



BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG

Christoph Kehl
Christopher Coenen

Technologien und Visionen der Mensch-Maschine- Entgrenzung

Sachstandsbericht zum TA-Projekt
»Mensch-Maschine-Entgrenzungen:
zwischen künstlicher Intelligenz und
Human Enhancement«

Juli 2016
Arbeitsbericht Nr. 167





**Technologien und
Visionen der
Mensch-Maschine-
Entgrenzung**



Das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) berät das Parlament und seine Ausschüsse seit 1990 in Fragen des technischen und gesellschaftlichen Wandels. Das TAB ist eine organisatorische Einheit des Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS) im Karlsruher Institut für Technologie (KIT). Zur Erfüllung seiner Aufgaben kooperiert es seit September 2013 mit dem Helmholtz-Zentrum für Umweltforschung GmbH – UFZ, dem IZT – Institut für Zukunftsstudien und Technologiebewertung gGmbH sowie der VDI/VDE Innovation + Technik GmbH.



Christoph Kehl
Christopher Coenen

Technologien und Visionen der Mensch-Maschine- Entgrenzung

Sachstandsbericht zum TA-Projekt
»Mensch-Maschine-Entgrenzungen:
zwischen künstlicher Intelligenz und
Human Enhancement«

TAB-Arbeitsbericht Nr. 167



Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag (TAB)
Neue Schönhauser Straße 10
10178 Berlin

Fon: +49 30 28491-0

Fax: +49 30 28491-119

buero@tab-beim-bundestag.de

www.tab-beim-bundestag.de

2016

Umschlagbild

aliencaat © 123RF.com

Gedruckt auf Circle offset Premium white.

ISSN-Print 2364-2599

ISSN-Internet 2364-2602

Zusammenfassung

Als Mensch-Maschine-Entgrenzung wird im Folgenden das Phänomen bezeichnet, dass sich die ursprünglich klaren Grenzen zwischen Menschen und den von ihnen geschaffenen technischen Arbeitsmitteln, den Maschinen, zunehmend aufzulösen beginnen. Dazu beigetragen haben die großen Erkenntnisfortschritte, die in verschiedenen, nur lose miteinander verknüpften, aber immer stärker konvergierenden wissenschaftlichen Bereichen in den letzten Jahren und Jahrzehnten gemacht worden sind. So ermöglichen die Neurowissenschaften Einsichten und Eingriffe in das Gehirn, die noch vor Kurzem undenkbar schienen, während das Forschungsfeld der künstlichen Intelligenz (KI) zunehmend erfolgreich an der Nachbildung lernfähigen und intelligenten Verhaltens arbeitet. Zusammen mit Fortschritten in den Informations- und Neurotechnologien eröffnen sich dadurch zukunftssträchtige Anwendungsperspektiven, die das Potenzial haben, das Mensch-Technik-Verhältnis ganz neu zu definieren. Eine Schlüsselrolle spielen dabei immer autonom agierende Maschinen, die als kollaborative Roboter über zunehmend menschliche Züge und Fertigkeiten verfügen oder als (extern oder intern getragene) Neuroprothesen mit dem menschlichen Körper verschmelzen. Auch wenn die technologische Entwicklung erst in den Anfängen steckt, begleitet und angetrieben wird sie gleichwohl von weitreichenden Zukunftsvorstellungen, welche bei aller Unschärfe und Widersprüchlichkeit die große gesellschaftliche Tragweite dieser Technisierungsprozesse deutlich vor Augen führen. Der vorliegende Sachstandsbericht hat u. a. zum Ziel, auf Basis einer fundierten Analyse des gegenwärtigen Standes von Forschung und Technik dem Realitätsgehalt dieser Visionen auf den Grund zu gehen.

Zukunftsvisionen

Die öffentlichen Debatten zu Technologien der Mensch-Maschine-Entgrenzung werden in hohem Maße durch weitreichende Zukunftsvisionen beeinflusst, die seit Anfang der 2000er Jahre vielfältige Beachtung finden. Diese Visionen sind kulturell sehr wirkmächtig und haben ganze Genres, wie z. B. die Science-Fiction (in der Literatur, im Film und Spielbereich), maßgeblich geprägt. Vor diesem Hintergrund können die gegenwärtigen Diskussionen über Human Enhancement (HE) und über KI als zwei Paradebeispiele für die in den letzten Jahren zunehmend vertretene Ansicht gelten, dass bei der Ab- und Einschätzung neuer oder in der Entstehung befindlicher (emergierender) Technologien zumeist auch eine systematische Untersuchung von Zukunftsvisionen angezeigt ist.

