Von der Unmöglichkeit der Testung des Vorhandenseins von Bewusstsein in künstlichen intelligenten Agenten

von Nikodem Jan Skrobisz

Hausarbeit zum Seminar:

Neuromorphe Hardware und ihre Bedeutung für autonome Agenten

Dozent: Antonio Bikić

Sommersemester 2021

Ludwig-Maximilians-Universität München

Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft

Inhalt

Nikodem Skrobisz

18.09.2021

1. Einleitung

Eine der größten Glaubensfragen des 21. Jahrhunderts ist die von der Möglichkeit der Existenz von Bewusstsein in Künstlichen Intelligenten Agenten, während diese in ihren verschiedenen Formen von Robotern und IoT Geräten bis hin zu reinen KI-Programmen immer mehr in den menschlichen Alltag integriert werden. Die zunehmend akzelerierenden Entwicklungen in den Bereichen der Robotik und Informatik beflügeln nicht nur in Gestalt von Science-Fiction die Fantasie der breiten Massen, sondern auch die von Intellektuellen, Entwicklern und Theoretikern. Es werden bereits Theorien gebaut, wie wir mit hypothetisch bewussten synthetischen Wesen ethisch umgehen sollen. Andere spekulieren über eine Intelligenzexplosion und die daraus folgende Unvermeidbarkeit einer technologischen Singularität oder ins Äußerste gehend von der zwangsläufigen Auslöschung der Menschheit durch die Geburt eine Maschinenzivilisation.

Eine diesen theoretischen Unterfangen voranzustellende, essentielle Frage in Hinblick auf die Weiterentwicklung von künstlichen intelligenten Agenten, insbesondere in Form von immer stärker werdenden Künstlichen Intelligenzen, ist die Frage, ob diese artifiziellen Agenten Bewusstsein haben können und werden. Diese Frage zu beantworten ist essentiell, denn das Vorhandensein von Bewusstsein in artifiziellen Agenten ist unter anderem fundamental für unsere ethischen Theorien und schließlich Entscheidungen zur Gestaltung nicht nur der künstlichen intelligenten Agenten selbst, sondern auch von menschlichen Gesellschaften und ihren Rechtssystemen, die zunehmend von immer komplexeren artifiziellen, intelligenten Systemen und Agenten bevölkert, reguliert und erweitert werden. Die Antwort hätte ebenso weitreichende Folgen für unser Verständnis davon, was es bedeutet ein Mensch zu sein – ist Bewusstsein doch das, wovon die meisten von uns intuitiv annehmen, dass es uns von toter Materie und Maschinen unterscheidet.

Unser großes Pech ist, dass diese Frage eine Glaubensfrage ist. Es ist eine Glaubensfrage, denn um sie beantworten zu können, müssen wir zuerst eine andere Frage positiv beantworten können: *Ist es möglich das Vorhandensein von Bewusstsein in einem künstlichen intelligenten Agenten zu testen oder anderweitig objektiv nachzuweisen?* Wie diese Arbeit aufzeigen will, ist die Antwort auf diese Frage jedoch ein *Nein*, denn es ist – zumindest zum aktuellen Stand - unmöglich sicheres Wissen über das Vorhandensein von Bewusstsein in einem künstlichen Agenten zu erlangen.

Um dies aufzuzeigen, soll in dieser Arbeit zuerst das allgemeine Problem der objektiven Untersuchung des Bewusstseins basierend auf den Ansätzen der Philosophen Thomas Nagel, Joseph Levine, David Chalmers und Colin McGinn erläutert werden. Im nächsten Teil wird dann darauf eingegangen, was es bedeutet bei einem künstlichen, intelligenten Agenten Bewusstsein zu testen, welche Kriterien solch ein Test erfüllen muss und wie es im Vergleich dazu um die Testung von Bewusstsein bei Menschen steht. Im Anschluss soll noch einmal exemplarisch an zwei bisher vorgeschlagenen Tests - dem Turing Test und dem Q3T - aufgezeigt werden, warum diese Bewusstsein nicht zuverlässig nachweisen können.

2. Das allgemeine Problem einer objektiven Untersuchung des Bewusstseins

2.1 Was ist Bewusstsein überhaupt?

Das subjektive Erlebnis des Bewusstseins machen wir Menschen in jedem unserer wachen Augenblicke: Wir erleben wie Emotionen, Reize, Erinnerungen und Gedanken in unser Bewusstsein eintauchen und wieder daraus verschwinden. Alles, was wir erleben, erleben wir in unserem Bewusstsein, womit es zentral für die menschliche Erfahrung ist. Obwohl Bewusstsein so alltäglich für uns ist, ist es notorisch schwer zu definieren, theoretisch zu erfassen und zu erklären, insbesondere da es häufig mit anderen oder damit konfundierenden Eigenschaften wie Intelligenz oder Gedanken verwechselt wird.

Der analytische Philosoph Thomas Nagel bietet in seinem 1974 erschienen Aufsatz „Wie ist es, eine Fledermaus zu sein?“ eine es auf das Essentielle herunterbrechende Definition an. Er beschreibt Bewusstsein als „den subjektiven Charakter von Erfahrung“ (Nagel, 2016, S. 9). Laut ihm lässt sich Bewusstsein, also die „Tatsache, dass ein Organismus überhaupt bewusste Erfahrung hat“ (Nagel, 2016, S.9) als „dass es irgendwie ist, dieser Organismus zu *sein*“ (ebd.) definieren. Das Bewusstsein ist das subjektive Erlebnis jemand zu sein aus der Erste-Person-Perspektive.

