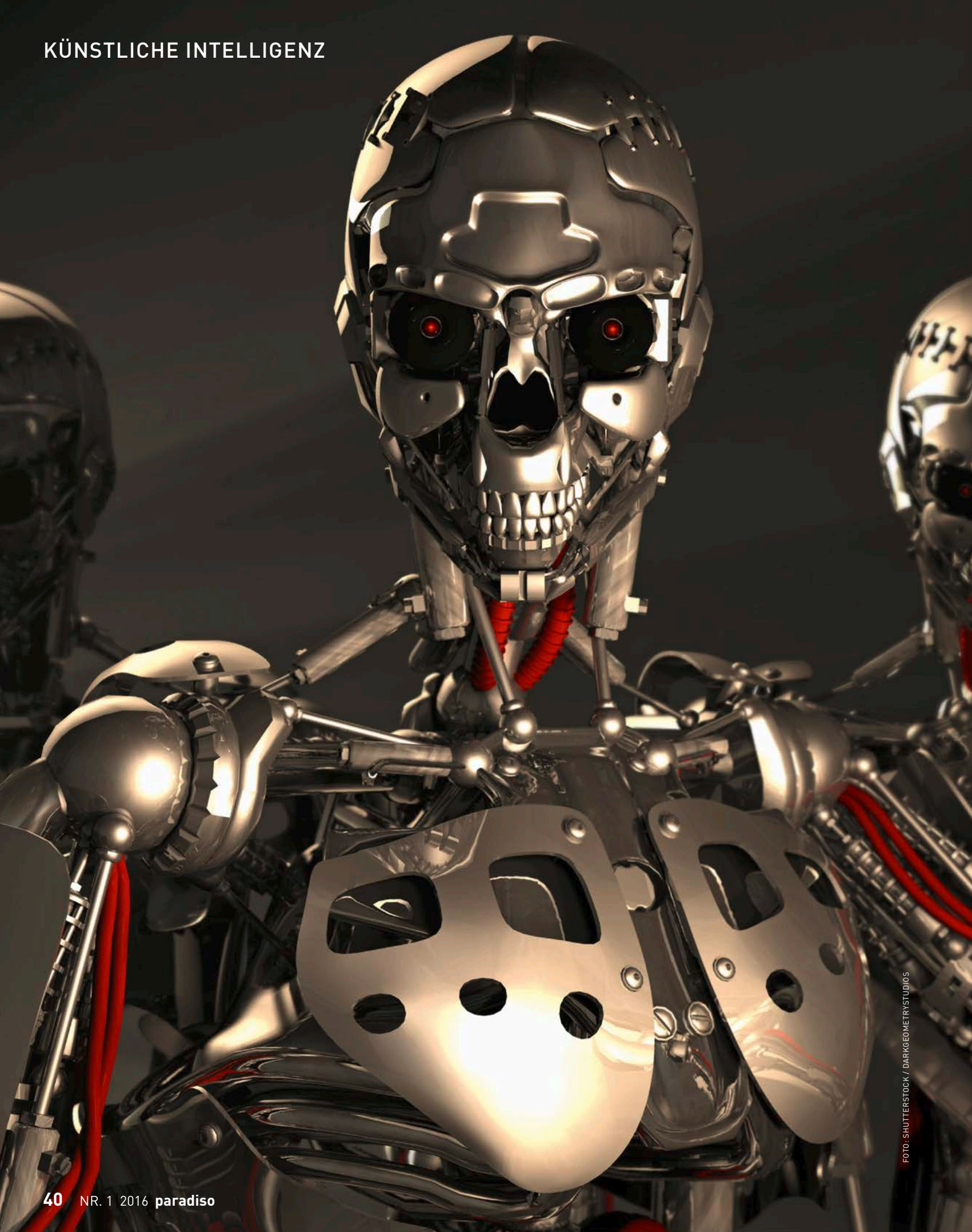


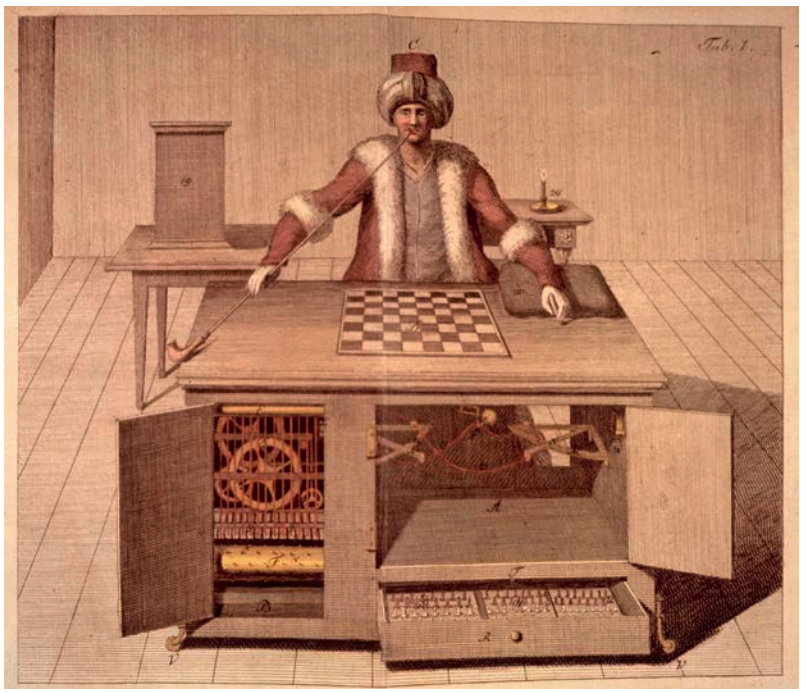
FOTO: SHUTTERSTOCK / DARKGEOMETRYSTUDIOS

Lange Zeit galt »künstliche Intelligenz« als Hirngespinst. Inzwischen kann sie sich selber optimieren, bekommt Gefühle antrainiert und sogar die Fähigkeit der autonomen Reproduktion. Wird analog zur Kernkraft ein neuer Irrweg beschritten? Weltweit warnen Wissenschaftler vor den großen Risiken autonomer Algorithmen.