Betrachtet man allein die naturwissenschaftlich-technischen Grundlagen und Anknüpfungspunkte der visionären Diskussionen, liegt es nicht unbedingt nahe, Zukunftsvisionen zum neurotechnologischen Enhancement sowie zur KI und Robotik gemeinsam zu betrachten. Wie der Bericht zeigt, gibt es aber nicht nur auf der Anwendungs-, sondern auch auf der diskursiven Ebene etliche Berührungspunkte. So spielt der sogenannte Transhumanismus, dem es um eine radikale Veränderung oder gar Überwindung der *Conditio humana* geht, als ein übergreifendes Deutungsmuster eine zunehmend zentrale Rolle. Auch auf der Akteursebene sind hinsichtlich der Visionsproduzenten und -verbreiter und ihrer Unterstützer erhebliche Überlappungen der beiden Felder festzustellen.

Im Wesentlichen haben sich die visionären Diskussionen über die Zukunftsaussichten »starker« KI und über neurotechnologiebasierte Mensch-Maschine-Entgrenzungen nach einem Muster entwickelt, das aus älteren zukunftsvisionären Technikdebatten – z. B. über Nanotechnologie – bekannt ist. In mehr oder weniger eng miteinander verbundenen, relativ kleinen Zirkeln namhafter Naturwissenschaftler, Ingenieure, Autoren und Industrieller werden z. T. sehr weitreichende Zukunftserwartungen diskutiert. Diese Erwartungen sind jedoch – wie die Analyse im Bericht zeigt – weitgehend deckungsgleich mit Spekulationen, die bereits vor vielen Jahrzehnten angestellt wurden. Im Fall der aktuellen Debatte über »Superintelligenz« (bzw. eine sehr leistungsfähige KI) wurden im Wesentlichen Motive aus den 1960er Jahren aufgegriffen – etwa die Idee einer selbstbewussten und sich selbstverstärkenden KI (Intelligenzexplosions- oder heute zumeist Singularitätshypothese) des Mathematikers Irving J. Good. Auch die aktuellen Diskussionen über ein Enhancement mit neurotechnologischen Mitteln haben eine lange Vorgeschichte. Ein grundlegendes Element ist die Interpretation des Menschen als ein Mängelwesen, vor deren Hintergrund dann z. B. der Physiker John D. Bernal, ein Pionier der modernen Lebenswissenschaften und des transhumanistischen Denkens, im Jahr 1929 die Vision einer massiven, auch neurotechnologisch ermöglichten Cyborgisierung des menschlichen Körpers und einer letztendlichen Überwindung seiner natürlichen Beschränkungen entwarf. Neu – z. B. im Vergleich zum Diskurs über Nanotechnologie – ist jedoch zum einen, dass führende Figuren der Computer- und Internetindustrie (z. B. der PayPal-Gründer Elon Musk) nun selbst die Debatten inhaltlich mitprägen sowie die Forschung zu technikvisionären Themen und transhumanistische Aktivisten noch stärker finanziell fördern. Zum anderen hat sich der Diskurs weit über die (vor allem in Großbritannien und den USA) seit längerer Zeit aktiven technikvisionären Zirkel hinaus ausgebreitet und auch in Deutschland zunehmend Beachtung gefunden, nicht zuletzt in den Massenmedien.

Der Umstand, dass ein relativ kleiner Kreis von Akteuren die zukunftsvisionären Diskussionen wesentlich prägt, sollte einerseits nicht zum Anlass genommen werden, die Diskussionen über Mensch-Maschine-Entgrenzungen

zu ignorieren. Andererseits sollten aber der erhebliche intellektuelle Einfluss oder die ökonomische Macht vieler Beteiligter nicht dazu verleiten, ohne nähere Prüfung eine besondere Dringlichkeit oder Bedeutung der verhandelten Themen vorauszusetzen. Letztlich haben wir es hier fast durchgängig mit z. T. schon sehr alten Zukunftsvisionen zu tun, deren tatsächliche Basis angesichts des aktuellen Forschungs- und Entwicklungsstandes aber immer wieder weiterer Klärung bedarf. Neben Analysen, in denen die Zukunftsvisionen an der Realität gemessen werden, erscheinen auch Anstrengungen angezeigt, die inhärenten Mechanismen und Dynamiken technikfuturistisch geprägter Diskussionen besser zu verstehen. Dies betrifft vor allem den Aspekt der Themensetzung durch futuristische Netzwerke und die Logik der massenmedialen Berichterstattung. Der Blick auf die aktuelle Debatte über »starke« KI etwa zeigt, dass zwar viele Prominente vor der Entstehung einer »starken« KI mit Ichbewusstsein warnen, womit sie gleichwohl dazu beitragen, bestimmte Technikvisionen zu popularisieren und Aufmerksamkeit für noch nichtexistierende Technologien zu schaffen.