2.2 Phänomenales Bewusstsein als Qualia

Die Extension von Thomas Nagels Beschreibung des Bewusstseins umfasst bei genauerer Betrachtung zwei Aspekte des Bewusstseins. Einmal das subjektive Erlebnis jemand zu sein aus der Erste-Person-Perspektive und zweitens das daran eng verknüpfte phänomenale Bewusstsein der Welt durch Qualia. Während Ich diesen Text schreibe, bin ich nicht nur da und fühle mich wie Ich, ich nehme auch durch das Bewusstsein das Licht des Bildschirms und die Schrift darauf wahr und fühle unter meinen Fingern die Tasten. Physikalisch beschrieben, existieren jedoch weder die Farben noch die Empfindungen, die ich spüre, tatsächlich. Meine Sinneszellen reagieren auf Reize der Außenwelt, z.B. die Zellen in meinen Augen auf die vom Bildschirm emittierten Photonen, und leiten über Nervenzellen elektrische Impulse in mein Gehirn, wo schließlich unter anderem im Visuellen Cortex der Input aus den Augen verarbeitet wird. Diese Verarbeitung führt in meiner Psyche zu mentalen Zuständen, die Ich in meinem Bewusstsein als subjektive Erlebnisse von Farben, Gerüchen und Geschmäckern erlebe. Zum Beispiel führt die Reflektion von Photonen von der Schale einer Tomate, die ich betrachte, zu dem subjektiven Erlebnis der Farbe *Rot*. Solche subjektiven Erlebnisse, dieses Wie-es-ist-als-Ich die Welt zu erleben aus meiner Erste-Person-Perspektive bezeichnen Philosophen in der Regel als Qualia. (vgl. Tye, 1997)

Qualia sind streng genommen nicht Bewusstsein, sondern die Erlebnisse, die das Bewusstsein macht. Allerdings sind Qualia nicht ohne ein Bewusstsein denkbar, da sie für ihre Existenz von einem Bewusstsein erlebt werden müssen. Somit ist die Erklärung und Beweisbarkeit von Qualia als Phänomenales Bewusstsein eng verwoben mit der des Bewusstseins.

2.3 Qualia und die Erklärungslücke

Qualia sind wissenschaftlich betrachtet etwas Bizarres, denn trotz der Versuche von unzähligen Generationen an Wissenschaftlern, können wir sie nicht restlos naturwissenschaftlich und objektiv erklären. Den materialistischen Paradigmen der Naturwissenschaften folgend, sollte die ganze Welt – inklusive von subjektivem Phänomen wie Qualia – auf physische Prozesse reduzierbar und damit objektiv erklärbar sein. Doch wie bereits Nagel in seinem Aufsatz „Wie ist es, eine Fledermaus zu sein?“ schreibt, wissen wir bei subjektiven Phänomenen nicht, „wie dies der Fall sein kann.“ (Nagel, 2016, S.39).

Egal wie viel wir über die physischen Prozesse im Gehirn, die Veränderungen von Neuronen, die involvierten elektrischen und chemischen Reaktionen lernen, wir können uns nicht erklären, wie aus diesen objektiven Zuständen die entsprechenden subjektiven Erlebnisse erwachsen. Egal wie viel wir in einem Gehirn herumstochern, der Geschmack *Salzig*, die Farbe *Rot* oder auch das qualitative Erlebnis von *Schmerz* lassen sich dort nicht finden, sondern nur Materie, deren Zustände sich ändert. Es erscheint mysteriös, wie diese Veränderungen in objektiver Materie zu subjektiven Erlebnissen in der Psyche führen. (vgl. Tye, 1996, Abschnitt 5)

Diese Lücke im Erklären von Qualia beschreibt Jospeh Levine in seinem Essay „Materialism and Qualia: The Explanatory Gap“ als Explanatory Gap beziehungsweise Erklärungslücke bei dem Versuch die Identität von physischen und psychischen Zuständen herzustellen. So ist zum Beispiel die physische Erklärung, dass Schmerz identisch ist mit dem physischen Zustand der Stimulation von C-Fasern im Körper, nicht ausreichend. Die Erklärung erfasst nämlich nicht die subjektive Empfindung, wie es ist, Schmerz qualitativ zu erleben. Es ist auch intuitiv denkbar, eine Stimulation von C-Fasern ohne Schmerzerlebnis sich vorzustellen oder ein Schmerzerlebnis ohne die Stimulation von C-Fasern. Die Stimulation von C-Fasern ist die kausale Erklärung dafür, was im Körper passiert, wenn das Bewusstsein Schmerz erlebt, aber es erklärt nicht, wie sich die Qualia *Schmerz* an sich konstituiert. Die Zustände C-Faser-Stimulation und Schmerz haben verschiedene Eigenschaften – einmal physikalische und einmal psychische -, sodass sie nicht identisch genug sind, um einander restlos erklären und definieren zu können. (vgl. Levine, 1983, S.2ff). Es bleibt eine epistemologische Lücke zwischen dem was wir wissenschaftlich als einen physischen Zustand im Gehirn beobachten können und dem, was wir subjektiv als Bewusstsein psychisch erleben.

