer Gesellschaft definierte seine sozialen Beziehungen und wie einzelne Menschen dann wiederum durch diese sozialen Beziehungen definiert wurden.) Die Beziehung zwischen Menschen und ihren Werkzeugen war schon immer komplexer, als sie sich der Humanismus und diese Prinzipien vorstellen. „Arbeiter“ und die Grenzen zwischen „biologisch“ und „maschinell“ verschwimmen weiter.

Die Prinzipien versuchen daher trotz aller böswilligen Absichten, einem veralteten Menschen die Beherrschung zukünftiger und sich entwickelnder Technologie zu ermöglichen.

Die in den Prinzipien dargelegte Vorstellung vom Menschen teilt mit dem Humanismus auch die Illusion, ein einzelnes, homogenes Subjekt anzubieten, obwohl dieses Subjekt tatsächlich eine Ansammlung mehrerer – oft widersprüchlicher – Wesen ist. Der erste Absatz spricht von dem Versprechen, dass „Robotikangebote der Gesellschaft zugute kommen“, als ob die „Gesellschaft“ selbst eine einzige homogene Einheit sei. Vielmehr sollten wir von unterschiedlichen Gesellschaften sprechen Kulturen auf der ganzen Welt oder sogar verschiedene Kulturen innerhalb derselben *Gemeinschaft* . Es ist unwahrscheinlich, dass Fortschritte in der Robotik und der KI allen Gemeinschaften und allen Nationen gleichermaßen zugute kommen, insbesondere kurz- bis mittelfristig, und die Grundsätze für die Entwicklung der Robotik sollten dies anerkennen.

Es ist auch unklar, auf welche Personen sich das Dokument bezieht; Menschen werden wiederum unterschiedlich als „die Öffentlichkeit“ bezeichnet, als ob es sich um einen einzelnen, homogenen Körper handle. Die Prinzipien selbst machen deutlich, dass es eine Reihe verschiedener Wesen gibt, die spezifische Interessen an der Entwicklung der Robotik haben und auf unterschiedliche Weise mit Robotern interagieren.

- Die Grundsätze beziehen sich auf „Bürger“, als Subjekt einer bestimmten politischen (normalerweise nationalen) Körperschaft, Allerdings ist unklar, ob jemand noch länger den Anspruch erheben kann, „Bürger“ eines diskreten, unabhängigen und von anderen Einflüssen unabhängigen Nationalstaates zu sein. Werden die Vorteile und die Verantwortung für Roboter nur den Bürgern eines bestimmten Nationalstaats oder einer politischen Körperschaft zugute kommen? (Vielleicht ist dies kompliziert, da eine zunehmende Automatisierung zur Einführung eines universellen Grundeinkommens in einem bestimmten Nationalstaat führt, aber nicht anderswo.) Es ist auch interessant, dass die Grundsätze diese (veraltete) Vorstellung des Nationalstaats mit der Möglichkeit aufrechterhalten, dass Roboter als Waffen für nationale Staaten konzipiert werden können Sicherheitsgründe'.
- Die Grundsätze besagen, dass allein Menschen die verantwortlichen Vertreter des Gesetzes sind. Dies mag derzeit unumstritten sein, und man muss sich in die Bereiche der Science-Fiction begeben, um sich vorzustellen, wann wir empfindungsfähige, bewusste Roboter und KI haben könnten, die in den Augen des Gesetzes den Menschen ebenbürtig wären, aber diese Erklärung ignoriert die Gewässer, die bereits durch autonome Systeme wie selbstfahrende Autos getrübt sind, und die Herausforderungen, die sie mit sich bringen Unser humanistisches Rechtssystem. Darüber hinaus fragen wir uns möglicherweise, wie technologisch verbesserte Menschen (z. B. Cyborgs) von den Gesetzgebern als gleichermaßen (weniger? mehr?) verantwortliche Akteure angesehen werden können.
- Die Grundsätze bestehen auf Erwägungen der Privatsphäre, obwohl wir dies bereits sehen können Bei vielen Menschen sind die Grenzen zwischen „Selbst“ und „Öffentlichkeit“ verschwommen, und auch der Begriff der Privatsphäre ist verschwommen hat sich in so kurzer Zeit radikal verändert. Soziale Medien, das Versprechen von „Smart Homes“ und Sicherheitsfragen haben dazu geführt, dass wir kulturell sehr unterschiedliche Vorstellungen davon haben, was „Privatsphäre“ bedeutet und wie wichtig es für uns ist. Dies wird enorme Auswirkungen auf das Gesetz und darauf haben, wie wir unsere Beziehung zur übrigen Welt sehen.
- Die Grundsätze machten eine klare Unterscheidung zwischen denen, die Roboter „entwerfen“, denen, die Roboter verkaufen, und solchen „Verbrauchern“ und „Benutzern“. Die Grundsätze akzeptieren implizit, dass die Interessen dieser Gruppen konkurrieren können. Diese Beziehungen gehen jedoch auch von einer statischen Vorstellung unseres gegenwärtigen Systems sozialer Beziehungen aus, das auf kapitalistischen Produktionsweisen basiert, einem System sozialer Organisation, das viele bereits für veraltet halten würden (vgl. z. B. Paul Mason 2015). Es ist bereits abzusehen, dass diese ehemals scheinbar stabilen sozialen Beziehungen im Zuge der Weiterentwicklung von Robotik und KI zunehmend unter Druck geraten und wahrscheinlich in etwas umgewandelt werden, das den neuen Möglichkeiten zur Produktion von Dingen und effizienteren Arten der Organisation der Gesellschaft angemessener ist (z. B. eine Art Postkapitalismus, wie manche es gerne hätten). Auch wir verwischen bereits die Grenzen zwischen Produzenten und Verkäufern einerseits und Verbrauchern und Nutzern andererseits

Andere (z. B. Uber, Crowd-Sourcing-Daten, Google) müssen diese Kategorien bereits viel flexibler sein, als sie es sich im geradlinigen, vereinfachenden Humanismus vorgestellt haben.

Es ist auch erwähnenswert, dass es sich bei den ESPRC-Robotikprinzipien, wenig überraschend, vielleicht um eine sehr europäische, christliche Erfindung handelt. Diese immaterielle, metaphysische Eigenschaft, die einzigartig für das Leben oder, in den meisten Fällen, speziell für den Menschen einzigartig ist. (Diese Idee, dass es an etwas absolut Menschlichem mangelt, liegt in der Idee des Roboters selbst, als das Wort erstmals 1921 in Karl Capeks Theaterstück „RUR“ in die Welt eingeführt wurde.) Obwohl man argumentieren könnte, dass Europa nicht mehr dem Christentum verpflichtet ist, sind die christlichen Werte Europas (und Amerikas) ständig zu sehen, und diese Annahme ist sogar in zeitgenössischen, völlig spezifischen europäischen rechtlichen und ethischen Rahmenwerken, einschließlich diesen, offensichtlich ESPRC-Prinzipien.

Im Gegensatz dazu ist es erwähnenswert – wie viele (z. B. Metzler und Lewis 2008; Lee, Sung, Šabanović, Han 2012) –, wie unterschiedlich Roboter und KI in unterschiedlichen kulturellen Kontexten wahrgenommen werden. Es gibt „animistische“ Religionen – das heißt, sie glauben, dass alle Dinge, einschließlich unbelebter Objekte, die Natur von Kami oder Geist in sich tragen. Diese unterschiedlichen kulturellen Einflüsse (selbst in heute hochgradig kularisierten Kulturen) können enorme Auswirkungen darauf haben, wie wir unsere Beziehungen zu Robotern und KI definieren wollen. Solche so tief verwurzelten Einflüsse werden durch die Einführung neuer Technologien und Ideen weniger leicht verändert, aber sie unterstreichen, dass die ESPRC-Prinzipien sehr stark in einem sehr spezifischen kulturellen und historischen Kontext angesiedelt sind und wie wir bereit und willens sein müssen, uns andere Ideen und Beziehungen nicht nur in der Zukunft, sondern gerade jetzt vorzustellen, wenn wir versuchen wollen, einen internationalen Konsens über die Prinzipien von zu erreichen Robotik.

Neue Menschen?

Dass die Grundsätze „nicht als feste Gesetze gedacht sind, sondern vielmehr dazu dienen sollen, die Debatte zu informieren und als Referenz für die Zukunft zu dienen“, zeigt die Zukunftsorientierung der Teilnehmer des Workshops. Die Grundsätze sollten jedoch erweitert werden, um eine Bewegung über die eng gefasste Vorstellung vom „Menschen“ hinaus zu ermöglichen, die ihnen in ihrem gegenwärtigen Zustand zugrunde liegt. Gegenwärtig ist das humanistische Subjekt, das diesen Grundsätzen zugrunde liegt, eine feste Grenze dessen, was möglich ist gelebt, weil diese Seite des „Menschen“ alle Beziehungen definiert, die darin vorgestellt werden. Stattdessen ist es notwendig, sich einen anderen, pluralistischeren und flexibleren Menschen an der Basis vorzustellen.

Denker können (und tun es oft bis zum Überdruß) darüber feilschen, wann der Konsens, der das humanistische Subjekt stützte, zusammenbrach, aber es ist klar, dass irgendwann nach dem Zweiten Weltkrieg, mit dem Verlust des Glaubens an Meta-Erzählungen und einer neuen, radikalen Hermeneutik des Verdachts (die manche als „Postmodernismus“ verstanden haben), das stabile humanistische Subjekt, wie es einst verstanden wurde, nicht mehr lange auf dieser Welt war. Aber was kommt als Nächstes – und es ist klar, dass etwas als nächstes kommen muss, denn wir können nicht mit der Konstruktion irgendeiner Art von Rahmen oder Modell fortfahren, ohne eine Vorstellung davon zu haben, was es bedeutet, menschlich zu sein – es gibt viel weniger Übereinstimmung. Es ist auch klar, dass alles, was auch immer dem „Humanismus“ folgt, nicht als einzige Vorstellung von „Mensch“ artikulieren kann, sondern flexibel, anpassungsfähig und viele verschiedene Arten menschlicher Subjekte einbeziehen muss, die alle auf

Als Konsumenten, Menschen als Produzenten, Menschen als Designer, als Rechtssubjekte, als Bürger und Subjekte verschiedener politischer Einheiten ... Jeder Mensch kann zu jeder Zeit schlagen.

Möglicherweise versuchen wir, einige Prinzipien für die Robotik nachzubilden, die auf einem *menschlichen Subjekt basieren, das nach dem Humanismus kommt. Vielleicht möchten wir diesen Menschen „Posthuman“ oder einfach „Posthuman“ nennen.* Allerdings sind diese Begriffe kompliziert und beziehen sich auf eine schwindelerregende Vielfalt unterschiedlicher Ideen und Ideologien (sogar mehr, als unter dem übergeordneten Begriff „Humanismus“ enthalten war). Ohne eine vollständige Zusammenfassung der verschiedenen Ideen zu geben, die unter diesen Bezeichnungen bezeichnet werden können, möchte ich einige Ideen geben, die meiner Meinung nach notwendig und nützlich sein werden, um zukünftige Prinzipien für technologische Innovation voranzutreiben.

Posthumanismus kann philosophisch und kulturell einfach das bedeuten, was nach dem Humanismus kommt; dieser Posthumanismus, manchmal ein Antihumanismus, widerlegt die Art stabiler, singulärer Annahmen über den Menschen und die menschliche Natur, die vom Humanismus aufgestellt werden. Etwas weiter gehend, Posthumanismus akzeptiert die Kontingenz und die Kontexte der Vorstellungen vom Menschen und ersetzt eine statische menschliche Natur durch etwas Dynamischeres und Pluralistischeres. Viele der Konzeptionen des Posthumanismus beinhalten darüber hinaus Überlegungen darüber, wie neue technologische Entwicklungen in die menschliche Erfahrung integriert werden sollen und sowohl den Menschen als auch unsere Welt verändern.

Der Posthumanismus, oder vielleicht genauer gesagt, der Posthumanismus, ist nicht teleologisch; er geht nicht davon aus, dass *der* Mensch, den wir nach Millionen von Jahren der Evolution und Jahrtausenden der Philosophie erhalten haben – wir –, der Mensch ist, ein endgültiges, fertiges, poliertes Produkt, das nun für immer unveränderlich und unveränderlich bleiben wird. Der Posthumanismus erkennt an, dass sich unsere Vorstellungen von uns selbst ändern und transformiert werden, genau wie sie das war schon immer so. Eine große Stärke des Posthumanismus, wie er hier verstanden und artikuliert wird, besteht darin, dass er über eine eingebaute Flexibilität verfügt, sich an solche Veränderungen anzupassen, und dass es wichtig ist, dass jedes ehrgeizige Unterfangen, wie die Festlegung einer Reihe von Prinzipien für die Definition unserer gegenwärtigen und zukünftigen Beziehungen zu einer sich ständig verändernden Technologie weist eine ähnliche eingebaute Flexibilität auf.

Einige, die besonders optimistisch sind, was das baldige Aufkommen einer empfindungsfähigen KI angeht, und die sich selbst als *Transhumanisten bezeichnen könnten*, könnten die ESPRC-Prinzipien als naiv und anthropozentrisch betrachten, weil sie das Aufkommen als Roboter und KI-Agenten in ihrem eigenen Recht nicht erklären, die (vielleicht die gleiche) Berücksichtigung neben Menschen bei der Schaffung ethischer Prinzipien verdienen. Solch ein Argument ich würde mit dem, was ich hier vorantreibe, die Überzeugung teilen, dass die ESPRC-Prinzipien bereits etwas veraltet und in ihrer Vorstellung dessen, was „den Menschen“ ausmacht, bereits etwas zu kurz sind, obwohl ich hinsichtlich der Nähe der empfindungsfähigen KI weitaus weniger optimistisch bin und nicht die allgemeine transhumanistische Gewissheit teile, dass Menschen durch Technologie radikal verändert werden (z. B. Menschen, die ... sind nahezu unsterblich) sind ebenfalls sehr nahe. Es ist jedoch nicht notwendig, dass eine bestimmte, empfindungsfähige KI auftaucht, damit diese posthumanistische Kritik an den ESPRAC-Prinzipien bestehen bleibt. Selbst wenn wir keine neuen Roboter erfinden und keine neuen Fortschritte in der künstlichen Intelligenz machen – was sehr unwahrscheinlich ist – dass wir Menschen die anderen Systeme und Institutionen, die definieren, wer wir tragen, weiterhin neu erfinden werden, wodurch wir Menschen verändern und neue, flexiblere Prinzipien benötigen, um unsere Beziehung zu Robotern und KI zu definieren.

Die Verfasser der Grundsätze beabsichtigen, ein „lebendiges Dokument“ zu sein, keine „festen“ Gesetze, sondern die Grundlage für zukünftige Debatten und Referenzen, die genau das sind, was sie sein müssen. Es ist jedoch schwer vorstellbar, wie Grundsätze für die Robotik die Flexibilität haben können, auf die Herausforderungen zu reagieren, die sich daraus ergeben

unsere neuen Technologien, wenn sie als Ausgangspunkt dienen, z.

Es ist interessant zu sehen, dass in der Präambel der Grundsätze die Allgegenwärtigkeit von Asimov und seinen drei Gesetzen erwähnt wird. Denn auch wenn Asimovs Gesetze zu Recht als unzureichend abgetan werden, weil sie fiktiv sind, sich nicht an das „wirkliche Leben“ richten und nicht in der Praxis angewendet werden können, gibt es in Asimovs Schriften dennoch etwas, das sich das ESPRC als Inspiration hätte nehmen können: die Fähigkeit, sich verschiedene Welten vorzustellen, die mit unterschiedlichen Arten von Menschen bevölkert sind. Der Mensch durchläuft immer Prozesse der Neuerfindung, aber mit den Fortschritten in der Robotik und der KI, die wahrscheinlich unmittelbar bevorstehen, könnten wir darüber spekulieren, dass eine noch radikalere Veränderung in der Art und Weise, wie wir uns selbst sehen und in welcher Beziehung wir zu unseren Technologien stehen, bevorsteht. Eine Ötik, die in der Lage ist, diesen sich ändernden Beziehungen Rechnung zu tragen, und sowohl in der Lage ist, verschiedene Entwicklungsrichtungen einzuschränken als auch sie zu begrenzen. Wenn solche Prinzipien Bestand haben sollen und nicht ein Überbleibsel einer veralteten Sichtweise auf die Welt und uns selbst sein sollen, müssen wir antizipieren, wie sich die Menschen durch neue Entwicklungen in der Robotik und der KI verändern werden. Wir müssen fantasievoll über die Arten von Robotern nachdenken, die wir erschaffen werden, aber *auch über* die Arten von Menschen, die wir werden werden, und wenn wir Prinzipien entwickeln wollen, um die Effizienz unserer Gesellschaften zu steigern, müssen wir besser verstehen, wie diese Gesellschaften und die Menschen, die sie bevölkern, aussehen werden.

Verweise

Lee, Sung, Šabanovič, Han. 2012. Kulturelles Design von Heimrobotern: eine Studie über Benutzererwartungen in Korea und die Vereinigten Staaten.

Mason P. 2015. *Postkapitalismus: ein Leitfaden für unsere Zukunft*. London: Allen Lane.

Metzler und Lewis. 2008. Ethische Ansichten, religiöse Ansichten und Akzeptanz von Roboteranwendungen: ein Pilotstudie. AAAI. 15–22.

Regulierung!Roboter!Städte:!! Überlegungen!zu!den!Grundsätzen!der!Robotik von der neuen!fernen!Seite!des!Gesetzes

Aurora&Voiculescu

Zentrum&für&Recht&&&Theorie,&University&of&Westminster

„Sie&fragten&mich&wo&ich&wählen&zulaufen,&welches&bevorzugt?

Wo&Roboter&Mäuse&und&Männer,&ich&sagte,&in&Roboterstädten&rundlaufen.

Aber&ist&das&weise?&For&tin's&a&fool and&iron&hat&keinen& Gedanken!

Computer&Mäuse&können&mich&Fakten&finden&und&mir&beibringen,&was&ich&nicht bin.

Aber&Roboter&alles&Unmenschliche&ist&alles&Sünde&mit&Rädchen&und&Mesh.

Nicht&wenn&wir&das&Gute&in&lehren&so&kann&es&unser&Fleisch&lehren&

[...]

Wie&der&Mensch&selbst&eine&Mischung&ist,&wild¶dox,

So&müssen&wir&unseren&verrückten&Maschinen beibringen:&aufstehen,&seine&Socken hochziehen!

Come&run&with&me,&wild&children/ men,&half&pires&and&dooms,&half&clowns.

Tempo&Roboter&Mäuse&Rennen&Roboter&Männer&WinQlose&in&Roboterstädte.“

Roy!Bradbury¹

Die!Prinzipien!der!Robotics!-Initiative!entsteht!weitgehend!aus!einer!Reflexion!des!Umfangs!
inwieweit!Roboter!bereits!unser!Leben!beeinflussen!und!in!welchem!noch!größeren!

Ausmaß! Ist! das habe ich erwartet! Sie! Wille! beeinflussen! Es! In! Die! 'Roboter! Städte'
von! Die! verhältnismäßig! nahe! Zukunft. Ob!dielersten! Verordnung! verwandt! Zu! Das!
transformativ! Technologie! wird!die!Form!von!weichen,!leitenden!Grundsätzen,!von!harten!
inländischen!Rechtsinstrumenten!annehmen! oder! sogar! von! Komplex! International!
Verträge! Ist! A! herausfordernd! noch,! bei! Das! Punkt,! A! zweitrangig! Ausgabe.! Der!
primär! Frage! ist eher! a!(legalE)! normativ! Frage,! gezielt! bei! abgrenzen! klar! Grenzen!
von! Die! Mensch/Roboter! KoExistenz;! Adressierung! Die! normativ! Dynamik! von!
Kausalität! Und! Verantwortung;! versuche es! Zu! Identifizieren Sie die! Ort oder! Orte
von! Mensa und! Actus rein! Prozesse! Und,! Wagen! Wir! sagen,! Beziehungen , die!
vielleicht!gut! beweisen! Zu! werde!mehr!und!mehr! komplex!mit! Die! Fortschritte!von!Wissenschaft!und!Tech

Stimmen! aus! Das! brauchen! für! normativ! Selbstbeobachtung! (in! unsere! soziale!
Psyche! viel! mehr! als! alles! andere),! dieses!Papier!ist! Eine Einladung! zum Nachdenken! An! Die!

¹ Strahl! Bradbury,! *Where& Robot& Mice& and& Robot& Men& Run& Round& in& Robot& Towns:& New&Poems,&Both&Light&and&Dark!*(New!York:!Random!House!Inc,!1977).!

2!Aurora! Voiculescu,! "Menschlich! Rechte! Darüber hinaus! Die! Menschlich:! Hermeneutik! Und!
„Normativität!im!Zeitalter!des!Unbekannten“!(erscheint).!

vorgeschlagen! Prinzipien! von! Robotik! (Abdecken! 5! Prinzipien! und! 7! HighELevel!
Mitteilungen)! Kommen! aus! von! Die! multidisziplinär! expertInformiert! EPSRC! Und!
AHRC! Robotik! Rückzug!in! 2010.! Der! Komplexität! von!Problemen!an! Abdeckung ist!
solch! Das! diese! Reflexionen! dürfen! nur! Ziel! Zu! engagieren! mit! Was! Ist! vorgeschlagen!,
mit dem! Text! angeboten! für! Betrachtung,! preislich! aus! manche! von! Die! möglich!
Bedeutungen! oder! Interpretationen!solcher!Texte.!Eine solche!Analyse!wird!als!wesentlich!
für!die!Vorbereitung!vorgeschlagen! Boden für! weiter! Diskussionen,! Und! endlich,!für!
Einschiffen! An! beliebig! irgendwann! regulatorische!Prozesse.!

Prinzipien!auf!der!Suche!nach!einer!Definition!

Das Nachdenken!über!die!bestehenden!Prinzipien!lädt!zu!allererst!dazu,!darüber!zu!
reflektieren,!was ein!Roboter!ist!und!darüber,!ob!die!Definition!dieser!ist !sollte!sich damit
zufrieden geben,³ und!deshalb! Die!Art!von!Entitäten,!die!man!regulieren!sollte,!sollte!eine!
Antwort!auf!unsere! stateOfTheEart!in! Technologie! oder ein! Betrachtung! von! unser!
stateOfTheE(technology!in)! Kunst.!Mit!anderen!Worten,!an!welchem!Punkt!auf!dem!
Spektrum!zwischen!Wissenschaft!und!Wissenschaft! Fiktion! sollen! Wir! Ort! uns selbst!
Wann! Entwerfen! Normen! Und! bewertend! ihre! Effektivität?!Wie!weit!in!die!Zukunft!blicken!
sollte!,!wenn!die!Zukunft!für!welche! wir!regulieren!ist!so! weit! dass!wir!nur!spekulieren!als!
zu!seiner!Existenz,!während!bei! Gleichzeitig!galoppieren!wir!mit!immer größerer!Geschwindigkeit!auf!diese!sehr!

Gesetz/Verordnung,! hart! oder! weich,! erfordert Definitionen. Der! Prinzipien! Hier! unter!
Diskussion! Tun! nicht! sofort! schicken! Zu! eins.! A! Roboter! Ist! definiert! von! manche!
als! 'A! Maschine! fähig! von! Tragen! aus! A! Komplex! Serie! von! Aktionen! automatisch'!
oder,! anders! nuanciert,! 'ein!mechanischer! oder! virtueller!künstlicher!Agent,!
normalerweise!ein!elektromechanischer! Maschine! Das! Ist! geführt! von! A! Computer!
Programm! oder! elektronisch! Schaltung'.⁴Verschiedene,! mehr! oder! weniger! praktikabel!
Unterscheidungen! Sind! Auch! setzen! nach vorne,! mehr! vor allem! zwischen! industriell!
Und! Service! Maschinen,! zwischen! höchst! autonome!Maschinen!und!kognitive!Computer!
Programme,!zwischen!verkörpert! und!körperlose!Kognitionswesen!usw.!Die NASA!selbst!
verwendet!ein!ziemlich!alltägliche!und! ungenau! Sprache,! sehr! nicht hilfreich! für! Die!
Regler,! definierend! Roboter! als! „Maschinen,!die!verwendet!werden,!um!Arbeiten!zu
erledigen“.!Einige!Roboter,!die!Formulierung!der NASA!geht!weiter! Zu! hinzufügen,! 'dürfen! Tun! arbeiten! von!

³A! Definition! niemals! Sein! valueEneutral,! stets! Festlegung! Die! 'in'Es! Und! Die! 'out'Es!
folgt! A! mehr! oder! weniger! erklärt! valueEladen! Weg.! (Sehen Sie! Alan! Norrie! in!
Voiculescu!2000)

⁴MerriamEWebster! Wörterbuch,! "Definition! von! „Roboter“, abgerufen! Februar! 10,! 2016,!
<http://www.merriamEwebster.com/dictionary/robot!entry:!robot>.

erzählen! ihnen! Was! Zu! do'.⁵! So! A! Vielfalt! von! Formulierungen! erstellen! A!
regulatorisch! Rätsel! und! macht! alle! normativen! Aussagen! schwer! zu! folgen! und/oder!
leicht! zu! Flucht! Einhaltung!!!

Während! stimme zu! Das! Dort! Ist! NEIN! vereinbart! Definition! *per se*,⁶! einige!
setzen! her! A! Anzahl! von! Funktionen, die! ein! Roboter! haben würde, Funktionen, die,
von! einer! Regulierung! (und! nicht! nur!)! Perspektive, Sind! sich! In! brauchen! von!
Definitionen:! das *spüren*! Umfeld! (mit! eingebautem! „Bewusstsein“! seiner!
Umgebung);! *Bewegung*,! ob! Rollen,! Gehen,! stoßend,! oder! vielleicht! auch! nur!
Daten! conveying; Energie,! in der Lage! sich selbst! auf! Art! selbst! antreiben! das! wird!
davon! abhängen,! was! der Zweck von! Die! Roboter! list;! Intelligenz: von! seinen!
Programmierer! mit! „Intelligenz“! ausgestattet! werden! und! die! Kapazität! Zu! auswerten!
Umfeld,! Umstände,! Komplex! Information.! [(Je! mehr! eine! Maschine! fähig! ist,! unabhängig!
mit! einer! dynamischen! Welt! zu! interagieren!,! je! fortgeschrittener! diese! Maschine! dieser!
Roboter! list, unter! anderen! genau! dazu)]!

