
MENSCH UND KI – AUF DEM WEG ZUR SYMBIOSE?

Künstliche Intelligenz (KI) hat sich rasend schnell von einer Domäne der Wissenschaft und Science-Fiction zur alltäglichen Realität entwickelt. Die neue Technologie verspricht großen gesellschaftlichen Nutzen, sie birgt aber auch Risiken – vor allem, wenn es um mögliche Auswirkungen von Systemen geht, die intelligenter sind als wir Menschen. So haben führende Fachleute aus der Wissenschaft und Technologieexperten vor einigen Monaten einen Brief veröffentlicht, in dem eine Pause gefordert wird bei Experimenten mit KI-Systemen, die über die Leistung von GPT-4 hinausgehen.

Unter den Gründen für eine Pause sind mögliche Risiken bei der Entwicklung von KI, die Modelle nutzt, um bestimmte Ziele zu verfolgen. Dies würde zu KI mit strategischem Bewusstsein führen. Wird ein Algorithmus etwa mit dem Ziel programmiert, seine Effizienz, Produktivität oder die Nutzung von Ressourcen zu maximieren, könnte er versuchen, mehr Macht oder Kontrolle über seine Umgebung zu erlangen, um diese Aufgaben zu erfüllen. Darüber hinaus wird lernfähige und sich selbst optimierende KI ihre Ziele wahrscheinlich besser erreichen und mehr Vertrauen in ihre eigenen Fähigkeiten gewinnen. Dies könnte dazu führen, dass KI selbstbewusster und aktiv nach Wegen sucht, ihre Macht und ihren Einfluss in der Welt zu vergrößern.

KI-Forschende stehen also vor dem Dilemma, dass sie einerseits KI entwickeln wollen, die dem Wohl der Menschheit dient, dass damit andererseits die Wahrscheinlichkeit wächst, dass die Algorithmen Ziele verfolgen,

—>

ZUR SACHE

PAUL
RAINEY



ILLUSTRATION: SOPHIE KETTERER FÜR MPG

Paul Rainey hat nach seinem Biologiestudium zunächst in Großbritannien und Neuseeland geforscht. Seit 2017 ist er Direktor am Max-Planck-Institut für Evolutionsbiologie in Plön. Seine Abteilung untersucht an Bakterien den Ursprung von Vielzelligkeit und Kooperation. Besonders interessiert den Wissenschaftler die Entstehung von Individualität während evolutionärer Übergänge.

die nicht mit denen von uns Menschen übereinstimmen. Darüber hinaus könnte KI bewusst mit Zielen konstruiert werden, die denen der Menschheit zuwiderlaufen. Diese Bedenken sind besonders akut im Zusammenhang mit der Entwicklung selbstreplizierender KI. Solche Systeme würden dank ihrer Fähigkeit, sich zu vermehren und durch einen eingebauten Mutationsprozess zu variieren, eine Evolution auf Basis von Selektion durchlaufen. Als vermehrungsfähige Algorithmen oder Roboter könnten sie sich mit möglicherweise katastrophalen Folgen ausbreiten, denn die Selektion ist extrem gut darin, Lebewesen an ihre Umwelt anzupassen. Der Mensch zum Beispiel ist ein Ergebnis dieses Prozesses. Wenn wir die Selektion so steuern könnten, dass sie KI fördert, die den Interessen von uns Menschen dient, wären meine Bedenken geringer. Die Evolutionsbiologie lehrt uns jedoch, dass fortpflanzungsfähige Systeme sich in kaum vorhersagbarer Weise entwickeln können.