Dies bedeutet nicht zwangsläufig, dass der Materialismus der Naturwissenschaften falsch ist. Es bedeutet lediglich, dass wir im Kontext von subjektiven Erlebnissen wie Qualia „[…] nicht einmal ansatzweise eine Konzeption davon [haben], wie er wahr sein könnte.“ (Nagel, 2016, S.37).

2.4 Die Irreduzibilität der subjektiven Erfahrung

Dass sich Bewusstsein und seine Aspekte wie Qualia aufgrund ihrer subjektiven Natur nicht rein objektiv und damit wissenschaftlich-materialistisch erklären lassen – oder zumindest wir nicht wissen wie – erläutert Thomas Nagel bereits in seinem Essay „Wie ist es, eine Fledermaus zu sein?“.

Thomas Nagel beschreibt darin den Erkenntnisprozess zur Objektivität, als „eine Richtung […], in die der Verstand schreiten kann“, Richtung der Unabhängigkeit von einer individuellen Perspektive (Nagel, 2016, S.27). Er illustriert dies anhand der Erkenntnis vom objektiven Charakter des Naturphänomens eines Blitzes. Wenn wir die objektive, physische Natur eines Blitzes erfassen, reduzieren wir es um unser subjektives Erlebnis der Erscheinung eines grellen Lichtes, auf die physische Realität dahinter. Die objektive Natur eines Blitzes als eine elektrische Entladung in der Atmosphäre ist unabhängig von unserer subjektiven Erfahrung. Sie wäre auch einem Alien

Wissenschaftler ohne Augen zugänglich, im Gegensatz zum subjektiven Phänomen. Objektive Tatsachen sind nämlich nicht an eine bestimmte Perspektive geknüpft. (vgl. Nagel, 2016, S.27)

Wenn also die Erkenntnis von der objektiven Natur von Dingen durch Reduktion darin besteht, sie von einer subjektiven und artspezifischen Perspektive loszulösen, stellt uns das vor ein Problem, wenn wir die objektive Natur eben solch einer subjektiven Perspektive erforschen wollen. Das Bewusstsein ist schließlich die subjektive Perspektive selbst und wenn wir davon die subjektive Perspektive abziehen, bleibt Nichts – unserer epistemischen Apparate tappen im Dunkeln. Oder wie Nagel schreibt: „Wenn der subjektive Charakter der Erfahrung nur von einer einzigen Perspektive aus ganz erfasst werden kann, dann bringt uns jeder Schritt hin zu größerer Objektivität, d.h. zu einer geringeren Verbindung mit einer Erlebnisperspektive, nicht näher an die wirkliche Natur des Phänomens heran: Sie führt uns weiter von ihr weg.“ (Nagel, 2016, S.31)

2.5 Die Unmöglichkeit Bewusstsein zu verstehen

Aufgrund der Schwierigkeiten Bewusstsein und Qualia zu erklären und verstehen, argumentieren Philosophen der Denkschule des *New Mysterianism* wie Colin McGinn in seinem 1989 erschienen Aufsatz „Can We solve the Mind-Body-Problem?“, dass die Lösung dieses schwierigen Problems des Bewusstseins uns Menschen unzugänglich ist und wir sie niemals finden werden. Auch wenn wir wissen, dass *de facto* das Gehirn die kausale Basis des Bewusstseins sein muss (vgl. McGinn, 1989 S.1), können wir nie herausfinden wie das funktioniert. Dies liegt daran, dass die Fähigkeit unseres Bewusstseins Konzepte vom Bewusstsein zu bilden und zu begreifen davon eingeschränkt ist, dass es selbst ein Bewusstsein ist – „one's form of subjectivity restricts one's concepts of subjectivity“ (McGinn 1989 S.9) und unserer räumliche Wahrnehmung der Welt nicht ausreichend ist, um das Bewusstsein zu erfassen. (vgl. McGinn, 1989, S.10) Selbst wenn wir unser eigenes Bewusstsein doch verstehen würden, wären wir spätestens beim Versuch andere Formen von Subjektivität zu verstehen - wie die des Erlebnisses von Fledermäusen oder eben Maschinen - an einer Grenze unserer Erkenntnismöglichkeiten angekommen. (vgl. McGinn, 1989, S.9)

2.6 Das P-Zombie Problem

Wir haben nun ein Grundverständnis davon, warum es – zumindest mit dem den Menschen zur Verfügung stehenden epistemischen Werkzeugen – nicht möglich ist, Bewusstsein restlos objektiv zu erfassen und zu verstehen. Aber lässt es sich von außen erkennen? Nein. Das Hauptproblem dabei ist, dass subjektives Erleben in einem Organismus von außen logisch nicht unterscheidbar ist von seiner Abwesenheit. Schein und Sein sind von außen empirisch identisch.