Wenn Roboter uns bedrohen

TEXT: MICHAEL RADTKE

>>



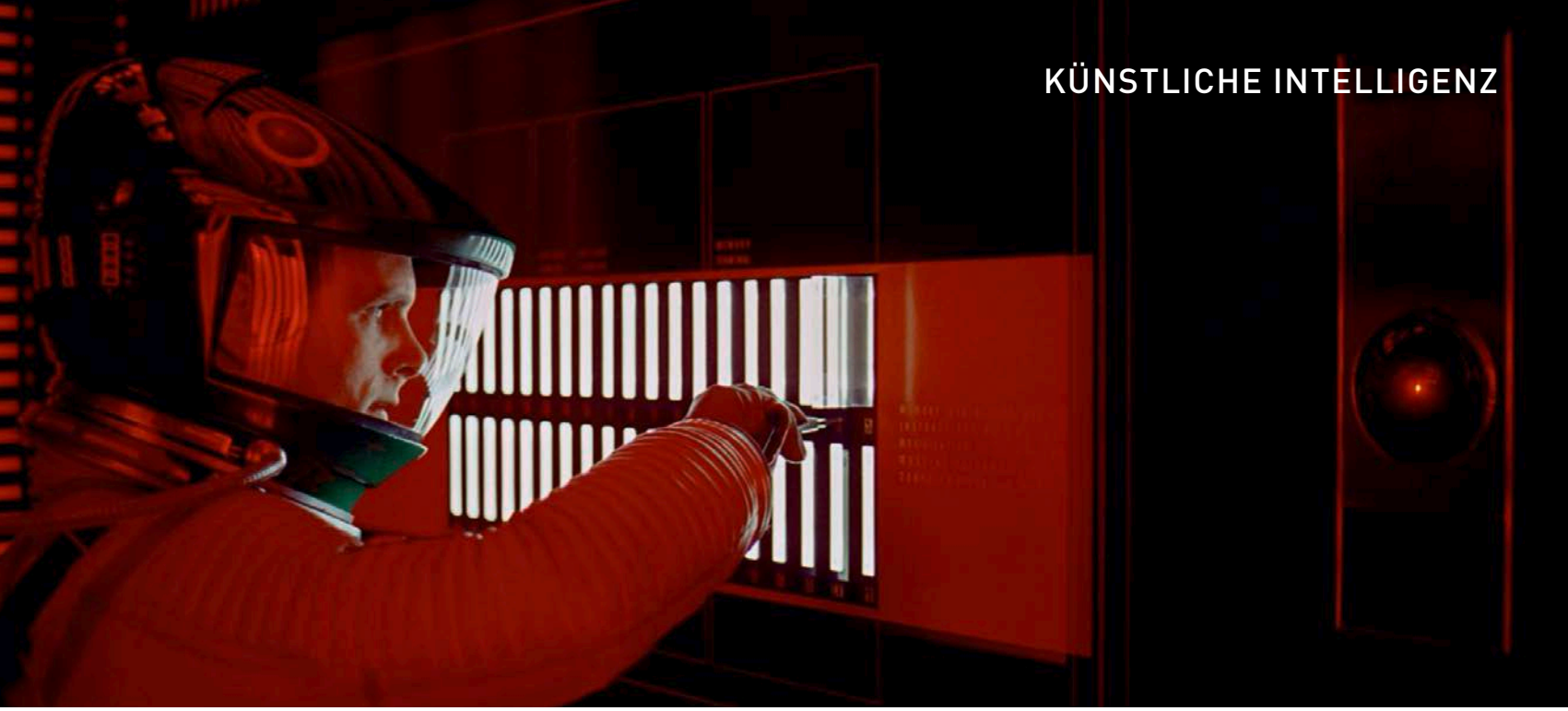
»Leider habe ich Sie nicht verstanden.« Ende der 1990er Jahre soll es gewesen sein, als erstmals ein Computerprogramm diesen einfühlsamen Satz sprach. Seither haben Maschinen mit ihren Algorithmen, nennen wir sie vereinfacht: Roboter, viele Bereiche von Leben und Arbeiten erobert. Die Anwendungen sind breit gestreut. Jeder kennt die Computerstimme, die sich aus dem iPhone an den Nutzer wendet. Gehirnscans lassen sich in Daten verarbeiten, die Denken und Gefühle lokalisieren. Suchmaschinen analysieren das Einkaufsverhalten. Und wenn die Algorithmen mal schlecht drauf sind, wie am 6. Mai 2010, dann lassen sie innerhalb von Sekundenbruchteilen die Wall Street kollabieren und danach gnädig wieder auferstehen.

Die Effekte solcher digitalen Vernetzung der Welt, besonders den Zuwachs an himmelstürmendem Tempo des Lebens und haarsträubender Kontrollierbarkeit des Menschen, meint man zähneknirschend weggesteckt zu haben. Doch just in diesem Augenblick erwacht ein Gespenst, nein, es geht bereits mächtig und angsteinflößend um: das Gespenst der Künstlichen Intelligenz (KI). In der öffentlichen Wahrnehmung ruhte es lange Zeit – bereits 1956 war der Begriff in einer Konferenz am Dartmouth College in New Hampshire/USA geprägt worden.