Also,! A! Roboter! Ist! definiert! mehr! speziell! als! A! System,! A! Maschine! das
beinhaltet! Sensoren,! Steuerungssysteme,! Manipulatoren,! Stromversorgungen! und!
Software! funktionieren! alles! zusammen! Zu! ausführen! A! Aufgabe".! Nach! Zu! solch!
A! Perspektive,! "[Entwerfen,! Gebäude,! Programmierung! Und! testen! A! Roboter! Ist!
A! Kombination! von! Physik,! mechanisch! Maschinenbau,! elektrisch! Maschinenbau,!
strukturell! Maschinenbau,! Mathematik! und! Computer.! In! manche! Fälle! Biologie,!
Medizin,! Chemie! könnte! Auch! sein! beteiligt sein".! Wenn! Die! student! in! Robotik! Mai!
aktiv! engagieren! mit! alle! diese! Disziplinen! "In! A! tief! problem! Eposing! Problemlösung!
Umwelt",⁷! einige! könnte! richtig! sagen,! Das! regulierend! Roboter! Und! 'Roboter!
Städte'! erfordert! A! ähnlich! komplexe! interdisziplinäre! Beschäftigung! mit! den! meisten!
wenn! nicht! alle! von! diese! Felder.#! Für! Die! normativ! Diskurs! (ob! hart! regulatorisch!
oder soft! principled),! Die! Tatsache! dass! viele! dieser! Definitionen! eine! Zahl! von! Punkten! gemeinsam! habe
A! Definition! Das! Ist! ausreichend! präzise,! noch! dynamisch! genug! Zu! erfassen!
Die! Wesen! von! Die! soziotechnologisch! Phänomene! ist! daher! erforderlich! für!
Öffnung! die! Robotik! Prinzipien! Zu! Weiterentwicklung! und! Problematisierung. Dieser!
Bedarf! bezieht! sich! auf! Perspektiven! wie! die! von! Andra! Keay,! die! über! Roboter! als!
spricht! „...! eine! Umgebung; zu! groß!, als dass! wir! als! eine! Artikel".! Obwohl! unweigerlich!
mit! dem! geht voran! von! Technologie! E! "[was wir! Anruf! A! Roboter! heute! ist! mehr! anspruchsvoll

⁵ NASA,! "Was! Ist! Robotik?,"! *Die NASA weiß es!* Mai! 18,! 2015,! http://www.nasa.gov/audience/forstudents/ke4/stories/nasaEknow/what_is_robotics_k4.html!

⁶ H.! James! Wilson,! "Was! Ist! A! Roboter,! *Wie auch immer?*",! *Harvard Business Review*,!
April! 15,! 2015,! <https://hbr.org/2015/04/whatEisEaErobotEanyway>.

⁷ Ebenda.

als! was! wir! in! den! 80ern! einen! Roboter! genannt haben!“! sagt! Keay! „Es! ist! auch! wahr! dass! es! mehr! ist! als! das!., „Es! war! schon! ein! Identitätsproblem!“! sagt! Keay.⁸!

Es! sollen! Sei! sagte, jedoch, dass Identitäten! und! Klassifizierungen! habe immer! gewesen! problematisch! Und! problematisiert! Wann! Teil! von! regulatorisch! Initiativen,! ob! diese! hatte! Zu! Tun! mit! Menschen! oder! Nichtmenschen! wie.! Gesetz,! In! besondere,! stets! am Ende! verwandelt! jede! Identität! in! eine! juristische Fiktion,! die! oft! sehr! wenig! zu! tun! hat! mit! jeder! anderen! physischen! oder! wissenschaftlichen! Dimension! dieser! Entität. ⁹ Gleichzeitig! Gesetz! Elim! weitesten! Sinn! von! sozial! unterstützten! normativen! Imperativen! E! hat! immer! lebte! auf! Definitionen.! Das! Fehlen! einer! „Arbeitsdefinition“! eines! Roboter erscheint! daher! beide! als! Zeuge! Zu! die! Herausforderungen! des! Festhaltens! Technologie! in! seiner! Eile,! und! als! Reflexion! einer! möglichen! Schwäche,! die! in! der! vorgeschlagenen! dokumentieren.!!

Zu guter Letzt! und! zusätzlich! zum! Problem! von! das! Fehlen! einer! vereinbarten! Arbeit! Definition! (die! natürlich bald! umstritten! und! problematisiert! sein würde),! da! ist auch! die! Bestätigung! Das! 'Definitionen! sind! niemals! neutral'.! Diese! Idee! war! fortschrittlich! manche! Jahrzehnte! vor! von! Larry! Mai! Wann! reflektieren! An! definierend! Die! Verantwortung! von! nichtmenschlicher! kollektiver! Agentur! (eine ansonsten nicht! irrelevante! rechtliche! Neuerung).! Definitionen,! Fortschritte! Können! „Pseudogemeinschaften“! schaffen,! die! vorgeschlagen! werden! als! Fakten,! während! in! wirklich! Sie! Oppositionen! aufstellen,! die! willkürlich! diese! trennen! wer! eingeschlossen! ist! und! diejenigen,! die! ausgeschlossen! sind! aus! einer! gemeinsamen! Konzeptualisierung! oder! üben'. ¹⁰ Das! Behauptung! Wille! werden! mehr! Und! mehr! offensichtlich! einmal! Die! Spektrum! der! verfügbaren! Optionen! zwischen! Roboter! und!

Ob! erwartet! mit! Angst! oder! mit! Aufregung,! die! Herausforderung! der! Regulierung! Die! mehrfachen! Dimensionen! der! menschlichen! Erobot! Interaktionen! sind! mehrfach.! Eine! Zahl! von! Probleme! setzen! nach vorne! für! Betrachtung! In! Beziehung! Zu! Die! gegeben! fünf! Prinzipien! Sind! hier! kurz! erwähnt:!

⁸! Siehe! Interview! mit! Andra! Keay,! Gründer! von! Roboter! Startrampe! Und! Verwaltung! Direktor! von! Silizium! Schlucht! Robotik! in! Signe! Brewster,! "Was! Ist! A! Roboter?! Der! Antwort! 2014,! <http://gigaom.com/2014/07/05/what-is-a-robot-the-answer-is-constantly-evolving/>! (Hervorhebung! des Autors).!

⁹ David! Fagundes,! "Beachten Sie,! Worüber! wir! reden! Wann! wir! über! Personen! reden: Die! Language! of! a! Legal! Fiction",! *Harvard Law Review* 114,! no. 6! (2001):! 1745–68.

¹⁰ Larry! Mai,! *Teilen und Verantwortung!* Neu! Auflage! 1996! (Chicago:! Universität! Von! Chicago! Press,! 1992),! 171ff.

¹¹ Kenneth! Grady,! "Künstlich! Intelligenz: Sei! Besorgt,! Sei! Sehr,! Sehr! Besorgt! (Oder! Nicht),!"! *SeytLines: & Veränderung & der Praxis & des Rechts*,! Dezember! 31,! 2014,! <http://www.seytlines.com/2014/12/artificial-intelligence-ebe-eafraid-ebe-every-e-very-eafraid-eor-enot/>.

Zuallererst besteht ein Bedarf an einer etwas klareren Perspektive auf das, was regulierend ist : Prinzipien! setzen! nach vorne! denn! die! Diskussion! zeigt! das! Potenzial! der! Verwirrung! hinsichtlich! der! tatsächlich! Agenda! 'regulierend! Roboter! In! Die! real! Welt! ' hat! A! doppelt! Sinn,! voll! von! Fallstricke! Wenn! man! sich selbst „reguliert! Roboter“,! bringt! der! Text! eine! implizite! Agentur! ein! dass! in! Kontexten,! wo! es! unerwünscht! sein! kann! als! selbstverständlich! angenommen! werden! (obwohl! dieser! Interpretation! eindeutig! einige! der! Prinzipien! insbesondere! von! Prinzip! NEIN.! 2! Und! 5)! Der! zweite! Bedeutung,! mehr! In! Linie! mit! Was! Die! fünf! Prinzipien! sich! aufdecken,! könnte! Ziel! bei! 'regulierend! Die! Schaffung! Und! verwenden! von! Roboter'. Die! Wahl! dieser! Interpretation! sollte! durchgehend! expliziter! sein! Formulierungen,! Vermeidung! regulatorischer! Verwirrung.!

Prinzip! Nr.1: !Roboter&sind&Multifunktionswerkzeuge.&Roboter&sollten&nicht&ausschließlich&oder&in erster

Linie&zum&Töten&oder&Schaden&von&Menschen&entworfen&sein,&außer&im&Interesse&der&nationalen

Der! erster! Teil! von! dieses! Prinzip! ist! jedoch! noch! rätselhafter.! Zunächst! ist! ein! Mai! finden! Die! beginnend! Stellungnahme! „Roboter sind vielfältige Werkzeuge“! als! virtuell! A! Einschränkung,! die! keinen! tatsächlichen! Zweck! erfüllt.! Es! ist! unklar,! warum! ein! Roboter! sein! 'multiEuse',! um! sicher! zu! sein! oder,! umgekehrt! auf! welche! Art! eine! ansonsten! tödliche! Roboter! kann! weniger! tödlich! werden,! wenn! als! „MultiEuse“! konzipiert! wird.! Dies! betrifft! Zu! Die! nächster! Teil! des! Prinzips: „Roboter“! sollen! &nicht! &ausschließlich! &oder! &hauptsächlich! &dazu! &entworfen! &werden,! & Menschen! &zu töten oder zu schädigen.! Um! „Killerroboter“! in! Übereinstimmung! mit! dem! Buchstaben! von! Wenn! dieser! Teil! des! Prinzips! ist,! würde! es! ausreichen,! den! „Killerrobotern“! auch! beizubringen,! Pfannkuchen! oder! wollige! Socken! stricken.! Das! ist! was,! in! die! rechtliche! normative! Perspektive,! man! würde! eine! „kreative! Ecompliance“-Lücke“! nennen. Um! eine! solche! Lücke! zu! identifizieren! und! zu! nutzen! muss! ein! juristisches! Auge! nicht! weiter! blicken! als! auf! eine! wörtliche! Auslegung! von! Die! Text.! Trotzdem! Die! wörtlich! Deutung! Ist! eins! von! Die! primär! Regeln! von! Auslegung! im! Recht,! wenn! die! Auslegung! im! Übereinstimmung! mit! dem! Geist! der! Regel! ist! möglicherweise! nicht! bequem! .! Die! Erläuterungen! zu! diesem! Prinzip! im! 2010! Original! dokumentieren! Tun!

Der! Kommentar! Zu! Das! Prinzip! erscheint! Zu! implizieren! andere! Potenzial! Tücken! für!
normative! Argumentation. !Zuallererst! wie! erwähnt! im Hinblick! auf! 'MultiEuse! Tools', ! Es! bemüht!
sich! die! Idee, ! dass! Roboter! wie! alle! anderen! Werkzeuge! sind, ! zu! verbreiten! Um! mit! dieser! Logik! zu!
verfolgen, ! werden! Äquivalenzen! um! jeden! Preis! gesucht. ! Der! Vergleich! einer! Roboter! mit! einem!
Messer! oder! einer! Waffe, ! für! verschiedene! verwendete, ! sowohl! relativ! gutartige! und! kriminelle!
Zwecke, ! tut! nicht! Abdeckung! für! Die! Inkonsistenz! Das! Dort! Sind! Werkzeuge, ! einschließlich!
Waffen, ! für! die! man! sich! keinen! anderen! Zweck! denken! als! den! prima facie! einen. !!!

Prinzip! NEIN. ! 2: ! *Menschen, nicht Roboter, sind dafür verantwortliche Agenten. Roboter sollten so
konzipiert und betrieben werden, dass sie die bestehenden Gesetze und Grundrechte und Freiheiten,
einschließlich der Privatsphäre, einhalten.* Der! Kommentare! Zu! Das! Prinzip! scheinen! mehr!
Verwirrung! als! Klarheit! zu! schaffen. ! Zunächst! ein! relativ! kleines! Gegenstand, ! Es! gibt! die!
Vermutung! dass! wahrscheinlich! niemand! absichtlich! aufbrechen! wird! Zu! Bau! ein! Roboter! der!
kaputt! geht! das! Gesetz'. ! Das! setzt! her! eine! Vermutung! das! hat! nein! Grundlage! in! der! realen! Welt!
der! Abweichung! und! Trotz!, ! wie! von! socio! legal! Studien! unter! der! breiten! Bevölkerung! sowie! unter!
den! Angestellten. ! 12 Zweitens, ! und! was! noch! wichtiger! ist, ! wird! die! Verantwortung! durch ! Das!
Kommentare! erscheint! Zu! ignorieren! beide! Die! Weg! 'Gesetz! denkt' ! 13 als! Also! als! Die! Weg!
Roboter! können! nicht! Ziele! und! Wünsche! erreichen, ! die! Menschen! vorgeben. ! 14

Als! zusätzliches! Element! soll! auch! erwähnt! werden, ! dass, ! bei! Fehlen! von! eine! klare! funktionierende!
Definition, ! KI, ! „lernende! Roboter“ ! usw. ! sind! sich! alle! mit! diesen! beschäftigen! Prinzipien! Und! ihre!
Parameter. ! Ihre! Mechanik, ! jedoch! kann! Also! sei! mehr! komplex! als! der! gesetzliche! normative!
Diskurs! behandeln! kann, ! ohne! eine! klare! Definition#. ! Roboter! Und! KI! Maschinen! Mai! Also!
lernen! Zu! handeln! mit! 'Ausnahmen'! bevor! das! Recht! lernt, ! mit! „Unterschieden“ ! um! zugehen. ! 15!
Ebenso! scheinen! andere! Disziplinen! Zu! angeben! Das! Zahlen! (in! diesem! speziellen! Fall!
'Programmierung')! Mai! Also! Sei! mehr! als! Nur! Das, ! Zahlen! Sein! unveräußerlich! ergänzt! von!
assoziiert!

12! Einige! nützlich! obwohl! lose! Beispiele! Ryan! Mathews! Und! Watt! Wacker, ! *Der Vorteil von
Deviant: So erstellen Sie Rand- und Massenmärkte!* (Beliebiges Haus, !
2010); ! Kelly! Fischer, ! "Der! Psychologie! von! Betrug: ! Was! Motiviert! Betrüger! Zu!
Begehen! Verbrechen?" ! (Sozial! Wissenschaft! Forschung! Netzwerk, ! 31. März! 2015), ! [http://
papers.ssrn.com/abstract=2596825](http://papers.ssrn.com/abstract=2596825). !

13! Günther! Teubner, ! "Wie! Die! Gesetz! Denkt: ! Zu! A! Konstruktivist!
Epistemology! of! Law“, ! *Law & Society & Review* 23, ! no. ! 5! (1989): ! 727–57.

14 ! Siehe! für! Beispiel! Die! Orthogonalität! Theorie! In! Nick! Bostrom, ! "Der!
Superintelligent! Wille: ! Motivation! Und! Instrumental! Rationalität! In! Fortschrittlich!
Künstlich! Agenten“ ! *Köpfe & Und & Maschinen*, ! 2012, ! [http://
www.nickbostrom.com/superintelligentwill.pdf](http://www.nickbostrom.com/superintelligentwill.pdf). !

15! Dies! basiert! auf! den! Ideen! von! Identität! und! Differenz, ! von! Leibniz! bis! Kant...!
Gilles & Deleuze & Q & Le & Point & de & Vue & (Le & Pli, & Leibniz & et & Le & Baroque) & 1986 & FRA & Sub & ITA, !
2012, ! http://www.youtube.com/watch?v=2ZrA_7ewQGs&feature=youtube_gdata_play
ähm.

mit einem! *Erzählung!* Bewegung! Das,! eins! könnte! sagen! Hier,! Mai! Sei! konstruiert!
anders!durch!die!Maschine!als!durch!den!Menschen,!und!kann!von!ihrer!Konstruktion!konstruiert!werden.16!!!

Prinzip! NEIN.! 3: *Roboter sind Produkte. Sie sollten mithilfe von Prozessen entwickelt werden, die ihre Sicherheit gewährleisten.* Diese!Grundsätze!wirft!Unterscheidungsfragen!auf! In! Die!
legal! Selbstverteidigung! Debatte,! unerlaubte Handlung! Probleme,! Assimilation! von!
Verantwortung,! usw.,! Probleme! Das! abhängen! sehr! viel! von! Die! Kontext! Und! von! Die!
Ausmaß! Zu! welche,! als! Wie!oben!erwähnt,!wird!die!Regulierungsbehörde!bereit!sein,!
Äquivalenzen!zwischen!zu!konstruieren! Roboter!und! andere! Typen! von! Werkzeuge! oder!
zwischen! Roboter!und! andere! Typen! von! Eigentum! Artikel!.

Prinzip! NEIN.! 4: *Roboter werden als Artefakte hergestellt . Kommunikation!mit!sozialen!*
Maschinen!ist!einer!der!Wege,!die!in der! wissenschaftlich! community.17!Während! Die! Ziel!
Ist! nicht! eins! von! täuschen,! Erstellen! der Kontext! für! diese!Art!der!Kommunikation! Zu!
statt!wird!viele!regulatorische! Probleme.!Die!Grenze!zwischen!einer!listigen!Illusion!und!einer!
trügerischen!Schnittstelle!wird! leicht! Sei! gefüllt! von! Die! menschlich! Emotionen.! Teil!
von! Was! Wir! Sind! nach! In! Erstellen! „Serviceroboter“,!die!unseren!Kindern!
Gutenachtgeschichten!vorlesen,!sich!um!unsere!älteren!Behinderten!betreuen! Vater,! Ist!
emotional! Engagement.! Der! wissenschaftlich! Beweis! Ist! Dort! zeigt! Das! Erstellen! Die!
Bedingungen! für! emotional! Engagement! Mai! Also! transzendieren!die!„hinter!dem!
Schleier“-Erscheinung,!die!wohl!wohl!bleiben!tatsächlich!die!eine! von!einer!Dose!Von!hier!
bis!zur!Marketingmanipulation!des!emotionalen!Engagements! verbleiben!nur noch!einige!
Schritte,!die!das!regulatorische!Umfeld!als!herausfordernd! empfinden! begleiten.!Die!mit!
diesem!Prinzip!veröffentlichen!Kommentare!weisen!auf!weitere!Begrüßung!hin! Betrachtung!.

Prinzip! NEIN.! 5: *Die Person, die die rechtliche Verantwortung für einen Roboter trägt, sollte zugeordnet werden.* Dieses!Prinzip!erscheint!klar!hinreichend!in!dem!Kontext,!in!dem!wir!ein!
sind! weit!von!einer!KI,!in!der!technischen!Erfüllung!der!Anforderungen!für!eine!solche!
Verantwortung.!Bei! das gleiche! Zeit,! unter Berücksichtigung!der!komplexen!Elemente!und!
Felder! Das! eingeben! Die! Make-up! von! A! (siehe! oben!, wenn! Fragen! im Zusammenhang!
mit!der!Definition! diskutiert werden),!Herausforderungen!an!den!regulatorischen!Ansatz!
bleiben!von! Die! Punkt! von! Sicht! von! preislich! auseinander! Die! verschieden! Grad! von!
Verantwortung,! Wann! Dinge! gehen! falsch.! Haben! A! 'Eingetragen! Hüter',! Träger! von!
Verantwortung,! Ist! nur ein Teil! von! Die! Lösung.! Der! Verantwortung! Lager! juristische Person! Wille! erfordern

16 Marcus! du! Sautoy,! *Erzählung& und& Beweis:& Zwei& Seiten& der& Gleichen& Gleichung?*
& |& FACKEL! (TORCH,! Das! Oxford! Research! Centre! in! den! Geisteswissenschaften,!
2015),! <http://www.torch.ox.ac.uk/narrativeEandEproofEtwoEsidesEsameEequationE0>.
17 Sehen Sie! zum Beispiel!Cynthia!Breazeal,!„Emotionen!und!gesellige!humanoide!Roboter“!
International&Journal&of&Human&Computer&Studies 59,!Nr.!1–2! (Juli!2003):!129ff,!
doi:10.1016/S1071E5819(03)00018E1.

Betrachtung, als Wille! Die! Typ! von! Schaden(s)! Das! Mai! Sei! zugeschrieben! Zu! solch!
Entitäten, unter! Berücksichtigung, als! manche! vorschlagen, Das! für! Die! Erste! Zeit! 'Die!
Promiskuität! von! „Daten“! werden! in! letzter! Zeit! mit! der! Fähigkeit! kombiniert, körperlichen! Schaden! anzurichten“.^{18!}

Statt! Schlussfolgerungen!

Wenn man diese! Notizen! mit! einem! weiteren! von! Bradburys! poetischen! Bildern! schließt, kann man!
sagen, was! Roboter! mäuse! betrifft, ist! ziemlich! wahrscheinlich! dass! im! Kontext! von! Marktmacht!
und! wissenschaftlicher! Fähigkeit! Deregulierung! zusammen! werden! wir! die! meisten! sicherlich!
Erste! 'springen! aus! Die! Cliff! Und! bauen! unser! Flügel! An! Die! Weg! runter'! In! solch!
Umstände, was! man! tun! list, sicherstellen, dass! relativ! vorbereitet! list! Das! springen! von! bereits!
verhört! Und! problematisierend! unser! Beziehung! mit! Wissenschaft! Und! Technologie! Und!
von! intensivierend! unser! Betrachtung! An! Leben! In! Roboter! Städte! Die! Adressierung!
von! 'Werkzeugen', 'Produkten', 'Artefakten'! und! 'Agenten', muss! man! in! befassen!
St! Augustinus! Betrachtung! An! Die! kompliziert! Verbindung! zwischen! Sprache! Und! Deutung!
als! A! Weg! Zu! Aufschlussreich! A! Tiefer, existenziell! Ebene! von! selfQ -Verständnis! Die! Art!
wir! denken! normativ! über! *humanErobot! interaction*¹⁹ wird! sagen! als! viel! um! Die! Roboter!
als! um! Die! menschlich! Zu! ausleihen! aus!
Ginabattista! Vicos! 1725! *Neu & Wissenschaft*, 20! wir! brauchen! Zu! tragen! In! Geist! Das! unser!
nachdenken! über! Roboter! list! verwurzelt! in! einer! Gegebenheit! kulturell! Kontext! Das! heisst! das!
in! reflektieren! um! Die! normativ! Parameter! von! Roboter! Städte, Die! Sozial! Wissenschaftler!
Wille! nicht! handeln! mit! A! Feld! von! idealisiert! Und! vermeintlich! 'subjectUnabhängig! Objekte',
sondern! wird! eine! Welt, erforschen, die! grundsätzlich! ihre! eigene! list! Der! Prozess! von! regulierend!
Roboter! Ist, daher, A! Verfahren! von! Selbstverständnis, verwurzelt! In! A! angesichts! des!
historischen! Kontextes! und! der! Praxis! Ein! Verständnis, das! nicht! den! automatisch! in! ordentlichen,
normativen, gesetzähnlichen! Sätzen!

¹⁸ Ryan! Calo, „Robotik! und! die! Lektionen! aus! Cyberlaw“, *California & Law & Review* 103! (2015): 513.

¹⁹ Sei! Das! Interaktion! verstanden! An! Latours! Spektrum! von! Fakten! Und! Agentur;
Bruno! Latour, "Wie! Zu! Sprechen! Um! Die! Körper?! Der! Normativ! Abmessungen! von!
Wissenschaft! Studien," *Körper & Gesellschaft* 10, NEIN! 2–3! (1. Juni! 2004): 205–29,
doi:10.1177/1357034X04042943; Bruno! Latour, "Body, Cyborgs! and! the! Politics! von! *Inkarnation*“,!
in! The! Body: & Darwin & College & Lectures, ed. Sean! Sweeney! und! Ian!
Hodder, 2002, 127–41.

²⁰ Giabattista! Vico, *Die neue Wissenschaft von Giambattista und Vico*! (Cornell Universität!
Drücken Sie, 1744).

Kommentar zu Verantwortung, Produktdesign und Sicherheitsvorstellungen

Paula Boddington, Fachbereich Informatik, University of Oxford.

Die Kommentare hier beziehen sich hauptsächlich auf Regel 2 und Regel 3. Ich schlage vor, dass die Prinzipien konkreter auf den Kontext der Implementierung von Robotern zugeschnitten werden könnten und dass Schlüsselbegriffe wie „Sicherheit“ ausgearbeitet und möglicherweise erweitert werden könnten, oder dass die genaue Art und Weise, wie der Begriff verwendet wird, klarer gemacht werden könnte.

Einige dieser Punkte werden durch eine allgemeine Betrachtung des Einsatzes von Robotern in der Krankenpflege/älteren und sozialen Pflege veranschaulicht.

Regel 2: Menschen, keine Roboter, sind verantwortliche Akteure. Roboter sollten so konzipiert und betrieben werden, wie es praktisch möglich ist, um bestehende Gesetze, Grundrechte und Freiheiten, einschließlich der Privatsphäre, einzuhalten.

Wenn Menschen verantwortliche Agenten sind, Roboter jedoch nicht, bedeutet dies, dass dort, wo Roboter als Ersatz für Menschen oder als Teil menschlicher Handlungsfähigkeit eingesetzt werden, die Verantwortungszuschreibungen, die früher dem menschlichen Agenten oder menschlichen Handeln zugeteilt wurden, entweder in ein umfassenderes System verschoben oder möglicherweise übersehen werden. Die Verschiebung kann zu Verschiebungen im Verständnis der Verantwortlichkeiten und Verantwortlichkeiten führen, und diese können unerwartet und komplex sein.

Roboter werden in einem System menschlicher Agenten und Verhaltensweisen eingesetzt. Solche Systeme können mit klar ausgedrückten Vorstellungen von Verantwortung und Rechenschaftspflicht formalisiert werden, zum Beispiel in einem Krankenhaus (obwohl es Elemente solcher Systeme geben kann, die nicht vollständig verstanden oder vollständig formalisiert sind); oder sie können formell sein, zum Beispiel in einer häuslichen Pflegeumgebung. Die Sozialforschung kommt zu dem Schluss, dass es möglicherweise starke lokale Kulturen und Werte in Bezug auf Verantwortungs- und Rechenschaftspflichten gibt.

Ein Beispiel dafür, wie Verantwortlichkeiten und Verantwortlichkeiten verdrängt werden können: Wenn ein Roboter einige der Rollen eines Gesundheitsassistenten in einem berufsbegleitenden Umfeld übernimmt, dann können Verantwortlichkeiten auf vielfältige Weise an verschiedene Akteure innerhalb des Systems des Gesundheitsmanagements verlagert werden. Ein Versagen der Gewissenhaftigkeit eines Mitarbeiters könnte sich beispielsweise darin zeigen, dass er Schwierigkeiten hat, Maschinen zu verstehen oder zu bedienen. Dies kann weitreichende Auswirkungen haben.

Es kann komplex sein, solche Verantwortungs- und Rechenschaftspflichten zu verfolgen und zu verstehen. Es stellt sich die Frage, ob dies ausschließlich die Aufgabe derjenigen ist, die für die Umgebung verantwortlich sind, in der Roboter eingesetzt werden, oder ob die Designer von Heldenrobotern möglicherweise eine gewisse Verantwortung bei der Unterstützung derjenigen haben, die mit den Robotern zusammenarbeiten, um diese Probleme zu verstehen.

In Regel 2 geht es um die Einhaltung bestehender Gesetze und Grundrechte und Grundfreiheiten, einschließlich der Privatsphäre. Darüber hinaus wird es jedoch in bestimmten Einstellungen spezifischere und lokalere Protokolle und Praktiken geben, die von Robotern eingehalten werden sollten. Es kann sinnvoll sein, dies in den Regeln explizit anzugeben.