Nicht nur können wir mit physischen, beobachtbaren Zuständen – wie Nagel und Levine aufzeigten – Bewusstsein nicht restlos erfassen und erklären; das Wissen über physische Zustände enthält auch keine logische Abhängigkeit mit dem Wissen über das Vorhandensein von Bewusstsein. Sämtliches beobachtbares, objektives Verhalten eines Wesen nach außen ist logisch ohne ein Bewusstsein denkbar und erklärbar, weil Bewusstsein offensichtlich keine conditio sine qua non, also keine notwendige Bedingung für irgendetwas zu sein scheint abgesehen vom subjektiven Erlebnis an sich.

Um sich dies zu veranschaulichen, helfen mehrere Gedankenexperimente wie das des *Chinesischen Zimmers* und das Gedankenexperiment des *Philosophischen* beziehungsweise des *Phänomenalen Zombies*, wie es der Philosoph David Chalmers in seinem 1996 erschienen Buch „The Conscious Mind“ konzipiert. Letzteres soll hier exemplarisch erläutert werden.

Man stelle sich vor, man hätte einen Zwilling. Dieser Zwilling ist physisch und funktional absolut identisch mit einem selbst. Sein Körper ist genau gleich aufgebaut, sein Nervensystem verarbeitet exakt die gleichen Informationen und er zeigt nach außen das exakt gleiche Verhalten, er spricht, er erzählt Witze und agiert wie ein normaler Mensch nach außen hin. „All of this follows logically from the fact that he is physically identical to me, by virtue of the functional analyses of psychological notions. He will even be 'conscious' in the functional senses described earlier—he will be awake, able to report the contents of his internal states, able to focus attention in various places, and so on. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie." (Chalmers 1996, S.95) Der einzige Unterschied zwischen einem normalen Menschen und diesem Zombie wäre also, dass er kein Bewusstsein hat. Das Licht wäre aus, es wäre niemand „da", der Erlebnisse hat.

Solch ein Zombie ist widerspruchsfrei logisch denkbar. Daraus schlussfolgert Chalmers, dass "consciousness fails to logically supervene on the physical" (Chalmers, 1996, S.97). Eine vollständige physikalische und funktionale Beschreibung, impliziert also noch nicht die Existenz von Bewusstsein und erklärt auch nicht wie physische Zustände kausal zu phänomenalen Erlebnissen führen und notwendig machen. Daher kann Bewusstsein nicht von reduktionistischen Erklärungen erfasst werden und eine Beschreibung und Erklärung der beobachtbaren, physischen Zustände eines Organismus von außen, nicht ausreichen, um das Vorhandensein von Bewusstsein sicherzustellen.

2.7 Das P-Zombie Problem für Menschen lösen

Wenn P-Zombies logisch denkbar sind und wir nicht logisch beweisen können, ob jemand tatsächlich Bewusstsein hat, woher wissen wir dann, dass die Menschen um uns herum Bewusstsein haben? Zumindest scheinen wir dieses Wissen ja intuitiv zu besitzen und dann könnte es doch sein, dass wir über ein ähnliches Verfahren auch das Bewusstsein bei Künstlichen Intelligenten Agenten erkennen könnten. Doch tatsächlich ist es nicht möglich zu wissen, ob andere Menschen tatsächlich ebenso Bewusstsein haben, wie Du selbst.

Der Computerwissenschaftler Jaron Lanier argumentiert sogar – vermutlich mehr sarkastisch als ernst – in seinem Paper „You can´t argue with a Zombie", dass Philosophen wie Daniel Dennett, die das Bewusstsein als restlos objektiv erklärbar abtun, P-Zombies sein müssen. „It turns out that it is possible to distingush a zombie from a person. A zombie has a different philosophy. […] Dennett is obviously a Zombie." (Lanier, 1995, S.1)

Rein logisch ist tatsächlich eine Position wie der Solipsismus, also dass Du das einzige tatsächlich existierende Bewusstsein bist und alles um dich herum nur ein Produkt dessen, logisch ohne innere Widersprüche denkbar. (vgl. Avramides, 2019) Der Grund, warum wir im Alltag davon ausgehen, dass andere Menschen auch Bewusstsein haben, ist recht simpel: Wir gehen schlicht pragmatisch davon aus, weil andere Menschen uns so ähnlich sind, wir mit ihnen Empathie empfinden und Gesellschaften aufbauend auf der Prämisse einer ähnlichen, kompatiblen Natur formen müssen.

Bei der Interaktion mit Künstlichen Intelligenten Agenten könnte jedoch so ein Pragmatismus und eine ohne gesunde Zweifel haltlose Empathie gefährlich werden - wie nicht wenige Science-Fiction Filme wie zum Beispiel Alex Garlands „Ex Machina" imaginieren – ganz abgesehen davon, dass künstliche intelligente Agenten nicht zur gleichen Spezies wie wir gehören und wir die Intuitionen, dass sie dennoch Bewusstsein haben könnten, nicht objektiv testen können wie der nächste Abschnitt aufzeigen soll.

3. Was es bedeutet Bewusstsein in einem Künstlichen Intelligenten Agenten zu testen

3.1 Was Testen bedeutet

Was bedeutet es überhaupt etwas zu testen? Der Duden definiert einen Test als „nach einer genau durchdachten Methode vorgenommener Versuch, Prüfung zur Feststellung der Eignung, der Eigenschaften, der Leistung o. Ä. einer Person oder Sache“ (Duden, 2021).

