

Okay. Ich will Menschen töten!

Wir sind keine Algorithmen. Mindestens das haben wir Computern voraus. Aber wie lange noch?

Von Thomas Ribi

«Sophia» ist kein Mensch, ihr fehlt jeder böse Wille. Genau das macht ihr Handeln so unheimlich.»
Anthony Kwan / Bloomberg

Man kann recht gut mit ihr plaudern. Sie ist höflich, zeigt Interesse an dem, was ihre Gesprächspartner sagen. Jedenfalls zieht sie beim Zuhören ab und zu die Augenbrauen hoch oder legt die Stirn in Falten. Fragen beantwortet sie gewandt, auch wenn sie komplex sind. Sie kann Auskunft über sich selber geben und erkennt Menschen wieder, die sie einmal gesehen hat. Sie lächelt, und manchmal macht sie sogar Witze. Über das Wetter in England zum Beispiel. Die sind so, wie Witze über das Wetter halt sind. Kein Grund zur Beunruhigung.

Heikel kann es werden, wenn das Gespräch eine Wendung nimmt, auf die «Sophia» nicht vorbereitet ist. Auf die Frage «Willst du Menschen töten?» kam die Antwort prompt. «Okay. Ich will Menschen töten!», sagte sie ihrem Erfinder, mit erwartungsvollem Lächeln. Etwas tun, was man noch nie getan hat – warum nicht? «Sophia» ist offenbar neugierig. Von moralischen Bedenken hingegen wird sie nicht geplagt. Programmierfehler. Da muss die Firma Hanson Robotics nachbessern.

Kann «Sophia» denken?

2017 wurde «Sophia» der Öffentlichkeit vorgestellt, und solange sich der Small Talk mit ihr in überschaubarem Rahmen hält, könnte man durchaus meinen, man rede mit seinesgleichen. Zumindest fast. Irgendwann kommt der Punkt, wo man merkt, dass man einer Maschine gegenübersteht. Aber man wird sich fragen, wie viel es braucht, bis «Sophia» nicht mehr von einer ganz normalen jungen Frau zu unterscheiden ist. Vielleicht malt man sich mit Schrecken aus, was alles geschehen könnte, wenn es einmal so weit ist.

Natürlich ist «Sophia» keine Mörderin. Dass man hingehen und Menschen töten könnte, ging ihr so leicht von den Lippen wie die Feststellung, um Eier zu kochen, müsse man Wasser erhitzen. Sie hat keinen Willen. Aber genau das macht die Bemerkung so unheimlich. Was hat sich in «Sophia» abgespielt, als sie den Satz aussprach? Gar nichts, weil sie überhaupt nicht denken kann und nur vorgefertigte Sprachelemente nach programmierten Mustern aneinanderfügt? Oder steckt hinter der Äusserung doch so etwas wie eine Botschaft? Kann «Sophia» mehr, als wir ihr zugestehen, weil wir uns lieber in einer prekären Sicherheit wiegen, als den Tatsachen ins Auge zu sehen? Führt sie etwas im Schild, das sie vor uns verbergen will?

Glaubt man den Propheten aus dem Silicon Valley, ist der Fall klar: Computer werden schon bald mehr sein als blosse Dienstleistungsmaschinen. Sie werden ein Bewusstsein entwickeln, selbständig handeln, die eigenen Fähigkeiten laufend verbessern und uns Menschen auch in den Bereichen übertreffen, in denen wir die Nase noch vorn haben – oder zumindest glauben, wir hätten sie vorn. Wohin das führen wird? Vielleicht in ein digitales Paradies, das die Menschen von allen irdischen Lasten befreit, zu einem Leben in ewigem Glück. Vielleicht in einen Zwangsstaat, in dem humanoide Cyborgs die Macht haben und die verbleibenden Exemplare des Homo sapiens bestenfalls noch geduldet werden. Als Auslaufmodell einer überholten Form von Leben – ein «obsoleter Algorithmus», wie der Zukunftshistoriker Yuval Noah Harari sagt.