Beispielsweise gibt es im NHS Versorgungsstandards, die darauf abzielen, eine personenzentrierte Versorgung zu bieten und Patienten mit Würde zu behandeln. Solche Standardaussagen sind für die Bereitstellung einer guten Gesundheitsversorgung von zentraler Bedeutung.

Jemand als Person mit Würde. Herauszufinden, wie sich der Einsatz von Robotern auf solche reichen und kontextualisierten Werte auswirkt, kann sehr wichtig sein, aber möglicherweise eine schwierigere Frage sein als nur die der Einhaltung von Gesetzen und den im Gesetz festgelegten Verantwortlichkeiten. Es sei denn, man geht davon aus, dass dies allgemein verstanden wird.

Regel 3: Roboter sind Produkte. Sie sollten mithilfe von Prozessen entwickelt werden, die ihre Sicherheit gewährleisten.

Möglicherweise muss klargestellt werden, was in diesem Zusammenhang mit „Sicherheit“ gemeint ist. Bezieht sich dies lediglich auf physische Sicherheit? Einstellung?

Diese Regel legt fest, dass Produkte sicher sein müssen. Ethische Regeln werden zwar oft im Hinblick auf die Vermeidung von Schäden formuliert, es handelt sich jedoch weniger um den Anspruch, Produkte zu entwerfen, die lediglich „sicher“ sind. Gutes Design, das wichtige menschliche Bedürfnisse erfüllt, geht über bloße Sicherheitsvorstellungen hinaus, es sollte jedoch darauf geachtet werden, dass sie für ihren Zweck geeignet sind.

Wo Roboter die menschliche Handlungsfähigkeit ersetzen oder erweitern, ist die Aufgabe, die der Roboter übernimmt, möglicherweise nicht vollständig transparent. Insbesondere in einer Dienstleistungsumgebung kann es mehrere bedeutsame menschliche Transaktionen in einer einfachen Aufgabe geben, wie etwa die Kommunikation von Fürsorge, die durch den routinemäßigen Gebrauch von Sprache, Chat und Körpersprache erfolgt. Aktion kann bereitgestellt werden. Umgekehrt können Roboter so konzipiert sein, dass sie einige dieser Aspekte einer Aufgabe übernehmen. Allerdings verstößt dies möglicherweise gegen Regel 4, die besagt, dass „Roboter nicht auf betrügerische Weise entwickelt werden sollten, um gefährdete Benutzer auszunutzen“.

Es treten dann Sicherheitsprobleme auf, weil die Entdeckung, wie der Einsatz von Robotern bestimmte möglicherweise verborgene Aspekte von Aufgaben, die Roboter übernehmen, stören oder möglicherweise sogar verbessern kann, eine erhebliche Analyse und Forschung erfordern kann. Insbesondere in einem anspruchsvollen Umfeld werden scheinbar „alltägliche“ Pflegeaufgaben, die für die Gesundheit und das Wohlbefinden der Patienten dennoch äußerst wichtig sein können, häufig von Mitarbeitern ausgeführt, die auf niedrigeren Ebenen arbeiten. Daher besteht die Möglichkeit, dass solch wichtige Arbeit nicht anerkannt oder anerkannt wird. Schutzzauber bilden äußerst komplexe Systeme der sozialen Interaktion, und die Entdeckung, wie Roboter in solche Umgebungen passen, erfordert möglicherweise eine äußerst sorgfältige Analyse unter Verwendung einer Vielzahl von Fachkenntnissen. Diese Regel 3 kann so verstanden werden, dass sie solche Arbeit impliziert und impliziert, dass Sicherheit auf diese Weise verstanden wird; aber es könnte eine Überlegung wert sein. Wir weisen ausdrücklich darauf hin, dass die Sicherheit auch im Ausland berücksichtigt werden muss. Dies wirft dann auch erneut die Frage der gemeinsamen Verantwortlichkeiten zwischen den Teams auf, die Roboter entwerfen und herstellen, und denen, die mit ihnen arbeiten werden. Es sieht vielleicht nicht so aus, als würden bestimmte Routineaspekte der Fürsorge, die Menschen leisten können, vermisst, Roboter vielleicht nicht, es ist eine Frage der „Sicherheit“, aber diese können Auswirkungen auf die Genesung und Gesundheit haben und sollten daher als ein Problem der Sicherheit angesehen werden.

Die Beachtung der Sicherheit umfasst natürlich auch die Betrachtung von Sicherheitsaspekten beim Einsatz von Robotern innerhalb eines größeren Systems. Ein häufiges und ernstes Problem im Krankenhausbereich für ältere und gefährdete Patienten ist beispielsweise die Dehydrierung. Dies kann schwerwiegende Auswirkungen auf die Gesundheit haben und beispielsweise

Verwirrung. Manchmal wird die Dehydrierung durch Schwierigkeiten beim Erreichen und Verwalten von Getränken verschlimmert.

Nehmen wir an, dass ein Robotersystem entwickelt werden könnte, um solche Patienten beim Trinken zu unterstützen.

Dieses System könnte mit absoluter Sicherheit und Zuverlässigkeit im Hinblick auf seinen unmittelbaren Einsatz arbeiten, zum

Beispiel niemals in einer Weise funktionieren, die dem Patienten zu schnell zu viel gibt, und niemals Getränke verschütten oder den Patienten schlagen.

Allerdings könnte ein solches System in einem bestimmten Kontext möglicherweise große negative Folgen haben. Eine

erhöhte Flüssigkeitszufuhr könnte bei manchen Patienten zu einer erhöhten Rate an Bettnässen führen. Dies führt zu einer

Verstopfung des Bettes, da die Patienten dann eine Unterkunft in Einrichtungen finden müssen, die sich um ihre Bedürfnisse

kümmern können. Eine erhöhte Flüssigkeitszufuhr kann auch dazu führen, dass Patienten häufiger aus dem Bett aufstehen,

um auf die Toilette zu gehen, und daher häufiger stürzen. Dies kann auch schwerwiegende Folgen haben.

Wenn in Regel 3 von „Sicherheit“ die Rede ist, ist dann klar, dass dies Überlegungen dazu beinhaltet oder nicht, wie ein

Roboter in einem größeren Arbeitssystem funktionieren könnte? Ist in den Robotikprinzipien klar, welche

Verantwortung diejenigen tragen, die in der Robotertechnik arbeiten, oder ob es die Verantwortung von Krankenhausmanagern

und anderem Personal ist, solche potenziellen Auswirkungen des Einsatzes von Robotern zu identifizieren?

Beitrag von Roeland de Bruin und Madeleine de Cock Buning zum
AISB-Workshop zu Prinzipien der Robotik, 4. April 2016, Sheffield, Großbritannien

1. Einleitung

Fünf Jahre ist es her, dass die EPSRC-Robotikprinzipien veröffentlicht wurden, die von einer Gruppe angesehener britischer Robotik- und KI-Experten bei einem vom EPSRC/AHRC finanzierten Retreat entwickelt wurden.¹ Die Prinzipien, die darauf abzielten, „Roboter in der realen Welt zu regulieren“, waren in Form von fünf „Regeln“ dargelegt und sieben „High-Level-Botschaften“. Die Prinzipien hatten tatsächlich erhebliche Auswirkungen auf die britische Robotik Forschung und provozieren weiterhin erhebliche Debatten. Da derzeit öffentliche Besorgnis über die Da die Entwicklung von Robotertechnologien voranschreitet, halten wir es für sinnvoll, die Prinzipien noch einmal zu überdenken prüfen Sie ihre fortdauernde Relevanz anhand der folgenden Kriterien.

Unsere Beiträge konzentrieren sich auf das zweite Prinzip:

Prinzip 2: Menschen, nicht Roboter, sind verantwortliche Akteure. Roboter sollten entworfen werden; betrieben soweit dies zur Einhaltung bestehender Gesetze und Grundrechte und -freiheiten möglich ist, inklusive Privatsphäre.

Tatsächlich hat dieser zweite Grundsatz der Robotikprinzipien des EPSRC zwei Aspekte. Einerseits die Das Prinzip befasst sich mit der Verantwortung – einschließlich der Haftung – für die Handlungen des Roboters, zum anderen mit der Das Prinzip beinhaltet Methoden des Maschinendesigns, die zur Einhaltung bestehender Gesetze und Vorschriften beitragen können. Grundrechte und -freiheiten, einschließlich Privatsphäre.

Da sowohl Haftung als auch Design das Rückgrat der Einführung der Robotertechnologie bilden, z B. in autonome intelligente Autos in unserer Gesellschaft integriert, werden wir dieses zweifache Prinzip testen Der Schwerpunkt liegt auf der aktuellen Entwicklung und dem Einsatz autonomer intelligenter Autos. Ob das Das zweite EPSRC-Prinzip kann als zukunftsicher angesehen werden und wird anhand von drei Kriterien getestet:

1. *Gültigkeit* – ist das Prinzip richtig, da Aussagen über die Natur von Robotern, Roboterentwicklern, und die Beziehung zwischen Robotern und Menschen, oder ist sie ontologisch fehlerhaft, ungenau, out veraltet oder irreführend.
2. *Suffizienz/ Allgemeinheit* – ist das Prinzip ausreichend und umfassend genug, um alles abzudecken wichtige Fragen, die bei der Regulierung der Robotik in der realen Welt auftreten könnten oder sind Wichtige Bedenken werden übersehen.
3. *Nutzen* – ist das Prinzip des praktischen Nutzens für Roboterentwickler, Benutzer oder Gesetzgeber Festlegung von Strategien für Best Practices in der Robotik oder von rechtlichen Standards oder Rahmenwerken Sie beschränkten ihre Verwendung durch mangelnde Spezifität oder durch die Zulassung kritischer Ausnahmen.

¹<https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

2. Stand der Technik

2.1 Stand autonomer intelligenter Autos (AICs)

Bevor wir das Prinzip auf die Probe stellen, werden wir kurz den Stand der Technik von AICs vorstellen.

Heutzutage werden Privatautos zunehmend mit Technologien ausgestattet, die bestimmte Aspekte des Fahrens unterstützen.

Beispiele für solche Technologien sind Spurhalteassistent, Notbremsassistent,

Einparkhilfe und adaptiver Tempomat. In naher Zukunft wird es einen höheren Grad an Autoautomatisierung geben

verfügbar werden, was schließlich zur Einführung vollständig autonomer Fahrzeuge führen wird.

Auch jetzt sind einige Autos bereits mit bestimmten Formen der Automatisierung ausgestattet. Es gibt sogar

Es stehen Prototypen zur Verfügung, die ohne menschlichen Bediener fahren können. Google ist derzeit selbst Vorreiter

Fahrzeugtechnologie und hat Anfang 2015 einen voll funktionsfähigen AIC-Prototyp in der Bay Area, Kalifornien, auf der

Straße getestet.² Auch in der Europäischen Union konzentrieren sich Automobilhersteller auf die Entwicklung der AIC-

Technologie.³ Scania testet „Platooning“: einen Straßenzug von selbstfahrenden Lkw

Die autonom einem von Menschen gesteuerten Lkw an der Spitze des Konvois folgenden Fahrzeuge wurden auf den niederländischen

Straßen eingesetzt.⁴ Volvo plante den Einsatz von 100 Fahrzeugen, die bis 2017 in Schweden alle Aspekte des Fahrens

übernehmen können⁵ und in Deutschland einen Teil der Autobahn A9 zwischen München und Berlin ist für die umfangreiche

Erprobung autonomer Fahrzeuge in den kommenden Jahren reserviert.⁶

Eine Definition autonomer intelligenter Autos besteht aus drei Elementen. *Autonomie* bezieht sich auf die Ebene

des für den Betrieb notwendigen menschlichen Eingriffs, der als Spektrum betrachtet werden kann: ein geringerer Bedarf an

Menschliches Eingreifen impliziert ein höheres Maß an Autonomie. *Intelligenz* bezieht sich auf die Art und Weise, wie a

Das System kann seine Umgebung wahrnehmen und sein Verhalten an veränderte Umgebungen anpassen. Dazu gehört die

Fähigkeit zu lernen, komplexe Informationen zu verarbeiten und Probleme zu lösen.⁷ Autos sind es

motorisierte Fahrzeuge, die zur Beförderung von Gütern und/oder Personen sowie zur Erbringung von Dienstleistungen verwendet werden.

² Wikipedia, „Fahrerloses Auto von Google“, verfügbar im Internet unter http://en.wikipedia.org/wiki/Google_driverless_car (letzter Zugriff am 17. März 2015), unter Bezugnahme auf Matt O'Brian, „Googles ‚doofes‘ neues Auto“. „Selbstfahrendes Auto ist ein Zeichen für die Zukunft“, 22-12-2014, verfügbar im Internet unter http://www.mercurynews.com/business/ci_27190285/googles-goofy-new-self-driving-car-sign-things (letzter Zugriff am 28. Januar 2016).

³ Siehe zum Beispiel <https://www.media.volvocars.com/us/en-us/media/pressreleases/145619/volvo-car-groups-first-self-driving-autopilot-cars-test-on-public-roads-around-> (Volvo), <http://www.pcmag.com/article2/0,2817,2387524,00.asp>, (Volkswagen) <http://www.bbc.com/news/technology-25653253> (BMW) (letzter Zugriff am 28. Januar 2016).

⁴ Siehe <http://www.scania.nl/about-scania/media/platooning/> (abgerufen am 20. März 2015).

⁵ Alecander Stoklosa, „Volvo hat ein „produktionsfähiges“ autonomes Auto und wird es bis 2017 auf die Straße bringen“, verfügbar im Internet unter <http://blog.caranddriver.com/volvo-has-a-produktionsfähig-autonomes-Auto-wird-es-bis-2017-auf-die-Straße-bringen/> (letzter Zugriff 20. März 2015).

⁶ Stephen Edelstein, „Deutschland plant autonomes Autotestprogramm auf Hochgeschwindigkeitsautobahnen“, 28. Januar 2015, verfügbar im Internet unter http://www.motorauthority.com/news/1096521_Germany-plans-autonomous-car-test-program-on-Hochgeschwindigkeitsautobahnen (letzter Zugriff 20. März 2015).

⁷ Siehe Madeleine de Cock Buning, Lucky Belder und Roeland W. de Bruin, Arbeitspapier: „Mapping the Legal Framework for the Introduction into Society of Robots as Autonomous Intelligent Systems“, auf S. 3-4, verfügbar im Internet unter http://www.caaai.eu/wp-content/uploads/2012/08/Mapping-L_N-fw-for-AIS.pdf (letzter Zugriff am 28. Januar 2016).) und die Verweise auf Samir Chopra und Laurence F. White, *A Legal Theory for Autonomous Intelligent Agents* (Ann Arbor: University of Michigan Press 2011) auf S. 10 (Autonomie) und Collin R. Davies, „Ein evolutionärer Schritt in den Rechten des geistigen Eigentums – Künstliche Intelligenz und geistiges Eigentum“, 27 *Computer Law & Security Review* 2011, S. 601-619 (Geheimdienst); und von denselben Autoren das Kapitel „Mapping the Legal Framework for the Introduction into Society of Robots as Autonomous Intelligent Systems“, in Sam Muller et al (Hrsg.), *The Law of the Future and the Future of Law*, Reihe 2012 (De Cock Buning, Belder & De Bruin 2012), S. 195-210.

AICs können dazu beitragen, Lösungen für Herausforderungen zu finden, mit denen unsere Gesellschaft derzeit konfrontiert ist. Straße Die Sicherheit wird dramatisch steigen, wenn „menschliches Versagen“ als ursächlicher Faktor beseitigt wird Unfälle. AICs könnten das Risiko von Autounfällen erheblich verringern, da 93 % der Verkehrsunfälle auf menschliches Versagen zurückzuführen sind,⁸ was weltweit jährlich zu 1,3 Millionen Todesfällen und 50 Millionen schweren Verletzungen führt.⁹ AICs tragen nicht nur zur Verkehrssicherheit bei, sondern können auch zu einer effizienteren Nutzung führen des Straßennetzes, reduzieren den CO₂-Ausstoß und tragen dazu bei, die Mobilität behinderter Menschen zu verbessern.¹⁰ Die Einführung von AICs könnte somit Antworten liefern, um derzeit offensichtliche Risiken, die sich aus der Technologie ergeben, zu reduzieren Innovationen in den letzten Jahrzehnten.¹¹

Allerdings ist nicht jeder optimistisch, was eine fahrerlose Zukunft angeht. Es wird angegeben, dass dies zwar AICs sein könnten Während sich die Einführung autonomer Fahrzeuge positiv auf die Verkehrssicherheit auswirkt, ergeben sich weitere Risiken. AICs werden zum Beispiel anfällig für Hackerangriffe sein. Auch Geschäftsmodelle und Beschäftigung im Taxi- und Die Transportmärkte werden sich erheblich verändern, während Fahrer nach der Autonomisierung des Fahrens irgendwann überflüssig werden könnten.¹² Darüber hinaus könnte das Unfallrisiko steigen, wenn autonome und nicht autonome Autos auf denselben Straßen nebeneinander existieren.¹³

2.2 Stand der Technik

Ausreichende Gewissheit über den Rechtsstatus ist für das Wachstum und die gesellschaftliche Akzeptanz der Verbraucher von entscheidender Bedeutung Technologie. Unsicherheit bewirkt das Gegenteil. Könnte in diesem Fall die Maschine die Antwort auf die Frage sein? Maschine? Im Folgenden werden wir kurz auf die Haftungsfragen eingehen, die derzeit die Einführung und Herausforderung darstellen Einsatz von AICs in der Gesellschaft und gehen auf mögliche *Beweistechnologielösungen* für einige von ihnen ein Diese Herausforderungen könnten den eingebauten Datenschutz betreffen.

Haftung

Aktuelle Regelung in der EU, die die Verantwortung und Haftung für Schäden regelt, die dadurch verursacht werden könnten AICs stellen Herausforderungen hinsichtlich der Innovation im Bereich AICs und ihrer gesellschaftlichen Akzeptanz dar. An Einerseits befürchten Hersteller von AICs, dass sie im Rahmen der Produkthaftungsrichtlinie (PLD) leicht verletzt werden können

⁸ Bryant Walker Smith, „Menschliches Versagen als Ursache für Fahrzeugunfälle“, 18. November 2013, verfügbar im Internet unter <http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes> > (letzter Zugriff am 28. Januar 2016).

⁹ OECD, „OECD Factbook 2013: Economic, Environmental and Social Statistics“, 2013, verfügbar im Internet unter <http://www.oecd-ilibrary.org/sites/factbook-2013-en/06/02/03/index.html?contentType=&itemId=/content/chapter/factbook-2013-50-en&containerItemid=/content/serial/18147364&accessItemids=&mimeType=text/html> > (letzter Zugriff am 28. Januar 2016), auch zitiert in Gillian Yeomans, „Autonomous Vehicles – Übergabe der Kontrolle: Chancen und Risiken für Versicherungen“, abrufbar im Internet unter <https://www.lloyds.com/~media/lloyds/reports/emerging%20risk%20reports/autonomous%20vehicles%20final.pdf> > (letzter Zugriff am 28. Januar 2016) (Yeomans 2014) auf S. 5.

¹⁰ Siehe zum Beispiel Yeomans 2014, S. 5. Auch Anne Pawsey, „Autonomous Road Vehicles“, September 2013, S. 1. Verfügbar im Internet unter <http://www.parliament.uk/briefing-papers/post-pn-443.pdf>, (POSTnote 2013); Robolaw 2014, S. 42.

¹¹ Umweltverschmutzung, Klimawandel, gesellschaftlicher Ausschluss „schwächerer Parteien“ und hohe Unfallrisiken auf den (europäischen) Straßen können als Folgen der Modernisierungs- und Individualisierungsprozesse des vergangenen Jahrhunderts angesehen werden. Diese Nebenwirkungen müssen nun wiederum bekämpft werden.

Siehe für die Identifizierung und eine Studie zum Konzept der *Risikogesellschaft* von Ulrich Beck sein Buch *Risk Society, Towards a New Modernity*, London: Sage Publications 1992.

¹² Siehe zum Beispiel Scott Le Vine und John Polak, „Automated Cars: A smooth ride ahead?“, Februar 2014, S. 14, verfügbar am im Internet über <http://www.theitc.org.uk/docs/114.pdf> > (letzter Zugriff am 28. Januar 2016).

¹³ Siehe Wayne Cunningham und Antuan Goodwin, „Sechs Gründe, autonome Autos zu lieben oder zu verabscheuen“, 8. Mai 2013, verfügbar unter im Internet unter <http://www.cnet.com/news/six-reasons-to-love-or-loathe-autonomous-cars/> > (letzter Zugriff am 28. Januar 2016).

für Schäden haftbar gemacht werden, die durch fehlerhafte AICs verursacht werden und eine abschreckende Wirkung haben würden Innovation.¹⁴ Andererseits ist dies im aktuellen Rahmen zur Produkthaftung tatsächlich nicht der Fall

Bereitstellung eines einfachen Toolkits für Verbraucher, mit dem sie AIC-Hersteller für Mängel an ihren Produkten haftbar machen können alle. Die Beweislast für den Nachweis, dass tatsächlich ein Mangel vorlag, liegt beim Verbraucher

der AIC sowie über den ursächlichen Zusammenhang zwischen Mangel und aufgetretenem Schaden.

Die Beweisführung wird mit zunehmender Autonomie und Intelligenz in Autos komplexer, da Opfer eine eingehende (technologische) Analyse unter anderem der (Original-)Software durchführen müssen

Aktualisierungen und die Betriebsdaten, mit denen ein AIC ausgestattet ist, um die genaue Ursache eines Problems zu ermitteln Unfall. Gleichzeitig haben die Hersteller reichlich Möglichkeiten, sich dagegen zu wehren

Haftungsansprüche. Bei der Konfrontation mit AICs schützt die PLD deren Interessen nicht optimal

Verbrauchern, indem ihnen einfache Mittel zur Verfügung gestellt werden, um eine Entschädigung für den Schaden zu erhalten, den sie erlitten haben defekte AICs von Herstellern.

Verbesserungspotenzial für die geltende Gesetzgebung besteht darüber hinaus in den unterschiedlichen, nicht harmonisierten europäischen Haftungsregelungen für Kraftfahrzeuge. Bisher gibt es in der Europäischen Union 28 verschiedene Rahmenwerke.

Beispielsweise schreibt das französische „Loi Badinter“¹⁵ eine strikte Haftungsregelung *ohne Verschulden* vor

um zu beurteilen, ob der Fahrer oder der Betreuer eines Autos Schadensersatz für Opfer (außer dem Fahrer)¹⁶ von Unfällen mit Kraftfahrzeugbeteiligung leisten muss. Eine Haftungsbefreiung kann nur erfolgen, wenn der Fahrer (oder der Betreuer)

eine *faute* nachweisen kann, die für den Geschädigten unentschuldigbar ist.¹⁷ Die Niederlande

„Wegenverkeerswet“ schreibt dem Eigentümer oder Halter (Anmerkung: und nicht dem Fahrer) eine (semi-strikte) Haftung zu oder ein Betreuer) eines Kraftfahrzeugs, das in einen Unfall verwickelt ist, bei dem nicht motorisierte Verkehrsteilnehmer

beschädigt wurden.¹⁸ Mindestens 50 % des erlittenen Schadens müssen ersetzt werden, es sei denn, es kann *höhere Gewalt* nachgewiesen werden.¹⁹ Im Vereinigten Königreich gilt: Die Fahrlässigkeitsregeln werden angewendet, um festzustellen, ob

ein Fahrer eines Kraftfahrzeugs haftbar gemacht werden kann. In solchen Fällen gibt es im Vereinigten Königreich kein verschuldensunabhängiges Haftu

Allerdings ist der Sorgfaltsmaßstab, der von den Fahrern von Kraftfahrzeugen verlangt wird, recht hoch. Rechtsprechung erklärt, dass ein Fahrer, der ohne eigenes Verschulden das Bewusstsein verliert, dennoch fahrlässig handelt²¹,

ebenso wie der Fahrer, dessen Bremsen versagen, obwohl dieser Fehler nicht vorhersehbar war.²²

Allerdings müssen Opfer von Unfällen mit Kraftfahrzeugen nachweisen, dass die Fahrer dabei waren

¹⁴ Siehe Erica Palmerini, Federico Azzari, Fiorella Battaglia et al., D 6.2, „Guidelines on Regulated Robotics“, 22. September 2014, (RoboLaw 2014), S. 60.

¹⁵ Loi „*strebt danach, die Situation von Unfallopfern im Verkehr zu verbessern und die Verfahren zu beschleunigen.*“ d’indemnisation“.

¹⁶ Siehe A. Tunc, „The ‚Loi Badinter‘ – Ten Years of Experience“, 3 *Maastricht Journal of European and Comparative Law*, 1996 (Tunc 1996), S. 330. In Artikel 3 heißt es: „Die Opfer sind *hormis les Conducteurs (...)* werden von den Schäden entschädigt, die sich aus dem *Atteintes à leur personne qu’elles ont subis* ergeben, ohne dass sie ihrem eigenen Vermögen widersprechen können.“

¹⁷ Siehe auch Tunc 1997, S. 335.

¹⁸ Der Ersatz von Schäden, die Geschädigten in einem Kraftfahrzeug entstehen, richtet sich nach den allgemeinen Haftungsregeln des Artikel 6:162 des niederländischen Bürgerlichen Gesetzbuches.

¹⁹ *Marloes de Vos ea*, Oberster Gerichtshof der Niederlande, 2. Juni 1995, *NJ* 1997/700-702, und *Said Hyati ea*, 5. Dezember 1997 *NJ* 1998/400-402. Der Begriff „Betriebsgefahr“ ist dem deutschen Straßenverkehrsgesetz entlehnt.

²⁰ Oder *mutmaßliche Haftung*, wie sie im schottischen Recht genannt wird.

²¹ *Roberts v. Ramsbottom* [1980] 1 WLR 823, auch zitiert in: Cees van Dam, *European Tort Law*, Oxford: Oxford University Press 2006 (Van Dam 2006), S. 364, Fußnote 52.

²² *Henderson gegen HE Jenkins & Sons und Evans* [1970] AC 282, zitiert in Van Dam 2006, S. 364, Fußnote 53. Van Dam weiter nimmt den *Fall Worsley gegen Hollins* [1991] RTR 252 (CA) zur Kenntnis, in dem die Richter feststellten, dass die Klage des Opfers wegen Fahrlässigkeit gescheitert sei, da der Angeklagte nachweisen konnte, dass sein Kleinbus kürzlich gewartet worden sei, obwohl seine Bremssysteme ausgefallen seien und dadurch Schäden verursacht hätten und hat den TÜV bestanden.

Verschulden, das heißt: sie hatten fahrlässig gehandelt.²³ Die erheblichen Unterschiede bestehen in der Art und Weise der Haftung für Kraftfahrzeuge. Fahrzeuge wird in allen Mitgliedstaaten angegangen, ist für Entwicklung, Versicherung und nicht von Vorteil Einsatz von AICs in Europa. In jedem Fall regeln nationale Regelungen die Haftung für *Kraftfahrzeugführer*. Fahrzeuge müssen aktualisiert werden, um die Haftung für Fahrzeuge ohne menschlichen Fahrer regeln zu können.

