

Verbessern Humanoiden die Welt? (von Christoph Hübenthal)

Klara, Tom, Adam und die alte Frage, ob der Mensch alles tun sollte, was er technisch kann – Eine Annäherung

Mit Humanoiden werden große Hoffnungen, teilweise sogar Heilserwartungen verbunden, aber auch große Befürchtungen, insbesondere die, eines Tages von Robotern beherrscht zu werden. Unter diesen Aspekten sollen Funktion und Entwicklung der Humanoiden in den Romanen „Klara und die Sonne“ von Kazuo Ishiguro, „Maschinen wie ich“ von Ian McEwan sowie dem Film „Ich bin dein Mensch“, Regie Maria Schrader, betrachtet werden. Welche Funktionen sind Klara, Adam und Tom, den Humanoiden in den genannten Werken, bei der Herstellung mitgegeben worden, wie entwickeln sie sich im Zusammenleben mit Menschen und wie ergeht es den Menschen dabei? Wird ihr Leben besser?

Klara ist eine KF, eine künstliche Freundin. Die Funktion der KFs wird im Buch schnell deutlich: sie sind als Unterstützung für Jugendliche vorgesehen, die wegen der herrschenden gesellschaftlichen Verhältnisse zum einen von Einsamkeit bedroht sind, zum anderen aus dem gleichen Grund aber auch in sozialem Verhalten eingeübt werden müssen. Klara ist von den Humanoiden in den drei Werken am weitesten von einer Menschengleichheit entfernt. Das beginnt beim „Körper“: Einerseits sieht sie „irgendwie französisch aus“, also wie ein reales Mädchen, insbesondere wegen ihrer Frisur¹. Andererseits ist Klara aber körperlich etwas unbeholfen und hat zumindest gelegentlich Wahrnehmungsstörungen. Hinsichtlich ihrer „geistigen“ Fähigkeiten erfährt der Leser schnell, dass Klara als lernende Maschine angelegt ist. Sie ist in der Lage, durch Beobachtung zu lernen und will das auch (15). Dies gilt auch und insbesondere für das Erkennen von Gefühlen² und die daraus folgende Entwicklung von Empathie³. Bei ihr ist also der Schritt von der künstlichen Intelligenz zur emotionalen künstlichen Intelligenz vollzogen⁴. Dies ist bei ihr aber nicht gekoppelt mit kognitiven Höchstleistungen, im Gegenteil: Klara weiß praktisch nichts von der Welt. Trotzdem ist sie für ihren eng umgrenzten Auftrag, eine jugendliche Person, Josie, beim Erwachsenwerden zu begleiten, hinreichend ausgestattet. Und sie erfüllt ihren Auftrag gut, Josies Sozialverhalten zu verbessern⁵. Sie rettet auch durch ihre Empathiefähigkeit⁶ die Freundschaft zwischen Josie und Rick, dem Nachbarjungen und einzigen Freund. Sie bringt sich selbst in Gefahr, weil sie

¹ So Josie, S. 21

² Im Fenster des Geschäfts stehend beobachtet sie die verschiedenen Stimmungen der sie betrachtenden Kinder und lernt, sie zu unterscheiden: „...nach unserem zweiten Tag im Fenster wusste ich den Unterschied schnell zu erkennen“, S. 17

³ „Als Josie vorhin nicht mitfahren konnte, habe ich große Traurigkeit gefühlt“, S. 117

⁴ Zu Stand und Bedeutung der Forschung der automatisierten Emotionserkennung s. Misselhorn: Künstliche Intelligenz und Empathie, Reclams Universalbibliothek, Ditzingen 2021, S. 20ff

⁵ So das Urteil der Mutter: „Trotzdem ist sie reifer geworden, seitdem du da bist. Sie nimmt mehr Rücksicht“, S. 127

⁶ Ob Klara empathisch ist oder nur Empathie imitieren kann, spielt dabei keine Rolle. S. hierzu die Beschreibung der „Behandlung“ von Afghanistan-Veteranen der US Armee durch eine sehr einfach konstruierte künstliche Intelligenz. Die Probanden empfanden diese Behandlung als hilfreicher als die gängigen Behandlungen durch Menschen. Catrin Misselhorn: Künstliche Intelligenz und Empathie, S. 73

glaubt, durch ein Opfer die Sonne dazu bewegen zu können, Josie zu heilen. So ist auch Josies Fazit in der Abschiedsszene sehr positiv: „Du warst toll, Klara. Ganz großartig“. (342)