Wenn wir einen Test durchführen, wollen wir den Nachweis über das Vorhandensein eines Zustandes oder Dings erhalten. In einem wissenschaftlichen Rahmen sollte dieser Test dabei drei Kriterien erfüllen: Validität, Reliabilität, Objektivität. Der vom Test erbrachte Nachweis sollte eine möglichst hohe Validität und Reliabilität haben, und damit einen möglichst schmalen Konfidenzintervall. Des Weiteren sollte er objektiv sein, das heißt, das Ergebnis sollte davon unabhängig sein, wer den Test „durchführt, auswertet und interpretiert“, sodass wir mit einer hohen Gewissheit davon ausgehen können, dass wenn wir den Test durchführen, das Ergebnis den objektiven Tatsachen entspricht und nicht von Subjektivität in Form von z.B. Vorurteilen verzerrt wird. (vgl. Hussy et al., 2013, S.86)

Bevor man jedoch einen Test für etwas entwirft, muss man die Sache, deren Vorhandensein man mit dem Test nachweisen will, als Variable operationalisieren. Das heißt, man macht diese Variable „der Beobachtung und Erfassung zugänglich“ (vgl. Hussy et al., 2013, S.39), indem man ihr empirische Sachverhalte, also „konkret mess- bzw. beobachtbare Größen“ zuordnet. (vgl. Hussy et al., 2013, S.39) Ohne eine Operationalisierung ist es nicht möglich einen Test zu gestalten und durchzuführen.

3.2 Wie Bewusstsein bei Menschen getestet wird

Wie in den vorherigen Abschnitten aufgezeigt haben, ist es uns Menschen unmöglich das Bewusstsein auf objektiv empirische Zustände zu reduzieren. Das liegt daran, dass „the property of consciousness itself (or specific conscious states) is not an observable or perceptible property of the brain.“ (vgl. McGinn, 1989, S.9). Behavioristisch also empirisch lassen sich Schein und Sein, wie die Gedankenexperimente vom *Chinesischen Zimmer* und vom *P-Zombie* zeigen, nicht unterscheiden. Daher ist es auch nicht möglich, Bewusstsein eindeutig zu operationalisieren. Es gibt keine allgemein anerkannte Operationalisierung für Bewusstsein.

Dennoch gibt es in der Medizin und Psychologie eine Reihe von Tests, mit denen versucht wird zu prüfen, ob ein Mensch in einem medizinischen Sinne bei Bewusstsein ist. Jedoch testen diese Tests nicht tatsächlich das Bewusstsein, sondern das Vorhandensein von beobachtbaren Zuständen, bei denen wir Menschen aus unserer eigenen Erfahrung davon ausgehen, dass ein Mensch Bewusstsein hat. Diese subjektive Basis führt jedoch nur zu einer ungenügenden Objektivität.

Das Standardverfahren für das Testen von Bewusstsein bei Menschen ist ein „Accurate Report“, also das Prüfen, ob ein Mensch seine subjektiven Erlebnisse akkurat wiedergeben kann, indem man ihn danach befragt. (vgl. Seth et al., 2005, S.119ff)

Eine weitere in der Medizin und Forschung genutzte Methode Bewusstsein bei einem Menschen vorgeblich zu testen, ist durch die Messung der Hirnaktivität. Dabei wird aus der Erfahrung, dass die Fähigkeit akkurat die eigenen subjektiven Erlebnisse wiederzugeben bei Menschen mit rohen EEG Werten in dem Bereich 20 – 70 Hz korreliert, abgeleitet, dass wenn eine EEG Messung Werte in diesem Bereich ergibt, Bewusstsein vorhanden sei muss. (vgl. Seth et al., 2005, S.122)

Diese Art von Testung hat jedoch ihre Grenzen, da sie eben nur das Vorhandensein von Zuständen misst, bei denen wir aus unserer eigenen Erfahrung und aus den Berichten anderer Menschen davon ausgehen, dass Bewusstsein vorhanden ist. Diese Ausgangsbasis des subjektiven Erlebnisses führt jedoch zu keiner restlos objektiven Möglichkeit Bewusstsein nachzuweisen.

So deuten die meisten Erfahrungsberichte darauf hin, dass ein Mensch während einer Narkose kein Bewusstsein hat – schließlich kann er währenddessen kein subjektives Erlebnis wiedergeben und für die meisten Patienten fühlt sich eine Vollnarkose tatsächlich wie ein An- und Abschalten des Bewusstseins an. Ebenso deuteten Messungen mit einem EEG darauf hin, dass ein Mensch während einer Narkose kein Bewusstsein haben sollte, weil ein EEG während einer Narkose nur noch eine elektrische Aktivität im Gehirn von rund 4 Hz misst, also deutlich unter dem Bereich von 20 – 70 Hz, der mit Bewusstsein assoziiert wird. (vgl. Seth et al., 2005, S.123)

Jedoch können wir dies tatsächlich nicht wissen. Es könnte auch schlicht sein, dass die Wirkung der Narkose beinhaltet, dass unser Gehirn keine Erinnerungen über die bewussten Erlebnisse während der Narkose speichern kann. Menschen könnten während einer Narkose durchaus bewusste Erfahrungen haben – es ist uns nur nicht möglich dies zu wissen, da die subjektive Erinnerung daran in der Regel fehlt. Wobei auch dies nur eine vage Regel ist – denn tatsächlich berichten Patienten immer wieder von Träumen während einer Narkose, die auch unabhängig von der Tiefe der Narkose aufzutreten scheinen. (vgl. Leslie et al., 2007, S.35) Da Träume ein subjektives Erlebnis des Bewusstseins sind, kann man dies als Indiz auffassen, dass Bewusstsein auch dann in Menschen vorhanden sein kann, wenn die elektrische Aktivität des Gehirns das Gegenteil suggeriert.

