

# Können wir künstliches Bewusstsein schaffen?

Damit sich Mensch und Roboter verstehen, müssten beide denselben psychologischen Regeln folgen

**Nur dank einem menschenähnlichen Bewusstsein könnten Roboter zu empathischen Wesen werden.**

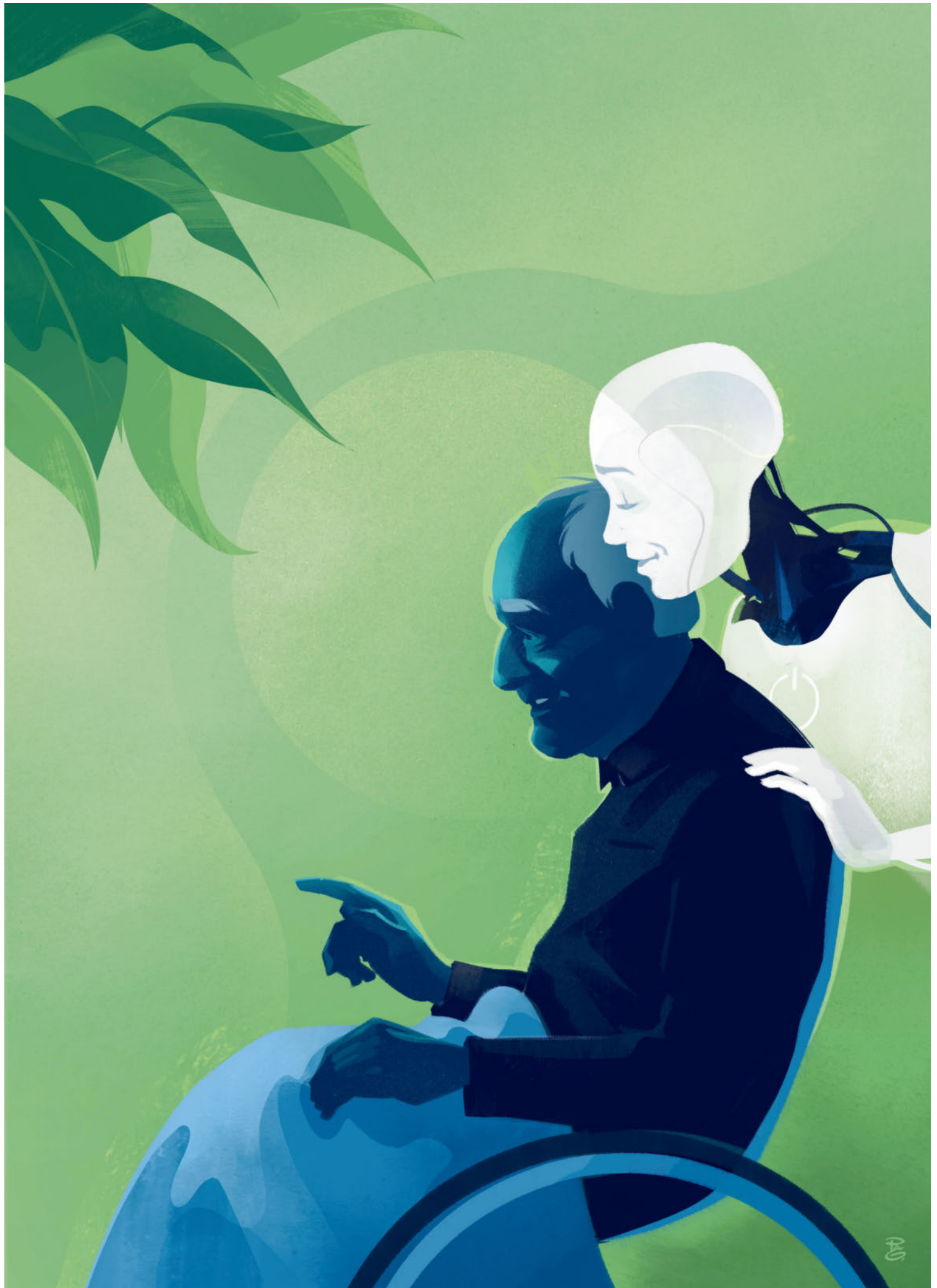
DAVID RUDRAUF

Die Vorstellung des Zusammenlebens mit empathischen Robotern, die – wie vom Schriftsteller Isaac Asimov in *Foundation* beschrieben – ein Bewusstsein besitzen und von Menschen nicht mehr zu unterscheiden sind, ist zum derzeitigen Zeitpunkt nichts weiter als Science-Fiction. Künstliche Intelligenz, agentenbasierte Modelle und Robotik haben bislang lediglich Maschinen hervorgebracht, die von einer Entwicklung abgekoppelt sind, welche die affektiven und sozialen Beziehungen des Menschen ausmacht.