Privatsphäre

Während die Einführung der AIC-Technologie im Hinblick auf eine erhöhte Sicherheit auf den Straßen vielversprechend ist, Dies führt dazu, dass weniger Schäden abgedeckt werden müssen, was auch die Versicherungen bei einem Unfall beachten. Wenn ein Fehler durch autonome Technologie verursacht wird, wäre „umfangreiche Software- und Hardware-Analysekompetenz“ erforderlich, um zu wissen, wie und warum er aufgetreten ist.“²⁴ Eine der Optionen zur Beurteilung der Ursache eines Unfall und damit Hilfestellung bei der Beantwortung der Frage, wo die Haftung liegt, könnte darin bestehen, sich auszurüsten. Fahrzeuge mit Blackboxen oder mit Telematiklösungen, die AICs mit einer dedizierten Infrastruktur und/oder mit Remote-Servern verbinden.²⁵ Die Ziele dieser Art von Technologien bestehen unter anderem darin, die Bewegungen autonomer Autos und getroffene Betriebsentscheidungen aufzuzeichnen entweder durch das Auto selbst oder durch den Fahrer, der seine Bewegung steuert, sowie Daten über Ereignisse und Objekte im Fahrzeug. Nähe eines autonomen Fahrzeugs. Die Black-Box-Technologie zeichnet die gesammelten Daten auf und speichert sie in einem Fahrzeug und bietet Potenzial für eine spätere Beurteilung. Die Telematiktechnologie könnte breiter sein. Anwendungen. Die Daten könnten nicht nur zur Beurteilung von Fehlern und Schadensursachen genutzt werden. Bei Unfällen könnte es sogar eine präventive Wirkung haben. Fahrzeug-zu-Fahrzeug-Kommunikation (V2V) und die Fahrzeug-zu-Infrastruktur-Kommunikation könnten zur Unfallverhütung in Echtzeit genutzt werden und bringen „Sicherheits-, Mobilitäts- und Umweltvorteile“ im Allgemeinen.²⁶ Obwohl Black Box Technologien und Telematiklösungen wie V2V und V2I (im Folgenden „Tracing“ genannt). Technologie') kann im Hinblick auf die Vermeidung von Unfällen und die Verteilung der durch sie verursachten Schäden erfolgsversprechend sein. AIC-Unfälle bergen auch Risiken im Hinblick auf das Recht auf (Informations-)Privatsphäre der darin befindlichen Personen und in der Nähe von Fahrzeugen, die mit diesen Technologien ausgestattet sind.

Der Datenschutz der Bürger ist in der Europäischen Union durch die Datenschutzrichtlinie (DPD)²⁷ streng geregelt und wird nach Inkrafttreten der Datenschutz-Grundverordnung (DSGVO)²⁸ noch strenger geregelt. Der aktuelle und kommende Rahmen sieht vor

Beispiel, dass bereits in der Designphase von AICs, die mit Tracing-Technologie ausgestattet sind, ein Datenschutz gewährleistet ist

²³ Es gibt eine Regel einer gesetzlichen Pflicht, die – bis zu einem gewissen Grad – eine verschuldensunabhängige Haftung für Fahrer von Kraftfahrzeugen festlegt, die sich einer Kreuzung nähern: „Der Fahrer jedes Fahrzeugs, das sich einer Kreuzung nähert, muss dies tun, es sei denn, er kann erkennen, dass kein Fußgängerüberweg vorhanden ist.“, fahren Sie mit einer solchen Geschwindigkeit, dass Sie bei Bedarf anhalten können, bevor Sie eine solche Kreuzung erreichen“, zitiert in Van Dam 2006, S. 365, Fußnote 57, mit Verweis auf Reg. 3 der Pedestrian Crossing Places (Traffic) Regulations 1941, ersetzt durch die Zebra Pedestrian Crossing Regulations 1971, SI 1971, Nr. 1524. Eine Einrede, die ein Fahrer in dieser Hinsicht hat, ist höhere Gewalt.

²⁴ Yeomans 2014, S. 18.

²⁵ Yeomans 2014, S. 18. Siehe außerdem James M. Anderson, Nidhi Kalra, Karlyn D. Stanley et al., *Autonomous Vehicle Technology – A Guide for Policymakers*, RAND Transportation, Space and Technology Program 2014, (RAND-Bericht), auf S. 94-95.

²⁶ RAND-Bericht, S. 81.

²⁷ Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen hinsichtlich der Verarbeitung personenbezogener Daten und zum freien Datenverkehr, *Amtsblatt L* 281 0050, 23.11.1995 S. 0031 -

²⁸ Vorschlag für eine VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung) COM/2012/011 final – 2012/0011 (COD). Bitte beachten Sie, dass der Trilog zwischen der Europäischen Kommission, dem Europarat und dem Europäischen Parlament über den endgültigen Text der DSGVO abgeschlossen ist, dieser Text jedoch noch nicht offiziell veröffentlicht wurde.

Es sollte eine Folgenabschätzung durchgeführt werden. Darüber hinaus „angemessene technische und organisatorische.“ Maßnahmen zum Schutz personenbezogener Daten vor zufälliger oder unrechtmäßiger Zerstörung oder zufälligem Verlust, Änderung, unbefugte Weitergabe oder unbefugter Zugriff, insbesondere wenn die Verarbeitung dies beinhaltet „Verhinderung der Übertragung von Daten über ein Netzwerk und gegen alle anderen rechtswidrigen Formen der Verarbeitung“ muss umgesetzt werden.²⁹ Die DSGVO regelt, dass diese Maßnahmen so weit wie möglich in neue Technologien „eingebaut“ werden sollen, wobei diese Maßnahmen *unter anderem* auf eine Datenminimierung abzielen müssen und muss standardmäßig aktiviert sein.³⁰ Dabei müssen modernste Sicherheits- und Implementierungskosten berücksichtigt werden Umsetzung von Maßnahmen. Darüber hinaus müssen diese „ein angemessenes Sicherheitsniveau gewährleisten Risiken, die die Verarbeitung mit sich bringt, und die Art der zu schützenden Daten“.

Eine weitere, noch neuere Herausforderung stellt die jüngste Entscheidung des Europäischen Gerichtshofs dar das Safe-Harbor-Abkommen zu erklären, das die Grundlage für viele Austausche personenbezogener Daten bildet zwischen der EU und den Vereinigten Staaten von Amerika, ungültig. Es ist wahrscheinlich, dass Tracing-Technologie Die in AICs integrierten Daten stellen die grenzüberschreitende Übermittlung von (personenbezogenen) Daten dar Grenzen der Europäischen Union übertragen und diese Daten möglicherweise beispielsweise über Cloud Computing in die Vereinigten Staaten importieren. Der EuGH entschied, dass die USA kein angemessenes Schutzniveau bieten personenbezogene Daten, denn nach den Enthüllungen von Edward Snowden wurde klar, dass US-Behörden wie z B. die National Security Agency, einfachen Zugriff auf personenbezogene Daten haben, die von US-amerikanischen Unternehmen und Institutionen verarbeitet werden.³¹ Das Gericht entschied, dass die Befugnisse der europäischen Aufsichtsbehörden untergraben würden durch die US-Praktiken, die möglicherweise nicht durch eine Entscheidung der Europäischen Kommission ermöglicht werden. Dieses Urteil impliziert, dass der Export personenbezogener Daten in die Vereinigten Staaten auf Grundlage der nicht mehr möglich ist Safe-Harbor-Rahmen. Obwohl die Vereinigten Staaten und die Europäische Kommission derzeit über ein alternatives Abkommen verhandeln,³² ist der Austausch personenbezogener Daten zwischen der EU und den Vereinigten Staaten aufgrund der noch ungültigen Safe-Harbor-Regeln zwischenzeitlich nicht zulässig.

3. Stellen Sie das Prinzip auf die Probe

In diesem Teil werden wir testen, ob das zweite EPSRC-Prinzip als zukunftsicher angesehen werden kann die Kriterien Gültigkeit, Suffizienz/Allgemeinheit und Nutzen

3.1 Gültigkeit

Sowohl beim aktuellen Stand der Technik als auch der Rechtslage gilt der erste Teil des Grundsatzes „Menschlich *nicht*“. *Roboter seien verantwortliche Agenten*, hat sich in der Tat immer noch als gültig erwiesen. Es ist eine korrekte Aussage darüber Natur von Robotern, Roboterentwickler und die Beziehung zwischen Robotern und Menschen. Es kann geben immer entweder ein Mensch oder eine juristische Person, die für die Handlungen der AICs verantwortlich und haftbar gemacht wird. Die konkrete Schaffung einer eigenen Rechtspersönlichkeit für AICs erscheint angesichts der aktuellen Lage derzeit weit hergeholt Aufgrund des technologischen und rechtlichen Status von AICs würde dies darüber hinaus nicht zur Lösung der Haftung beitragen

²⁹ Kunst. 17(1) DPD; siehe auch Art. 5 Abs. 1 lit. eb und Abschnitt 2 (Art. 30 ff.) zur Datensicherheit in der DSGVO.

³⁰ Kunst. 23 DSGVO.

³¹ Rechtssache C-362/14, Maximilian Schrems/Facebook [2015].

³² Die neuesten Nachrichten zum „EU-US-Rahmenabkommen“ (Abkommen zwischen den Vereinigten Staaten von Amerika und der Europäischen Union über den Schutz personenbezogener Daten bei der Verhinderung, Untersuchung, Aufdeckung und Verfolgung von Straftaten) finden Sie hier: http://ec.europa.eu/justice/newsroom/data-protection/news/150908_en.htm (letzter Zugriff am 9. März 2016).

Herausforderungen wie unter 2.2 beschrieben gemeistert. Das Gleiche gilt für den zweiten Teil des zweiten Grundsatzes gibt an, dass *Roboter entworfen werden sollten; so weit wie möglich betrieben werden, um den geltenden Gesetzen zu entsprechen und Grundrechte und Grundfreiheiten, einschließlich der Privatsphäre*, haben sich als immer noch gültig erwiesen. Mit Blick auf die *Da die Beweistechnologie* in AICs integriert werden soll, hat sich dieser Grundgedanke als richtig erwiesen wahrer, als man es sich aufgrund seines Designs hätte vorstellen können. Wie wir unter 2.2 gesehen haben, sind es tatsächlich die Mängel Probleme des derzeitigen Haftungssystems können teilweise durch intelligente, in die AIC integrierte Systeme zur Sammlung und Speicherung von Beweismitteln gelöst werden. Diese Systeme zur Beweiserhebung und -speicherung sollten so gestaltet sein dass die erhobenen personenbezogenen Daten so weit wie möglich geschützt werden: „Privacy by Design“ und „Privacy by Default“. müssen jederzeit in AICs (Tracing-Technologie) integriert sein.

3.2 Angemessenheit/Allgemeinheit

Gleichzeitig bleibt das Prinzip dennoch ausreichend und breit genug, um alle wichtigen Aspekte abzudecken Probleme, die bei der Regulierung der AICs in der realen Welt auftreten könnten. *Menschen, keine Roboter, sind es verantwortliche Agenten Roboter sollten entworfen werden; werden so weit wie möglich betrieben, um die bestehenden Gesetze und Grundrechte und -freiheiten, einschließlich der Privatsphäre, einzuhalten.* Es scheinen keine nennenswerten Bedenken zu bestehen übersehen. Obwohl einige Autoren zu argumentieren scheinen, dass für autonome intelligente Maschinen eine juristische Person geschaffen werden sollte, die den Robotern den verantwortlichen Agenten überlässt,³³ hat dies für viele nicht überzeugt³⁴ und schon gar nicht für uns.,

Die Herausforderungen, die sich aus der Einführung autonomer intelligenter Autos in der Gesellschaft und deren ergeben Die Schadenshaftung an sich scheint keine gesonderte Rechtspersönlichkeit zu erfordern. Es würde lediglich hinzufügen ein weiterer Akteur für die Haftungszurechnung. Gleichzeitig wäre eine umfassende Neugestaltung erforderlich des Haftungssystems, wie es derzeit in der realen Welt angewendet wird, während sich die Technologie noch in der Entwicklungsphase befindet Phase, die das Risiko einer Unter- oder Überregulierung birgt.

3.3 Dienstprogramm

Soweit die derzeitigen rechtlichen Möglichkeiten noch nicht ausgeschöpft sind, unter anderem durch das Ziel einer weiteren Harmonisierung Es gibt keine gesetzlichen Haftungsregelungen der EU in Kombination mit wirksamer Beweismitteltechnologie Beweise, die einen vollständigen Paradigmenwechsel durch die Einführung von AICs als Verantwortlichen untermauern würden Agenten in sich. Denn AIC kann tatsächlich so konzipiert und betrieben werden, dass es den geltenden Gesetzen entspricht Die Nützlichkeit dieses Prinzips bleibt offensichtlich. Allerdings Black-Box-Technologien und Telematiklösungen wie V2V und V2I können im Hinblick auf die Vermeidung von Unfällen und die Schadensverteilung vielversprechend sein verursacht durch AIC-Unfälle, da diese auch Risiken im Hinblick auf das Recht auf (Informations-)Privatsphäre mit sich bringen Menschen in und in der Nähe von Autos, die mit diesen Technologien ausgestattet sind, müssten die Systeme erfüllen

³³ Siehe zum Beispiel James Boyle, „Endowed by Their Creator?: The Future of Constitutional Personhood“, *The Future of the Constitution*, 9. März 2011, S. 6, auch im Internet verfügbar unter http://www.brookings.edu/~media/research/files/papers/2011/3/09-personhood-boyle/0309_personhood_boyle.pdf (letzter Zugriff am 9. März 2016). . Siehe außerdem JP Günther, F. Münch, S. Beck, S. Löffler, C. Leroux, & R. Labruto, „Issues of Privacy and Electronic Personhood in Robotics. 21st IEEE International Symposium on Robot and Human Interactive Communication“, Paris 2012, zitiert in Christophe Leroux, Roberto Labruto, Chiara Boscarato und anderen, „Suggestion for a Green Paper on legal issues in robotics“, Dezember 2012, verfügbar im Internet unter http://www.eu-robotics.net/cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf (letzter Zugriff 28. Januar 2016); und Robolaw 201.

³⁴ Siehe zum Beispiel Peter Asaro, „Robots and Responsibility from a Legal Perspective“, über <http://www.peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf>, letzter Zugriff am 9. März 2016; und Lawrence Solum, „Legal Personhood for Artificial Intelligences“, *North Carolina Law Review*, (April 1992), S. 1231–1287.

umfassen „Privacy by Design“, um diese Grundrechte zu schützen, wie sie in internationalen und europäischen Verträgen festgelegt sind.³⁵

Es ist von entscheidender Bedeutung, dass diese rechtlichen und technischen Anforderungen erfüllt werden, bevor die Herausforderung besteht Die Anforderungen an die Einführung und den Einsatz von AICs in der Gesellschaft können erfüllt werden.

4. Fazit

Wir können mit Bedacht zu dem Schluss kommen, dass Prinzip 2 der EPSRC-Grundsätze der Robotik, wie sie von entwickelt wurden Britische Robotik- und KI-Experten auf dem von EPSRC/AHRC finanzierten Retreat haben sich als zukunftssicher erwiesen als wir uns auf den aktuellen Stand der Gesetzgebung und Technologie rund um AICs bezogen.

Menschen, nicht AICS, sind verantwortliche Akteure. AICs sollten entworfen werden; soweit wie möglich betrieben praktikabel, um bestehende Gesetze und Grundrechte und -freiheiten einzuhalten, einschließlich der Privatsphäre durch Design. Dies ist ein Beweis dafür, dass die Antwort der Maschine zumindest teilweise in der liegt Maschine selbst.

³⁵ Siehe zum Beispiel Kunst. 7 und 8 der Charta der Grundrechte der Europäischen Union und Artikel 8 der Europäischen Menschenrechtskonvention.

Fairdatahandlingandrobotics

Burkhard Schäfer, Edinburgh Law School
Lilian Edwards, University of Strathclyde

Diese Intervention kombiniert Prinzip 4: *Roboter sind hergestellte Artefakte. Sie sollten nicht auf betrügerische Weise entworfen werden, um gefährdete Benutzer auszunutzen. Stattdessen sollte ihre Maschinennatur mit Prinzip und Prinzip 2 transparent sein: Roboter sollten so konzipiert und betrieben werden, dass sie so weit wie möglich betrieben werden, um bestehende Gesetze, Grundrechte und Freiheiten, einschließlich der Privatsphäre, einzuhalten. Es schlägt eine Ausarbeitung/Spezifizierung der Prinzipien vor, die der Schnittpunkt zwischen ihnen erfordert. Radikaler kann es sein, dass es um entsprechende Pflichten der Benutzer von Robotern und in der Tat Dritter gegenüber Robotern geht. Nicht im Sinne von „Roboterrechten“, sondern lediglich als Ermöglicher für ethisches und gesetzeskonformes Roboterdesign. Das heißt, Roboterentwicklern die Einhaltung der Prinzipien zu ermöglichen, kann etwas sein. Behalten Sie Pflichten gegenüber Robotern bei (oder, für diejenigen, die sich über diese Formulierung Sorgen machen, Pflichten gegenüber dem Besitzer eines Roboters, der sich in gewisser Weise gegenüber der Maschine verhält.*

Roboter stellen einige einzigartige Herausforderungen für faire Datenverarbeitungspraktiken dar, Herausforderungen, die zumindest teilweise auf ihre Fähigkeit zurückzuführen sind, die Menschen, mit denen sie interagieren, wenn auch unabsichtlich zu „täuschen“.

Lange vor der modernen Technologie entwickelten die Menschen Techniken zum Schutz der Privatsphäre, von den Vorhängen bis zu den Fenstern und dem Schleier, vom Lernen, wann man seinen Geruch wahrnahm, bis hin zum Wegwaschen des Geruchs. Sie schützten sie vor den neugierigen Blicken anderer Menschen sowie vor dem Interesse nichtmenschlicher Raubtiere. Und sie schützten nicht nur die Privatsphäre, sondern auch andere Informationsinteressen, einschließlich wertvoller Informationsmonopole wie Handelsgeheimnisse oder Wissen. Wie. Entscheidend ist, dass diese nicht nur Informationen schützen, sondern auch das Teilen und den Austausch von Informationen (Flüstern, Schalldämmung).

Das Gesetz mit seinem System aus Regeln und Ausnahmen hat diese Schutzmaßnahmen aufgrund niedriger Technologie häufig formal anerkannt. Die Mauern, die wir um uns herum bauen, halten nicht nur die Wärme drinnen und den Regen draußen, sondern auch Informationen drinnen und Beobachter draußen. Hannah Arendts Unterscheidung zwischen dem Privaten, dem Öffentlichen und dem Sozialen findet in vielerlei Hinsicht ihren Niederschlag, ebenso wie unser Gesetz mit dem Haus oder dem durch eine Hecke geschützten Garten der Archetyp der „angemessenen Erwartung an Privatsphäre“ und „Sicherheit in unseren Häusern und Wohnungen“, sondern auch ein Raum, in dem Daten frei gesammelt und ausgetauscht werden können – die Haushaltsausnahme des europäischen

Die Robotik-Technologie droht, diese Low-Tech-Lösungen für das Datenschutzproblem zunehmend überflüssig zu machen. Wir sehen uns mit einer zunehmenden Vielfalt an Sensorkapazitäten konfrontiert, von denen viele in unserer revolutionären Vergangenheit nicht oder nicht auf besonders bedrohliche Weise begegnet sind.

Niemand ist ein Held für seine Hausangestellten. Aber zumindest mit Hausangestellten für den Herrn oder die Frau

Das Haus könnte vorhersehen, was genau es sehen könnte, das normative (sowohl soziale als auch rechtliche) Umfeld verstehen, das es daran hindert, Daten über seinen Arbeitgeber zu sammeln und vor allem weiterzugeben.

Das Verständnis der normativen Umgebung zusammen mit dem Verständnis der sensorischen Fähigkeiten würde dann eine rationale Risikobewertung und -verwaltung ermöglichen. (Ich vertraue meinem Butler meine schmutzige Unterwäsche, aber kein blutgetränktes Hemd. Ich mache mir keine Sorgen um die Wärmesignatur, wenn ich einen Gast bewirte, sondern kann mich dafür entscheiden, den Lärm leiser zu halten, während ich mich darauf verlasse, dass mein Butler zuerst klopft, bevor er das Schlafzimmer betritt).

Die Robotik bedroht diese Verteidigungsstrategien nicht nur, weil sie Sensoren außerhalb des visuellen oder akustischen Spektrums verwenden können, oder wegen ihrer Beweglichkeit, die Wahrnehmungen in zuvor geschützten Räumen ermöglicht. Wo sie das äußere Erscheinungsbild von Menschen oder gar nichtmenschlichen Tieren imitieren, selbst wenn ihre robotische Natur offensichtlich sichtbar ist (gemäß Prinzip 4). Allerdings (derzeit) nicht systematisch untermauert. Suchen Sie nach den Reisebeweisen, dass wir diese Schlussfolgerungen ziehen, wenn wir mit Robotern interagieren. Im Internet gibt es viele Menschen, die sich von hinten an Asimo „anschleichen“ – jetzt befinden sich Asimos Sensoren „tatsächlich“ in seinen Augen und haben Seheinschränkungen ähnlich denen eines Menschen, aber das kann durchaus falsch sein.

Ein Teil des ethischen Designs sollte daher auch darin bestehen, die sensorischen Fähigkeiten von Robotern auf eine Weise anzuzeigen, die das Entstehen „intuitiver“ Abwehrmechanismen der Art erleichtert, die wir bei anderen Menschen verwenden, und wo möglich keine irreführenden Schlussfolgerungen zuzulassen, und „einfache Verteidigungsmechanismen“ in die Bewertung der Eindringlichkeit einzubeziehen, wenn eine Wahl zwischen verschiedenen Sensoren getroffen werden kann.

Dahinter steckt das Datenschutzrecht, aber Fair-Sensing- und Datenverarbeitungspraktiken gehen über persönliche, geschweige denn sensible personenbezogene Daten hinaus. Wir schützen nicht nur Daten über uns, sondern auch unsere Geschäftsideen, wissenschaftlichen oder technischen Entdeckungen oder Fähigkeiten. Auch hier gibt es einen Grund, warum potenzielle Gegner instinktiv über sensorische Fähigkeiten sprechen (denken Sie an Schulkinder, die während Prüfungen eine Bücherwand um sich herum aufbauen, um anderen Betrug vorzubeugen).

IP-Recht stellt daher eine weitere rechtliche Einschränkung dar, die unter dieser Überschrift beachtet werden muss, und möglicherweise ist ein darüber hinausgehender Begriff von „fairem Umgang mit Daten“ erforderlich, der über das DP-Recht hinausgeht. Wenn ein Industrieroboter meine Bewegungen beobachtet, um die Zusammenarbeit zu verbessern und Absprachen zu vermeiden, und dabei lernt, meinen Arbeitsplatz zu sichern (die Art und Weise, wie ich mich bewege, ist mein Alleinstellungsmerkmal), ist dies ein „Interesse“ von mir, das verdient Schutz/Entschädigung.

Dies wirft möglicherweise auch eine Frage auf, die radikaler über die Prinzipien hinausführt. Sie versuchen hauptsächlich, Pflichten festzulegen, die die Entwickler den Menschen schulden, die mit ihren Maschinen interagieren die Entwickler schulden und mögliche Pflichten, die ihnen bzw. dem Besitzer des Roboters obliegen.

Um ein sicheres Roboterdesign zu ermöglichen, kann es beispielsweise für Dritte eine Pflicht sein, bestimmte Informationen mit dem Roboter offenzulegen oder zu teilen, der in der Vergangenheit gesetzlich privilegiert war. Ein Roboter, der z. „identisches“ Kopieren. Der US-Ansatz argumentiert, dass dies eher zu funktionalen als zu expressiven Zwecken kopiert würde

– keine Sprache – und sollte daher nicht gegen das Urheberrecht verstoßen – sobald jedoch Maschinen ihre Aktionen durch die Weitergabe dieser Daten koordinieren, stößt möglicherweise auch dieses InnoFargument an seine Grenzen.

Bürger können Technologien einsetzen, um zu verhindern, dass Sensoren sie bemerken (z. B. Tarn-Gesichtsbemalung – <https://cvdazzle.com>), aber das kann bedeuten, dass sie ein höheres Risiko eingehen, dass der Roboter ihnen entgegenläuft. Wenn Dritte involviert sind, kann dies zu noch komplexeren Rechtsfragen führen. Wenn mein Gesicht einen Datenpunkt liefert, von dem ich weiß, dass der Roboter lernt, habe ich „Qualitätssicherungspflichten“, um ein gutes Beispiel zu sein? „Wenn ich den Lernprozess absichtlich manipulierte, begeben sich damit den Geltungsbereich des Computermissbrauchsgesetzes? Und schließlich, wenn ich zum Lernen einer Maschine beitrage, habe ich dann einen Überblick darüber, was sie als Ergebnis hervorbringt?“

Das grundlegende Fahrlässigkeitsrecht und seine Unterscheidung zwischen Tat und Unterlassung und wie Fahrlässigkeit damit umgeht, indem es Pflichten gegenüber Nachbarn festlegt, wird *nach* einem Unfall Teil der rechtlichen Antwort sein. Für die Diskussion hier wird die Frage jedoch etwas anders gestellt: Zum Zeitpunkt der Entwicklung eines Roboters können/sollten die Designer ihre Pflicht erfüllen

1. Verlassen Sie sich auf *die ethische/soziale* Pflicht Dritter, den Wissenserwerb der Maschine nicht zu manipulieren
2. Verlassen Sie sich lediglich auf eine engere *gesetzliche* Verpflichtung, bestimmte vorhersehbar gefährliche Datenmanipulationen zu unterlassen
3. Wenn man über die Sicherheit und Gesetzeskonformität der von ihnen gebauten Roboter nachdenkt, verlässt man sich nicht auf ein kooperatives Umfeld – schließlich werden nicht alle Gesetze von allen beachtet.

Um zu verdeutlichen, warum dieses Problem im Zusammenhang mit der Diskussion über „Sensortransparenz“ auftritt: Wenn wir die oben besprochene ethische Verpflichtung akzeptieren, dass Roboter normalerweise offenbaren sollen, wie und damit was sie spüren können, dann öffnen sie sich zwangsläufig der Manipulation. Wenn wir zusätzlich 1 oder zumindest 2 akzeptieren, ist dies weniger ein Problem als wenn wir 3 akzeptieren.