Als Zwischenergebnis kann man also festhalten, dass Klara mit der Erfüllung ihres begrenzten Auftrags die im Roman dargestellte Welt tatsächlich verbessert.⁷

Eine neue Situation entsteht durch den Wunsch der Mutter, dass Klara Josie nach ihrem Tod „fortsetzen“ soll.⁸ Erst das führt den Roman zu einer Kernfrage der KI-Forschung: Stimmt es, dass – in Capaldi's Worten – „nichts in Josie [ist], das die Klaras dieser Welt nicht fortsetzen könnten“ (242)? Den Beweis für diese These muss Capaldi im Roman nicht antreten, da Josie gesund wird. Bei genauer Betrachtung besteht die These von Capaldi aus zwei aufeinander aufbauenden Fragen: Er glaubt, dass eine KI eine konkrete, bereits entwickelte menschliche Persönlichkeit, Josie, nach ihrem Tod ersetzen („fortsetzen“) kann. Hierzu müsste die KI die komplette Persönlichkeit einer konkreten Person übernehmen. Sie müsste Josie sein – und nicht „nur“ eine beliebige Tochter, die mit anderen Eigenschaften als Josie z.B. bei ihrer „Mutter“ Muttergefühle erzeugt. Dies geht weit über die vorgeschaltete Frage hinaus, ob eine KI einen Menschen in einer bestimmten Rolle abstrakt ersetzen kann, also nicht mit den Eigenschaften einer bestimmten, verstorbenen Person ausgestattet ist, sondern wie Tom im Film ein Lebenspartner oder wie Adam im Roman von McEwan ein Freund sein soll, deren Persönlichkeiten kein reales Vorbild haben. Klara gibt am Ende des Romans von Ishiguro eine Antwort, die nur verneint, dass eine konkrete Person „fortgesetzt“ werden kann: „Mr. Capaldi glaubt, es gebe nichts Besonderes in Josies Innerem, das sich nicht fortsetzen ließe...Aber ich glaube jetzt, er hat am falschen Ort gesucht. Es gab nämlich sehr wohl etwas Besonderes, aber nicht in Josie. Es war in denen, die sie geliebt haben.“ (348)

Wenn nach dieser Wertung also der Versuch, eine konkrete Person durch eine KI zu ersetzen, zum Scheitern verurteilt ist, kann man dann wenigstens ohne den Bezug zu einer konkreten Persönlichkeit, Menschen durch Humanoide ersetzen?

Adam, die KI aus „Maschinen wie ich“⁹ und Tom aus „Ich bin Dein Mensch“ sind im Gegensatz zu Klara „Superhirne“. Sie sind nicht nur in der Lage, blitzschnell absurde Rechenaufgaben zu lösen¹⁰, sondern haben auch Zugang zum gesamten Wissen im Internet.¹¹ Auch äußerlich sind

⁷ Dabei soll die Frage, inwieweit derselbe technische Fortschritt, vor allem in der Arbeitswelt und bei der Fähigkeit zu Genmanipulationen, die Notwendigkeit einer Unterstützung der Jugendlichen beim Erwachsenwerden erst produziert hat, außer Betracht bleiben, weil sie den Rahmen sprengen würde.

⁸ Abgesehen von (nachgemachter) Körperlichkeit ist diese Vision schon (fast) Realität: Gaspard Koenig beschreibt in „Das Ende des Individuums“ die App Replika, mit der die Unternehmensgründerin ein reales Leben tatsächlich „fortsetzt“: Sie ist „mit der Herstellung eines Roboters bekannt geworden, der auf den Daten eines verstorbenen Freundes beruht, mit dem sie seither „virtuell“ kommunizieren kann. Wenn Sie die Spuren Ihres gesamten Lebens auf Facebook hinterlassen, steht nichts einer Verlängerung nach dem physischen Tod im Wege, indem Ihre Art, sich auszudrücken, Ihre Gewohnheiten, Ihre Denkweisen wiederverwendet werden“ (S. 54). Interessant auch die Beschreibung der App in z.B. dem Google-Appstore und die Bewertungen durch Nutzer!