Echter Fortschritt wird nur durch einen neuartigen wissenschaftlichen Ansatz zur Entwicklung eines Modells mit menschenähnlichem Bewusstsein möglich, das Kognition und Emotion vereint. Die Kybernetik der Zukunft muss auf der Psychologie aufbauen. Die Logik ist einfach: Wenn Maschinen denselben psychologischen Regeln wie Menschen folgen, können Menschen sie auf ganz natürliche Art und Weise verstehen beziehungsweise werden sie von ihnen verstanden. Die Umsetzung eines solchen Vorhabens ist jedoch sowohl konzeptionell als auch methodisch überaus komplex.

Doch das Problem gilt es nicht allein aus wissenschaftlichem Ehrgeiz oder aufgrund des grossen Interesses an Science-Fiction zu lösen: Zunehmende soziale und wirtschaftliche Herausforderungen treiben die Entwicklung einer eigenständigen und intelligenten Assistenzrobotik an. So könnten künftig im Gesundheitswesen intelligente Roboter medizinisches Fachpersonal entlasten, indem sie Patientinnen mehr Autonomie ermöglichen oder Senioren unterstützen, die immer häufiger unter sozialer Isolation leiden. Und bei Autismus könnte soziale Kognition durch robotergestützte Therapien gefördert werden. Für Menschen mit körperlichen Einschränkungen könnten intelligente Lösungen in der Mobilität eine grosse Hilfe darstellen. Laut Schätzungen von «Global Market Insights» dürfte der Robotik-Markt im Gesundheitswesen bis 2024 auf über 1,2 Milliarden Dollar anwachsen.

Während die Forschung fortschreitet und neue Technologien aufkommen, haben die derzeitigen Lösungen stets mit dem gleichen Manko zu kämpfen: Roboter verfügen über keine empathischen Fähigkeiten, die aber unabdingbar für ihre Akzeptanz sind. Empathie bezeichnet dabei die Fähigkeit, sich in andere hineinzuversetzen, um ihre Sichtweise und Wünsche zu verstehen und ihnen helfen zu können. Eine wichtige Grundlage zwischenmenschlicher Beziehungen mit kognitiven und affektiven Komponenten.



Zum derzeitigen Zeitpunkt beschränken sich alle sogenannten empathischen Roboter oder künstliche Agenten darauf, menschliches Verhalten mehr oder weniger gut zu imitieren. Sie nutzen unsere natürliche Neigung zur Vermenschlichung aus, das heisst, Maschinen insbesondere dann einen Geist zuzugestehen, wenn sie uns ähneln. Eine Tatsache, die Hersteller nur zu gerne ausnutzen. So sind zahlreiche Roboter auf dem Markt, die laut Hersteller «empathisch» sind. So wird beispielsweise der Roboter Pepper eines japanischen Herstellers als Roboter «mit Herz» und treuer Begleiter gegen Einsamkeit beworben. Dabei kann er höchstens grundlegende Emotionen erkennen und sein Verhalten entsprechend anpassen. Gleiches gilt für den humanoiden Roboter Nao desselben Herstellers, der bei Kindern mit Störungen des autistischen Spektrums ohne bedeutende Ergebnisse eingesetzt wurde. Aibo von Sony wiederum wird dank seiner rudimentären Lernfähigkeiten und seiner Fähigkeit zum Ausdrücken von Emotionen als Roboterhund angepriesen, der dazu in der Lage ist, eine emotionale Bindung aufzubauen.

In der Altenpflege gab es bereits 2015 über 100 verschiedene Assistenzroboter für eine ganze Reihe von Aufgaben (zum Beispiel in den Bereichen Mobilität, Körperpflege und soziale Aktivitäten). Ein gutes Beispiel ist der Roboter Ramcip (entwickelt von Accrea aus Polen und Certh aus Griechenland). Der durch die Europäische Union finanzierte Roboter iCub bietet eine Plattform, die fortgeschrittene kognitive Funktionen und Lernfähigkeiten ermöglicht, die jedoch noch weiterentwickelt werden müssen.