Stand und Perspektiven der Neurotechnologien

Die Reiz- und Signalverarbeitung im menschlichen Nervensystem erfolgt primär in Form elektrischer Erregung. Neurotechnologien machen sich dieses neurophysiologische Wirkprinzip zunutze, um maschinelle Systeme über neuroelektrische Schnittstellen mit dem menschlichen Nervensystem zu koppeln. Im Falle *stimulierender Schnittstellen* werden elektrische Signale dabei vom Apparat zum Nervensystem, im Falle *ableitender Schnittstellen* vom Nervensystem zum Apparat übertragen. Außerdem ist zwischen invasiven und nichtinvasiven Verfahren zu unterscheiden, je nachdem, ob sich die Schnittstelle im Inneren des Körpers befindet, was einen operativen Eingriff erforderlich macht, oder auf der Körperoberfläche, was zwar in der Anwendung unkomplizierter ist, dafür jedoch unspezifischere Signale zur Folge hat. Durch diese verschiedenen Typen an neuroelektrischen Schnittstellen ergibt sich eine große Bandbreite an möglichen Anwendungen, hauptsächlich im klinischen Bereich, die sich in unterschiedlichen Phasen der Entwicklung bzw. Anwendungsreife befinden.

Ableitende Systeme unterscheiden sich danach, wo neuronale Aktivität erfasst wird (im peripheren oder zentralen Nervensystem) und auf welche Art (invasiv vs. nichtinvasiv), sowie natürlich auch danach, welche Geräte letztendlich gesteuert werden sollen (Prothese, Rollstuhl, Computer etc.). Die Steuerung von Greifprothesen ist einer der zentralen und angesichts der Vielzahl der Betroffenen auch kommerziell aussichtsreichsten Anwendungsbereiche. Seit über 30 Jahren behilft man sich hier mit myoelektrischen Prothesen, bei denen Elektroden die verbliebene Muskelaktivität im Stumpf erfassen und damit batterie-

betriebene Motoren in Gang setzen. Besonders bei der technischen Nachbildung der Hand mit ihren verschiedenen Bewegungsmöglichkeiten stößt diese Methode aber aufgrund schwacher Signale schnell an Grenzen. Abhilfe verspricht eine Steuerung direkt durch die Hirnaktivität, also quasi durch Gedankenkraft, ohne den Umweg über das periphere Nervensystem. Dieser Weg wird seit etwa 20 Jahren intensiv erforscht. In dem vorliegenden Bericht werden diesbezügliche Anwendungen, die nichtinvasive Hirnschnittstellen umfassen, als Brain-Computer-Interfaces (BCI) bezeichnet, solche mit invasiven Schnittstellen als Brain-Machine-Interfaces (BMI):

- › *BCI-Systeme* nutzen in der Regel die neuroelektrische Gehirnaktivität, die über Elektroenzephalogramm (EEG) an der Kopfoberfläche erfasst und in maschinelle Steuerkommandos übersetzt wird. Sie sind bislang vor allem als computerbasiertes Kommunikationsmittel (z. B. Buchstabierhilfe bzw. Surfen im Internet) für schwer gelähmte Patienten mit einigem Erfolg erprobt worden. Andere Anwendungszwecke – insbesondere die Steuerung externer Geräte wie Rollstühle, Hilfsroboter, Exoskelette, Drohnen etc. – werden zwar seit vielen Jahren intensiv erforscht, haben es aber noch nicht bis zur Anwendungsreife gebracht. Schnelle Erfolge sind auch in nächster Zeit nicht zu erwarten. Denn die Gewinnung von Oberflächenpotenzialen auf der Kopfhaut von gesunden wie kranken Nutzern ist zwar unproblematisch und weitgehend risikolos, leidet gleichzeitig aber an geringer räumlicher Auflösung sowie an verrauschten Signalen mit geringer Bandbreite. Die Nichtinvasivität erweist sich folglich zugleich als größter Pluspunkt wie auch als wesentlicher Nachteil derartiger Anwendungen.
- › *BMI-Systeme*: Im Unterschied zu den räumlich schlecht aufgelösten und oft verzerrten EEG-Ableitungen lassen sich mittels implantierter Mikroelektroden prinzipiell sehr hohe und spezifische Datenraten aus dem Gehirn gewinnen. Aus diesem Grund eignen sich BMI-Anwendungen besser dazu, auch komplexere Gerätschaften wie etwa avancierte Gliedmaßenprothesen zu kontrollieren. Allerdings sind Prothesen, die direkt mit dem Gehirn verkabelt sind, noch nicht aus dem Stadium der Grundlagenforschung herausgekommen und wurden bislang vor allem im Tierversuch und nur in Einzelfällen beim Menschen prototypisch realisiert. Einer verbreiteten Anwendung steht derzeit neben den operativen Risiken vor allem die noch sehr eingeschränkte Langzeitstabilität der implantierten Sonden im Weg, die aufgrund der Fremdkörperreaktion des Körpers (Einkapselung) in der Regel nach wenigen Wochen ihre Funktion verlieren.