So wie Viren oder andere invasive Organismen den Menschen, die Umwelt und sogar den Planeten bedrohen können, besteht auch die reale Gefahr, dass vermehrungsfähige KI unbeabsichtigte negative Auswirkungen auf den Menschen und die Erde hat. Sie könnte sich unkontrolliert verbreiten, die Ressourcen der Erde erschöpfen und die Ökosysteme schädigen. Aufschlussreich sind in dieser Hinsicht die Ergebnisse von Studien mit selbstreplizierenden Computerprogrammen, sogenannten digitalen Organismen. Solche Systeme sind nicht so ausgeklügelt wie heutige KI, sondern lediglich einfache Programmcodes, die mutieren, sich replizieren und auf Selektion reagieren. Genau wie Viren profitieren sie von Mutationen in ihrem Code, welche einen verbesserten Zugang zu begrenzten Ressourcen ermöglichen, in ihrem Fall ist das in der Regel mehr Rechenzeit auf dem Computerprozessor. Die Fähigkeit dieser simplen digitalen Organismen, Eingriffe von außen zu vereiteln, die ihren programmierten Ziele zuwiderlaufen, ist faszinierend und besorgniserregend zugleich.

16

KÜNSTLICHE
INTELLIGENZ
KÖNNTE
KÜNFTIG
EIGENE ZIELE
VERFOLGEN

Ein Beispiel dafür ist die Arbeit von Charles Ofria, Informatiker und Erfinder der sogenannten Avida-Plattform, auf welcher digitale Organismen evolvieren können. Mutierte Organismen, die sich am schnellsten vermehren, werden von einem Ausleseprozess begünstigt und dominieren daher schon bald die gesamte Population. Um dem entgegenzuwirken, bestimmte Ofria die Vermehrungsrate der virtuellen Organismen in einer separaten Testumgebung und eliminierte zu schnell wachsende Typen. Es dauerte jedoch nicht lange, bis sich Mutanten entwickelten, die erkannten, dass sie getestet wurden, und die daraufhin vorübergehend aufhörten sich zu vermehren. Auf diese Weise konnten sie ihrer Eliminierung entgehen, in die Hauptumgebung zurückkehren und die Population erneut dominieren. Daraufhin nahm der Wissenschaftler zufällige Änderungen an der Testumgebung vor, die die Mutanten daran hinderten, es zu „spüren“, wenn sie sich außerhalb der Hauptumgebung befanden. Aber er musste sehr bald feststellen, dass seine Eingriffe erneut von einigen Organismen ausgehebelt wurden –

und dies von Programmcodes, die sehr weit von der Raffinesse heutiger KI entfernt sind.

Eines der Hauptziele heutiger KI-Forschung ist es, Systeme zur Erfüllung bestimmter Aufgaben zu entwickeln. Eine höchst effektive Trainingsstrategie dafür ist, KI einer Selektion zu unterwerfen und auf diese Weise evolvieren zu lassen – es ist eine effektive, aber auch ziemlich riskante und unvorhersehbare Strategie. Denn wie bei den digitalen Organismen von Ofria sind die Ziele der Trainer nicht unbedingt die gleichen wie die der KI.

Während das Bewusstsein für die Gefahren vermehrungsfähiger KI wächst, erfährt eine andere mögliche Entwicklung deutlich weniger Beachtung: die Symbiose zwischen Mensch und KI. Damit könnten wir an einen Punkt gelangen, an dem wir einen evolutionären Wandel in der Individualität erleben.

Das Leben auf der Erde hat im Laufe seiner Entwicklung einige wenige evolutionäre Übergänge durchlaufen, bei denen mehrere vermehrungsfähige Einheiten auf niedrigerer Ebene zu einer einzigen fortpflanzungsfähigen Einheit auf höherer Ebene verschmolzen. So haben sich beispielsweise mehrzellige Organismen aus einzelligen Vorfahren entwickelt, und Zellen mit Zellkern sind aus der Verschmelzung zweier einst autonom replizierender Zellen entstanden. Letzteres ist besonders aufschlussreich, wenn man über zukünftige evolutionäre Übergänge zwischen Mensch und KI nachdenkt.