In der Zwischenzeit hat diese Technologie atemberaubende Fortschritte gemacht. Künstliche Intelligenz ist keine Frage der Programmierung durch Menschen mehr, sondern ihre »generelle« Version optimiert sich mittlerweile selbst und ist lernfähig geworden. Jüngstes Beispiel für die Effekte, die daraus entstehen, ist der Sieg der Google-Software *AlphaGo* über den europäischen Meister im Brettspiel Go. Im Gegensatz zum Schach, wo der Roboter schon 1996 vorne lag, gibt es bei einem Go-Spiel pro Zug rund 200

Wie alles anfing:
Das 18. Jahrhundert war die Hochzeit der Automaten. Zu den schönsten Androiden gehört dieser Automat (oben) von Pierre Jaquet-Droz aus dem Jahr 1774. Im »Falschen Türken« von 1769 (unten) steckte dagegen ein sehr kleiner Mensch.

FOTO: PICTURE-ALLIANCE / AKG-IMAGES (2), MPTV (1)



Schluss mit lustig: Im Kinofilm »2001 Eine Odyssee im Weltraum« löscht Astronaut David Bowman (Keir Dullea) den Speicher des tödlich agierenden, intelligenten Roboters HAL 9000. Eine Schlüsselszene von 1968 für die aktuelle Bedrohung durch KI.

