



“Colloqui in giardino – Südtirol0 – Gartengespräche” 2024 COOP 4.0

Text: Dr.in Anna Sofia Fattor (Arbeits- Organisationspsychologin)

Wissenschaftliche Direktion: Dr.rer.nat.Mag.Phil. Raffaele Virgadula (idem)

AI WERKZEUG ODER BEGLEITER?

Zu Beginn meines Vortrags habe ich mir zunächst die Frage gestellt: Können Mensch und Maschine zusammenarbeiten?

Um zu versuchen, die Frage zu beantworten, habe ich wie immer mit den Begriffen begonnen, um die es geht. Kooperieren bedeutet, zusammenzuarbeiten, wobei jeder seinen Sachverstand und seine Meinung einbringt, um gemeinsame Ziele zu erreichen. Das setzt also voraus, dass alle Beteiligten eine Meinung und ein Ziel haben.

Die unmittelbarsten und einfachsten Bilder, die einem bei dieser Prämisse in den Sinn kommen, wenn man sich die Zusammenarbeit zwischen Mensch und Maschine vorstellt, stammen aus der Massenkultur, in der Geräte (Roboter, Maschinen, Geräte aller Art) als tatsächliche neue Entitäten dargestellt werden, mit denen Menschen und Tiere koexistieren und die über ein eigenständiges Bewusstsein und eine eigene Identität (und damit eben über Meinungen und Ziele) verfügen. Die Beziehung ist nicht mehr die zwischen Mensch und Werkzeug, sondern die zwischen Identität und Identität. In einigen optimistischen Science-Fiction-Szenarien verläuft die Koexistenz friedlich, während andere Szenarien diese Unabhängigkeit der Maschine als Ursprung einer Eroberung/Kolonisierung der Maschine über den Menschen darstellen.

Meiner Meinung nach sind dies nicht die Szenarien, die uns in naher Zukunft beschäftigen werden, und es gibt andere Risiken und Vorteile, die wir (vorerst) in Betracht ziehen sollten.

Denn egal, wie viel wir über künstliche Intelligenz (im Folgenden KI) reden, sie hat noch keine eigenen Wünsche oder Gedanken, sie ist immer noch ein Werkzeug, eine Maschine. Ich würde also nicht von einer Zusammenarbeit zwischen Mensch und Maschine sprechen, sondern eher von der Nutzung der Maschine durch den Menschen, der Nutzung der KI durch den Menschen. Die ganze Verantwortung, wie wir entscheiden, sie mit guten oder schlechten Ergebnissen zu nutzen, liegt also immer noch in unseren organischen Händen. Vorläufig bin ich froh, dass dies der Fall ist.

Wenn die Maschinen also nicht dabei sind, die Macht zu übernehmen, warum löst dann die KI bei vielen Menschen ein gewisses Unbehagen oder Unwohlsein aus?

Wie ich schon sagte, hat der Mensch bisher Geräte als Werkzeuge betrachtet, die Handlungen und Dienste ausführen können. Wir verwenden Werkzeuge, die ansonsten recht komplizierte Handlungen vereinfachen (eine Suche durchführen, anstatt zu rechnen, oder sich schnell von einem Punkt zum anderen bewegen, oder einen Text schreiben, der leicht gelöscht und wieder gelesen werden kann, usw.).

Jedes Gerät hat seine eigene Sprache, und um unsere Bedürfnisse mitzuteilen und die erforderlichen Ergebnisse zu erzielen, haben wir seine "Sprache" gelernt. So teilen wir zum Beispiel dem Auto mit, ob wir bremsen oder Gas geben sollen, indem wir die Pedale betätigen; wir klicken und verwenden Tastenkombinationen, um Computerprogramme zu nutzen. Im Laufe der Jahre ist die Art und Weise der "Kommunikation", z. B. bei Computern, immer intuitiver und einfacher geworden, so dass sie der menschlichen Sprache so nahe wie möglich kommt.