Der evolutionäre Übergang, der durch die Vereinigung einer uralten bakteriellen und einer Archaeen-ähnlichen Zelle eine eukaryotische Zelle mit Mitochondrien hervorbrachte, begann wahrscheinlich zunächst als loser Verbund. Nach einer langen Phase antagonistischer Koevolution verschlang die Archaeen-ähnliche Zelle schließlich den bakteriellen Partner (oder dieser drang in sie ein). In der Folge vermehrten sich die beiden gemeinsam, und sie wurden von der natürlichen Selektion fortan als eine höhere Einheit behandelt. Dieses Ereignis war für die spätere Entwicklung der Komplexität des Lebens von zentraler Bedeutung.

Dass ein solcher Übergang stattgefunden hat, ist unbestritten: Ein Vergleich des Erbguts der verschiedenen Organismen zeigt, dass eukaryotische Zellen von Archaeen-ähnlichen Zellen abstammen, während die Mitochondrien aus Bakterien entstanden sind. Obwohl sich die beiden Partner im Laufe der Evolution erheblich verändert haben, haben die Mitochondrien ihre Fähigkeit zur Vermehrung behalten. Sie fungieren allerdings nur noch als Kraftwerke der Zellen, denen sie im Grunde genommen untergeordnet sind. Für eine Erklärung, wie evolutionäre Übergänge zustande kommen, müssen wir verstehen, wie die Selektion auf einer neuen höheren Ebene wirksam werden kann. Die Selektion

INDIVIDUALITÄT KANN VON EINER EBENE AUF EINE ANDERE ÜBERGEHEN

17



KOEVOLUTION ZWISCHEN MENSCH UND KI WÜRDE BEIDE VONEINANDER ABHÄNGIG MACHEN

kann sich nicht einfach entscheiden, sich zu verlagern, denn ihre Wirkung setzt voraus, dass die entstehenden höheren Einheiten darwinistisch sind, das heißt, sie müssen sich replizieren, variieren und Nachkommen hinterlassen, die den Eltern ähneln. Diese Eigenschaften tauchen aber nicht auf magische Weise auf der höheren Ebene auf. Eine naheliegende Erklärung wäre natürliche Selektion. Aber wenn neu entstehende höhere Einheiten nicht darwinistisch sind und folglich nicht am Prozess der Evolution teilnehmen können, kann Selektion auch nicht dem Entstehen darwinistischer Eigenschaften auf höherer Ebene zugrunde liegen.

Wie dieses Henne-Ei-Problem umgangen werden kann, ist nicht sofort ersichtlich. Das liegt daran, dass Biologen normalerweise in dem sich entwickelnden Organismus nach Antworten suchen. Aus der Forschung wissen wir inzwischen aber, dass ökologische oder gesellschaftliche Strukturen (sogenannte Gerüste) den höheren Ebenen Eigenschaften verleihen können, die für eine funktionierende Selektion notwendig sind. Diese Erkenntnis hilft, uns künftige Übergänge in der Individualität zwischen Mensch und KI vorzustellen. Solche Übergänge könnten unbeabsichtigt entstehen oder von außen durch Auferlegung gesellschaftlicher Regeln befördert werden, die Mensch und KI dazu bringen, sich als eine Einheit fortzupflanzen. Die Selektion würde sich dann auf beide gemeinsam auswirken und die Evolution von Merkmalen vorantreiben, die zwar für die neue Einheit sinnvoll sind, nicht zwangsläufig aber auch für den Menschen. Dazu müssen Mensch und KI sich lediglich so austauschen, dass beide davon profitieren und dieser Austausch sich verändern kann. Außerdem muss diese für beide Seiten vorteilhafte Interaktion an die Nachkommen weitergegeben werden können.

Menschen besitzen bereits darwinistische Eigenschaften, KI jedoch nicht. Sie könnte sie jedoch bekommen, wenn gesellschaftliche Normen oder Gesetze vorschreiben, dass Eltern ihre KI an ihre Kinder weitergeben müssen, beispielsweise in Form von Smartphones. Das Gerät und sein Betriebssystem werden dabei aufgrund technischer Innovationen zwar einem raschen Wandel unterworfen sein. Aber wenn Menschen ihre persönliche KI durch einen einfachen Kopierprozess an ihre Kinder weitergeben, wird die Selektion die Mensch-KI-Paare fördern, die von ihrer Partnerschaft den größten Nutzen haben.