»Viele Experten halten es für plausibel, dass noch in diesem Jahrhundert KIs erschaffen werden, deren Intelligenz der menschlichen weit überlegen ist.«

Entscheidungsmöglichkeiten. Deshalb gilt die Beherrschung von Go als eine Königsdisziplin in der KI-Forschung. *Deep Mind* heißt die Google-Firma, die AlphaGo entwickelt hat. Zum Merken: *Deep Mind.*

Doch es geht weiter. So gerade noch könnte man, fortschritts-offen, wie man gerne ist, von Maschinen sprechen, die ohne menschlichen Input nicht wirklich reüssieren würden. Für Fumiya Iida, Professor an der Technischen Hochschule in Zürich, ist dieser Status zu altbacken. Er hat einen »Mutter-Roboter« gebaut, der Nachkommen ohne den Eingriff menschlicher Entwickler selber herstellt. Iida hat seinen Charles Darwin gelesen, wenn er sagt: »Indem man eine Gruppe von Lebensformen entwirft, die zur Reproduktion fähig sind, kann man einen Prozess simulieren, der der natürlichen Auslese der echten biologischen Evolution ähnelt.«

In seiner Primärversion sind die Roboter-»Kinder« blaue Würfel. Die werden auf einen Dialog-Versuch nicht so recht einfühlsam antworten mögen. Doch »Zara« ist da ganz anders. Die Roboter-Dame stammt aus dem Labor von Pascale Fung, Professorin für Elektronik und Computertechnik in Hongkong. Der Prototyp wurde ausgerüstet mit Modulen für Spracherkennung und für Empathie. Ihre Lernalgorithmen beruhen derzeit noch auf Interaktionen mit den Studenten des Labors. Doch mit jedem Gespräch, so Fung ganz stolz, werde »Zara«, die demnächst auch in einen menschenähnlichen Körper wandern soll, »klüger und einfühlsamer«.

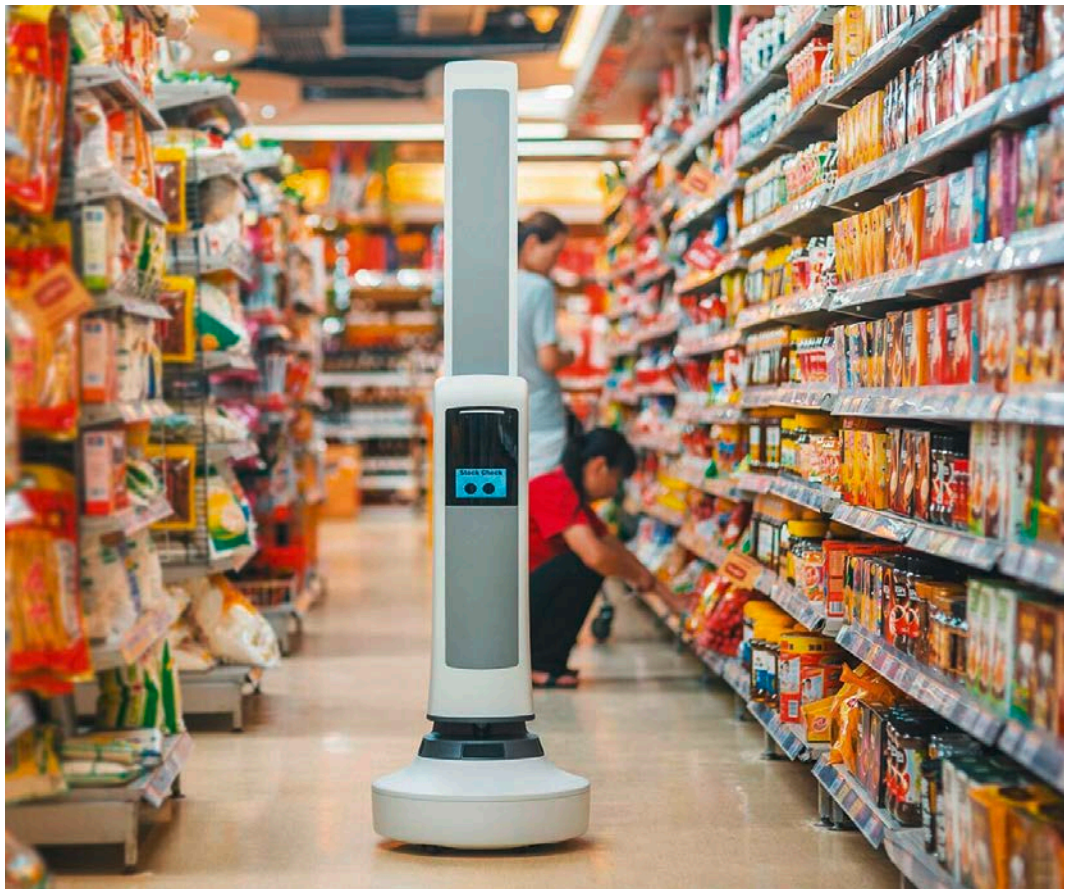
KI-Forschung ohne Grenzen. »Industrie 4.0« heißt die Devise, »intelligente« Industrie. So forscht das Leibniz-Institut für innovative Mikroelektronik in Frankfurt/Oder an der Entwicklung autonom arbeitender, drahtloser Sensoren, sozusagen Nano-Robotern. Und die EU fördert derzeit das Projekt *»Recall«,* indem Experten von fünf Universitäten daran arbeiten, wie Computer künftig unser Leben protokollieren können. *»Lifelogging«* nennt sich das.