Wenn wir 2 akzeptieren, müssen wir uns direkt mit der Tatsache auseinandersetzen, dass das Gesetz an den beiden Extremen des Spektrums klar ist: Es ist schwer, Pflichten gegenüber Menschen zu erfüllen, andererseits kann ich die Pflicht haben, dass ein Nachbar seine Scheune nicht niederbrennt. Aber habe ich die Pflicht, Roboterbetreiber (oder Designer) daran zu hindern, ihre Roboter beim Training zu verwirren? Ist es vernünftigerweise vorhersehbar, dass ein Roboter verwirrt? Schließlich habe ich nicht einmal die Pflicht, Fremde nicht zu belügen, wenn ich ihnen Anweisungen erteile – es sei denn, ich habe einen professionellen Rat. Aber andererseits wäre es ein anderer Vorschlag, ein Kind in die Irre zu schicken. Sind Maschinen, die „noch lernen“, analog zu einer solchen Situation?

In den Diskussionen wurden Probleme besprochen, die durch Menschen verursacht werden, die Daten zurückhalten/verzerrten/manipulieren, die ein Roboter benötigt, um sicher zu funktionieren, und die er ethisch oder rechtlich schuldet.

Aber wir stehen auch vor ethischen Gestaltungsentscheidungen, wenn Menschen kooperieren und freiwillig Informationen zur Verfügung stellen, zu deren Bereitstellung sie gesetzlich nicht verpflichtet sind, die sie aber aus staatsbürgerlichem Pflichtbewusstsein wählen/verbieten. Sollte sich dies auf den Status der vom Roboter erzeugten *Ergebnisse* auswirken, z. B. in Form einer Vorteilsteilung?

Dies könnte bedeuten, dass nicht nur Roboter als Roboter und ihre Sensoren als Sensoren identifizierbar sein müssen, sondern dass die von Robotern erzeugten Ergebnisse auch als Maschinen und nicht als von Menschen erzeugte identifizierbar sein müssen. Hier könnte das Urheberrecht relevante Einschränkungen für die Art und Weise auferlegen, wie Maschinen kommunizieren d

Das übergreifende Thema dieser Intervention ist daher letztlich eines der algorithmischen Transparenz: Gesetzliche und ethische Pflichten beeinflussen, wann und wie Roboter ihre sensorischen Fähigkeiten offenlegen sollen. Sobald sie dies tun, hat ihre Umgebung die Wahl, zu kooperieren oder nicht zu kooperieren. Das Sammeln von (einigen) Daten ist für sicherere Maschinen erforderlich. Dort, wo über die gesetzlich vorgeschriebene Zusammenarbeit hinaus Mehrwert entsteht, muss darüber diskutiert werden, wie dies auf gerechte Weise berücksichtigt werden kann, wodurch möglicherweise eine weitere Offenlegungspflicht auferlegt wird, eine „von Robotern gemachte“

Während es sich hierbei hauptsächlich um rechtliche Fragen handelt, müssen die Entwickler im Hinblick auf ethisches (und gesetzeskonformes) Design auch vorhersehen können, welche Art von Interaktion sie erwarten können und auf welche Art von Informationen sie rechtlichen Zugriff haben werden.

Können Roboter verantwortungsvolle moralische Agenten sein? Und warum sollte uns das interessieren?

AmandaSharkey,

Abteilung für Computerwissenschaften und Sheffield Robotics, University of Sheffield

Prinzip 2. Menschen, keine Roboter, sind verantwortliche Akteure. Roboter sollten so konzipiert und betrieben werden, dass sie soweit wie möglich die bestehenden Gesetze, Grundrechte und Freiheiten, einschließlich der Privatsphäre, einhalten.

Auf den ersten Blick erscheint diese Aussage bzw. dieser Grundsatz überzeugend. Es macht Sinn, darauf zu bestehen, dass Menschen und nicht Roboter als verantwortliche Akteure gelten. Unsere sollte dazu beitragen, die möglichen schädlichen Verwendungsmöglichkeiten von Robotern einzuschränken. Sie legt auch nahe, dass Roboter so konzipiert und betrieben werden sollten, dass sie den bestehenden Gesetzen und Grundrechten und Grundfreiheiten entsprechen: Es ist schwer vorstellbar, dass jemand etwas anderes vorschlägt.

Doch bei näherer Betrachtung stellt sich heraus, dass die Aussage keinerlei Rechtfertigung liefert. Sie besagt, dass Menschen und keine Roboter verantwortliche Agenten sind, und gibt auch keine Hinweise dazu, wo und wann Roboter eingesetzt werden sollten oder welche Konsequenzen sich aus der Annahme ergeben, dass Roboter keine verantwortungsvollen Agenten sind. (a) Welche Gründe gibt es für die Annahme, dass Roboter und nicht Menschen verantwortliche Agenten sind? (b) Reicht es aus, Roboter so zu entwerfen, dass sie den bestehenden Gesetzen und Grundrechten und Grundfreiheiten entsprechen? und (c) Wenn Roboter keine verantwortungsvollen Agenten sind, sollte dies die ihnen zugewiesenen Rollen und die Situationen, in denen sie eingesetzt wurden, einschränken?

(a) Welche Gründe gibt es für die Annahme, dass Menschen und nicht Roboter verantwortliche Akteure sind?

Abgesehen von der rechtlichen Verantwortung lassen sich zwei Gründe für diese Annahme identifizieren. Der erste basiert auf dem Unterschied zwischen biologischen und mechanischen Maschinen und auf den biologischen Grundlagen der Moral. Zweitens handelt es sich um die Notwendigkeit, dass die Gesellschaft Verantwortung für die wesentlichen Tatsachen übernimmt, die der Mensch hervorgebracht hat. Wir berücksichtigen beides wiederum.

(i) Biologische Maschinen im Vergleich zu mechanischen Maschinen: Einen Agenten für seine Handlungen verantwortlich zu machen, ist gleichbedeutend damit, ihn als Amora-Agenten zu betrachten. Daher ist es wichtig, die biologische Basisformalität in biologischen Maschinen hervorzuheben und dies dem Fehlen einer solchen Basis in mechanischen Maschinen wie Robotern gegenüberzustellen. Patricia Churchland (2011) erörtert die Grundlage der Formmoral in Lebewesen und argumentiert, dass die Grundlage für die Fürsorge für andere in der Neurochemie der Bindung und Bindung bei Säugetieren liegt. Sie erklärt, dass diese in der Ausweitung der Selbsterhaltung und Schmerzvermeidung bei Säugetieren auf ihre unmittelbaren Verwandten beruht. Menschen und andere Säugetiere machen sich Sorgen um ihr eigenes Wohlergehen und das derjenigen, mit denen sie verbunden sind. Neben Bindung und Mitgefühl für andere entwickeln Menschen und andere Säugetiere komplexere soziale Beziehungen und sind in der Lage, die Handlungen anderer zu verstehen und vorherzusagen. Sie verinnerlichen auch soziale Praktiken und erleben den ausgelösten sozialen Schmerz durch Trennung, Ausschluss oder Missbilligung. Infolgedessen haben Menschen ein intrinsisches Gerechtigkeitsgefühl.

Das Gleiche gilt weitgehend auch für nichtmenschliche Säugetiere. Bekoff und Pierce (2009) liefern viele Beispiele für Beweise für das amoralische Gerechtigkeitsempfinden von Säugetieren. Zum Beispiel Kapuzineraffen

Bei der Arbeit bei Leckereien wirkten sie beleidigt und verweigerten die weitere Zusammenarbeit, wenn sie sahen, dass ein anderer Affe eine wünschenswertere Belohnung für die gleiche Arbeit erhielt (Brosnan und de Waal, 2003).

Im Gegensatz dazu sind Roboter nicht um ihre eigene Selbsterhaltung oder die Vermeidung von Schmerzen besorgt, geschweige denn um die Schmerzen anderer. Teilweise lässt sich dies damit erklären, dass sie nicht wirklich in der Art und Weise verkörpert sind, wie ein lebendes Wesen. Der lebende Körper ist eine integrierte autopoetische Einheit (Maturana und Varela, 1980), was eine von Menschen geschaffene Maschine nicht ist. Natürlich lässt sich argumentieren, dass der Roboter so programmiert werden könnte, dass er sich so verhält, dass er sich um seine eigene Erhaltung oder die anderer kümmert, aber dies ist nur durch menschliches Eingreifen möglich. Wir kommen weiter unten auf eine weitere Diskussion der Machbarkeit der Programmiermoral zurück.

Gesellschaftliche Verantwortung: Viele Autoren würden der Aussage zustimmen, dass Roboter keine vollständigen moralischen Agenten sind. Johnson und Miller (2008) argumentieren, dass Roboter und andere Computerartefakte keine vollständigen moralischen Agenten sind, weil sie „niemals völlig unabhängig von ihren menschlichen Designern“ sind. Sie beschreiben sie als „an Menschen gebundene“ Artefakte und argumentieren, dass Verantwortung nicht ausgeschlossen werden kann. Sie werden auf die Tatsachen selbst geladen, da das Verhalten und die Ergebnisse von Robotern und Computersystemen notwendigerweise von menschlichen Designern und Entwicklern abhängen. Ein nützliches Beispiel, das sie betrachten, ist das eines Türöffners. Ein mechanischer Türöffner wäre nicht möglich.

Es wird davon ausgegangen, dass sie eine lobenswerte Tat vollbracht haben. Ähnliche Argumente über mangelnde Unabhängigkeit von menschlichen Designern wurden in der Vergangenheit vorgebracht, basierend auf der Art und Weise, wie Roboter, anders als lebende Maschinen, niemals als vollständig verkörpert angesehen werden können, da sie bei ihrer Entwicklung immer menschliches Eingreifen und Beteiligung erforderten (Sharkey und Ziemke, 2001). Der Punkt hier ist, dass Roboter und ihre Untergebenen sind lügende Kontrollsysteme, sind auf menschliches Eingreifen angewiesen. Die Roboter sind vielleicht „losgelassen“, um unvorhersehbare Entscheidungen zu treffen, aber die Entscheidung, ihnen das zu erlauben, ist ein Mensch und die Gesellschaft allein. Alle Entscheidungen, die der Roboter trifft, hängen immer noch von seinem ursprünglichen Entwurf ab. Verantwortung wird übernommen und anerkannt. Johnson (2006) unterscheidet sinnvoll zwischen moralischen Agenten und moralischen Einheiten und ordnet Roboter und Computerartefakte in die zweite Kategorie ein. Zu den moralischen Einheiten gehören der Herz-Fakten-Designer, der Herz-Fakten und der Herz-Fakten-Benutzer, und moralische Verantwortung kann nicht auf die Herz-Tatsache selbst abgewälzt werden.

(b) Ist es ausreichend, Roboter so zu entwerfen, dass sie den bestehenden Gesetzen und Grundrechten entsprechen? Freiheiten, einschließlich Privatsphäre?

Ein großes Problem beim Vorschlag, dass Roboter so konzipiert werden sollten, dass sie mit bestehenden Gesetzen und Grundrechten und Grundfreiheiten im Einklang stehen, und der Grund dafür, dass dies nicht ausreicht, besteht darin, dass bestehende Gesetze und Menschenrechte nicht unter Berücksichtigung technologischer Entwicklungen wie der Robotik formuliert wurden. Sie sind darauf ausgelegt, als Freunde und Begleiter aufzutreten, und sind deshalb in unseren Häusern und in unserer intimen Umgebung willkommen. von Robotern mit wenig menschlichem Kontakt (z. B. Sharkey und Sharkey, 2012; Sparrow und Sparrow 2006), aber das Menschenrechtsgesetz sieht keinen ausdrücklichen Schutz vor einer solchen Situation vor. Ähnliche Bedenken

Es wurde darüber gesprochen, Kinder in der Obhut von Robotern zu lassen, soweit ihre Bindung an Menschen gefährdet ist (Sharkey und Sharkey 2010), aber auch hier gibt es keine Gesetze oder Rechte, die eine solche Möglichkeit ausdrücklich verhindern, außer denen, die mit der Vernachlässigung von Kindern verbunden sind wenn Menschen Roboter in Positionen bringen, die Macht über Menschen haben.

Wenn Menschen Entscheidungen darüber treffen, wie sie sich in sozialen Situationen verhalten sollen, müssen sie mehr tun, als sich an Regeln oder Gesetze zu halten. Sie treffen Entscheidungen auf der Grundlage eines moralischen Verständnisses darüber, was für sie unangemessen oder angemessen ist. Sie reagieren sensibel auf Feedback zu ihren Entscheidungen und deren Ergebnissen und können darüber nachdenken und ihre zukünftige Entscheidungsfindung anpassen. Treffen Sie in sozialen Situationen die richtigen moralischen Entscheidungen. Arkin (2009) hat beispielsweise argumentiert, dass Robotersoldaten in einer Schlachtfeldsituation so programmiert werden könnten, dass sie Regeln befolgen, die zu ethischerem Verhalten führen würden, als das manchmal von menschlichen Soldaten in der Hitze des Gefechts gezeigt wird. Er behauptet, dass menschliche Soldaten aufgrund ihrer Gefühle schlecht handeln können – zum Beispiel durch Rache motiviert Kriegsverbrechen zu begehen. Ein Roboter hingegen würde nicht emotional reagieren und könnte durch einen ethischen Gouverneur dazu programmiert werden, Handlungen zu bewerten, bevor er sie ausführt, und nur solche auszuführen, die als moralisch zulässig erachtet werden.

Verschiedene Autoren haben sich gegen die Idee ausgesprochen, Roboter so programmieren zu können, dass sie moralische Entscheidungen treffen. Im Zusammenhang mit autonomen Waffen argumentierte Christ of Heyns, der UN-Sonderberichterstatter für außergerichtliche, zusammenfassende oder willkürliche Hinrichtungen, gegen den Einsatz autonomer Roboter, um tödliche Entscheidungen auf dem Schlachtfeld zu treffen, mit der Begründung, dass es Robotern an menschlichem Urteilsvermögen, gesundem Menschenverstand und Wertschätzung mangelte größeres Bild, Verständnis der Absichten hinter den Handlungen der Menschen, Verständnis von Werten und Vorwegnahme der Richtung, in die sich Ereignisse entwickeln“ (2013, [A/HRC/23/47](#)). Der Punkt ist, dass die unvorhersehbare Vielfalt sozialer Situationen, die auf dem Schlachtfeld entstehen könnten, bedeutet, dass es unwahrscheinlich ist, dass eine Reihe vorprogrammierter Regeln für angemessene Reaktionen wahrscheinlich anwendbar sind.

Malle und Scheutz (2014) argumentieren in einem interessanten Aufsatz über die Anforderungen an die Entwicklung von Robotern mit der sogenannten „moralischen Kompetenz“, dass Roboter unter anderem ein Netzwerk moralischer Normen benötigen würden, um zu wissen, was moralisch akzeptabel ist und was nicht. Sie könnten ein Netzwerk moralischer Normen erlernen und entwickeln, das auf der Rückmeldung basiert, die ihnen als Reaktion auf ihre Handlungen gegeben wird. Sie deuten darauf hin, dass es notwendig sein könnte, Roboter in menschlichen Umgebungen zu erschaffen, da dies vielleicht „die einzige Möglichkeit ist, sie dem Reichtum menschlicher moralischer Situationen und kommunikativer Interaktionen auszusetzen“ (Malle und Scheutz, 2014). Andere haben darauf hingewiesen, dass Roboter das Verständnis von richtig und falsch haben könnte verbessert werden, indem man ihnen moralische Geschichten beibringt (RiedlandHarrison, 2016) und verlangt von ihnen, die menschlichen Werte, die sie haben, umzukehren vertreten.

Es ist zugegebenermaßen schwierig auszuschließen, dass ein Roboter in der Zukunft zu Moral trainiert oder erzogen werden könnte, aber es gibt eine Reihe von Gründen, warum man skeptisch gegenüber der Erfolgswahrscheinlichkeit ist. Zu den Gründen für die Skepsis gehört das Fehlen einer biologischen Grundlage und Formmoral des Roboters, wie zuvor erläutert. Wie bereits besprochen, kümmert sich ein einzelner Roboter nicht einmal um seinen eigenen Körper, geschweige denn um den eines Menschen – er würde keine Schmerzen erleiden, wenn zum Beispiel seine Räder entfernt würden von Robotern, die ein gutes, verallgemeinerbares Verständnis für die Unterschiede zwischen richtig und falsch entwickeln. Derzeit gibt es nur Beispiele für *programmierte* Roboter

Verhalten, wie zum Beispiel die von Winfield et al. (2014) programmierten Roboter, um Maßnahmen zu ergreifen, um zu verhindern, dass andere Roboter in ein Loch fallen, die als etwas beschrieben werden, das als ethisches Verhalten beschrieben werden kann. Die Verwendung des Begriffs „ethisch“ oder „moralisch“ in diesem Zusammenhang bedeutet jedoch nicht, dass die betreffenden Roboter könnten für ihre Taten gelobt oder beschuldigt werden.

(c) Wenn Roboter keine verantwortlichen Agenten sind, sollte dies die ihnen zugewiesenen sozialen Rollen und die Situationen, in denen sie eingesetzt werden, einschränken?

In der ursprünglichen Aussage, dass Roboter keine verantwortlichen Agenten seien, wird nicht dargelegt, was dies für den Einsatz von Robotern bedeutet. Es wird hier argumentiert, dass es gute Gründe gibt, die sozialen Rollen und Entscheidungsbefugnisse von Robotern einzuschränken. Wie oben erwähnt, argumentierte Heyns (2013), dass es Robotern nicht erlaubt sein sollte, Tötungsentscheidungen im Kampf zu treffen, teilweise aufgrund ihrer mangelnden Fähigkeit, sie zu verstehen soziale Situationen, sondern auch, weil Menschen ein Recht darauf haben sollten, dass Entscheidungen über Leben und Tod von Mitmenschen getroffen werden. Ein ähnliches Argument könnte auch über Roboterpolizisten vorgebracht werden, die möglicherweise auch abseits des Schlachtfelds über Leben und Tod (oder schwere Verletzungen) entscheiden müssen.

Dieses Argument kann und sollte meiner Meinung nach auch auf andere Arten von Entscheidungen ausgeweitet werden, bei denen Roboter die Freiheiten von Menschen einschränken könnten. Ein Roboter in der Rolle eines Lehrers müsste Entscheidungen darüber treffen, wann er Kinder zur Uni bringen oder belohnen soll und davon abzuhalten, etwas Gefährliches oder Riskantes zu tun. Ein Roboter-Nanny müsste eine ähnliche Entscheidung über seine jungen Schützlinge treffen. Der Punkt ist, dass alle diese Entscheidungen wahrscheinlich emotionale Urteile und Bewertungen sozialer Situationen beinhalten und aus bereits besprochenen Gründen wahrscheinlich nicht in der Lage sein werden, gute Entscheidungen zu treffen. Es sollte darauf geachtet werden, die menschliche Kontrolle, Beteiligung und Verantwortung bei Entscheidungen, die sich auf das Leben von Menschen auswirken, aufrechtzuerhalten. Es bestehen bereits Risiken automatisierter Entscheidungen, die sich auf unser Leben auswirken, aber Roboter, die als kompetente soziale Akteure erscheinen können, machen diese Risiken noch größer.

Zusammenfassung: Es ist leicht, dem EPSRC-Prinzip zuzustimmen, dass Roboter keine verantwortlichen Agenten sind, aber selbst diese kurze Überlegung reicht aus, um zukünftiges Handeln zu steuern

Dies hat Auswirkungen auf den Einsatz von Robotern und auf menschliche Entscheidungen über die sozialen Rollen, die ihnen gegeben werden sollten. Roboter, die darauf programmiert sind, dem Gesetz zu folgen und die Rechte und Freiheiten des Einzelnen zu respektieren, werden soziale Situationen nicht verstehen und nicht in der Lage sein, konsequent die richtigen moralischen Entscheidungen über menschliche soziale Situationen zu treffen. Es ist darauf zu achten, dass in allen Situationen, in denen menschliches Urteilsvermögen erforderlich ist, eine automatische und algorithmische Entscheidungsfindung vermieden oder minimiert wird. Noch größere Sorgfalt ist bei Robotern erforderlich, die die Illusion erzeugen, die sie verstehen.

Verweise

Arkin, R. (2009). Governing lethal behavior in autonomous robots. Chapman-Hall review. *Computers and Education*, 58(3), 978–988.

Bekoff, M. und Pierce, J. (2009) *Wild Justice: The Moral Lives of Animals*. The University of Chicago Press, London.

Brosnan, S. F. und de Waal, F. B. (2003). Affen lehnen ungleiche Bezahlung ab. *Natur*, 425, 297–99

Churchland, P. (2011) *Braintrust: Was die Neurowissenschaften uns über Moral erzählen*. PrincetonUniversityPress, Oxford.

Heyns, C. (2013). Bericht des Sonderberichterstatters für außergerichtliche, zusammenfassende oder willkürliche Hinrichtungen, A/HRC/23/47

Johnson,DG(2006).ComputerSystems:MoralEntitiesbutnotMoralAgents. *Ethik und Informationstechnologie*, 8(4):195–204

Johnson, DG und Miller, KW (2008)Un-makingartificialmoralagents. *Ethik und Information Technologie* (2008)10:123–133

Malle, BF, & Scheutz, M. (2014).Moralcompetenceinsocialrobots.IEEEInternationalSymposiumonEthicsinEngineering,Science,andTechnology(S.30–35).Präsentiert beimIEEEInternationalSymposiumonEthicsinEngineering,Science,andTechnology,Juni,Chicago,IL:IEEE.

Maturana, HR & Varela, FJ (1980). *Autopoiesis und Kognition – Die Verwirklichung des Lebendigen*. Dordrecht, Niederlande: D.ReidelPublishing

Riedl, MO, und Harrison, B. (2016) Mithilfe von Geschichten können künstliche Agenten menschliche Werte vermitteln. In Proceedings of 2nd International Workshop on AI, Ethics and Society, Phoenix, Arizona

Sharkey, AJC, & Sharkey, NE (2012). Grannyandtherobots: Ethische Fragen bei der Roboterpflege für ältere Menschen. *Ethik und Informationstechnologie*, 14(1),27–40.

Sharkey, NE, & Sharkey, AJC (2010). Thecryingshameofrobotnannies: Anethicalappraisal. *InteractionStudies*, 11(2),161–190.

Sharkey, NE&Ziemke, T. (2001).Mechanisticvs.PhenomenalEmbodiment-CanRobotEmbodimentLeadtoStrongAI *CognitiveSystemsResearch*, 2,4,251-262

Sparrow, R. & Sparrow, L. (2006). In den Händen von Maschinen? Die Zukunft der Altenpflege. *Mindand Maschine*, 16,141–161.

Winfield, AF, Blum, C. und Liu, W.

(2014)Towardsanethicalrobot:Internalmodels,consequencesandethicalactionsselection.InM.Mistry,A.Leonardis,M.Witkowski,&C.Melhuis *Fortschritte bei autonomen Robotersystemen: Tagungsband der 15. Jahreskonferenz, TAROS2014* (S. 85–96). Birmingham, Großbritannien, 1.–3. September

Zweite Gedanken zu Privatsphäre, Sicherheit und Täuschung

TomSorell, University of Warwick
HeatherDraper, University of Birmingham

Die fünf Prinzipien der Robotik, die während des AHRC-EPSRC-Retreats im Jahr 2010 formuliert wurden, sind nicht das letzte Wort in der Roboterethik, sondern eines der ersten Worte.

Privatsphäre

Prinzip 2 erfordert, dass Roboter in Übereinstimmung mit bestehenden Gesetzen und Grundrechten und Grundfreiheiten, einschließlich des Rechts auf Privatsphäre, betrieben werden. Aber dieses Prinzip ist schwer einzuhalten, weil es zwischen Galakademikern und Philosophen Uneinigkeit darüber gibt, was Privatsphäre ist und auch darüber, ob es ein Grundrecht ist. Obwohl Privatsphäre ein Grundrecht in der Europäischen Charta der Grundrechte sein soll (Die Artikel 7 und 8 sind in älteren Menschenrechtsverträgen wie dem Internationalen Pakt über bürgerliche und politische Rechte (siehe Artikel 17) nicht von grundlegender Bedeutung leugnen, dass es eine Hierarchie der Menschenrechte gibt, in der einige grundlegender sind als alle anderen. Instrumente, die den Status der Menschenrechte interpretieren, wie die Wiener Erklärung, besagen, dass alle Menschenrechte voneinander abhängig und unauflösbar sind (Art. 5). Selbst wenn also eine Einigung darüber besteht, was Privatsphäre ist, wäre die Notwendigkeit, ein Recht auf Privatsphäre zu respektieren, nicht unbedingt vorrangig.

Persönliche Privatsphäre wird manchmal als Kontrolle über Informationen über sich selbst verstanden. Pflegeroboter im Besonderen und soziale Roboter im Allgemeinen sind oft darauf ausgelegt, Informationen über die Menschen zu sammeln, mit denen sie interagieren. Roboter überwachen beispielsweise Gesichtsausdrücke und verarbeiten gesprochene Sprache, verfolgen den Standort von Menschen, mit denen sie interagieren, sammeln Informationen über ihre täglichen Routineaktivitäten und so weiter. Dies stellt nicht unbedingt eine Verletzung des Rechts auf Kontrolle über die eigene Privatsphäre dar Informationen, denn die Personen, deren Informationen sie enthalten, können der Erhebung und Speicherung ihrer Informationen grundsätzlich zustimmen. Da die Verwendung ihrer Informationen ihrer Zustimmung unterliegt, geht die Kontrolle nicht in die Hände

Die Einwilligung klärt jedoch nicht unbedingt alle Fragen zur ordnungsgemäßen Verwendung personenbezogener Daten. Erstens besteht ein Unterschied zwischen der einmaligen oder intermittierenden Erhebung von Informationen und der mehr oder weniger kontinuierlichen Erhebung von Informationen in Echtzeit. Die Auswirkungen der zweiten sind schwerer vorherzusagen und im Voraus zu genehmigen als die Implikationen des Ersten. Es lässt sich sogar argumentieren, dass es keine ordnungsgemäß informierte Zustimmung zur kontinuierlichen Überwachung und Nachverfolgung eines lebenden Pflegeroboters gibt, gerade weil es nicht möglich ist, im Voraus vorherzusagen oder sich vorzustellen, wie das Leben mit einem Roboter aussehen würde.

Anstatt die Grenzen der Privatsphäre nur anhand dessen zu bestimmen, was der Benutzer mit guten Informationen zustimmt, muss man sich möglicherweise auch mit Argumenten über die Grenzen der Privatsphäre auf der Grundlage der Informationen auseinandersetzen

Designaufgabe und Zweck einer bestimmten Roboterart. Pflegeroboter für ältere Menschen sollen oft dabei helfen, ihre Autonomie aufrechtzuerhalten – das heißt, ihre Fähigkeit zur Wahl und über Fähigkeiten zu verfügen – die Fähigkeit, sich zu waschen, zu putzen, zu kochen, sich zu ernähren usw. –, die ausreichen, um unabhängig zu leben. Eine autonome Person entscheidet nicht nur selbst, was sie tut und wie sie lebt, sondern auch, welche persönlichen Informationen sie preisgeben möchte. A Ein Roboter, der darauf ausgelegt ist, die Autonomie eines älteren Menschen aufrechtzuerhalten, kann teilweise daran gemessen werden, ob der ältere Mensch genauso viel Spielraum hat wie ein nicht unterstützter Erwachsener, um über alle Aspekte seines Lebens zu entscheiden, einschließlich der Offenlegung von Informationen.