Letzendlich können wir bei Menschen nur intuitiv und pragmatisch, aber nicht objektiv davon ausgehen, dass sie Bewusstsein haben, solang ihre Gehirne noch am Leben sind, also Aktivität zeigen. Und auch dies nur von den Prämissen unserer eigenen unzuverlässigen Erfahrung und des Materialismus ausgehend, da noch kein Mensch nachweislich von den Toten wiederauferstanden ist, um zu berichten, wie es sich mit dem Bewusstsein nach dem Tod verhält und so diese Prämissen anzuzweifeln. Wobei auch hier der ein oder andere Anhänger einer Religion oder Analyst von Nahtoderlebnissen behaupten könnte, dass dies bereits der Fall war – aber damit bewegen wir uns nicht nur von einer objektiven Untersuchung weg und hin zu Spekulationen, sondern auch weg vom eigentlichen Thema dieser Arbeit.

3.3 Was ein künstlicher intelligenter Agent ist

Der Begriff *Intelligenter Agent* kommt aus der Informatik, konkreter aus ihrem Teilbereich Künstliche Intelligenz. Ein *Agent* ist dabei, wie von den Informatikern Stuart Russel und Peter Norvig in ihrem Buch „Artificial intelligence. A modern approach“ definiert, jedes System, welches „perceives and acts in an environment.“ (Russel & Norvig, 2016, S.59) Ein einfacher, die Umwelt wahrnehmender und darauf reagierender künstlicher Agent in diesem Sinne kann bereits ein klassisches Thermometer sein, welches auf die Temperaturveränderung in seiner Umgebung reagiert, indem entsprechend der Quecksilberspiegel darin steigt oder sinkt. (vgl. Russel & Norvig, 2016, S.15) Als *künstlich* sei alles bezeichnet, 1) was selbst kein Mensch ist und 2a) von mindestens einem Menschen direkt erschaffen wurde oder 2b) indirekt durch Dinge wie zum Beispiel Maschinen, die 2ba) selbst von mindestens einem Menschen erschaffen oder 2bb) von einem anderen künstlichen Ding erschaffen wurden. Ein *intelligenter Agent* ist ein Agent, welcher durch seine Komplexität nicht nur seine Umwelt wahrnimmt und darauf reagiert, sondern auch in der Lage ist Wissen über die Welt zu speichern und darauf basierend auf Probleme intelligent mit der rationalsten Lösung zu reagieren. (vgl. Russel & Norvig 2016, S.26ff, S.30)

künstliche intelligente Agenten variieren stark in ihrer Form, Komplexität und ihrer Intelligenz: sie reichen von Robotern bis hin zu Künstlichen Intelligenzen in Softwareform, von Spamfiltern, über ChatBots bis hin zu komplexen Systemen, die im Sinne einer starken Künstlichen Intelligenz in der Lage sind mit einer generellen Intelligenz auf eine Reihe unterschiedlicher Probleme zu reagieren und daraus zu lernen. (vgl. Russel & Norvig, 2016, S.27)

3.4 Warum künstliche intelligente Agenten überhaupt auf Bewusstsein testen?

Ob künstliche intelligente Maschinen in Form von Computern etc. Bewusstsein haben können, ist nicht nur eine zentrale Frage für ihre technische Weiterentwicklung seitdem Alan Turing sie in seinem Aufsatz „Können Maschinen denken?“ (vgl. Turing, 2021, S.7) aufstellte – auch wenn Turing sie selbst als irrelevant für ihre Entwicklung einstufte. Die Frage, ob wir künstlich Bewusstsein erzeugen können, impliziert existentialistische und ethische Probleme für uns Menschen, für unser Selbstbild, für unser Überleben, für unsere Sicht auf die Welt, unsere Gesellschaft und Politik. (vgl. Russel & Norvig, 2016, S.1042) Bereits 1863 formulierte Samuel Butler in seinem Artikel “Darwin Among the Machines” die Vision, dass am Ende der Evolution durch natürliche Selektion “the ultimate development of mechanical consciousness” (vgl. Butler, 1863) stehen würde und dass die immer schnellere Weiterentwicklung von Maschinen dazu führen würde, dass die Menschheit von diesen ersetzt, versklavt und ausgelöscht wird. Konsequenterweise ruft er am Ende seines Artikels zu einem menschlichen Vernichtungskrieg gegen die Maschinen auf (vgl. Butler, 1863) – deren höchster Entwicklungsstand zu der Zeit wohlgemerkt noch Dampfmaschinen waren.