⁹ Alle folgenden Seitenangaben in Klammern beziehen sich auf Ian McEwan, Maschinen wie ich, Diogenes Verlag, Zürich 2019

¹⁰ So testet gleich zu Beginn des Films Alma Tom

¹¹ Adam lüftet auf diese Weise das Geheimnis von Miranda, Charlies Freundin, und Tom findet die argentinische Forschungsarbeit, die Almas Arbeit zunichtemacht

die beiden nahezu perfekte „Menschen“.¹² Sie sollen aber keine konkreten Personen ersetzen oder fortsetzen. Es liegt also nahe, nach einer Antwort auf die Frage, ob eine KI einen Menschen in einer abstrakten Rolle ersetzen kann, möglicherweise sogar der „bessere Mensch“ ist, im Film und im Roman von McEwan zu suchen.

Toms Funktion im Film ist eindeutig: er soll, wenn auch nur als Test¹³, Alma einen (menschlichen) Partner ersetzen.^{14,15} Die dieser Funktion zugrundeliegenden Programmierungen erscheinen im Film im Einklang mit den allgemein akzeptierten Werten im menschlichen Zusammenleben. Sein Werben um Alma ist durchaus sozialadäquat, er ist sogar in der Lage, die Gelegenheit verstreichen zu lassen, mit Alma Sex zu haben, als diese ihn dazu durchaus offensiv, aber in betrunkenem Zustand, auffordert. Der Film gibt mehrere Antworten auf die Frage, ob ein Humanoid ein (besserer) Partner für einen Menschen sein kann. Almas Urteil im Gutachten ist klar negativ. Sie stellt die Frage, ob es nicht gerade die unerfüllten Sehnsüchte, das Streben nach Glück sind, was uns zum Menschen macht. Sie fürchtet, es entstehe eine Gesellschaft von Abhängigen, „satt und müde von der permanenten Erfüllung ihrer Bedürfnisse“. Die Menschen hätten keinen Antrieb, sich selbst zu hinterfragen, Konflikte auszuhalten, sich zu verändern. „Es steht zu befürchten, dass jeder, der länger mit einem Humanoiden gelebt hat, unfähig sein wird zum normalen menschlichen Kontakt“. Der Film bleibt bei dieser Antwort allerdings nicht stehen, sondern liefert weitere Aspekte und lässt die Frage damit letztlich offen. Zum einen durch den zumindest offenen Schluss. Alma beendet zunächst aus den im Gutachten genannten Gründen die Beziehung zu Tom. Ihre durch Tom ausgelösten Erinnerungen an eine unerfüllte Jugendliebe, die bezeichnender Weise Thomas heißt, veranlassen sie, nach Dänemark zu reisen, wo sie Thomas kennengelernt hat. Dort trifft sie Tom wieder, der das vorhergesehen hat.¹⁶ Alma steckt in einem Dilemma: Ihr Verstand sagt ihr, dass Humanoide nicht als Partner zugelassen werden sollten. Dabei gerät sie aber mit ihren Gefühlen in Widerstreit. Sie beschreibt den Widerspruch so, dass sie durch die Erfahrung mit Tom als Partner für herkömmliche Partnerschaften mit unperfekten Männern verdorben ist und ihr weiteres Leben immer eins „ohne dich“ sein wird – ihr also etwas fehlen wird, was sie ohne die Erfahrung mit Tom gar nicht vermisst hätte. Wie sie sich entscheidet, bleibt offen.

Einer, der sich entschieden hat, ist Dr. Stuber, der immer daran gescheitert ist, normale Beziehungen zu Menschen aufzubauen (sinngemäße Aussage: alle haben mich immer aus irgendwelchen nicht erklärbaren Gründen abgelehnt). Er ist restlos glücklich mit seiner KI Chloé. Bedenken, wie sie Alma geäußert hat, hat er offenbar nicht.

¹² Tom hat zwar gleich zu Beginn des Films einen technischen „Aussetzer“ und muss deshalb gewartet werden, wird aber überall als Mensch wahrgenommen, z.B. auch von Almas Ex-Partner und in Lokalen. Adam kauft sich eigenständig Kleidung und ist stolz darauf, als Mensch behandelt worden zu sein: „Ich wurde mit „Sir“ angeredet“ (269)

¹³ Alma und Tom sind Teil eines Projekts, das beantworten soll, ob Humanoiden heiraten dürfen.