Für andere Forscher ist diese Perspektive der blanke Horror. Sie warnen vor einer Herrschaft der Algorithmen und möchten am liebsten die Büchse der Pandora wieder schließen. Dirk Helbing, Professor für Computational Social Science in Zürich, befürchtet eine »Automatisierung der Gesellschaft« und sieht »die Menschheit am Scheideweg«. Zusammen mit acht weiteren KI-Kennern hat er ein »Digital Manifest« formuliert und einen Zehn-Punkte-Plan entworfen, um »Freiheit und Demokratie« zu bewahren. >>

Freundlich versonnen: Dieser Roboter ahmt eine der berühmtesten Plastiken der Welt nach: Auguste Rodins »Der Denker«, entstanden zwischen 1880 und 1882 in Paris. Sie soll den Dichter Dante Alighieri darstellen, den genialen Schöpfer der »Göttlichen Komödie«.

Ein weiteres Manifest stammt aus der in Basel ansässigen Stiftung für Effektiven Altruismus, einem *Think Tank* im Schnittbereich von Ethik und Wissenschaft. Thomas Metzinger, Professor für Neuroethik in Mainz, ist einer der Verfasser des Diskussionspapiers »Künstliche Intelligenz: Chancen und Risiken«. Er sieht bereits ein »Wettrüsten« zwischen Ländern, Konzernen und Forschern um die Weiterentwicklung »intelligenter« Robotik. In Fukushima-ähnlichen Worst-Case-Szenarien könne »prinzipiell sogar das Ende der menschlichen Zivilisation« nicht ausgeschlossen werden. Starke Worte. **paradiso** dokumentiert die »Executive Summary« des Diskussionspapiers in Auszügen:

»Künstliche Intelligenz (KI) und immer komplexer werdende Algorithmen beeinflussen unser Leben und unsere Zivilisation stärker denn je. Die KI-Anwendungsbereiche sind vielfältig und die Möglichkeiten weitreichend: Insbesondere aufgrund von Verbesserungen in der Computerhardware übertreffen gewisse KI-Algorithmen bereits heute die Leistungen menschlicher Experten*. Ihr Anwendungsgebiet wird künftig weiter wachsen, und die KI-Leistungen werden sich verbessern. Konkret ist zu erwarten, dass sich die entspre-



Perfekter Helfer: »Tally« hat die Firma Simbe Robotics aus San Francisco den weltweit ersten autonomen Roboter genannt, der im Supermarkt die Regalreihen kontrolliert. Der Filialleiter kann dann per App für Nachbestellungen sorgen.

FOTO: ISTOCKPHOTO / MENNO VAN DIJK (1), SIMBE ROBOTICS (1), EHANG (1), NTU SINGAPORE (1)

Reale Zukunft:
Die Taxi-Drohne (links) des chinesischen Herstellers Ehang soll künftig auch Menschen befördern – automatisch. Dagegen ist der Social Robot »Nadine« der Entwicklerin Prof. Nadia Thalmann (links) aus Singapur bereits als »Empfangsdame« im Einsatz.