Die von Butler formulierten Ängste finden heute ihr intensiviertes Echo in Schriften wie „Superintelligence: Paths, Dangers, Strategies“ von Nick Bostrom oder „Fanged Noumena“ von Nick Land, aber auch in der Kultur, in Filmen wie „Matrix“ oder Büchern wie Philip K. Dicks „Träumen Androiden von elektrischen Schafen?“. Auch wenn das mögliche Vorhandensein von Bewusstsein nur ein kleines Element all der Ängste und Visionen ist, die die Entwicklung von immer komplexeren und intelligenteren künstlichen Agenten bei Menschen auslöst, so ist es doch der mit den mitunter stärksten psychologischen und ethischen Implikationen.

Eine zuverlässige Methode, Bewusstsein – sowohl bei Menschen als auch bei künstlichen Agenten – zu testen, würde uns zumindest einige der existenziellen Probleme und Kränkungen, wenn nicht schon nehmen, so aber durch ihre Bestätigung zumindest greifbar und bearbeitbar machen.

Solch eine Methode existiert allerdings nicht. Wenn Vertreter des *New Mysterianism* wie Colin McGinn Recht haben und wir Bewusstsein niemals werden objektiv verstehen, operationalisieren und erklären können, wird es solche Methoden auch niemals geben und damit auch keine Möglichkeit Bewusstsein objektiv zu testen.

Dies hält dennoch Wissenschaftler und Philosophen nicht davon ab, verschiedene Möglichkeiten vorzuschlagen und auszuprobieren. Im Folgenden sollen zwei davon exemplarisch dargelegt und ihre Schwachstellen aufgezeigt werden.

4. Bisherige Testansätze

4.1 Warum der Turing Test nicht für Bewusstsein testet

Der berühmte Turing Test, konzipiert von Alan Turing, testet nicht für Bewusstsein. Auch wenn Turings Ausgangsfrage „Können Maschinen denken?“ des gleichnamigen Aufsatzes eigentlich die Frage nach Bewusstsein impliziert, da Gedanken auch Qualia sind, beantwortet der von ihm vorgeschlagene Test diese Frage nicht. Er substituiert die Frage lediglich nach der Frage, ob eine Maschine Menschen erfolgreich vortäuschen kann, selbst ein Mensch zu sein. Er besteht daraus, dass ein menschlicher Richter schriftlich sowohl mit einer Maschine als auch einem Menschen kommuniziert, die beide ihn zu überzeugen versuchen der echte Mensch zu sein. Wenn der Maschine es gelingt einen Menschen ausreichend zu imitieren, um den menschlichen Richter von ihrer Menschlichkeit zu überzeugen, reiche dies laut Turing aus, um zu urteilen, dass die Maschine tatsächlich denkt. (vgl. Turing, 2021, S.7ff) Dieser Test ist zu einem weder objektiv, da sein Ergebnis von der subjektiven Perspektive der beteiligten Menschen abhängt, zu einem kann er nicht zwischen Sein und Schein, zwischen einem P-Zombie und einem genuin bewussten Wesen unterscheiden.

4.2 Warum der Total Turing Test for Qualia Q3T auch nicht für Bewusstsein testet

In dem Aufsatz „Could there be a Turing Test for Qualia?” schlägt Paul Schweizer einen Total Turing Test für Qualia vor – abgekürzt Q3T -, aufbauend auf dem Total Turing Test - 3T - von Steven Harnard. (vgl. Schweizer, 2012, S.2) Dieser hypothetische Test basiert wie der Turing Test auf einem Gespräch zwischen Menschen und künstlichen intelligenten Agenten bzw. einem Roboter, allerdings ersetzt er das Imitationsspiel des ursprünglichen Turing Tests durch ein intensives Gespräch über das phänomenale Erlebnis des Roboters. Wenn der Roboter unerschöpflich darüber erzählen kann, *Wie-es-ist-er-zu-sein* und dabei Qualia beschreibt, also so wie ein bewusster und wacher Mensch es kann, so gilt der Q3T als bestanden. (Schweizer, 2012, S.7)

Schweizer argumentiert, dass wenn ein künstlicher intelligenter Agent in solch einem Gespräch genauso bewusstes Erleben von Qualia wie ein Mensch beschreibt, wäre deren Vorhandensein genau sicher bewiesen, wie das Vorhandsein von phänomenalem Bewusstsein von Qualia bei unseren Mitmenschen. Er argumentiert weiter, dass dann das Vorhandensein von Qualia in dem Roboter, der den Q3T besteht, nur vernünftig bestritten werden könnte durch ein Bestreiten des Funktionalismus, also mit dem Argument, dass es einen essentiellen Unterschied gibt zwischen biologischen, menschlichen Gehirnen und funktionsgleichen künstlichen Maschinen. (vgl. Schweizer, 2012, S.7)

Auch wenn ein Q3T tatsächlich ein starkes Indiz für das Vorhandensein von Bewusstsein liefern könnte, so schließt auch er nicht die grundlegende epistemologische Lücke. Er kann nicht zwischen einem P-Zombie und dem genuinen Vorhandensein von Bewusstsein unterscheiden – etwas, was Schweizer in seinem Aufsatz selbst zugibt: „[…]even the success of the Q3T robot could conceivably be explained without invoking P-consciousness per se, and so it still fails as a sufficient condition for attributing full blown qualia to computational artefacts.” (Schweizer, 2012, S.8)