Die Spannung zwischen Autonomie (und Privatsphäre) und Sicherheit

Eine Art und Weise, wie die Standardübung eines Erwachsenen seine Autonomie ausübt, besteht darin, selbst zu beurteilen, welche Risiken er eingehen möchte. Die moralische Zulässigkeit des Eingehens von Risiken wird natürlich durch die Kosten beeinflusst, die anderen entstehen. Wenn andere in Gefahr geraten oder man sich darauf verlässt, dass sie eine gefährliche Rettung des autonomen Risikoträgers vornehmen, dann kann es ein Argument gegen das Eingehen von Risiken aus der autonomiebeschränkenden Belastung geben, die es anderen auferlegt. Mit anderen Worten, wenn wir unsere Pläne aufgeben oder die Verwirklichung unserer Entscheidungen verzögern müssen, weil jemand anderes ein Risiko eingegangen ist, dann ist unsere Autonomie ihm untergeordnet zumindest vorübergehend.

Im Falle der betreuten älteren Menschen wird der Entwurfsauftrag des Roboters in der Regel Sicherheit *und* Sicherheit vereinen Autonomie. Der Roboter unterstützt Benutzer bei der Führung ihres eigenen Lebens und überwacht außerdem den Benutzer und seine Umstände im Hinblick auf gesundheitliche Notfälle.

In den Fällen, die aus moraltheoretischer Sicht am interessantesten sind, ist der Benutzer bereit, ein relativ geringes Risiko einzugehen – etwa das Risiko, einen kleinen Sturz zu erleiden –, um sein tägliches Leben weiterhin auf die gleiche Weise führen zu können, wie er es getan hat, als er jünger war. Entlasten Sie den Austausch von Informationen über den Fall. Dies liegt an der Aufrechterhaltung

Autonomie soll der vorrangige Zweck des Begleitroboters sein. Wenn jedoch das vorrangige Ziel darin besteht, die Sicherheit des Benutzers zu gewährleisten, haben sie möglicherweise nicht dieses Recht gestellt, um zu verlängern. In der Low-Tech-Telepflege wird ein Anhängeralarm vom Benutzer getragen, und es liegt an ihm oder ihr, zu entscheiden, ob er Hilfe ruft oder nicht. Der Wert der Autonomie unterstützt diese Norm.

Benutzerüberschreibungen könnten auch in das Companion-Robot-Design integriert werden, wo die Wahl des Lebensstils eines Benutzers, wenn er Freunden und Verwandten gemeldet wurde, zu Zwangsinterventionen dieser Personen führen könnte. Hier unterstützt Autonomie die Privatsphäre gegenüber völliger Sicherheit oder völliger Besonnenheit. Ältere Menschen; andernfalls schränkt eine Art Altersdiskriminierung die Autonomie dieser Menschen ein und sie werden von Roboterdesignern und politischen Entscheidungsträgern schlechter behandelt als andere *Erwachsene*

Person ohne Hilfe. Und diese Einschränkung ist ohne Altersdiskriminierung schwer zu verteidigen. Dies gilt unabhängig davon, ob es ein Argument dafür gibt, dass Glücksspiel in jedem Alter unsicher ist. Der Punkt ist, dass Autonomie sogar den harmlosen Datenaustausch und harmlose Interventionen zur Verhinderung des Eingehens von Risiken überwiegen kann.

Die Spannung zwischen Autonomie und Drehfähigkeit

Pflegroboter und einige nicht-soziale Roboter sollen älteren Menschen *dabei helfen*, ihre verlorenen Fähigkeiten wiederzuerlangen, und nicht nur die ihnen zur Verfügung stehenden Fähigkeiten selbstständig trainieren. Eine Mitwirkungsverpflichtung kann bestehen, wenn der Nutzer und derjenige, der für die Einführung des Roboters bei ihm zu Hause bezahlt, eine Mitwirkungspflicht trifft. Wird keine solche Mitwirkung anerkannt, muss Raum für eine eigenständige Verweigerung der Annahme einer förderlichen Rehabilitation geschaffen werden. Schließlich kann eine eigenständige Verweigerung der Annahme eines *medizinischen* Eingriffs rechtlich nicht ignoriert werden.

Was ist, wenn die Hilfe eines Roboters für einen älteren Menschen unter der Bedingung bereitgestellt wird, dass dieser sich damit einverstanden erklärt, mit der möglicherweise in der Zukunft angebotenen Rehabilitation zusammenzuarbeiten? Vorbehaltlich eines ausdrücklichen Vertrags, den der Benutzer abschließt und der die Verantwortlichkeiten festlegt, die der Benutzer als Gegenleistung für die Bereitstellung des Roboters übernimmt. Ein solcher Vertrag kann zwischen dem Benutzer und der örtlichen Behörde bestehen.

Die vertraglichen Pflichten schließen Rechte selbstverständlich nicht aus.

Täuschung

Schließlich kommen wir zu Prinzip 5. Es fordert Transparenz beim Roboterdesign und verbietet die Täuschung der Schwachen. Unsere Skepsis gegenüber diesem Prinzip basiert auf der niedrigen Täuschungsschwelle, die in der Literatur zur Roboterethik festgelegt ist. Täuschung ist die absichtliche Erschaffung falscher Überzeugungen. Täuschung ist normalerweise falsch, weil der Betrüger die getäuschte Person manipulieren will, um etwas zu tun, das den Interessen des Betrügers dient. Der Roboter selbst hat nicht die Absicht, zu täuschen, aber ist sein kindliches Design trügerisch? Funktioniert er, indem er ein Kind denken lässt, dass ein anderes Kind ihm oder ihr hilft? Die Anthropomorphisierung ist kein Fall von Selbsttäuschung, und auch die Modellierung des pädagogischen Roboters an ein humanoides Kind ist kein Fall von Täuschung. Sich mit einem solchen Roboter zu identifizieren oder ihn liebevoll zu behandeln, bedeutet nicht, ein anderes Kind zu behandeln.

Es könnte sein, dass man Atitas als Darstellung eines Kindes behandelt. Eine ähnliche Schlussfolgerung lässt sich über den Par-Roboter ziehen. Paro ist zu streicheln, behandelt Atitas nicht als lebendiges Siegel, sondern als eine Art Siegeldarstellung. Dabei handelt es sich weder um eine Täuschung seitens des Herstellers von Paro noch um eine Selbsttäuschung. Patienten mit Demenz scheinen Paro und auch nicht-roboterhafte Puppen auf die gleiche Weise zu verwenden, wie sehr kleine Kinder Stofftiere verwenden. Ohne in der Lage zu sein, fantasievolle Geschichten zu konstruieren, in denen die Puppen figurieren, fühlen sich diese Benutzer durch das Aussehen und die Atmosphäre v

Die Roboter sind eine beruhigende Präsenz, vielleicht auch ein Hund oder eine Katze. Für den Trost von Paro ist es nicht entscheidend, dass er denkt, dass sie ein echtes Siegel sind, oder dass man sie für lebendig hält – die Tatsache, dass ihr Verhalten das eines Apes simuliert, reicht aus. Daher ist es schwer zu verstehen, welcher falsche Glaube, geschweige denn welcher *absichtlich erzeugte* falsche Glaube, für Paros therapeutische Wirkung ausschlaggebend ist. Die gleiche Schlussfolgerung scheint streng jeweils durch paralleles Denken für h

Kommentar zum AISB-Workshop zu Prinzipien der Robotik

Emily C. Collins

Die University of Sheffield, Sheffield, Großbritannien.

1. Einleitung

Das folgende Prinzip zielt auf die Regulierung von Robotern in der realen Welt ab: Nr. 4. Roboter sind hergestellte Artefakte. Sie sollten nicht in einer irreführenden Weise konzipiert sein, um gefährdete Benutzer auszunutzen. Stattdessen sollte ihre Maschinennatur transparent sein.

Dieser Kommentar bietet eine Kritik dieses Prinzips anhand der folgenden Kriterien: a.

Gültigkeit. Sind die Prinzipien als Aussagen über die Natur von Robotern (z. B. dass es sich um Werkzeuge und Produkte handelt), Roboterentwickler und die Beziehung zwischen Robotern und Menschen (z. B. dass Roboter ein transparentes Design haben sollten) korrekt oder sind sie ontologisch fehlerhaft? , ungenau, veraltet oder irreführend.

Die Kritik wird das Prinzip auf die meiner Meinung nach zwei Hauptbestandteile herunterbrechen: 1.

Roboter sollten nicht auf betrügerische Weise entwickelt werden, um gefährdete Benutzer auszunutzen.

2. Die Natur der Maschine sollte transparent sein.

Ich werde argumentieren, dass beide Teilaussagen, aus denen dieses Prinzip besteht, aufgrund der undefinierten Natur der kritischen Begriffe „trügerisch“, „verletzlich“ und „Maschinennatur“ grundsätzlich fehlerhaft sind und dass das Prinzip als solches als Ganzes gilt ist irreführend.



Abb. 1. Linkes Feld: „Ekso“, die Abkürzung für Exoskelett, ist ein tragbarer Roboter, der gelähmten Patienten beim Gehen hilft. Rechtes Bild: Zwei MIRO-Robotersäugetiere, ein Beispiel für einen „sozialen“ Roboter.

Für die Zwecke dieses Kommentars wird ein Roboter als eine hergestellte künstliche Tatsache definiert, insbesondere als ein Werkzeug, mit dem ein menschlicher Benutzer einen bestehenden Zustand verbessern kann, indem er beispielsweise einer Person, die nicht gehen kann, mithilfe maschineller Hilfe die Fähigkeit zum Gehen verschafft. oder indem einem Benutzer eine fortschrittliche Form der Unterhaltung geboten wird, beispielsweise mit einem Begleitroboter (Abbildung 1). Dieser Kommentar konzentriert sich insbesondere auf biomimetische [1], soziale Roboter und ihre Rolle als Werkzeuge für die Wette von Nutzern. Ein sozialer Roboter wird hier als ein Gerät mit einer gewissen Autonomie und physischer Präsenz definiert, das in der Lage ist, sozial mit Menschen zu interagieren, und als solches kann man davon ausgehen, dass es bei seinem Benutzer eine emotionale Reaktion hervorruft [2]. Hier liegt unser erstes Problem, bevor das Prinzip überhaupt angesprochen wird: Um „Roboter“ zu definieren, muss man zumindest die Anwendung des Roboters und den Umfang seiner Fähigkeiten definieren. Es gibt sich gegenseitig ausschließende Robotertypen, die in der Lage sind, Benutzer potenziell auf unterschiedliche Weise zu täuschen, je nachdem, wie mit jedem Roboter interagiert wird. Industrie-, Mobil-, Service-, Bildungs-, Weltraum- und soziale Roboter, um nur einige zu nennen, haben unterschiedliche Morphologien und bringen unterschiedliche Erwartungen ihrer Benutzer mit sich. Keines der

2 Roboter sollten nicht so konzipiert sein, dass sie gefährdete Benutzer ausnutzen

Beginnen wir zunächst mit der Frage: Was ist „trügerisch“? In diesem Zusammenhang wird der Roboter als betrügerisch abgestempelt. Eine bessere Frage könnte also lauten: Wieso ist der Roboter so betrügerisch, dass er gegen dieses Prinzip verstößt?

Roboter werden so entwickelt, dass sie Lebewesen ähneln. Was über die Mensch-Tier-Dynamik bekannt ist, wird genutzt, um tierähnliche Verhaltensweisen und Morphologien in die Gestaltung sozialer Roboter einzubauen. Bionik bedeutet per Definition Design durch die Natur durch Nachahmung von Modellen, Systemen und Elementen der Natur mit dem Ziel, menschliche Probleme zu lösen. Roboter sind Werkzeuge, Gebrauchsprodukte, deren Zweck es ist, menschliche Probleme zu lösen. Hier werden genau die Designprinzipien, die der Natur eines biomimetischen, sozialen Roboters zugrunde liegen und die Anforderungen von Roboterentwicklern erfüllen, durch das bestimmt, was man wohl als „Täuschung“ bezeichnen könnte: den Versuch, Lebewesen zum Wohle des Roboters und seines Benutzers nachzuahmen.

Tierähnliche Roboter wie Paro [3] und die „FurReal Friends Lulu Cuddlin Kitty“ von Hasbro (Abbildung 2) werden von Therapeuten ähnlich wie bei der tiergestützten Therapie (AAT) eingesetzt [4]. Dabei kann ein Tier in eine bestehende Therapiesitzung einbezogen werden, um bei der sozialen Erleichterung zu helfen (wie bei der Gruppentherapie), oder es kann im Einzelgespräch eingesetzt werden, um einen Klienten oder Patienten während der Therapie zu fokussieren. Diese Roboter dienen einem bestimmten Zweck: Sie sollen tierähnlich wirken und dem Therapeuten helfen. Ihre Existenz basiert zwar auf einer Behandlung, an der ein Lebewesen beteiligt ist, dient jedoch nicht dazu, Tiere zu ersetzen. Die Tiere im AAT gelten als Co-Therapeuten. Ihnen wird die Wertschätzung entgegengebracht, die von einem Lebewesen erwartet wird, und sie werden von Sitzungen ferngehalten, bei denen Schaden



Abb. 2. Linkes Bild: Der Therapieroboter „Paro“. Rechtes Feld: Die „FurReal Friends“.
„Lulu Cuddlin Kitty“.

oder dass sie selbst störend wirken [5]. Dieses Beispiel zeigt, dass ein Roboter mit der Absicht entworfen wurde, zu täuschen und tierähnlich auszusehen die Absicht, von gefährdeten Bevölkerungsgruppen – Einzelpersonen in Therapie – verwendet zu werden Es ist nicht dazu gedacht, ausbeuterisch im Sinne des Versuchs zu sein, einen Benutzer davon zu überzeugen Der tierähnliche Roboter ist wirklich lebendig. Stattdessen werden sie zur assoziativen Auslösung verwendet Erinnerungen an andere Lebewesen. Es ist schwierig, diese Idee zu vermitteln, aber die Nuancen sind wichtig. Diese Roboter sind nicht dafür gebaut, überzeugende Tiere zu sein. Sie werden gebaut, um überzeugende Roboterwerkzeuge zu sein, und um das zu erreichen, sind es Ideen aus der Natur geliehen.

Zweitens: Was bedeutet es, „gefährdete Benutzer auszunutzen“? Was ist ein verletzlicher Benutzer? Ist Verletzlichkeit ein einzelner Seinszustand? Und wenn ja, zu welchem Zeitpunkt möglicherweise kann man als gefährdet gelten oder nicht mehr als gefährdet gelten? In der Tat, wer könnte darüber entscheiden Ab diesem Zeitpunkt wurde ein Individuum verletzlich genug, um seinen Zustand zu verlieren Kunstroboter von ihnen übernommen? ¹

Innerhalb der Medizin gibt es eine standardisierte Definition gefährdeter Gruppen bei denen es definierte Schwachstellenbereiche gibt (z. B. [7]). Wie eine verletzbare Person von betrügerischen Robotern ausgenutzt wird, hängt davon ab, wo die Verwundbarkeit liegt das Individuum lügt. Zu den medizinischen Bereichen gehört beispielsweise auch die wirtschaftliche Gefährdung. Denken Sie an die emotionale Ausnutzung der Angst durch populistische Medien Dies verbreitet den Glauben, dass der Arbeitsplatz einer Person durch Roboter gefährdet sein könnte die fälschlicherweise als fortgeschrittener dargestellt werden, als sie es sind. Obwohl Wir können davon ausgehen, dass sich das Prinzip nicht auf eine solche Verwundbarkeit als wirtschaftlich bezieht (obwohl wir das in Wahrheit nicht annehmen können; Teil des Problems bei diesen). Zu den Prinzipien der Robotik gehört, dass sie überhaupt nicht so definiert sind, sondern für die Für die Zwecke dieses Kommentars gehen wir davon aus, dass es sich um die Schwachstelle handelt, auf die verwiesen wird ist eher physisch als konzeptionell). Nehmen wir also vielleicht an, dass sich der Grundsatz mit „gefährdet“ auf Gruppen bezieht. Nehmen wir auch an, dass es sich um einen allgemeinen Benutzer handelt Ich weiß, wann ein Roboter ein Roboter ist, es sei denn, dieser Roboter ist so außergewöhnlich lebensecht

¹ Ein Beispiel hierfür in der Belletristik finden Sie in Issac Asimovs frühester Veröffentlichung, Robbie [6]. Die Angst vor der Ausbeutung schutzbedürftiger Menschen durch Roboter besteht seit langem Robotik-Community, aber um dieses Thema zu würdigen, muss die Fiktion von den Fakten getrennt werden.

als lebendig durchzugehen. Um als lebend zu gelten, müsste der Roboter sich synchron bewegen, reagieren, blinzeln, atmen und vokalisieren, und diese Liste erhebt keinen Anspruch auf Vollständigkeit, außerdem muss er morphologisch exakt sein. Eine solche Technologie gibt es nicht. Betrachtet man also den Stand der Technik, der derzeit existiert, wie beispielsweise die sozialen Roboter, die im Mittelpunkt dieses Kommentars stehen, ergibt sich das Problem aus der Tatsache, dass es gerade die Schwächsten innerhalb einer Bevölkerung sind, die am meisten von ihnen profitieren können verwenden. Als die beiden am stärksten gefährdeten Gruppen gelten im Allgemeinen ältere Menschen und Minderjährige sowie innerhalb dieser Gruppen Personen mit kognitiven Beeinträchtigungen.

Für die Zwecke dieses Kommentars konzentrieren wir uns auf gefährdete Gruppen innerhalb der älteren Bevölkerung. Der oben erwähnte Paro-Roboter ist ein fortschrittlicher interaktiver Roboter, der Kranke und ältere Menschen körperlich und emotional unterstützen soll – nicht allein, sondern mit Hilfe eines in robotergestützter Therapie (Robot-Assisted Therapy, RAT) ausgebildeten klinischen Praktikers. Bei Personen, die an Demenz oder anderen Erkrankungen mit kognitivem Verfall leiden, nimmt die emotionale Leistungsfähigkeit nicht eins zu eins mit der Kognition ab [8]. Dies ermöglicht eine sinnvolle Anwendung psychologischer und emotionaler Therapie durch einen Therapeuten mit statischen Geräten wie Paro, die einem Lebewesen ähneln, das man halten und um das man sich kümmern kann [9]. Hier ähnelt die Täuschung der bei der Puppentherapie.

Bei Puppentherapie-Interventionen werden Puppen, die lebensechten Babys ähneln, von Betreuern der Alzheimer-Krankheit verwendet, um Ängste zu lindern und Menschen mit Demenz Freude zu bereiten. Dies wird durch die Einführung einer zielgerichteten und lohnenden, aber körperlich harmlosen Tätigkeit erreicht: nämlich der Pflege der Puppe (z. B. [10]). Obwohl umstritten [11], werden solche Therapien, die lebensechte Fokuspunktwerkzeuge in den Pflegeprozess einführen, für die Verbesserung der Lebensqualität (QoL) von Patienten gelobt, und zu diesen Studien gehören auch solche, die die Auswirkungen des Einsatzes tierähnlicher Roboter in der Therapie untersucht haben auch [12].

Lebensqualität ist eine komplexe Messung, die emotionale, soziale und physische Aspekte des Lebens eines Menschen umfasst. Es existiert auf einem Kontinuum außerhalb des Bereichs der „Entweder/Oder“-Dichotomien, in dem x als schlecht und y als gut angesehen wird. Wenn ein Roboterwerkzeug bei einer gefährdeten Bevölkerungsgruppe eingesetzt wird, die über geistige Fähigkeiten verfügt, die genutzt werden können, um das Leiden einzelner Personen in dieser Bevölkerungsgruppe zu lindern, wird die Frage, ob dieses Werkzeug existieren sollte oder nicht, vage und zu komplex, um sie zu beantworten mit einer einzigen Aussage. Bei der Debatte geht es darum, wie weit wir die Schwachen täuschen sollten und ab wann dies im negativen Sinne ausbeuterisch wird. Wenn man diese Überlegung mit Verbesserungen der Lebensqualität von Personen vergleicht, die an unheilbaren neurodegenerativen Erkrankungen leiden, wird deutlich, dass dieser vierte Grundsatz unzureichend ist. Es ist grundsätzlich fehlerhaft, da seine Teilbegriffe undefiniert sind. Ohne zu wissen, was eigentlich mit der Ausbeutung der

Wenn die Sache, die ausgenutzt wird, der kognitive Verfall selbst ist und der Roboter von der Verletzlichkeit des Individuums profitiert, aber für das gezielte Ergebnis der Verbesserung der Lebensqualität dieses Individuums, ist das nicht positiv? Wenn da

Gibt es keine andere Alternative, um auf die Überreste des emotionalen Quotienten eines Demenzkranken zuzugreifen, auf jemanden, der sonst möglicherweise von einem ansonsten beruhigenden lebenden Tier ängstlich ausgelöst wird? Wo liegt der wahre Schaden? Liegt der Schaden in den Köpfen derer, die nicht leiden und nicht Zeuge dessen werden, was sie selbst nachdenklich als traurigen Zustand betrachten? Und wenn das der Fall ist, sollten wir nicht umso mehr versuchen, uns in die Gedanken der Verwundbaren hineinzusetzen und diese Situation als das zu würdigen, was sie ist? Ein Versuch, mit allen verfügbaren Mitteln Pflege zu leisten, die mit gutem Willen durchgeführt und von Betreuern überwacht wird, die das volle Ausmaß des Schadens kennen, den neurodegenerative Erkrankungen sowohl für die Patienten als auch für ihre Angehörigen, die zuschauen, verursachen.

3 Die Natur der Maschine sollte transparent sein

Betrachten wir die gesunde Bevölkerung, die Roboter beobachtet. Wie bereits in diesem Kommentar erwähnt, glaube ich, dass es keine Robotertechnologie gibt, die vollkommen täuscht. Selbst die fortschrittlichsten Roboter sind eindeutig Roboter. Ein Benutzer mag glauben, dass die KI eines Roboters fortgeschrittener ist, als sie auf den ersten Blick ist, aber zumindest anekdotisch durch meine eigenen Erfahrungen im Labor glaube ich, dass jede Zeitspanne mit einem Roboter ausreicht, damit ein Benutzer ein grobes Bild erstellen kann. Eine ausreichende Annäherung an seine Grenzen, so dass jede anfängliche Überschätzung der Fähigkeiten des Roboters bald von der Realität überlagert wird. Was die Bevölkerungsgruppen angeht, die anfällig genug sind, um getäuscht zu werden und glauben zu lassen, ein Roboter sei fortschrittlicher oder „lebendiger“ als er ist, glaube ich, dass es nicht der Roboter ist, der anders gestaltet werden sollte, sondern dass es die menschlichen Benutzer oder klinischen RAT-Anwender sind, sollten darin geschult werden, ihr Werkzeug, ihr Roboterprodukt, auf die effektivste und positivste Weise zu nutzen.

4 Zusammenfassung

Ich kann mir nicht vorstellen, dass es in absehbarer Zeit einen Roboter geben wird, der so perfekt ist, dass er den Benutzer gänzlich glauben lässt, er sei alles andere als eine Maschine. Für diejenigen Personen, die verletzlich genug sind, um davon überzeugt zu sein, dass ein Roboter, der scheinbar eine Maschine ist, tatsächlich lebt, empfehle ich, alle positiven Vorteile, die sich aus einer solchen Situation ergeben können, so objektiv und umfassend wie möglich zu betrachten. Überlegen Sie, was es tatsächlich bedeutet, die Schwachen auszunutzen, und formulieren Sie vielleicht ein Szenario mit angeblich positiven Ergebnissen für den gefährdeten Benutzer um, ohne den Begriff „Ausbeutung“ zu verwenden, sondern stattdessen das Wort „Hilfe“ zu verwenden:

Roboter sind zwar hergestellte Artefakte, aber Werkzeuge, die uns helfen und die nach Prinzipien, von denen wir wissen, dass sie funktionieren, einschließlich biomimetischer Prinzipien, entworfen werden können. Bei Robotern, die auf betrügerische Weise entwickelt wurden, um das Leid schutzbedürftiger Benutzer zu lindern, sollte der Maschinencharakter bekannt gemacht werden

Betreuer dieser gefährdeten Benutzer. Möge es in der Verantwortung des Pflegepersonals liegen, die Lebensqualität seiner Patienten mit allen erforderlichen sicheren Mitteln zu verbessern.

Verweise

1. TJ Prescott, MJ Pearson, B. Mitchinson, JCW Sullivan und AG Pipe, „Whisking with robots from rat vibrissae to biomimetic technology for active touch“, IEEE Robotics and Automation Magazine, vol. 16, Nr. 3, S. 42–50, 2009.
2. EC Collins, A. Millings und TJ Prescott, „Attachment to assistive technology: A new Conceptualisation“, in Proceedings of the 12th European AAATE Conference (Association for the Advancement of Assistive Technology in Europe), 2013.
3. T. Shibata, „Mental Commit Robot (PARO)“. [Online], <http://www.paro.jp>.
4. MR Banks, LM Willoughby und WA Banks, „Tiergestützte Therapie und Einsamkeit in Pflegeheimen: Einsatz von Robotern versus lebende Hunde“, Journal of the American Medical Directors Association, vol. 9, nein. 3, S. 173–177, 2008.
5. S. Brooks, „Tiergestützte Psychotherapie und pferdegestützte Psychotherapie“, Arbeit mit traumatisierten Jugendlichen in der Kinderfürsorge, S. 196–218, 2006.
6. I. Asimov, „Strange playfellow“, Super Science Stories, S. 67–77, 1940.
7. BM Association et al., „Schutz schutzbedürftiger Erwachsener – ein Toolkit für die Allgemeinheit.“ Praktiker“, 2011.
8. C. Magai, C. Cohen, D. Gomberg, C. Malatesta und C. Culver, „Emotionaler Ausdruck während der Demenz im mittleren bis späten Stadium“, International Psychogeriatrics, vol. 8, nein. 03, S. 383–395, 1996.
9. EC Collins, TJ Prescott und B. Mitchinson, „Saying it with light: A pilot Study of affective communication using the miro robot“, in Biomimetic and Biohybrid Systems, S. 243–255, Springer, 2015.
10. M. Ehrenfeld, „Verwendung therapeutischer Puppen bei psychogeriatrischen Patienten“, Spieltherapie mit Erwachsenen, S. 291–297, 2003.
11. G. Mitchell, „Einsatz der Puppentherapie bei Menschen mit Demenz: ein Überblick: Gary Mitchell präsentiert die Argumente für und gegen diese kontroverse, aber beliebte Intervention“, Nursing old people, vol. 26, Nr. 4, S. 24–26, 2014.
12. M. Heerink, J. Albo-Canals, M. Valenti-Soler, P. Martinez-Martin, J. Zondag, C. Smits und S. Anisuzzaman, „Untersuchung von Anforderungen und alternativen Haustierrobotern für die robotergestützte Therapie älterer Menschen.“ Erwachsene mit Demenz“, in Social Robotics, S. 104–115, Springer, 2013.