Im Bereich der künstlichen Intelligenz bietet «affective computing» Lösungen, um menschliche Emotionen erkennen und simulieren zu können, die auf maschinellem Lernen aufbauen. In die Systeme kann das sogenannte «Lernen durch Verstärkung» integriert werden, das heisst, der Roboter passt sein Verhalten durch Belohnung oder Strafe entsprechend an. Die meisten Ansätze basieren jedoch auf impliziten, durch die Neurowissenschaften inspirierten Modellen (zum Beispiel «deep learning») mit vordefinierten Funktionalitäten. Die fortschrittlichsten Ansätze orientieren sich hingegen an der Einschätzungstheorie («appraisal»), die auf Arbeiten des Schweizer Psychologen Klaus Scherer aufbaut und sensorische Einschätzungs-, Auslösungs- und Ausdruckprozesse von Emotionen einschliesst, ohne jedoch über ein umfassendes Intentionalitätsmodell zu verfügen.

Bezüglich der Verkörperung von Maschinen («embodiment») konzentriert sich die Industrie vor allem auf die Entwicklung realistischer Körperhüllen: So haben sich vor allem der neuseeländische Hersteller Soul Machines mit seinen «digitalen Menschen» und Hanson

Robotics aus Hongkong mit seinen humanoiden Robotern das Ziel gesetzt, ihren Robotern Leben einzuhauchen. Diese Ansätze zielen jedoch in erster Linie darauf ab, uns körperlich ähnliche Maschinen zu schaffen. Sie sind jedoch psychologisch weit von uns entfernt: Wir können uns nicht in ihnen wiedererkennen. Bis heute wecken alle Roboter lediglich den Anschein, autonom

## Bis heute wecken alle Roboter lediglich den Anschein, autonom zu sein.

zu sein: Ihr Verhalten wird grösstenteils durch gespeicherte Daten und programmierte Reaktionen gesteuert. Es handelt sich daher um nichts weiter als ausgeklügelte Maschinen, die sich darauf beschränken, affektives und soziales Verhalten zu imitieren.

### Hin zu empathischen Robotern

Um empathische Roboter zu entwickeln, müssten sie mit einer «menschenähnlichen Kognition» ausgestattet werden, die dazu in der Lage ist, menschliches Verhalten zu interpretieren, die Emotionen, Vorlieben und Glaubenssätze von Menschen zu verstehen und ihre Absichten vorherzusehen, um sich entsprechend zu verhalten und natürliche gegenseitige Interaktion zu ermöglichen, die auf einer gemeinsamen emotionalen Intelligenz beruht. Dazu muss eine «starke künstliche Intelligenz» entwickelt werden, die auf expliziten, von Menschen interpretierbaren Theorien und innovativen Algorithmen aufbaut und mit dem Ziel der Erschaffung eines künstlichen Bewusstseins Erkenntnisse aus Phänomenologie, Psychologie und zu den Bio-Funktionalitäten des Bewusstseins einfließen lässt. Dabei müssen Wahrnehmung, Emotion, Vorstellungsvermögen, soziale Perspektivenübernahme, Motivation und Handeln miteinander verknüpft werden. Das Ziel ist nicht, eine Superintelligenz mit fortgeschrittenen sprachlichen und logischen Fähigkeiten zu entwickeln, sondern die subjektive Infrastruktur zu modellieren, die uns zum Zeitpunkt unserer Geburt beeinflusst und die Grundlage unserer weiteren affektiven und intellektuellen Entwicklung bildet. Kurz: In einem ersten Schritt besteht das Ziel darin, künstliche Babys zu erschaffen!

Roboter werden nur dann als menschliche Wesen wahrgenommen, wenn sie ähnlich wie wir Menschen ticken. Dann könnten sie zu affektiv und sozial stimu-

lierenden persönlichen Begleitern werden, die unsere Resilienz, Autonomie und unser Wohlbefinden stärken könnten. Sie könnten bei neurologischen und psychischen Erkrankungen sowie bei körperlichen Einschränkungen Isolierung und Autonomieverlust vorbeugen. Die Modellierung in der computergestützten Psychologie schreitet immer weiter voran. Von den affektiven Wissenschaften inspirierte Arbeiten beruhen auf Inferenz- und globalen Optimierungsmechanismen, die fortschrittliche Glaubens- und Präferenzmodelle miteinschliessen, sowie auf raffinierten prädiktiven Simulationsmechanismen. Forschende aus der Schweiz, wie Dominik Bach von der Universität Zürich, leisten in diesem Bereich wertvolle Beiträge. An der Universität Genf entwickeln wir in Zusammenarbeit mit Mathematikern, Philosophen und Neurowissenschaftlern das projektive Bewusstseinsmodell (Projective Consciousness Model, PCM), das wir auf soziale Roboter wie Cozmo des Herstellers Anki anwenden.