5. Schlussfolgerung

Prinzipiell stoßen wir bei dem Versuch des Nachweises und der objektiven Erfassung von Bewusstsein bei Künstlichen Intelligenten Agenten auf die gleichen epistemischen Grenzen, die es uns bereits bei Menschen unmöglich machen Bewusstsein restlos logisch und objektiv zu erfassen, zu testen und zu verstehen. Die Antwort auf die Frage nach der Möglichkeit des Vorhandenseins von Bewusstsein in Künstlichen Intelligenzen bleibt daher die Demut der sokratischen Aporie; dass wir es schlicht nicht wissen können, da die Antwort jenseits unserer epistemischen Horizonte liegt.

„Consciousness remains a mystery.“ (Russel & Norvig, 2016, S.1040)

Wie wir angesichts der unauflösbaren Unsicherheit im Hinblick auf das Vorhandensein von Bewusstsein in künstlichen intelligenten Agenten handeln sollen, ist eine sich hier auftuende Frage, der es weitere Untersuchungen zu widmen gilt.

6. Quellen

Avramides, Anita (2019): Other Minds. Online verfügbar unter https://plato.stanford.edu/entries/other-minds/, zuletzt geprüft am 14.09.2021.

Butler, Samuel (1863): Darwin Among the Machines. Online verfügbar unter http://nzetc.victoria.ac.nz/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html zuletzt geprüft am 14.09.2021.

Chalmers, David (1996): The Conscious Mind, Oxford University Press

Duden (2021): Test. Online verfügbar unter https://www.duden.de/rechtschreibung/Test, zuletzt aktualisiert am 13.09.2021, zuletzt geprüft am 13.09.2021.

Hussy, Walter; Schreier, Margrit; Echterhoff, Gerald (2013): Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor. 2., überarbeitete Auflage. Berlin, Heidelberg: Springer Berlin Heidelberg (Springer-Lehrbuch).

Lanier, Jaron (1995), You can´t argue with a Zombie. Online verfügbar unter: http://www.jaronlanier.com/zombie.html, zuletzt geprüft am 11.09.2021

Leslie, Kate; Skrzypek, Hannah; Paech, Michael J.; Kurowski, Irina; Whybrow, Tracey (2007): Dreaming during anesthesia and anesthetic depth in elective surgery patients: a prospective cohort study. In: Anesthesiology 106 (1), S. 33–42. DOI: 10.1097/00000542-200701000-00010.

Levine, Joseph (1983): MATERIALISM AND QUALIA: THE EXPLANATORY GAP. In: Pacific Philosophical Quarterly 64 (4), S. 354–361. DOI: 10.1111/j.1468-0114.1983.tb00207.x.

McGinn, Colin (1989): Can We Solve the Mind--Body Problem? In: Mind 98 (391), S. 349–366. Online verfügbar unter http://www.jstor.org/stable/2254848. zuletzt geprüft am 14.09.2021.

Nagel, Thomas (2016): What is it like to be a bat? Englisch/Deutsch = Wie ist es, eine Fledermaus zu sein? Hg. v. Ulrich Diehl. Stuttgart: Reclam (Was bedeutet das alles?, Nr. 19324)

Russell, Stuart J.; Norvig, Peter (2016): Artificial intelligence. A modern approach. Unter Mitarbeit von Ernest Davis und Douglas Edwards. Third edition, Global edition. Boston, Columbus, Indianapolis: Pearson

Schweizer, Paul (2012): Could There be a Turing Test for Qualia? In: Revisiting Turing and his Test: Comprehensiveness, Qualia, and the Real World (AISB/IACAP Symposium), S. 41–48. Online verfügbar unter https://www.research.ed.ac.uk/en/publications/could-there-be-a-turing-test-for-qualia. zuletzt geprüft am 14.09.2021.

Seth, Anil K.; Baars, Bernard J.; Edelman, David B. (2005): Criteria for consciousness in humans and other mammals. In: Consciousness and cognition 14 (1), S. 119–139. DOI: 10.1016/j.concog.2004.08.006.

Shoemaker, Sydney (1991): Qualia and Consciousness. In: Mind 100 (4), S. 507–524. Online verfügbar unter http://www.jstor.org/stable/2255008. zuletzt geprüft am 14.09.2021.

Turing, Alan (2021) Computing Machinery and Intelligence / Können Maschinen denken?: Englisch/Deutsch. [Great Papers Philosophie] (Reclams Universal-Bibliothek)

Tye, Michael (1997): Qualia. Online verfügbar unter https://plato.stanford.edu/entries/qualia/#Explangap, zuletzt geprüft am 12.09.2021.

Tzafestas, Spyros G. (2016): An Introduction to Robophilosophy. Aalborg: River Publishers (River Publishers Series in Automation, Control and Robotics). Online verfügbar unter http://gbv.eblib.com/patron/FullRecord.aspx?p=4653786. zuletzt geprüft am 14.09.2021.

van Gulick, Robert (2004): Consciousness. Online verfügbar unter https://plato.stanford.edu/entries/consciousness/#FirPerThiPerDat, zuletzt geprüft am 12.09.2021.