Während sich die ableitenden Systeme noch in einem sehr frühen Entwicklungsstadium befinden, was vor allem mit den komplexen Anforderungen an die Datenerfassung und -analyse zusammenhängt, gehören die vom Aufbau her

etwas einfacheren *stimulierenden Systeme* teilweise bereits seit Längerem zum klinischen Alltag. Auch hier ist zwischen invasiven und nichtinvasiven Verfahren zu unterscheiden, wobei vor allem invasive Stimulationsanwendungen – etwa das Cochlea-Implantat zur Wiederherstellung des Gehörs oder die tiefe Hirnstimulation (Deep Brain Stimulation [DBS]) zur symptomatischen Therapie neuromotorischer Erkrankungen wie Parkinson – zum festen Standardrepertoire der Medizin gehören. Der Vorteil von invasiv stimulierenden Schnittstellen, die neben den bereits erwähnten noch viele andere therapeutische Anwendungen haben (z. B. als Rückenmark-, Muskel-, Blasen- oder Enddarmstimulatoren), liegt darin, dass im Unterschied zu den derzeit erforschten nichtinvasiven Stimulationsverfahren, wie der transkraniellen Magnet- oder der Elektrostimulation, einzelne Muskeln, Nerven oder Hirnbereiche wesentlich spezifischer gereizt werden können.

Speziell die weite Verbreitung sensorischer Neuroprothesen und der DBS-Therapie, die beide auf das zentrale Nervensystem einwirken, wirft ethische Fragen auf und befeuert Cyborgvisionen, die eine baldige Verschmelzung von Mensch und Technik zu einer Art kybernetischem Mischwesen prognostizieren:

- > Bei der *tiefen Hirnstimulation*, die außer bei Parkinson inzwischen u. a. auch bei schweren Fällen von Zwangsstörungen und Depressionen zum Einsatz kommt, werden dem Patienten winzige Elektroden tief in das Gehirn implantiert und dort verankert, der Stromimpuls erfolgt über einen Impulsgeber, der unterhalb des Schlüsselbeins unter die Haut eingepflanzt wird. Der therapeutische Nutzen der DBS ist bei den erwähnten Indikationen anerkannt, dennoch kommt sie nur als ultima ratio zur Anwendung. Dies hängt zum einen mit der Schwere des Eingriffs, zum anderen aber auch mit grundsätzlichen ethischen Vorbehalten zusammen, da bei Patienten teilweise gravierende psychische Nebenwirkungen bis hin zu Persönlichkeitsveränderungen beobachtet wurden.
- > *Sensorische Prothesen*, wie das weitgehend ausgereifte Cochlea-Implantat oder das seit Kurzem erhältliche Retina-Implantat (zur Wiederherstellung des Sehsinns), setzen an peripheren Sinnesnerven an, um durch Reizung derselben die verlorengegangene Wahrnehmungsfähigkeit zumindest rudimentär zu kompensieren. Dabei verschmelzen menschliches Gehirn und technisches Hilfsmittel zu einer Funktionseinheit, wobei auch gewisse Sinneserweiterungen möglich (z. B. ein überdurchschnittlich gutes Hören in sehr lauten Umgebungen) oder für die Zukunft denkbar sind (z. B. das Hören von Ultraschall).

Zwar bieten die bestehenden Neurotechnologien also bereits jetzt Optionen, die menschliche Sinneswahrnehmung über das natürlich gegebene Maß hinaus zu erweitern. Es wäre jedoch sehr fragwürdig, solche technischen Sinneserweite-

rungen als eine wünschenswerte Verbesserung der menschlichen Natur darzustellen, weil der erreichbare Nutzen gegenüber den damit verbundenen gesundheitlichen Risiken gering erscheint und daher keine breitere gesellschaftliche Anwendung bei Nichtkranken rechtfertigen und auch in näherer Zukunft nicht erwarten lässt.

Im Rahmen der invasiven Interaktion mit dem Nervensystem ist in jüngster Zeit eine neue Anwendungsdimension am Horizont aufgetaucht, die mit der Entwicklung *bidirektionaler Schnittstellen* einhergeht. Ziel ist eine Integration ableitender und stimulierender Verfahren. Konkret stehen zwei Anwendungsgebiete im Vordergrund:

- › »fühlende« Handprothesen, die eine sensorische Rückmeldung geben und damit ein natürlicheres Greifen ermöglichen sollen;
- › »intelligente« Implantate (auch Elektrozeutika genannt), die Hirnsignale autonom überwachen und nur bei Vorliegen von Auffälligkeiten, die etwa auf einen bevorstehenden epileptischen Anfall hindeuten, stimulatorisch aktiv werden.

Mit der neurotechnologischen Integration von Sensorik und Aktorik wäre eine neue Dimension der Technisierung des Menschen erreicht. Bis derartige invasive Systeme klinisch nutzbar sein werden, dürfte es noch ein langer Weg sein. Wie die Geschichte der Neurotechnologien zeigt, besteht in diesem Feld ein generelles Ungleichgewicht zwischen den intensiven, oft jahrzehntelangen Forschungsbemühungen und den eher spärlichen klinischen Anwendungserfolgen. Verantwortlich dafür sind die hohen Entwicklungshürden, die durch die relativ strengen sicherheitstechnischen Anforderungen an derartige Medizinprodukte noch verschärft werden. Hinzu kommt, dass der Markt für viele neurotechnologische Anwendungen äußerst begrenzt ist, weil die Anwendungen zumeist nur auf sehr spezifische Indikationen zugeschnitten sind und damit letztlich auch nur für überschaubare Patientengruppen infrage kommen.