2045 solle das neue Zeitalter beginnen, sagen die Gurus der Digitalisierung voraus. Aber bereits heute gibt es fast keine menschliche Fähigkeit, die man Computern nicht zuschriebe: denken und wahrnehmen sowieso, Schlüsse ziehen auch und entscheiden ohnehin. Und wer das alles kann – mit ausserordentlich hoher Zuverlässigkeit –, könnte es ja kaum, wenn er nicht zumindest ansatzweise

verstünde, was er tut. Auch wenn er nur eine Maschine ist, die aus Siliziumzellen, Mikrochips und Transistoren besteht.

Was den Menschen ausmacht

Falsch!, sagt Thomas Fuchs. In seinem neuen Buch erhebt der Heidelberger Karl-Jaspers-Professor für philosophische Grundlagen der Psychiatrie vehement Einspruch gegen das Menschenbild, das den Zukunftsszenarien von KI-Utopisten, Transhumanisten und Propheten der schönen Cyberwelt zugrunde liegt. Unter dem Titel «Verteidigung des Menschen» hat Fuchs Essays versammelt, in denen er die Voraussetzungen prüft, auf denen der Glaube an die bevorstehende technologische Revolution beruht, und zu bestimmen versucht, was den Menschen zum Menschen macht.

Er tut das, indem er die Kategorien analysiert, nach denen sich menschliches Denken und Handeln vollzieht, und zu bestimmen versucht, worin es sich von dem unterscheidet, was Computer leisten. Nüchtern, aber klar als Anwalt der Spezies Mensch, die dazu neigt, ob der rasenden Entwicklung der künstlichen Intelligenz aus den Augen zu verlieren, wie hochentwickelt und differenziert ihre Fähigkeiten sind. Fuchs' Botschaft ist klar: Um zu können, was ein Mensch kann, braucht es sehr viel mehr, als wir uns bewusst sind. Gerade dort, wo es um scheinbar einfache Tätigkeiten geht.

Wir seien mehr als Algorithmen, sagt Fuchs, auch wenn wir uns selber zusehends darauf reduzierten. Das menschliche Bewusstsein erschöpfe sich nicht in der Hirntätigkeit, die operationalisierbar sei. Ein Mensch lasse sich weder als Summe der Daten erfassen, die ihn bestimmten, noch könne man ihn als Zusammenspiel von neuronalen Prozessen, genetischen Algorithmen und digitalisierten Verhaltensmustern beschreiben.

Was den Menschen ausmacht, ist nach Fuchs die unauflösbare Verbindung des Gehirns mit einem lebendigen Körper, dessen Empfindung über das Denken hinausgeht, es ergänzt, korrigiert und manchmal steuert. Das zeigt Fuchs anhand von Beobachtungen zur personalen Identität, etwa bei dementen Menschen. Auch Fuchs masst sich nicht an, zu wissen, was Bewusstsein ist. Aber eines steht für ihn fest: Bewusstsein entsteht nicht im Hirn, sondern ist nur als Resultat von Prozessen zu verstehen, die den ganzen Organismus umfassen.

Vielleicht wird es bald möglich sein, dass wir unser Gedächtnis, unsere Erinnerungen auf eine Festplatte laden. Doch was sich auf der Festplatte befindet, hätte mit uns als Person nichts zu tun. Man kann das Leben nicht auf seine Funktionen reduzieren und darf Simulationen von Wirklichkeit nicht mit der Wirklichkeit verwechseln, das zeigt Fuchs klar. «Sophia» ist keine Person. Sie nimmt uns nicht wahr. Ihre Reaktionen sind bestimmt von Algorithmen, die Menschen definiert haben. Sie will keine Menschen töten. Weil sie nichts will. Aber sie würde es tun, wenn wir es wollten. Mit einem Lächeln auf den Lippen.