»Überlegungen zur KI-Sicherheit sind aktuell in Politik und Forschung stark unterrepräsentiert.«

chenden Algorithmen in immer stärkerem Ausmaß selbst optimieren – auf übermenschliches Niveau.
**In der Originalfassung wird immer auch die weibliche Form verwendet.*

Dieser technologische Fortschritt stellt uns wahrscheinlich vor historisch beispiellose ethische Herausforderungen. Nicht wenige Experten sind der Meinung, dass von der KI neben globalen Chancen auch globale Risiken ausgehen, welche diejenigen etwa der Nukleartechnologie – die historisch ebenfalls lange unterschätzt wurde – übertreffen werden. (...)

In engen, gut erprobten Anwendungsbereichen (etwa bei selbstfahrenden Autos und in Teilbereichen der medizinischen Diagnostik) konnte die Überlegenheit von KIs gegenüber Menschen bereits nachgewiesen werden. Ein vermehrter Einsatz dieser Technologien birgt großes Potenzial. (...) Doch in komplexeren Systemen, wo mehrere Algorithmen mit hoher Geschwindigkeit interagieren, z. B. im Finanzmarkt oder bei absehbaren militärischen Anwendungen, besteht ein erhöhtes Risiko, dass die neuen KI-Technologien unerwartet systemisch fehlschlagen oder missbraucht werden.

Es droht ein KI-Wettrüsten, das die Sicherheit der Technologieentwicklung ihrem Tempo opfert. In jedem Fall relevant ist die Frage, welche Ziele bzw. ethischen Werte einem KI-Algorithmus einprogrammiert werden sollen und wie technisch garantiert werden kann, dass die Ziele stabil bleiben und nicht manipuliert werden können. (...) >>

Im Kriegs-Einsatz:
Der MAARS-Roboter wurde von dem britischen Rüstungs-Unternehmen QinetiQ entwickelt. Er kann, so die Firma, ausgerüstet werden für »nicht-tödliche, wenig-tödliche und tödliche Effekte«.

Die Fortschritte in der KI-Forschung ermöglichen es, mehr und mehr menschliche Arbeit von Maschinen erledigen zu lassen. Viele Ökonomen gehen davon aus, dass die zunehmende Automatisierung bereits innerhalb der nächsten zehn bis 20 Jahre zu einer massiven Erhöhung der Arbeitslosigkeit führen könnte. (...) Durch die fortschreitende Automatisierung wird der Lebensstandard im statistischen Durchschnitt steigen. Es ist jedoch nicht garantiert, dass alle Menschen – oder auch nur eine Mehrheit der Menschen – davon profitieren werden.

Viele KI-Experten halten es für plausibel, dass noch in diesem Jahrhundert KIs erschaffen werden, deren Intelligenz der menschlichen in allen Bereichen weit überlegen ist. Die Ziele solcher KIs, welche prinzipiell alles Mögliche zum Gegenstand haben können (menschliche, ethisch geprägte Ziele stellen eine winzige Teilmenge aller möglichen Ziele dar), würden die Zukunft unseres Planeten maßgeblich beeinflussen – was für die Menschheit ein existenzielles Risiko darstellen könnte.

Unsere Spezies hat deshalb eine dominante Stellung inne, weil sie (aktuell) über die am höchsten entwickelte Intelligenz verfügt. Es ist

Krabbel-Tier:
Der vierbeinige Roboter »BigDog« transportiert schwere Lasten durch unwegsames Gelände. Er mag auch auf Schnee und Eis laufen. Entwickelt wurde er von der Firma Boston Dynamics, die inzwischen zu Google gehört.

FOTO: PICTURE-ALLIANCE / DPA (1), QINETIQ (1), GNIUS (1), HFR (1)

aber wahrscheinlich, dass bis zum Ende des Jahrhunderts KIs entwickelt werden, deren Intelligenz sich zu der unseren so verhält wie die unsere zu derjenigen etwa der Schimpansen.