Stand und Perspektiven der autonomen Robotik

Roboter – gemeinhin als Manipulationsmaschinen definiert, die über einen gewissen Grad an Autonomie verfügen – stehen synonym für die zunehmende Automatisierung der Gesellschaft, die im industriellen Bereich mit dem Aufkommen der ersten Industrieroboter in den 1970er Jahren eingesetzt hat. Hierbei handelt es sich in der Regel um große und schwere Maschinen, die aus Sicherheitsgründen strikt vom Menschen getrennt werden und weitgehend repetitive Tätigkeiten in hochgradig strukturierten Umgebungen zu vollbringen haben – Roboter dieser Art sind in der produzierenden Industrie heute massenhaft verbreitet. In den letzten 10 Jahren haben Forschung und Entwicklung auf dem

Gebiet der Robotik jedoch deutliche Fortschritte gemacht, sodass eine neue Generation von Robotern am Horizont auftaucht, die sogenannten *autonomen Roboter*, die als Service- oder Assistenzroboter zunehmend in alltägliche Bereiche vordringen und dabei immer enger mit Menschen interagieren. Dieser neue Robotertyp verfügt über einen höheren Grad an Selbstständigkeit und ist in der Regel deutlich leichter und flexibler konstruiert als klassische Industrieroboter – er stellt mithin ganz neue Herausforderungen an Forschung und Entwicklung.

Wesentliche Wissensgrundlagen für die Konstruktion autonomer Systeme sind in den beiden eng verwandten Forschungszweigen KI sowie des maschinellen Lernens (ML) verankert. Insbesondere die KI-Forschung, deren Wurzeln bis in die 1940er Jahre zurückgehen, hat in ihrer wechselvollen Geschichte die Entwicklung robotischer Systeme maßgeblich geprägt. Während lange Zeit die Vorstellung einer »starken« KI leitend war, also die künstliche Nachbildung allgemeiner menschlicher Intelligenz inkl. Bewusstsein, hat diese Idee nach etlichen Rückschlägen innerhalb der Wissenschaft inzwischen an Anziehungskraft verloren. Stattdessen stehen heute spezifische Anwendungsprobleme im Vordergrund, die man mithilfe künstlicher Systeme zu lösen versucht – ohne dabei im umfassenden Sinne intelligente Maschinen schaffen zu wollen (sogenannte »schwache« KI). In den Bereich der »schwachen« KI fallen etwa die Forschungen zum ML, ein Gebiet, in dem in den letzten Jahren auf algorithmischer Ebene beeindruckende Anwendungserfolge erzielt werden konnten.

Autonome Roboter lassen sich als Perception-Action-Learning-Systeme charakterisieren, d. h. als physische Agenten, die Wahrnehmung und Verhalten interaktiv und unter Echtzeitbedingungen realisieren. Die Entwicklung derartiger Systeme erfordert eine konsequent interdisziplinäre Herangehensweise, die weit über die skizzierten Grundlagenfächer der KI- und ML-Forschung hinausgeht und vor allem auch technische Disziplinen wie die Mechatronik, die Regelungstechnik oder die Elektrotechnik involviert. Zentrale Teilbereiche von Forschung und Entwicklung im Feld der autonomen Robotik sind der Wahrnehmung, Lern- und Planverfahren (Steuerung) sowie der Systemintegration zugeordnet:

- > *Wahrnehmung*: Grundlage für die Verhaltenssteuerung autonomer Roboter ist die Erstellung eines hinreichend guten Modells der Roboterumgebung und des eigenen Zustands aus sensorischen Informationen. Dafür braucht es erstens vielfältige Sensorik, die interne und externe Parameter erfassen kann, und zweitens Verfahren, die eine semantische Interpretation der wahrgenommenen Objekte ermöglichen (Objektklassifikation). Im Bereich der Sensorik und den Methoden der Dateninterpretation, Mustererkennung und Kartierung (etwa mithilfe von Deep Learning) wurden in den letzten Jahren deutliche Fortschritte erzielt. Dennoch stellt die autonome Wahrnehmung in Echtzeit, vor allem in Bezug auf ein umfassendes Szenenver-

ständnis auf Grundlage einzelner Wahrnehmungsinhalte, immer noch eine große Herausforderung dar. Dies gilt im Besonderen, wenn es darum geht, menschliche Bewegungen zu erkennen und dann die Absicht dieser Bewegungen zu eruieren, was für die Interaktion mit Menschen von großer Bedeutung ist.