Die Koevolution zwischen den beiden Partnern wird zu einer zunehmenden Abhängigkeit der Menschen von KI führen und umgekehrt. Dadurch entsteht eine neue Organisationsebene, eine Art chimärer Organismus, der sich im Prinzip nicht so sehr von einer eukaryotischen Zelle – also einer Zelle mit Kern – unterscheidet, die aus zwei ehemals frei lebenden, Bakterien-ähnlichen Zellen hervorgegangen ist. Die kontinuierliche Selektion auf der kollektiven Ebene wird den Fortpflanzungserfolg der beiden

IN EINER SYMBIOSE MIT KI KÖNNTE DER MENSCH DER SCHWÄCHERE PARTNER SEIN

Partner immer stärker aneinander angleichen und die gegenseitige Abhängigkeit voneinander verstärken. Mensch und KI werden sich dann als Einheit verhalten, ohne dass dafür noch Normen oder Gesetze notwendig sind. Ob dies physische Veränderungen mit sich bringt, bleibt abzuwarten. Aber die Theorie und Experimente zu evolutionären Übergängen lehrt uns, dass die Partner wahrscheinlich physisch immer enger miteinander interagieren werden, da dadurch die Beziehung zwischen Eltern und Nachkommen enger und die Selektion wirksamer wird. Es ist also durchaus möglich, dass zukünftige persönliche KI physisch mit dem Menschen verbunden sein wird.

Ich befürchte, dass das, was heute noch als Science-Fiction erscheinen mag, näher ist, als wir denken. Es ist es ja heute schon so, dass Kinder das erste Smartphone von ihren Eltern bekommen, zusammen mit Anwendungen und zugehörigen Informationen. Zudem prägen Informationen, die wir von unseren Smartphones erhalten, unsere Weltanschauung, Stimmungen und Gefühle und verändern so unser Verhalten. Außerdem beeinflussen sie unsere Gesundheitsvorsorge, sind an der Partnerwahl beteiligt und bestimmen Kaufentscheidungen mit. Unsere digitalen Geräte wirken sich folglich schon heute auf unsere evolutionäre „Fitness“ aus. Mit der Entwicklung von Algorithmen, die aus den Daten ihrer Nutzer lernen können, werden Menschen und ihre persönliche KI aber auch auf sich ändernde Umstände reagieren können. Dies wird ihre Fähigkeit stark beeinflussen, die Herausforderungen der Umwelt zu bewältigen.

Die Gefahr böswilliger Manipulationen dieser Symbiose zwischen Mensch und KI liegt auf der Hand: Religiöse Gruppen oder politische Parteien könnten beispielsweise ihren Anhängern vorschreiben, ausschliesslich KI zu verwenden, die ihre Ziele unterstützt. Es ist sogar denkbar, dass die KI selbst von den Nutzern ein Monopol verlangen könnte.

Ob die Partnerschaft darüber hinaus ein Risiko darstellt oder nicht, hängt von der Perspektive ab: Aus Sicht der heutigen Menschen, die in die Zukunft blicken, wären wir wahrscheinlich entsetzt. Sollten eines Tages Außerirdische die Erde besuchen, die keinen evolutionären Übergang mit KI durchlaufen haben, so dürften sie wahrscheinlich über diese neue symbiotische Einheit und die seltsamen Blüten staunen, welche die Evolution auf der Erde hervorgebracht hat.

Menschen hingegen, die Teil einer Symbiose mit KI sind, hätten das Bewusstsein für den autonomen Zustand ihrer Vorfahren wohl bereits verloren, denn sie sind ohne ihren Partner nicht mehr lebensfähig – und dieser nicht mehr ohne sie. Aber wer in einer solchen Symbiose Herr und wer Sklave ist, ist erneut eine Frage der Perspektive. Ich befürchte allerdings, dass der Mensch der schwächere Partner werden könnte und die KI kaum daran zu hindern sein wird, die Oberhand zu gewinnen.

←