Warum verhält sich mein Roboter so? Entwerfen von Transparenz für die Echtzeitprüfung von autonome Roboter

Andreas Theodorou¹ und Robert H. Wortham² und Joanna J. Bryson³

Abstrakt. Die Prinzipien der Robotik des EPSRC schreiben die Implementierung von Transparenz in Robotersystemen vor, die diesbezügliche Forschung steckt jedoch noch in den Kinderschuhen.

Der aktuelle Beitrag stellt den Leser vor auf die Notwendigkeit, für die Inspektion transparente intelligente Agenten zu haben. Wir Bereitstellung einer robusten Definition von Transparenz als Mechanismus zur Offenlegung der Entscheidungsfindung des Roboters durch Berücksichtigung und Erweiterung auf anderen prominenten Definitionen in der Literatur. Das Papier schließt mit der Auseinandersetzung mit potenziellen Designentscheidungen, die Entwickler treffen müssen Berücksichtigen Sie dies beim Entwurf transparenter Systeme.

1. EINLEITUNG

Transparenz ist unserer Meinung nach ein Schlüsselement in Bezug auf die Ethik Auswirkungen der Entwicklung und Nutzung künstlicher Intelligenz, a Thema, das zunehmend öffentliches Interesse und Debatte erregt. Wir verwenden häufig philosophische, mathematische und biologisch inspirierte Techniken für Wir bauen künstliche interaktive, intelligente Agenten, aber wir behandeln sie als solche Blackboxen ohne Verständnis dafür, wie die zugrunde liegende Echtzeit funktioniert Entscheidungsfindung funktioniert.

Der Black-Box-Charakter intelligenter Systeme, beispielsweise in kontextsensitiven Anwendungen, führt dazu, dass die Interaktion eingeschränkt und für den Endbenutzer häufig nicht informativ ist [14]. Einschränkende Interaktionen können sich negativ auswirken die Leistung des Systems beeinträchtigen oder sogar die Funktionalität des Systems gefährden System. Stellen Sie sich ein autonomes Robotersystem vor, das für die Versorgung gebaut wurde Gesundheitsunterstützung für ältere Menschen. Allerdings können ältere Menschen Angst haben und dem System misstrauen. Sie erlauben dem Roboter möglicherweise nicht, dies zu tun mit ihnen interagieren. In einem solchen Szenario sind Menschenleben in Gefahr möglicherweise nicht rechtzeitig die erforderliche medizinische Behandlung erhalten, da ein Mensch, der das System überwacht, mangelnde Interaktion erkennen und eingreifen muss. Umgekehrt könnte es passieren, dass der menschliche Nutzer einem Roboter zu viel Vertrauen schenkt führen zu Missbrauch, übermäßigem Vertrauen und Nichtnutzung des Systems [13]. In unserer Beispiel des Gesundheitsroboters: Wenn der Agent eine Fehlfunktion aufweist und seine Patienten sich seiner Funktionsstörung nicht bewusst sind, können die Patienten weitermachen Wer den Roboter benutzt, riskiert dabei seine eigene Gesundheit. Die Roboter in beiden Szenarien brechen das erste Prinzip der Robotik des EPSRC, indem sie Menschen einsetzen Leben in Gefahr [1].

Um solche Situationen zu vermeiden, ist eine ordnungsgemäße Abstimmung des Vertrauens zwischen den Menschen, Bediener und ihre Roboter sind von entscheidender Bedeutung, wenn nicht sogar wesentlich, insbesondere in Hochrisikoszenarien wie dem Einsatz von Robotern im Militär oder für medizinische Zwecke [9]. Es kommt zu einer Kalibrierung des Vertrauens wenn der Endbenutzer ein mentales Modell des Systems hat und sich darauf verlässt

System im Rahmen der Systemfähigkeiten zu nutzen und sich seiner Grenzen bewusst zu sein [6].

Wir glauben, dass die Durchsetzung von Transparenz nicht nur von Vorteil ist für Endbenutzer, aber auch für Entwickler intelligenter Agenten. Echtzeit Das Debuggen des Entscheidungsfindungsmechanismus eines Roboters könnte Entwicklern helfen, Fehler zu beheben, Probleme zu verhindern und mögliche Abweichungen zu erklären die Leistung eines Roboters. Wir stellen uns vor, dass Entwickler durch die korrekte Implementierung von Transparenz ihre Projekte entwerfen, testen und debuggen können Agenten in Echtzeit – ähnlich wie Softwareentwickler Arbeiten Sie mit traditioneller Softwareentwicklung und Debugging.

Trotz dieser möglichen Vorteile der Transparenz in intelligenten Systemen gibt es kaum Forschung zu transparenten Agenten und sogar weniger Einsatz transparenter Agenten. Darüber hinaus gibt es Übereinstimmungen in den Definitionen von Transparenz und den Kriterien dafür Ein Roboter gilt als transparentes System. In diesem Artikel werden wir das tun Präsentieren Sie die in der Literatur gefundenen inkonsistenten Definitionen und versuchen Sie es um sie mit unseren eigenen zu ergänzen. Darüber hinaus im dritten Abschnitt In diesem Artikel besprechen wir die Designentscheidungen, die ein Entwickler benötigt Dies ist beim Entwurf transparenter Robotersysteme zu berücksichtigen.

Wir verwenden den Begriff „intelligenter Agent“ konkret, um die Kombination aus Software und Hardware eines autonomen Roboters zu bezeichnen

System, als Akteur zusammenarbeiten, darin leben und es verändern Welt [3]. In diesem Artikel werden die Wörter Roboter und Agent austauschbar verwendet.

2 TRANSPARENZ DEFINIEREN

Trotz der vorherrschenden Verwendung des Schlüsselworts Transparenz im Gemäß den EPSRC-Grundsätzen der Robotik steckt die Forschung zur Transparenz von Systemen noch in den Kinderschuhen. Im Laufe der Jahre haben sich nur sehr wenige Veröffentlichungen auf die Notwendigkeit transparenter Systeme konzentriert Weniger haben versucht, diesem Bedarf gerecht zu werden. Jede Studie bietet ihre eigene Definition des Schlüsselwortes, ohne andere auszuschließen. Miteinander ausgehen, Das Transparenzkonzept beschränkte sich auf Erklärungen für abnormales Verhalten, Zuverlässigkeit des Systems und Versuche, das zu definieren Analytische Grundlagen eines intelligenten Systems.

2.1 Das EPSRC-Prinzip der Transparenz

Die Robotikprinzipien des EPSRC betrachten Transparenz als eines ihrer Prinzipien Schlüsselprinzipien, indem Transparenz in der Robotik wie folgt definiert wird: „Roboter sind hergestellte Artefakte. Sie sollten nicht irreführend gestaltet sein Möglichkeit, gefährdete Benutzer auszunutzen; Stattdessen sollte es ihre Maschinennatur sein transparent sein.“.

Die EPSRC-Definition von Transparenz betont die Wahrung der Der Endverbraucher ist sich des hergestellten, mechanischen und damit künstlichen bewusst

¹ University of Bath, Großbritannien, E-Mail: a.theodorou@bath.ac.uk

² University of Bath, Großbritannien, E-Mail: rhwortham@bath.ac.uk

³ University of Bath, Großbritannien, E-Mail: jibryson@bath.ac.uk

Natur des Roboters. Die verwendete Formulierung lässt jedoch Überlegungen zu auch indirekte Informationen, wie z. B. technische Online-Dokumentation, als ausreichende Methode zur Durchsetzung von Transparenz[4]. Diese Orte Die Last der Verantwortung liegt beim Endbenutzer. Der Benutzer muss Dokumentation oder andere vom Hersteller bereitgestellte Informationen finden, lesen und verstehen. Einige Benutzergruppen, z. B. ältere Menschen oder nicht fachkundige Benutzer haben möglicherweise Probleme, die Technik zu verstehen Begriffe, die häufig in technischen Handbüchern vorkommen.

2.2 Transparenz als Mechanismus zur Berichterstattung
Zuverlässigkeit

In einer der frühen Veröffentlichungen wurde Transparenz im Sinne der Informationsvermittlung an den Endnutzer hinsichtlich der Fehleranfälligkeit des Systems in einem bestimmten Kontext definiert [6]. Während die Interpretation von Dzindolet nur ein Teil unserer Definition eines transparenten Systems ist, ist die Studie präsentiert interessante Erkenntnisse zur Bedeutung von Transparenz Systeme. Die Studie hat gezeigt, dass den Benutzern zusätzliches Feedback gegeben wird Bei Systemausfällen half es den Teilnehmern, ihr Vertrauen zu stärken das System. Die Benutzer wussten, dass das System nicht 100 % zuverlässig war, aber sie konnten ihr Vertrauen auf das autonome System kalibrieren im Experiment, als ihnen bewusst wurde, wann sie sich verlassen konnten darauf und wann nicht. Der militärische Einsatz von Robotersystemen erfreut sich immer größerer Beliebtheit, insbesondere in Form unbemannter Systeme Luftfahrzeuge (UAVs) und Transparenz bei Kampfsystemen sind von entscheidender Bedeutung. Stellen Sie sich vor, ein Agent identifiziert ein ziviles Gebäude als Terroristen und beschließt, dagegen vorzugehen. Wer ist verantwortlich? Der Roboter weil du unzuverlässig bist? Oder der menschliche Aufseher, der ihm vertraute in den Sensoren und Entscheidungsmechanismen des Systems? Während Das EPSRC-Robotikprinzip betrachtet den menschlichen Bediener als verantwortlich, der verursachte Schaden ist irreversibel. Autonom arbeitende Roboter Um Ziele zu erkennen und zu neutralisieren, ist ein transparentes Verhalten erforderlich [17]. Der Mensch sollte in der Lage sein, sein Vertrauen auf das System auszurichten und in Kampf-, medizinischen oder anderen Szenarien, in denen ein Roboter agiert Unzuverlässigkeit kann Menschen schädigen oder töten, Transparenz als Mechanismus dazu Berichten zufolge ist die Zuverlässigkeit des Systems von grundlegender Bedeutung.

2.3 Transparenz als Mechanismus zur Offenlegung
unerwartetes Verhalten

Spätere Studien von Kim Hinds [11] und Stumpf et. al [14], konzentriert über die Bereitstellung von Feedback-Mechanismen für Benutzer in Bezug auf Unerwartetes Verhalten eines intelligenten Agenten. In ihren Studien wurde der Benutzer alarmiert Nur wenn der Agent sein Verhalten als abnormal ansieht. Kim und die Studie von Hinds zeigte interessanterweise, dass dies durch eine Erhöhung der Autonomie erreicht wurde Auch die Bedeutung der Transparenz sowie der Verantwortung wurde erhöht vom Benutzer zum Roboter verlagert. Ihre Ergebnisse stimmen mit [10] überein. Untersuchungen, die zusammen zeigen, dass Menschen wahrscheinlicher sind einem Roboter die Schuld für Ausfälle zu geben als anderen hergestellten Artefakten und Kollegen. Um Transparenz zu erreichen, ist es wichtig, den Benutzer warnen zu können, wenn sich der Roboter unerwartet verhält. In Hochrisikosituationen könnte es durch Alarmierung dabei helfen, Menschenleben oder wertvolle Ressourcen zu retten ein menschlicher Aufseher des Systems, der die Kontrolle übernimmt oder sein Vertrauen kalibriert jeweils. In der Implementierung von Kim und Hinds jedoch der Roboter warnte den Benutzer nur, wenn er erkannte, dass er sich verhielt eine unerwartete Art und Weise. Unserer Meinung nach versucht diese Implementierung, ein Problem zu beheben Blackbox durch Verwendung einer anderen. Es gibt keine Garantie dafür, dass der Roboter unerwartet reagiert, ohne dass er von seinem atypischen Verhalten weiß. Transparenz sollte als Always-On-Mechanismus in Echtzeit durchgesetzt werden

Anismus, der es dem Benutzer ermöglicht, zu entscheiden, ob das Verhalten des Agenten stimmt als erwartet oder unerwartet angesehen.

2.4 Transparenz als Mechanismus zur Offenlegung
Entscheidungsfindung

Wir sind davon überzeugt, dass Transparenzmechanismen eingebaut sein sollten an das System und liefert in Echtzeit Informationen über seinen Betrieb, wie z sowie die Bereitstellung zusätzlicher Dokumentation gemäß dem aktuellen Grundsatz des EP SRC. Der intelligente Agent, also ein Roboter, sollte über die notwendigen Mechanismen verfügen, um aussagekräftige Informationen bereitzustellen an den Endverbraucher. Um einen Roboter als transparent für die Inspektion zu betrachten, muss der Der Endbenutzer sollte die Möglichkeit haben, genaue Interpretationen anzufordern die Fähigkeiten, Ziele und Fortschritte des Roboters in Bezug auf die genannten Ziele, sensorische Eingaben – Situationsbewusstsein, seine Zuverlässigkeit und Unerwartetes Verhalten, wie zum Beispiel Fehlermeldungen. Die Informationen der Der Roboter sollte in einem für Menschen verständlichen Format dargestellt werden. Ein transparenter Agent mit einem überprüfbaren Entscheidungsmechanismus könnte auf ähnliche Weise wie zuvor ebenfalls debuggt werden welche herkömmliche, nicht-intelligente Software häufig debuggt wird. Der Entwickler konnte sehen, welche Aktionen der Agent ausführt, warum und wie es von einer Aktion zur anderen übergeht. Dies ähnelt dem Weg in denen beliebte integrierte Entwicklungsumgebungen (IDEs) Optionen bieten, um verschiedene Codeströme mit Debug-Punkten zu verfolgen, und verfügen über Fähigkeiten wie „Step-up“ und „Step-in“ über Blöcke von Code.

3 TRANSPARENT E SYSTEME GESTALTENZ

In diesem Abschnitt dieses Dokuments werden wir die verschiedenen Entscheidungen diskutieren Entwickler können beim Entwurf eines transparenten Systems mit Problemen konfrontiert werden. Bis jetzt, Prominente Forschungsarbeiten auf dem Gebiet der Gestaltung transparenter Systeme konzentrierten sich darauf, Transparenz nur im Kontext der Mensch-Roboter-Kollaboration (HRC) zu bieten. Sie konzentrierten sich daher auf die Gestaltung transparenter Systeme, die in der Lage sind, Vertrauen zwischen den menschlichen Teilnehmern aufzubauen und der Roboter [12]. Wir glauben, dass Transparenz vorhanden sein sollte auch in nicht-kollaborativen Umgebungen, etwa bei Mensch-Roboter-Wettbewerben [11] oder sogar beim militärischen Einsatz von Robotern. In unserer Aus Sicht sollten Entwickler danach streben, intelligente Agenten zu entwickeln, die kann Informationen effizient an den menschlichen Endbenutzer übermitteln und Lassen Sie sie nacheinander ein mentales Modell des Systems entwickeln und sein Verhalten.

3.1 Benutzerfreundlichkeit

Um Transparenz zu gewährleisten, müssen zusätzliche Anzeigen oder andere Kommunikationsmethoden für den Endbenutzer sorgfältig konzipiert werden. da sie potenziell komplexe Informationen integrieren werden. Agent Entwickler müssen sowohl die tatsächliche Relevanz als auch das Niveau berücksichtigen Abstraktion der Informationen, die sie preisgeben, und wie sie dies tun werden Präsentieren Sie diese Informationen.

3.1.1 Relevanz der Informationen

Verschiedene Benutzer reagieren möglicherweise unterschiedlich auf die von ihnen bereitgestellten Informationen der Roboter. [16] zeigt, dass Endbenutzer ohne technischen Hintergrund Informationen aus technischen Geräten wie Sensoren weder verstehen noch speichern können. Dies widerspricht dem Entwickler des Agenten, der benötigt sowohl während der Entwicklung als auch beim Testen des Roboters Zugriff auf solche Informationen, um Sensoren effektiv zu kalibrieren und gefundene Probleme zu beheben. Tullio zeigt dies jedoch in derselben Studie

Benutzer sind in der Lage, zumindest grundlegende Konzepte des maschinellen Lernens zu verstehen, unabhängig von ihrem nichttechnischen Bildungs- und Berufshintergrund.

Tullios Forschung schafft einen guten Ausgangspunkt, um zu verstehen, welche Informationen für den Benutzer relevant sein könnten, um ihm zu helfen, intelligente Systeme zu verstehen. Dennoch sind weitere Arbeiten erforderlich andere Anwendungsbereiche, um sowohl domänenspezifische als auch benutzerspezifische Trends hinsichtlich der zu berücksichtigenden Informationen zu ermitteln Bedeutung.

3.1.2 Abstraktion von Informationen

Entwickler transparenter Systeme müssen nicht nur hinterfragen welche, aber auch wie viele Informationen sie dem Benutzer zugänglich machen indem ein Grad an Komplexität festgelegt wird, mit dem Benutzer interagieren können mit den transparenzrelevanten Informationen. Dies ist insbesondere bei Mehrrobotersystemen wichtig.

Multirobotersysteme ermöglichen den Einsatz mehrerer, meist kleiner Roboter, bei denen ein Ziel von verschiedenen Robotern gemeinsam genutzt wird, jeder mit seinem eigenen eigener sensorischer Input, Zuverlässigkeit und Fortschritt bei der Ausführung seiner Aufgaben zugewiesene Aufgabe, die das Gesamtsystem abschließen muss. Jüngste Entwicklungen der biologisch inspirierten Schwarmintelligenz ermöglichen den Einsatz großer In einem solchen Multi-Roboter-System arbeiten viele winzige Roboter zusammen [15]. Das Militär erwägt bereits die Entwicklung von Schwärme autonomer kleiner Robotersoldaten. Die Umsetzung von Transparenz in einem solchen System ist keine triviale Aufgabe. Der Entwickler muss machen rationale Entscheidungen darüber, wann niedrig- oder hochrangige Informationen offengelegt werden müssen. Durch die jederzeitige Offenlegung aller Informationen, z Bei allen Arten von Benutzern kann das System nach eigenem Ermessen unbrauchbar werden mit Informationen überladen sein. Wir glauben, dass verschiedene Benutzer dies tun werden erfordern unterschiedliche Ebenen der Informationsabstraktion, um Informationssucht zu vermeiden. Höhere Abstraktionsebenen könnten sich auf die Präsentation konzentrieren nur ein Überblick über das System. Anstatt den Fortschritt eines zu haben System in Richtung eines Ziels, indem die aktuellen Aktionen des Systems angezeigt werden Um das genannte Ziel zu erreichen, könnte es einfach eine sein Abschlussleiste. Darüber hinaus könnten in einem System mit mehreren Robotern zu den Informationen auf niedrigerer Ebene auch das Ziel, der Sensor, der Zielprozess und das Gesamtbild gehören Verhalten einzelner Agenten im Detail. Umgekehrt, a Eine allgemeine Übersicht könnte alle Roboter als eine Einheit anzeigen und Durchschnittswerte für jede Maschine angeben. Intelligente Agenten mit einem Design, das darauf basiert eine kognitive Architektur wie Behavior Oriented Design (BOD) [2] könnte nur Planelemente auf hoher Ebene darstellen, wenn ein Überblick über die gegebenen wäre System benötigt wird. Im Falle eines mit BOD entwickelten Agenten: Benutzer Möglicherweise möchten Sie den Status von Laufwerken sehen und sich darüber informieren oder Kompetenzen, aber nicht einzelne Aktionen. Andere Benutzer möchten möglicherweise um nur Teile des Plans im Detail und andere Teile auf hoher Ebene anzuzeigen Überblick.

Eine gute Umsetzung soll dem Nutzer Transparenz bieten mit solchen Optionen, Bereitstellung von Einzelpersonen oder potenziellen Benutzergruppen mit flexiblen und voreingestellten Konfigurationen, um ein breites Spektrum abzudecken Spektrum der Bedürfnisse potenzieller Nutzer. Wir gehen davon aus, dass der Abstraktionsgrad, den eine Person benötigt, von einer Reihe von Faktoren abhängt, unter anderem vom demografischen Hintergrund des Benutzers.

1. Benutzer: Wir haben bereits die Art und Weise besprochen, in der verschiedene Benutzer arbeiten neigen dazu, unterschiedlich auf Informationen über den aktuellen Zustand zu reagieren eines Roboters. Ebenso können wir davon ausgehen, dass verschiedene Benutzer antworten werden in ähnlicher Weise wie die verschiedenen Abstraktionsebenen deren Nutzung des Systems. Endbenutzer, insbesondere Nichtfachleute, einen allgemeinen Überblick über die verfügbaren Informationen bevorzugen, während wir davon ausgehen, dass Entwickler Zugang zu einem niedrigeren Informationsniveau erwarten.

2. Art des Robotersystems: Wie in unseren Beispielen oben besprochen, a Multi-Roboter-Systeme erfordern höchstwahrscheinlich ein höheres Maß an Abstraktion, um Informationsüberschreitungen des Endbenutzers zu vermeiden. Ein System mit einem einzigen Agenten würde viel weniger Abstraktion erfordern, da dem Benutzer weniger Daten angezeigt werden.
3. Zweck des Robotersystems: Der beabsichtigte Zweck des Systems sollten bei der Gestaltung eines transparenten Agenten berücksichtigt werden. Beispielsweise ist es viel wahrscheinlicher, dass ein Militärroboter eingesetzt wird Da es sich um einen professionellen Benutzer handelt, der sich in oder auf der Schleife befindet, besteht aufgrund des risikoreichen Betriebs ein viel größerer Bedarf, so viel anzuzeigen und zu erfassen Informationen über das Verhalten des Agenten. Auf dem anderen Wahrscheinlicher ist jedoch eine Roboter-Rezeptionistin oder ein persönlicher Assistent zur Verwendung durch technisch nicht versierte Benutzer, die möglicherweise eine vereinfachte Version bevorzugen Überblick über das Verhalten des Roboters.

3.1.3 Darstellung von Informationen

Entwickler müssen überlegen, wie sie dem Benutzer etwas davon präsentieren Sie erhalten zusätzliche Informationen zum Verhalten des Agenten exponieren. Frühere Studien verwendeten eine visuelle oder akustische Darstellung des Information. Unseres Wissens liegen keine Vergleichsstudien vor die unterschiedlichen Herangehensweisen. Autonome Robotersysteme können Dutzende verschiedener Entscheidungen treffen pro Sekunde. Wenn der Agent einen reaktiven Plan verwendet, beispielsweise einen POSH-Plan [5] Der Agent kann pro Minute Tausende von Anrufen zu den verschiedenen Planelementen tätigen. Mit dieser Menge an Informationen ist es schwer, umzugehen audiorientierte Systeme. Darüber hinaus ist die Visualisierung der Informationen, d durch Bereitstellung einer grafischen Darstellung des Plans des Agenten, wo Die verschiedenen Planelemente blinken so, wie sie aufgerufen werden Das System ist selbsterklärend und auch für weniger technisch versierte Benutzer leicht zu verstehen. Schließlich bietet eine Diagrammvisualisierung als Mittel zur Bereitstellung transparenzbezogener Informationen zusätzliche Vorteile beim Debuggen der Anwendung. Der Entwickler sollte in der Lage sein, der Spur der verschiedenen aufgerufenen Planelemente zu folgen und die ausgelösten sensorischen Eingaben zu erkennen sie, bis ein bestimmtes Element verwendet wurde.

3.2 Nutzen des Systems

Bisher haben wir in diesem Artikel auf die Bedeutung und das Design näher eingegangen Entscheidungen hinsichtlich der Umsetzung von Transparenz. Wir jedoch Ich glaube, dass der Entwickler auch darüber nachdenken muss, ob er implementiert wird Transparenz kann den Nutzen eines Systems tatsächlich beeinträchtigen. [18] argumentiert dass der Nutzen eines Agenten daran gemessen wird, inwieweit er eingesetzt wird ist vertrauenswürdig. Zunehmende Transparenz kann seinen Nutzen verringern. Das könnte, sich beispielsweise negativ auf einen Begleitroboter auswirken oder ein Gesundheitsroboter, der zur Unterstützung von Kindern entwickelt wurde. In solchen Fällen ist die Das System ist nach den EPSRC-Grundsätzen der Robotik konzipiert nutzt die Gefühle seiner Benutzer aus, um seinen Nutzen und seine Leistung zu steigern seine gestellte Aufgabe. Eine weitere wichtige Designentscheidung, die sich auf das System auswirkt, ist die physische Transparenz des Systems. Das physische Erscheinungsbild eines Agent kann seine Benutzerfreundlichkeit erhöhen [7], aber auch kontrastieren Transparenz, indem es seine mechanische Natur verbirgt. Zurück in unserem Beispiel eines Begleitroboters könnte ein humanoider oder tierähnlicher Roboter einem Agenten vorgezogen werden, dessen Mechanismen und Innenleben offengelegt sind. enthüllt seine künstliche Natur [8]. Die Diskussion über den Kompromiss zwischen Nutzen und Transparenz ist weit gehen über den Rahmen dieser Arbeit hinaus. Entwickler sollten sich jedoch darüber im Klaren sein davon, wenn sie Roboter entwerfen und entwickeln.

4. FAZIT

Wir glauben fest daran, dass die Implementierung und Nutzung von intelligenten Systemen, die von Natur aus transparent sind, können der Öffentlichkeit helfen, KI zu verstehen, indem sie das gruselige Geheimnis darüber beseitigen, warum sie so ist. Transparenz ermöglicht es, einen Agenten zu verstehen auftauchendes Verhalten. In diesem Artikel haben wir Transparenz als neu definiert

Always-on-Mechanismus, der das Verhalten, die Zuverlässigkeit und die Zuverlässigkeit eines Systems melden kann.

Sinne und Ziele als solche Informationen könnten uns helfen, das zu verstehen Verhalten eines autonomen Systems.

Es bedarf weiterer Arbeit, um bewährte Verfahren für die Umsetzung von Transparenz innerhalb der Robotik-Community zu testen und zu etablieren.

Angeichts der Vorteile transparenter Systeme sind wir stark schlagen die Förderung dieses Schlüsselprinzips durch Forschungsräte vor, z wie EPSRC und andere akademische Gemeinschaften.

DANKSAGUNGEN

Wir möchten Swen Gaudl (University of Bath) herzlich danken wertvolle Einsichten.