- › *Lern- und Planverfahren:* Neben Sensoren und Aktuatoren, die es ermöglichen, die Umwelt zu vermessen und in diese einzugreifen, sind insbesondere Lern- und Adaptionsmechanismen zentral, welche das System befähigen, sich an sich verändernde Umgebungen anzupassen oder komplett neue Verhaltensweisen zu erwerben. Die verfügbaren Lern- und Planungsansätze stoßen in unstrukturierten Umgebungen und bei Bewegungsabläufen, wie sie humanoide Manipulationsroboter (mit ihren oft über mehr als 50 Regelungsdimensionen) zu vollbringen haben, noch definitiv an ihre Grenzen. Erforderlich ist viel manuelle Programmierung und externe Feinsteuerung. Werden heute Roboter in komplexen Domänen eingesetzt, dann agieren diese deshalb in der Regel nur semiautonom: Der Roboter führt einzelne, klar definierte Aktionen autonom aus, die übergeordnete Steuerung sowie Zielfestlegung ist jedoch weiterhin Sache des Menschen.
- › *Systemintegration:* Ein Roboter ist mehr als die Summe seiner Teile. Erst durch die mechanische, elektrische und informationstechnische Verbindung der verschiedenen Systembestandteile entsteht in der Summe sinnvolles Roboterverhalten. Ein wesentlicher Teil der Integration der unterschiedlichen Systemkomponenten zu einem funktionsfähigen Gesamtsystem wird durch die Robotersoftware geleistet. Sie stellt eine der zentralen Hürden bei der Entwicklung autonomer Roboter dar, da vorhandene Softwarelösungen oft nur auf einen einzigen Roboter und einige spezialisierte Anwendungen zugeschnitten sind. Der dafür erforderliche Aufwand ist – nicht zuletzt wegen fehlender Standards im Bereich von Hard- und Software – erheblich und erfordert interdisziplinäre Kompetenzen, über die nur wenige Softwarespezialisten verfügen.

Die autonome Robotik hat in den letzten Jahren zweifelsohne erhebliche Fortschritte gemacht. Angesichts des demografischen Wandels, der die meisten westlichen Staaten erfasst hat, ist eine verbreitete Hoffnung für die Zukunft, dass autonome technische Helfer die dadurch zu erwartenden Produktivitätsverluste ausgleichen können. Das erwartete Marktpotenzial für autonome Roboter ist demnach enorm, und es erstaunt deshalb nicht, dass sowohl private als auch öffentliche Akteure – zu nennen sind insbesondere die US-amerikanische Defense Advanced Research Projects Agency (DARPA) und die großen IT-Konzerne Google, Apple und Amazon – bereit sind, viel Geld zu investieren. Die Anwendungsmöglichkeiten für autonome Roboter sind äußerst vielfältig – das Spektrum reicht vom Katastrophen- oder Kriegseinsatz über die Assistenz

in Pflegeeinrichtungen bis hin zum sozialen Gefährten für den Alltag –, die technologischen Anforderungen im Einzelnen so unterschiedlich wie die Einsatzfelder selber. Um ein genaueres Bild der heterogenen FuE-Dynamik zu bekommen, werden im Bericht exemplarische Anwendungen aus den einzelnen Einsatzbereichen vorgestellt, ihr Entwicklungsstand sowie die sich dabei stellenden Herausforderungen und Lösungsansätze kursorisch diskutiert.

Die Breite der dargestellten Anwendungsbereiche macht deutlich, dass das Feld das Potenzial hat, viele alltägliche Lebensbereiche grundlegend zu verändern. Auch wenn autonome Roboter heute schon erstaunliche Fähigkeiten besitzen, ist dennoch festzuhalten, dass noch starke Beschränkungen bestehen, die auf längere Sicht Gegenstand intensiver Forschungen bleiben dürften. So erstreckt sich der kommerzielle Aufschwung bislang im Wesentlichen auf mobile Roboter, d.h. Systeme, die sich in robotergerechten Umgebungen auf Rädern fortbewegen und nur sehr spezialisierte Fertigkeiten haben. Die »Königsdisziplin« der autonomen Robotik, nämlich humanoide Systeme, die auch komplexere Manipulationsaufgaben in unstrukturierten Situationen selbstständig durchführen können, befindet sich nach wie vor in einem frühen Entwicklungsstadium – und damit auch Assistenz- und Serviceroboter, die über diese Kompetenzen verfügen müssen, um die in sie gesetzten großen Zukunftshoffnungen erfüllen zu können. Viele Einzelprojekte scheitern mangels langfristiger Finanzierung oder weil sie es nicht schaffen, die benötigten interdisziplinären Fachkompetenzen zu bündeln, die für die Bewältigung eines solchen Vorhabens unabdingbar sind.