¹⁴ Dabei ist die Erwartung der Hersteller, dass die Maschine ein besserer Partner ist als ein Mensch es sein könnte.

¹⁵ Anders als Klara soll er also nicht eine existierende Person ersetzen, sondern nur die Rolle übernehmen, die ansonsten ein (beliebiger) Mensch hätte.

¹⁶ Man kann auch spekulieren, ob Alma ihrerseits gehofft hat, Tom dort zu treffen.

Hier spaltet sich die Frage, ob KI das Leben besser machen, noch einmal auf: Hier das individuelle, subjektive Glück eines einzelnen Menschen, dort die Frage, was bedeutet das für die Menschheit? Anders gefragt: welche Nebenwirkungen hat das künstlich geschaffene Glück in der Partnerschaft? Nach Almas Statement in ihrem Gutachten wären die Folgen einer Zulassung von Humanoiden als Partner:innen verheerend. Eine Menschheit aus satten, nicht mehr veränderungsfähigen Menschen wäre wohl zum Untergang verurteilt. Dass man neben dem individuellen Glück auch das der Menschheit im Auge haben muss, ist auch in Toms Programmierung angelegt. Er versteht Almas Verzweiflung darüber, dass argentinische Forscher ihr mit einer Veröffentlichung zugekommen sind und damit ihre Arbeit wertlos gemacht haben, überhaupt nicht. Er bezeichnet die Tränen, die Alma vergießt, als „egoistische Tränen“. Entscheidend sei, dass die Erkenntnisse die Menschheit bereichern, nicht, wer die Erkenntnisse gewinnt. Tom ist insoweit als gespaltene Persönlichkeit konstruiert: Stellt er einerseits mit seiner Perfektheit als Partner das individuelle Glück über das der Menschheit, stellt er andererseits den Erkenntnisgewinn für die Menschheit über das individuelle Leid der „geliebten“ Partnerin.

Eine weitere kleine Szene legt schließlich noch die Frage nahe, ob Tom nicht nur der bessere Partner, sondern auch der bessere Mensch sein soll. Er versteht Schadenfreude nicht. Im Café beobachten Tom und eine andere Besucherin Jugendliche, die recht krasse Videos über slapstickartige Unfälle betrachten und sie sehr lustig finden. Nachdem die Frau auf Toms Frage auch nicht erklären kann, was daran lustig ist, bringt Tom sie schließlich mit der Frage, ob auch tödliche Unfälle lustig seien, zum Nachdenken und sie verneint die Frage.

Im Film wird nur die relativ harmlose menschliche Schwäche „Schadenfreude“ in Toms Programm korrigiert, in McEwans Roman ist das Thema der ethischen Prägung und des ethischen Handelns ein ganz zentrales¹⁷. Charlie ist offenbar skeptisch: „Adam war mir angeblich¹⁸ moralisch überlegen“. (123)

Es stellt sich daher die Frage, ob Adam tatsächlich mit seiner ethischen Handlungsfähigkeit die Welt verbessert. Zu betrachten ist vor allem, dass Adam der Polizei sämtliche Beweise dafür liefert, dass Miranda Gorringer zu Unrecht beschuldigt hat, sie vergewaltigt zu haben und Gorringer dafür zu Unrecht verurteilt wurde. Adams Anzeige führt zu einer Verurteilung von Miranda zu einer Gefängnisstrafe.

Die rechtliche Beurteilung von Mirandas Tat durch Adam ist (jedenfalls nach deutschem Recht) korrekt. Daran ändert auch der Umstand nichts, dass Gorringer Mirandas Jugendfreundin vergewaltigt hat und dafür nicht bestraft wurde. Das (deutsche) Recht kennt allerdings keine Pflicht zur Anzeige. Adams Verhalten geht über die Einhaltung von Rechtsvorschriften deutlich hinaus. Seine Programmierung sah in der Situation also offenbar eine moralische Verpflichtung zur Anzeige¹⁹. Verbessert die Annahme einer moralischen Pflicht zur Anzeige in

¹⁷ Der Roman erläutert nicht, wie Adam zu seinen moralischen Urteilen und daraus folgenden Handlungen kommt. Zum Stand der Forschung und den verschiedenen Typen moralischer Akteure s. Cathrin Misselhorn, *Maschinenethik*, 4. Auflage Reclams Universalbibliothek, Ditzingen 2018, 2019, S. 70 ff.