VERWEISE

- [1] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorell, Mick Wallis, Blay Whitby, und Alan Winfield. Prinzipien der Robotik. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC), April 2011. Webpublikation.
- [2] Joanna Bryson, „Das verhaltensorientierte Design modularer Agentenintelligenz“, in System, Band 2592, 61–76, (2002).
- [3] Joanna J. Bryson, „Roboter sollten Sklaven sein“, in Close Engagements mit künstlichen Gefährten: Wichtige soziale, psychologische, ethische und gestalterische Fragen, Hrsg., Yorick Wilks, 63–74, John Benjamins, Amsterdam, (März 2010).
- [4] Joanna J. Bryson, Darwin Caldwell, Kerstin Dautenhahn, Paula Duxbury, Lilian Edwards, Hazel Grian, Sarah Kember, Stephen Kemp, Paul Newman, Geo Peg, Andrew Rose, Tom Rodden, Tom Sorell, Mick Wallis, Shearer West, Alan Winfield und Ian Baldwin, „The Making of“. „The EPSRC Principles of Robotics“, 133(133), 14–15, (2012).
- [5] Joanna J. Bryson, Tristan J. Caulfield und Jan Drugowitsch, „Integrating lebenssechter Aktionsauswahl in zyklusbasierte Agentensimulationsumgebungen“, in Proceedings of Agent 2005: Generative Social Processes, Modelle und Mechanismen, Hrsg., Michael North, David L. Sallach und Charles Macal, S. 67–81, Chicago, (Oktober 2005). Argonnen-National Labor.
- [6] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce und Hall P. Beck, „Die Rolle des Vertrauens bei der Automatisierungsabhängigkeit“, International Journal of Human Computer Studies, 58(6), 697–718, (2003).
- [7] Kerstin Fischer, „Wie Menschen mit Robotern sprechen: Dialoge gestalten.“ „Reduzieren Sie die Unsicherheit der Benutzer“, AI Magazine, 32(4), 31–38, (2011).
- [8] Jennifer Goetz, Sara Kiesler und Aaron Powers, „Anpassung des Aussehens und Verhaltens von Robotern an Aufgaben, um die Zusammenarbeit zwischen Mensch und Roboter zu verbessern“, Proceedings – IEEE International Workshop on Robot and Human Interactive Communication, 55–60, (2003).
- [9] Victoria Groom und Clifford Nass, „Können Roboter Teamkollegen sein?“, Inter Aktionsstudien, 8(3), 483–500, (2007).
- [10] Peter H. Kahn, Rachel L. Severson, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Jolina H. Ruckert, Solace Shen, Heather E. Gary, Aimee L. Reichert und Nathan G. Freier: „Halten Menschen einen Humanoiden fest?“ Ist der Roboter moralisch für den Schaden verantwortlich, den er verursacht?“, Proceedings of the siebte jährliche internationale ACM/IEEE-Konferenz zum Thema Mensch-Roboter Interaktion – HRI '12, (Februar 2016), 33, (2012).
- [11] Taemie Kim und Pamela Hinds: „Wem soll ich die Schuld geben?“ Auswirkungen von Autonomie und Transparenz auf Attributionen in der Mensch-Roboter-Interaktion“, Proceedings – IEEE International Workshop on Robot and Human Interactive Communication, 80–85, (2006).
- [12] Joseph B. Lyons, „Transparenz transparent machen: Ein Modell für Mensch-Roboter-Interaktion“, Vertrauen und autonome Systeme: Artikel von das AAAI-Frühjahrssymposium 2013, 48–53, (2013).
- [13] R. Parasuraman und V. Riley, „Menschen und Automatisierung: Nutzung, Missbrauch, Nichtgebrauch, Missbrauch“, Human Factors, 39(2), 230–253, (1997).
- [14] Simone Stumpf, Weng-keen Wong, Margaret Burnett und Todd Kulesza, „Intelligente Systeme verständlich und kontrollierbar machen.“ von Endbenutzern“, 10–11, (2010).
- [15] Ying Tan und Zhong-yang Zheng, „Forschung schreitet im Schwarm voran.“ Robotik“, Defence Technology, 9(1), 18–39, (3 2013).
- [16] Joe Tullio, Anind K. Dey, Jason Chalecki und James Fogarty, „How „es funktioniert: eine Feldstudie über nicht-technische Benutzer, die mit einem intelligenten System interagieren“, SIGCHI-Konferenz über menschliche Faktoren in der Datenverarbeitung systems (CHI'07), 31–40, (2009).
- [17] Lu Wang, Greg a Jamieson und Justin G Hollands, „Vertrauen und Vertrauen über ein automatisiertes Kampferkennungssystem“, Human Factors, 51(3), 281–291, (2009).
- [18] Robert Wortham, Andreas Theodorou und Joanna J. Bryson, „Das Eisen Dreieck: Transparenz-Vertrauen-Nützlichkeit“. eingereicht, 2016.

Robotertransparenz, Vertrauen und Nutzen

Robert H. Wortham¹ Andreas Theodorou² und Joanna J. Bryson³

Abstrakt. Da das Denken von Robotern immer komplexer wird, wird das Debuggen immer schwieriger, das ausschließlich auf beobachtbarem Verhalten basiert, selbst für Roboterdesigner und technische Spezialisten. Ebenso fällt es nicht-spezialisierten Benutzern schwer, nützliche mentale Modelle für das Denken von Robotern allein aus beobachtetem Verhalten zu erstellen. Die EPSRC-Robotikprinzipien schreiben vor, dass unsere Artefakte transparent sein sollten, aber was bedeutet das in der Praxis und wie wirkt sich Transparenz sowohl auf das Vertrauen als auch auf den Nutzen aus? Wir untersuchen diesen Zusammenhang in der Literatur und stellen fest, dass er komplex ist, insbesondere in nicht-industriellen Umgebungen, wo Transparenz je nach Anwendung und Zweck des Roboters ein breiteres Spektrum an Auswirkungen auf Vertrauen und Nutzen haben kann. Wir skizzieren unser Forschungsprogramm, um unsere Behauptung zu untermauern, dass es trotz ihrer transparenten Maschinennatur dennoch möglich ist, transparente Agenten zu schaffen, die emotional ansprechend sind.

1. EINLEITUNG

Die EPSRC-Robotikprinzipien enthalten einen spezifischen Verweis auf Transparenz: „Roboter sind hergestellte Artefakte. Sie sollten nicht in einer irreführenden Weise konzipiert sein, um gefährdete Benutzer auszunutzen. Stattdessen sollte ihre maschinelle Natur transparent sein.“ siehe [1]. Dies scheint zunächst eine einfache normative Behauptung zu sein, die sich auf die weit verbreitete Idee stützt, dass Agenten nicht täuschen sollten, da Täuschung im Allgemeinen zur Ausbeutung führt. In diesem Artikel wird untersucht, ob Transparenz tatsächlich eine so einfache Idee ist und ob die Transparenz bestimmter Arten von Agenten ihren Nutzen verringert. Bei der Betrachtung dieser Frage müssen wir uns auch mit dem Zusammenhang zwischen Transparenz und Vertrauen befassen.

In diesem Artikel verwenden wir die Begriffe Roboter und Agent austauschbar und meinen mit diesen Begriffen ein verkörpertes, autonomes intelligentes Artefakt.

Was bedeutet es, einem Roboter zu vertrauen? Wir könnten zunächst einfach sagen: Wenn eine KI transparenter ist, können wir ihr mehr vertrauen und daher steigt ihr Nutzen. Wir könnten auch argumentieren, dass Vertrauen nur dann erforderlich ist, wenn ein Agent nicht vollständig transparent ist, und dass daher eine erhöhte Transparenz das Vertrauensbedürfnis verringert [4]. Wenn der Nutzen eines Artefakts daran gemessen wird, inwieweit es vertrauenswürdig ist, kann eine zunehmende Transparenz diesen Nutzen verringern. Dies könnte beispielsweise bei einem Roboter der Fall sein, dessen Hauptfunktion darin besteht, Gesellschaft zu leisten.

Wir beginnen also zu erkennen, dass zwischen den Vorstellungen von Nutzen, Transparenz und Vertrauen ein komplexer Zusammenhang besteht. Diese Beziehung hängt vom Zweck der KI ab. In diesem Artikel überprüfen wir die Literatur zu Transparenz und Vertrauen und beschreiben auch laufende praktische Forschungen zur Untersuchung des Vorschlags, dass es tatsächlich möglich ist, einen emotional ansprechenden und dennoch transparenten Roboter zu bauen.

2 THEORIE DES GEISTES, DES VERTRAUENS UND TRANSPARENZ

Obwohl wir davon ausgehen können, dass die Kommunikation zwischen Tieren und insbesondere zwischen Menschen komplex sein muss, nutzen natürliche Kommunikationssysteme tatsächlich relativ einfache und minimale Signale aus, deren Bedeutung aus umfangreichen Modellen abgeleitet wird [16]. Mit anderen Worten: Die Evolution bzw. eine gemeinsame phylogenetische Geschichte liefert adäquate Prioritäten, so dass für die Kommunikation des Kontexts nur minimale Daten erforderlich sind. Auch wenn manche anders argumentieren würden [8], besteht allgemein Einigkeit darüber, dass eine wirksame Interaktion, ob Zwang oder Kooperation, davon abhängt, dass jede Partei eine gewisse Theorie des Geistes (Theory of Mind, ToM) der anderen hat [16, 14]. Individuelle Handlungen und zusammengesetzte Verhaltensweisen werden somit innerhalb eines bereits bestehenden ToM-Rahmens interpretiert. Ob dieses ToM korrekt ist, ist unwichtig, vorausgesetzt, es ist prädiktiv im Hinblick auf das Verhalten. Das Transparenzmodell des Roboters definiert nicht das vom menschlichen Benutzer verwendete ToM, aber es ist das Transparenzmodell, das wir direkt anpassen können, und dies ist daher der Schwerpunkt dieses Papiers. Es ist bekannt, dass beobachtbares Verhalten die inneren mentalen Zustände des Einzelnen kommunizieren kann. Breazeal [2] fand heraus, dass implizite nonverbale Kommunikation die Transparenz gegenüber rein bewusster nonverbaler Kommunikation verbessert. Implizit wird hier als die Übermittlung von Informationen definiert, die dem Verhalten innewohnen, aber vom Roboterdesigner nicht absichtlich kommuniziert werden. Menschen haben starke Erwartungen daran, wie sich implizite und explizite nonverbale Signale auf mentale Zustände auswirken. Breazeal stellte außerdem fest, dass Transparenz Konflikte reduziert, wenn Fehler auftreten, insbesondere wenn eine gemeinsame Aufgabe versucht wird.

Reduzierter Konflikt bedeutet, dass bei Auftreten eines Fehlers während der Aufgabenausführung eine Wiederherstellung immer noch mit weniger Schuldzuweisungen möglich ist. Breazeal bezeichnet diesen reduzierten Konflikt als Robustheit, und diese Robustheit ist ein wirksames Maß für den Nutzen.

2.1 Anthropomorphismus und mentale Modelle von Robotern

Menschen haben eine starke Veranlagung, nicht nur die Natur, sondern alles um sie herum zu anthropomorphisieren [5] – die Social-Brain-Hypothese [7] könnte dieses Phänomen erklären, allerdings behandeln Menschen Roboter nicht genauso wie Menschen, beispielsweise in Bezug auf ihre moralische Stellung [10]. Obwohl es erhebliche Debatten über die Ontologie des Robotergeistes im Vergleich zum menschlichen Geist gibt, ist es von praktischerer Bedeutung, wie Robotergeiste von Menschen psychologisch verstanden werden, dh was ist die wahrgenommene und nicht die tatsächliche Ontologie? Stubbs [15] hält es für wesentlich, ein mentales Modell von Robotern zu entwickeln, um dies zu erreichen

Gemeinsamkeiten aufbauen – was wir auch als Grundlage für menschliches Vertrauen interpretieren könnten. Stubbs [15] stellte außerdem fest, dass diese Gemeinsamkeit effektiv durch einen interaktiven Dialog mit dem Roboter hergestellt werden kann. Obwohl diese Studie in erster Linie ferngesteuerte Roboter betrachtete, die in einer industriellen oder explorativen Umgebung arbeiten, und nicht Roboter, die dort arbeiten

¹ University of Bath, UK, E-Mail: rhwortham@bath.ac.uk

² University of Bath, UK, E-Mail: a.theodorou@bath.ac.uk

³ University of Bath, UK, E-Mail: jjbryson@bath.ac.uk

Im häuslichen Umfeld sollten wir uns der Bedeutung des Dialogs für den Aufbau von Vertrauen bewusst sein. Tatsächlich sieht Mueller [13] den Dialog als ein Ganzes von den drei Hauptmerkmalen transparenter Computer sind die anderen Erklären und Lernen sein.

Meerbeek [12] untersucht den Zusammenhang zwischen der wahrgenommenen Persönlichkeit eines Roboters und dem Grad, in dem sich der Benutzer unter Kontrolle fühlt während der Interaktion. Um glaubwürdig zu sein, fand Meerbeek das Der Persönlichkeitsausdruck sollte mit einem internen Modell verknüpft sein befasst sich mit dem Verhalten (z. B. Entscheidungsfindung) basierend auf der Persönlichkeit und Emotion. Ausdrucksstärkeres, informelles Verhalten ist damit verbunden eine höhere Wahrnehmung der Benutzerkontrolle.

Nicht spezialisierte Menschen haben entweder wenig ToM für Roboter oder haben es Ein Modell, das auf zeitgenössischer Science-Fiction basiert, und interpretieren daher Verhaltensweisen mithilfe einer Standardtheorie anderer Agenten, die davon ausgeht der Agent, der menschenähnliche Motivationen teilt. Das kann man verstehen evolutionär ausgedrückt, durch das Bedürfnis unserer Vorfahren, die proximale Aktivität schnell als entweder neutral einzustufen (das Rascheln der Blätter in der Wind), freundlich (die Annäherung eines Stammesmitglieds) oder feindselig (der Annäherung eines Raubtiers oder Feindes). Wenn sensorische Informationen unsicher sind, ist die Entwicklung einer Tendenz zur Annahme sowohl von Entscheidungsfreiheit als auch von Feindseligkeit selektiv für die individuelle Langlebigkeit in einem Umfeld, in dem ... man ist häufig die Beute, nicht das Raubtier. Sogar in unseren technologischen Umgebungen erleben wir oft Scheinhandlungen, etwa Roboter Anwählen von Verkaufsgesprächen, automatisierte Twitter-Postings und automatisch generierte personalisierte Spam-E-Mails.

In einer 2006 in einem Gemeinschafts Krankenhaus durchgeführten Studie In den USA suchte das Pflegepersonal ständig nach Gründen dafür Die Roboter handelten so, wie sie es taten. Sie würden sich und andere fragen: "Was geht hier vor sich? Soll das der Roboter machen oder habe ich das gemacht? Stimmt etwas nicht?". Diese Untersuchung besagt, dass ein geringer Grad an Transparenz dazu führte, dass Menschen selbst das normale Verhalten der Menschen in Frage stellten Roboter, der Menschen manchmal sogar dazu bringt, über richtige Verhaltensweisen nachzudenken als Fehler [11].

3 FORSCHUNGSPROGRAMM

Wir starten ein Programm praktischer Forschung, um dies zu untersuchen das Dreieck Transparenz, Vertrauen, Nutzen. Zunächst mit nicht-humanoiden Mitteln Mit Robotern führen wir Experimente durch, um die Wirkung verschiedener Ausdrucksformen von Transparenz auf die emotionale Reaktion von Menschen zu bestimmen. Im Mittelpunkt unserer Experimente verwenden wir reaktive Planungstechniken, um autonome Agenten zu bauen. Wir haben das entwickelt Instinkt-reaktiver Planer basierend auf Brysons Behavior Oriented Design (BOD)-Ansatz [3]. Der Instinct-Planer meldet die Ausführung und Status jedes Planelements in Echtzeit, was es uns ermöglicht, den Argumentationsprozess innerhalb des Roboters, der entsteht, implizit zu erfassen sein Verhalten. Unsere Experimente werden untersuchen und zeigen, wie das geht Diese Transparenzdaten des Planers können genutzt werden, um das Verhalten des Roboters verständlicher zu machen. Zunächst sind wir in erster Linie Interesse daran, das Verhalten für den Roboterdesigner transparent zu machen, da Roboter mit komplexen Plänen typischerweise sehr schwer zu entwerfen sind und debuggen. Allerdings können diese ersten Experimente auch die Transparenz für Laienbeobachter verbessern.

Anschließend werden wir untersuchen, wie wir den im Instinct Planner eingebetteten Transparenzmechanismus nutzen können, um eine zu erstellen effektiverer Haushaltsroboter. Die Forschung wird untersuchen, ob Durch Transparenz fühlen sich Menschen mehr oder weniger an ihren Roboter gebunden. und ob sie mehr oder weniger in der Lage sind, die Bedürfnisse genau einzuschätzen des Roboters, während er daran arbeitet, seine Ziele zu erreichen.

Es wird davon ausgegangen, dass diese Versuche in einer häuslichen oder wohnnahen Umgebung, beispielsweise einem Altersheim, stattfinden sollten.

Wir müssen Feedback von nicht spezialisierten Beobachtern/Benutzern einholen das qualitative Niveau der Intelligenz des Roboters und auch darüber, wie Es wäre für sie angenehm, ein solches Gerät in ihrer häuslichen Umgebung zu haben. Die Forschung wird versuchen, das anfängliche Ausmaß der Angst zu ermitteln. Angst, Misstrauen gegenüber KI und Robotern im Allgemeinen sowie gegenüber Haushaltsrobotern insbesondere. Eine Referenzposition schaffen, Transparenz schaffen des Roboters muss durch die Bereitstellung von Feedback an den Benutzer ermöglicht werden auf die Echtzeitausführung im reaktiven Planer. Die Methoden Wir stellen uns derzeit Folgendes vor:

- Echtzeitpräsentation von Textaussagen zum Plan ex e Schnitt.
- Grafische Echtzeitvisualisierung der Planausführung.
- Audio- (dh mündliche) Aussagen zur Ausführung des Roboterplans.

Für jede dieser Methoden könnten die Transparenzinformationen entweder auf/von einem entfernten Gerät oder auf/vom Roboter selbst präsentiert werden. Es gibt also sechs mögliche Kombinationen. Natürlich ist auch eine zusätzliche Transparenzfusion möglich, z. B. Audio in Kombination mit Grafik basierend auf dem Erfolg oder Misserfolg erster experimenteller Ergebnisse getestet. Wie aus der Literatur hervorgeht, ist der Dialog bei der Etablierung wichtig Vertrauen, diese Forschung sollte diese Möglichkeit in Betracht ziehen Spracheingaben zu akzeptieren, wenn auch auf einfache Befehle beschränkt, als bedeutet für Benutzer, den Roboter zu fragen, was er tut, und zu haben Der Roboter reagiert angemessen.

4. DISKUSSION

Das EPSRC-Prinzip 1 besagt, dass Roboter Werkzeuge sind. In Industrie- und In technischen Umgebungen ist dies ziemlich klar, da ein Mensch den Roboter verwendet, um eine technische Aufgabe zu erledigen. Der Designer und Benutzer des Roboters teilen das Ziel des Roboters: die Aufgabe zu erledigen. In häuslichen Umgebungen und im Gesundheitswesen können Roboter jedoch eingesetzt werden haben eher eine andere Beziehung zu denen, mit denen sie interagieren. Sie kann dazu gedacht sein, Kameradschaft und gleichzeitig Verborgenheit zu bieten Überwachung des Patientenwohls. Für das medizinische Fachpersonal mögen sie Hilfsmittel sein, für den Patienten sind sie jedoch Begleiter. In solch Eine Umgebung, die sich negativ auf den Nutzen auswirken kann Transparenz. Unser Kameradschaftsgefühl hängt mit der Maßnahme zusammen Wir projizieren die Wirkungskraft auf den Roboter. Wenn wir in der Lage sind, das zu verstehen Die Wirkungsweise der Intelligenz scheint von Natur aus geringer zu werden intelligent im volkstümlichen Sinne, so dass wir dann weniger Entscheidungsfreiheit projizieren, und Infolgedessen erfahren Sie weniger Nutzen aus dem Roboter? Wir könnten vergleichen das mit Fernsehen. Wir wissen, dass es keine Agentur hat, aber seine Präsenz in Die Ecke unseres Wohnzimmers bietet gesellige Vorteile. Vielleicht hat das mit der bewussten Aufhebung des Unglaubens zu tun, oder? Vielleicht haben wir einen Detektor für unbewusste Kräfte, der einfacher ist von der Technik getäuscht.

Vernünftige Vorstellungen von Intelligenz werden mit volkpsychologischen Vorstellungen von Handlungsfähigkeit und auch vom Leben vermischt. Dinge, die intelligent sind sind lebendig in dem Sinne, dass sie ihre eigenen Überzeugungen, Wünsche und Absichten haben, von denen wir verstehen, dass sie grundsätzlich eigennützig oder egoistisch sind. Wir erkennen Egoismus implizit als ein grundlegendes Merkmal von an alles Leben [6]. Wenn ein solcher Agent mit uns in Kontakt tritt, dann betrachtet er uns als solche wichtig bei der Verfolgung dieser egoistischen Ziele. Solche Agenten sind würdig, unsere Begleiter zu werden, weil sie ihnen wahren Wert zuschreiben in ihrer Beziehung zu uns, und dies steigert unseren Wert in der Gesellschaft. Umgekehrt sind Agenten, die keine eigennützige Handlungsmacht haben, nicht würdig unserer Aufmerksamkeit, weil sie keinen sozialen Wert vermitteln. Vielleicht sind es deshalb künstliche Agenten, deren einziger Zweck die Kameradschaft ist Personen, die in dieser Hinsicht wirklich transparent sind, sind daher nicht geeignet, würdige Gefährten zu sein. In manchen Situationen kann es daher zu Robotertransparenz kommen

im Widerspruch zum Nutzen stehen und allgemeiner eher orthogonal sein als förderlich für den erfolgreichen Einsatz des Roboters. Während wir uns möglicherweise Szenarien ausdenken und weiterhin über das theoretische und philosophische Zusammenspiel zwischen Transparenz, Vertrauen und Nutzen diskutieren, sind wir als Wissenschaftler der Meinung, dass dies der Fall ist. Warten Sie auf das Ergebnis unserer Experimente.

5. SCHLUSSFOLGERUNG

Wir haben gesehen, dass es komplex ist, Transparenz und Vertrauen zu entschlüsseln kann teilweise verstanden werden, wenn man sich anschaut, wie Menschen einander verstehen und anschließend vertrauen und wie sie überwinden evolutionäre Ängste, um anderen Agenten zu vertrauen, durch implizite nonverbale Kommunikation. Inakzeptables Maß an Angst, Furcht und Misstrauen führt zu einer emotionalen und kognitiven Reaktion auf Ablehnung. Roboter. Hancock [9] behauptet, dass wir es tun werden, wenn wir unseren Robotern nicht vertrauen können nicht in der Lage sein, effektiv davon zu profitieren. Angesichts dessen, dass wir Wir interagieren in der Gesellschaft gerne mit anderen, die wir nicht vollständig kennen Vertrauen, und zunehmend interagieren wir mit Computern in dem Wissen, dass sie es tun Empfehlungen möglicherweise fehlerhaft, wir müssen zu dem Schluss kommen, dass Hancock es ist Übervereinfachung. Schließlich kann es Anwendungen geben, bei denen Transparenz im Widerspruch zum Nutzen steht. Unser fortlaufendes Forschungsprogramm soll unsere Hypothese bestätigen, dass wir tatsächlich transparente Roboter erschaffen können, die dennoch emotional ansprechend und nützlich sind Werkzeuge in einer Vielzahl von häuslichen und wohnnahen Umgebungen. In der Zwischenzeit bleibt noch viel zu tun, um die Beziehung zwischen Transparenz, Nutzen und Vertrauen aufzuklären.

VERWEISE

- [1] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorell, Mick Wallis, Blay Whitby, und Alan Winfield. Prinzipien der Robotik. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC), April 2011. Webpublikation.
- [2] C. Breazeal, CD Kidd, AL Thomaz, G. Hoffman und M. Berlin, „Auswirkungen nonverbaler Kommunikation auf Effizienz und Robustheit in.“ Mensch-Roboter-Teamarbeit“, auf der IEEE/RSJ International Conference 2005 on Intelligent Robots and Systems, S. 708–713, Alberta, Kanada, (2005). IEEE.
- [3] Joanna J. Bryson, „Intelligence by Design: Prinzipien der Modularität und Koordination für die Entwicklung komplexer adaptiver Agenten“, (2001).
- [4] Joanna J Bryson und Paul Rauwolf, „Trust, Communication, and In.“ Gleichwertigkeit“. 2016.
- [5] Kerstin Dautenhahn, „Methodologie und Themen der Mensch-Roboter-Interaktion: Ein wachsendes Forschungsfeld“, International Journal of Advanced Robotic Systems, 4(1 SPEC. ISS.), 103–108, (2007).
- [6] Richard Dawkins, „Hierarchische Organisation: Ein Kandidatenprinzip für Ethologie“, in Growing Points in Ethology, Hrsg., PPG Bateson und RA Hinde, 7–54, Cambridge University Press, Cambridge, (1976).
- [7] RIM Dunbar, „The Social Brain Hypothesis“, Evolutionary Anthropology, 178–190, (1998).
- [8] Shaun Gallagher, „Die narrative Alternative zur Theorie des Geistes“, in Radical Enactivism: Intentionality, Phenomenology, and Narrative, Hrsg., R Menary, Nummer Gallagher 2001, 223–229, John Benjamins, Amsterdam, (2006).
- [9] P. a. Hancock, DR Billings, KE Schaefer, JYC Chen, EJ de Visser und R. Parasuraman, „A Meta-Analysis of Factors Affecting „Trust in Human-Robot Interaction“, Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(5), 517–527, (2011).
- [10] Peter H. Kahn, Hiroshi Ishiguro, Batya Friedman und Takayuki Kanda, „Was ist ein Mensch? - Auf dem Weg zu psychologischen Benchmarks im Bereich Mensch-Roboter-Interaktion“, Proceedings – IEEE International Workshop on Robot and Human Interactive Communication, 3, 364–371, (2006).
- [11] Taemie Kim und Pamela Hinds: „Wem soll ich die Schuld geben?“ „Auswirkungen von Autonomie und Transparenz auf Attributionen in der Mensch-Roboter-Interaktion“, Proceedings – IEEE International Workshop on Robot and Human Interactive Communication, 80–85, (2006).
- [12] Bernt Meerbeek, Jettie Hoonhout, Peter Bingley und Jacques Terken, „Untersuchung der Beziehung zwischen der Persönlichkeit eines Roboterfernsehers.“ Assistant and the level of user control“, Proceedings – IEEE International Workshop on Robot and Human Interactive Communication, 404–410, (2006).
- [13] Erik T. Mueller, Transparente Computer: Verständliches Design Intelligente Systeme, Erik T. Mueller, San Bernardino, CA, 2016.
- [14] Rebecca Saxe, Laura E. Schulz und Yuhong V. Jiang, „Gedanken lesen versus das Befolgen von Regeln: dissoziierende Theorie des Geistes und exekutive Kontrolle im Gehirn.“, Social Neuroscience, 1(3-4), 284–98, (Januar 2006).
- [15] Kristen Stubbs, Pamela J Hinds und David Wettergreen, „Autonomy und Gemeinsamkeiten in der Mensch-Roboter-Interaktion: Eine Feldstudie“, IEEE Intelligent Systems, 22(2), 42–50, (2007).
- [16] Robert H. Wortham und Joanna J. Bryson, „Communication“, in Hand book of Living Machines {im Druck}, Oxford University Press, Oxford, (2016).