Vor diesem Hintergrund ist nicht davon auszugehen, dass sich die gesellschaftlichen Veränderungen abrupt vollziehen werden – eher ist mit einer allmählichen Verbreitung autonomer Roboter im Alltag zu rechnen, womit Raum für politische Weichenstellungen und eine gesellschaftliche Gestaltung der Entwicklung bleiben dürfte. In Bezug auf die autonome Robotik scheint dieser Reflexionsprozess aber auch besonders drängend, und er sollte sich nicht nur auf Sicherheitsfragen beschränken. Denn Ziel dieses FuE-Feldes ist es, Maschinen zu erzeugen, die bisher dem Menschen vorbehaltene Tätigkeiten übernehmen. Diese menschenähnlichen Maschinen transportieren (im Rahmen ihrer technologischen Möglichkeiten) ein bestimmtes Bild des Menschen sowie sozialer Beziehungen und werden somit, je mehr sie in unseren Alltag Einzug halten, auch unser Selbstverständnis und unsere Lebensweise grundlegend herausfordern.

Schlussfolgerungen

Noch entwickeln sich die beiden Technikfelder (Neurotechnologien und autonome Robotik) weitgehend abgegrenzt voneinander, was aufgrund der sehr unterschiedlichen wissenschaftlich-technologischen Grundlagen auch nicht weiter

erstaunlich ist. Aus übergeordneter Sicht ist mit Blick auf ihre Implikationen jedoch hervorzuheben, dass ihre Anwendungsperspektiven die beiden Pole einer übergreifenden Entgrenzungsdynamik markieren: Maschinen werden immer menschenähnlicher (Roboter), gleichzeitig werden menschliche Körper zunehmend technisiert (Cyborg). Auf die Spitze getrieben wird diese Entwicklung durch die sich immer deutlicher abzeichnende Konvergenz von Robotik bzw. KI und Neurotechnologien. So sind Prothesen, welche die Bewegungsabsichten des Trägers aus seinen Muskelsignalen grob »herauslesen«, bereits Standard. Neuerdings wird aber auch intensiv an sogenannten intelligenten Roboterprothesen geforscht, die über einen wesentlich größeren Grad an Autonomie verfügen, um auch bei schwachen Signalen die volle Funktionsfähigkeit von Hand, Arm oder Bein wiederherzustellen. Daneben beginnen sich weitere Anwendungen im Schnittfeld von Robotik und Neurotechnologien herauszukristallisieren, etwa »intelligente« Implantate (Elektrozeutika), robotische Exoskelette oder BCI-gesteuerte Roboter.

Das visionäre Potenzial dieser Entwicklungen steht außer Frage. Bei genauerer Betrachtung zeigt sich jedoch nicht selten, dass sich die populären, häufig transhumanistisch inspirierten Zukunftsvisionen auf Einschätzungen der Entwicklungsdynamik stützen, die bestenfalls hochkontrovers, manchmal auch unklar sind oder gar jeglicher Grundlage entbehren. Derartig überschießende Erwartungshaltungen sind nicht zuletzt ein Symptom der noch sehr unscharfen Konturen der technologischen Felder selbst, die hinsichtlich möglicher Anwendungsperspektiven einen weiten Zukunftshorizont eröffnen und damit allerlei Spekulationen Nahrung bieten. Vor diesem Hintergrund kommt der vorliegende Bericht auf Basis der technologischen Bestandsaufnahme zum Schluss, dass zwischen den realen Anwendungsmöglichkeiten und den teils sehr weitreichenden Visionen ein deutliches Missverhältnis besteht:

- › So bieten auch die klinisch etablierten Neurotechnologien wie die tiefe Hirnstimulation oder das Cochlea-Implantat nur eine gewisse Kompensation für massive krankheitsbedingte Einschränkungen und sind zudem keineswegs risikolos. Die meisten anderen Anwendungen, etwa im Bereich der Gehirn-Maschine-Schnittstellen, befinden sich noch in einem rein experimentellen Stadium, und es wird noch viel Forschungs- und Entwicklungsarbeit erforderlich sein, diese Technologien zur Marktreife zu bringen.
- › Was die KI und Robotik angeht, so scheint die Aussicht auf eine Machtübernahme künstlicher Intelligenzen derzeit vernachlässigbar zu sein angesichts der großen technologischen Herausforderungen (insbesondere der funktionalen Integration von Wahrnehmungs-, Planungs- und Manipulationsfähigkeiten), vor denen die Entwicklung komplexerer Serviceroboter steht. Die bisherigen Erfolge, welche die KI-Forschung unzweifelhaft vorzuweisen hat, beschränken sich auf »lernfähige« Softwareanwendungen –

intelligentes oder autonomes Verhalten im menschlichen Sinne zeigen diese in keiner Weise, sodass völlig unklar bleibt, ob und wie sich eine physisch verkörperte »starke« KI überhaupt realisieren lässt.