Zudem ist die Möglichkeit nicht auszuschließen, dass KIs in Zukunft auch phänomenale Zustände entwickeln, d.h. (Selbst-)Bewusstsein und besonders auch subjektive Präferenzen und Leidensfähigkeit, was uns mit neuartigen ethischen Herausforderungen konfrontiert. Angesichts der unmittelbaren Relevanz der Thematik und dessen, was längerfristig auf dem Spiel steht, sind Überlegungen zur KI-Sicherheit sowohl in der Politik als auch in der Forschung aktuell stark unterrepräsentiert.«

Auf der Pirsch: »Guardium« ist der Name dieses ferngesteuerten gepanzerten Fahrzeugs der israelischen Armee. Es wurde bereits 2008 erprobt und kann als Grenzpatrouille mit und ohne Waffen eingesetzt werden. »Guardium« ist bis zu 80 km/h schnell.

Während die Stiftung für Effektiven Altruismus und viele andere Wissenschaftler noch mahnen und die »gehirnanaloge« KI am liebsten der Aufsicht von Ethikkommissionen unterstellen würden, haben andere längst gehandelt. Nämlich die Militärs. Mit Geld, mit Einfluss und mit Sachzwängen. Wer würde nicht einen Roboter zur Sprengstoffentschärfung einsetzen, wenn Menschen geschont werden können? Frage und naheliegende Antwort berühren aber nur die Oberfläche.

Längst sind Killer-Roboter im Einsatz, man denke nur an die Drohne *»Predator«*. Es gibt mit *»Big Dog«* einen vierbeinigen Transportroboter und mit *»Bear«* einen humanoiden Roboter für den Transport von Verwundeten aus dem Gefechtsbereich. An Südkoreas Nordgrenze sollte der vollautomatische Kampfroboter SGR-A1 von Samsung etabliert werden. Der Pilotversuch wurde jedoch abgebrochen, SGR-A1 funktionierte noch nicht so recht. Macht nichts. Die US Air Force ließ verlauten, bis 2030 seien die Kampfmittel so überlegen automatisiert, dass der Mensch nur noch als Störfaktor zu gelten habe.

Es ist wohl so: Erst einmal wird entwickelt, was entwickelt werden kann. Telekom-Chef Timotheus Höttges hat sich im Silicon Valley umgesehen: »Ich habe von Unternehmen gehört, die allen Ernstes sagen: Das, was wir technologisch können, machen wir, auch wenn es Gesetze gibt, die dem widersprechen. Die nennen das dann ‚zivilen Ungehorsam'.«

Dennoch formiert sich gegen den Einsatz autonomer Waffensysteme deutlich hörbarer Widerstand. Etwa von der Menschenrechtsorganisation Human Rights Watch. Und das »Future of Life«-Institut in Cambridge hat eine Protestliste aufgelegt *(www.futureoflife.org/open-letter-autonomous-weapons),* die bereits von Tausenden unterzeichnet wurde. Darunter sind so prominente Namen wie Wissenschaftler Stephen Hawking, Microsoft-Researcher Eric Horvitz, der Unternehmer Elon Musk (»Tesla«) oder der berühmte Intellektuelle Noam Chomsky. Selbst der aktuelle Berliner Koalitionsvertrag setzt sich für eine völkerrechtliche Ächtung vollautomatisierter Waffensysteme ein.

Drastisch fasst die Schlüsselszene in Stanley Kubricks weitsichtigem Film »2001: Eine Odyssee im Weltraum« von 1968 die aktuell kulminierenden Ängste, Sorgen und Proteste zusammen. Sie ist zugleich eine handfeste Handlungsanweisung. Da schwebt Keir Dullea als Astronaut David Bowman in den rotglühenden, mit angeblich perfekter künstlicher Intelligenz ausgestatteten Rechner HAL 9000 hinein und vernichtet dessen Software. Mit einem Schraubenzieher. Zuvor hatte der unfehlbare HAL Bowmans Mission torpediert und dessen Mit-Astronauten getötet. Der Rechner bittet Bowman um Gnade: »Ich habe Angst!« Dann memoriert er mit ersterbender Stimme das Kinderlied »Hänschen klein«. Doch auch das Empathie-Modul für diese finale Gefühlsduselei hilft ihm nicht mehr. Bowman schaltet ihn aus.