¹⁸ Hervorhebung von mir

¹⁹ Oder Adam ist bereits – in der Terminologie von James H. Moor, zitiert nach Misselhorn aaO. – vollumfänglicher moralischer Akteur, den es in der Realität als Maschine nicht gibt. Dies spielt aber für die hier behandelte Frage keine Rolle.

solchen Fällen die Welt? Es gibt gute Gründe, die rechtliche Regelung solcher Sachverhalte auch moralisch als abschließend zu betrachten. Es erscheint wenig erstrebenswert, unter moralischen Gesichtspunkten die gesamte Menschheit zu Hilfsheriffs zu machen. Das Gesetz geht auch an anderen Stellen davon aus, dass es Situationen gibt, in denen es dem einzelnen Menschen nicht zumutbar ist, zur Durchsetzung von – moralisch gerechtfertigten – Gesetzen aktiv beizutragen. Man denke an diverse Zeugnisverweigerungsrechte.²⁰

Daneben hat sich Adam des Geldes bemächtigt, dass er im Auftrag von Charlie bei Börsenspekulationen erwirtschaftet hat und überwiegend für wohltätige Zwecke verwendet. Natürlich kann man durchaus darüber streiten, ob es moralisch (und rechtlich) korrekt war, dass Charlie zunächst das gesamte Geld behalten hat. Hier hat Charlie offensichtlich selbst Zweifel: Als Adam eigenmächtig über einen kleinen Teil des Geldes für den Kauf von Kleidung verfügt und sich dafür rechtfertigt („Außerdem bin ich zu der Überzeugung gelangt, dass mir auch was von dem Geld zusteht, das du unterm Bett aufbewahrst“.) ändert sich Charlies Haltung zumindest etwas: „Okay“, sagte ich. „Da ist was dran“ (269). Ein Einverständnis, dass Adam über das gesamte Geld verfügt, liegt darin sicher nicht. Es handelt sich daher zumindest bei der Aneignung der Teilsumme, die Charlie zustand, rechtlich um Diebstahl.

Adams Verhalten führt also in beiden Fällen zu rechtlich und/oder moralisch zweifelhaften Ergebnissen. Seine Programmierung zum „moralisch Überlegenen“ ist misslungen und beruht (noch?) auf Hybris. McEwan lässt Turing als moralische und wissenschaftliche Instanz im Buch konstatieren: „Der Geist [des Menschen] gibt der Wissenschaft nach wie vor nichts als Rätsel auf.“ (399) Und noch deutlicher: „Maschinenlernen aber hat seine Grenzen. Ein künstlich geschaffener Verstand braucht notwendig Regeln, nach denen er funktioniert... Wer schreibt den Algorithmus...?“ (400). McEwan vermittelt also einen skeptischen Blick auf Nutzen und Gefahren bei der Entwicklung künstlicher Intelligenzen²¹.

Er legt damit überzeugend dar, dass der Weg zu einer moralischen Verbesserung der Welt durch Maschinen lang und gefahrenreich ist²².

²⁰ §§ 52 ff StPO für Angehörige von Beschuldigten, aber auch für bestimmte Berufsgruppen, wo das Zeugnisverweigerungsrecht im Interesse der Allgemeinheit liegt.

²¹ In einer Mischung aus Spielfilm und Dokumentation behandelt der sehenswerte Film „K.I. – Die letzte Erfindung“ von Regisseur Christian Twente aus 2020/2021 diese Thematik in Verbindung mit einem Überblick über den Stand der Entwicklung.

²² Nur angerissen wird die interessante Frage, welche ethischen Regeln für die Menschen im Umgang mit hochentwickelten Humanoiden gelten sollen. Charlie hat Adam zerstört. Durfte er das? Turing verneint das vehement: „Ich hoffe, eines Tages wird man das, was Sie Adam angetan haben, als schwere Straftat verurteilen“ (401). Turing geht dabei davon aus, dass Adam eine Persönlichkeit und Gefühle hatte (ebenda). Es wäre lohnend, auch diesem Thema nachzugehen, würde hier aber zu weit führen.