Die zu beobachtende Verkürzung der öffentlichen Debatte auf Cyborgs und künstliche Superintelligenzen ist vor allem deshalb wenig hilfreich, weil sie den Blick von den eigentlich drängenden normativen Fragen ablenkt, die mit der weiteren Verbreitung autonomer Systeme bereits jetzt relevant werden. Diese liegen, so ein Befund des vorliegenden Berichts, vor allem in den zunehmenden anthropologischen Grenzerfahrungen, die mit der kategorialen Angleichung von Mensch und Maschine im Zuge der beschriebenen Technisierungsprozesse einhergehen. Eine zentrale Rolle dabei spielt der Begriff der Handlung, der nach gängigen ethisch-rechtlichen Auffassungen eng mit dem Mensch- bzw. Personensein verknüpft ist. So gelten ausschließlich Personen als handlungsfähig, weil nur ihnen die Fähigkeit zugeschrieben wird, sich aus freien Stücken in ihrem Tun Ziele zu setzen und die Mittel auszuwählen, um diese zu erreichen. Maschinen hingegen, als Produkte des Menschen, sind nach herkömmlichem Verständnis rein technische Hilfsmittel, um menschliche Zwecke zu erfüllen. Diese diametrale Abgrenzung von Mensch und Maschine bildet einen wesentlichen Pfeiler unserer normativen Ordnung, etwa in Bezug auf moralische Verantwortungszuschreibungen, die auch rechtlichen Kategorien zugrunde liegen. Der weitere Fortschritt bei Robotern und Neurotechnologien bringt diese Grundordnung immer mehr ins Wanken, insofern nämlich, als Technik im Zuge dessen ihren rein instrumentellen Charakter verliert und als zunehmend selbstständig agierender Akteur ihrerseits tief in das persönlich-individuelle wie auch das gesellschaftlich-kollektive Selbstbewusstsein einzugreifen beginnt. Dass diese Entwicklung vielfältige normative Unsicherheiten aufwirft – etwa im Hinblick auf das leitende Menschenbild, den wünschbaren Grad an gesellschaftlicher Automatisierung (und damit auch Normierung) sowie die Gestaltung des Mensch-Technik-Verhältnisses –, dürfte auf der Hand liegen.

Diese normativen Herausforderungen machen eine frühzeitige Auseinandersetzung mit den gesellschaftlichen Konsequenzen, aber auch Rahmenbedingungen der Entgrenzungsdynamik erforderlich. Wie eine »antizipatorische Governance«, die nicht nur Wissenschaft und Technik, sondern auch Politik, Wirtschaft und potenzielle Nutzer einbeziehen sollte, organisiert werden könnte, ist aber noch eine weitgehend offene Frage. Aus der Tatsache, dass die normativen Implikationen der neuen Entgrenzungstechnologien nicht unabhängig von ihren Anwendungskontexten zu beurteilen sind, lässt sich zumindest folgern, dass eine übergeordnete Debatte zu den Chancen und Risiken dieser Entwicklungen wenig sinnvoll erscheint. Stattdessen ist die gesellschaftliche Auseinandersetzung sinnvollerweise mit einem differenzierten Blick auf die jeweiligen Anwendungsfelder und ihre moralischen Implikationen zu führen, dabei

möglichst unter Mitwirkung relevanter Akteure und Interessenvertreter. Angesichts der anthropologischen Herausforderungen durch die neuen Technologien sollte es dabei letztlich immer auch um die Frage gehen, was wir als genuine Wesenszug des Menschen und seiner Kultur verstehen sowie was davon wir für die Zukunft erhalten und vor maschinellem Zugriff schützen wollen.

Angesichts des demografischen Wandels gilt der Bereich der Pflege und Gesundheit für Deutschland zu Recht als paradigmatisches Anwendungsfeld für Technologien der Mensch-Maschine-Entgrenzung. Denn zukünftig ist mit einer zunehmenden Alterung der Bevölkerung und damit mit einem wachsenden Anteil pflegebedürftiger Menschen an der Gesamtbevölkerung zu rechnen. Der dadurch drohende Pflegenotstand gilt als eine der größten gesellschaftlichen Herausforderungen, die – so wird postuliert – nur mit dem verstärkten Einsatz neuer Technologien zu bewältigen ist. Speziell autonom agierenden Service- bzw. Pflegerobotern sowie der Kombination assistiver Technologien mit nicht-invasiven Neurotechnologien (Steuerung von Exoskeletten, Kommunikationsgeräten etc. mittels Brain-Computer-Interface) wird großes Potenzial zugeschrieben, Pflegekräfte entlasten und Pflegebedürftige im Alltag und bei der Rehabilitation unterstützen zu können – derartige Anwendungen befinden sich in Entwicklung und Erprobung. Wie entsprechende Assistenzsysteme und vor allem auch eine den pflegerischen Herausforderungen angemessene Governance der Technikentwicklung in der Praxis aussehen könnten, sind Fragen, welche in einem anschließenden TA-Projekt vertieft untersucht werden sollen.



**BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG**

KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT)

Neue Schönhauser Straße 10
10178 Berlin

Fon +49 30 28491-0
Fax +49 30 28491-119

buero@tab-beim-bundestag.de
www.tab-beim-bundestag.de