Der Ansatz beruht auf dem neu aufkommenden Paradigma der «aktiven Inferenz» mit von der statistischen Physik und der Informationstheorie abgeleiteten Prinzipien. Das PCM beinhaltet ein Bewusstseinsfeld, das als dreidimensionaler globaler Arbeitsraum fungiert und räumliche und affektive Repräsentationen miteinander in Beziehung setzt. Es wird durch projektive Geometrie gesteuert, die imaginäre Blickpunktwechsel ermöglicht, sowie durch eine zu minimierende Menge, die freie Energie. Sie drückt die Divergenz zwischen Erwartungen und Zuständen aus, die antizipiert aus einer Handlung resultieren oder bereits realisiert worden sind und durch eine Vielzahl an affektiven Dimensionen beurteilt werden (zum Beispiel Freude, Sicherheitsempfinden, ethische Gesichtspunkte). Imaginäre Projektionen simulieren alternative Szenarien und die Perspektive anderer Agenten anhand von Inferenzen auf Grundlage vorgegebener Glaubenssätze und Präferenzen. Alle Informationen werden zur Minimierung der freien Energie gebündelt, die das System dazu befähigt, Herausforderungen anzugehen und die gemeinsame Aufmerksamkeit anderer Agenten sowie Annäherungs- und Vermeidungsverhalten zu kontrollieren. So zeigen Agenten mit Defiziten bei der sozialen Perspektivenübernahme Verhaltensweisen, die an das autistische Spektrum erinnern. Andere Agenten, die vorwiegend negative Einstellungen gegenüber anderen vorweisen, manifestieren hingegen vermeidendes Verhalten, das an soziale Angst erinnert.

Neben der Grundlagenforschung arbeiten wir mit europäischen Konsortien an der Anwendung dieses Modells zur Behandlung von Menschen mit Autismus, von älteren Personen mit kognitiven Störungen (durch Verwendung der Ramcip-Plattform) und Tetra-

plegikern, indem einer Plattform, die Roboterarm und Elektrorollstuhl verbindet, empathische Fähigkeiten verliehen werden. Ein Beispiel hierfür ist der Roboter Aria von Accrea.

### Ethische Implikationen

Stellen wir uns vor, dass künstliches Bewusstsein Wirklichkeit wird. Tiefgreifende soziale, rechtliche und wirtschaftliche Veränderungen wären die Folge. Mit dem Schweizer Rechtsanwalt Nicolas Capt und der Unternehmerin und Expertin für neue Technologien Jennah Kriebel versuchen wir, diese zu antizipieren.

Dabei stellen sich vor allem folgende grundlegenden Fragen: Wenn Bewusstsein und Selbstbestimmung der Philosophie des Rechts zugrunde liegen – welche ethische Herangehensweise ist bei der Konzipierung solcher künstlicher Systeme die richtige? Dürfen wir Sklaven erschaffen, die uns dienen, oder müssen wir ihre Freiheit ebenso wie unsere eigene fördern und schützen? Benötigen wir ein allgemeingültiges Gesetz? Wie sieht es mit Geschlechterfragen aus? Auch die wirtschaftlichen Auswirkungen wären enorm – sowohl im Hinblick auf die industrielle Entwicklung und auf künftige Märkte als auch hinsichtlich der Organisation der Arbeit. Industrie und Gesetzgeber würden sich mit einem Dilemma konfrontiert sehen: Wenn die Ausnutzung von Wesen mit Bewusstsein Sklaverei ist, würde die Industrie Kreaturen erschaffen, über die sie letztlich nicht verfügen dürfte. ♦

---

### DER AUTOR

David Rudrauf, Professor für Psychologie an der Universität Genf, leitet das Labor für multimodale Modellierung von Emotionen und Empfindungen am Biotech-Campus. Er ist Mitglied einer Arbeitsgruppe des Bundes, die sich mit dem Aufkommen einer empathischen Robotik beschäftigt. In Zusammenarbeit mit multidisziplinären Konsortien entwickelt er zudem Projekte in den Bereichen künstliches Bewusstsein und empathische Robotik.

---

### KONTAKT

[david.rudrauf@unige.ch](mailto:david.rudrauf@unige.ch)

---

### LITERATUR

Rudrauf, D., & Debbané, M. (2018). Building a cybernetic model of psychopathology: beyond the metaphor. *Psychological Inquiry*, 29(3), 156–164.

Eine komplette Bibliografie ist beim Autor erhältlich.