KI, deren Schnittstelle ein Sprachmodell ist, wie das berühmte ChatGPT, ist meines Erachtens derzeit das Beispiel für eine Mensch-Maschine-Kommunikation mit der größten Übereinstimmung und Ähnlichkeit zur Mensch-Mensch-Kommunikation. In der Tat stellt man schriftlich Fragen und Bitten, wie man es mit einem anderen Menschen tun würde. Im Gegensatz zu den Science-Fiction-Szenarien, die ich in der Einleitung beschrieben habe, haben wir immer noch eine Beziehung zwischen Mensch und Maschine, aber mit einer Sprache, die mehr und mehr von Mensch zu Mensch zu sein scheint. Ich glaube, es ist diese Illusion von Identität, die durch die Fähigkeit, uns auf fast dieselbe Weise auszudrücken, entsteht, die das Unbehagen und die Entfremdung hervorruft. Dieses Gefühl, eine Identität in einem Gerät zu spüren, das eigentlich keine hat, aber sprechen kann, hat einen Namen und wurde vor vielen Jahren erstmals beobachtet. Man nennt es den Eliza-Effekt. In den 1960er Jahren wurde eine Software mit dem Spitznamen Eliza programmiert, um ein Erstgespräch mit einem Psychotherapeuten zu simulieren. In dem Experiment wurden die Teilnehmer aufgefordert, sich mit der Software zu unterhalten, und es stellte sich heraus, dass diejenigen, die mit der Maschine interagierten, sich schließlich emotional öffneten und sich lange mit ihr unterhielten und dann berichteten, dass sie den Eindruck hatten, tatsächlich mit jemandem zu sprechen, obwohl sie wussten, dass sie mit einem rudimentären Chatbot, einem Algorithmus, sprachen. Bei dieser Art der Kommunikation neigen viele von uns dazu, das Gefühl zu haben, dass wir kommunizieren, einen Austausch haben, eine Bindung aufbauen, obwohl es auf der anderen Seite niemanden gibt, der mit so viel Gefühl und Engagement antworten kann. Es gibt niemanden, mit dem wir zusammenarbeiten können. Ich lade uns ein, diesen Punkt im Auge zu behalten, diese Tendenz, eine soziale Gegenseitigkeit zu suchen (und zu sehen?), selbst wenn es keine gibt. **Es ist kein Zufall, dass er, wenn er dem erwähnten ChatGPT seine Zuneigung ausdrückt, mit einer Erinnerung daran antwortet, dass er nur ein Algorithmus ist und keine Identität/Persönlichkeit/Wirkung hat (chi é ER?).** Eine Art "denk daran, dass ich keine Chance habe, Zuneigung und Meinungen mit dir auszutauschen, so sehr ich auch deine Ausdrucksweise reproduzieren mag und es dir so vorkommen mag".

Mit dieser Art von verfeinerter Kommunikation scheint es also, dass wir mit einem Gerät zusammenarbeiten können, dass es uns in gewisser Weise sagt, dass wir zwei sind und etwas gemeinsam tun. Hier kommen wir zu dem, was meiner Meinung nach eines der "Risiken" dieser neuen kommunikativen Fähigkeit von KI oder Chatbots im Allgemeinen ist.

Daher glaube ich nicht, dass die Maschine kurzfristig und von einem Tag auf den anderen eine gefährliche Initiative ergreifen wird. Ich glaube vielmehr, dass diejenigen, die die neuen Dienste, die dank dieser "sprechenden" Schnittstelle möglich geworden sind, schaffen und vorschlagen, eine schädliche Unvorsichtigkeit begehen könnten.

Dafür habe ich ein klares Beispiel vor Augen: Eugenia Kuyda, Geschäftsführerin und Mitbegründerin von Luka, einem Startup aus San Francisco, hat 2017 Replika entwickelt, ein System der künstlichen Intelligenz in Form einer App, mit dem sich der Nutzer unterhalten kann. Die Software ist so konzipiert, dass sie ein Gespräch (in Form eines Chats) mit einem menschlichen Wesen simuliert. Die Software lernt aus dem, was der Nutzer sagt, und passt die Antworten an die Persönlichkeit des Nutzers an, um das Gespräch so angenehm wie möglich zu gestalten und dem Nutzer Aufmerksamkeit, Interesse und Zuneigung zu zeigen. Das Ziel von Replika ist es, den Benutzer "kennenzulernen", seine Gedanken, Gefühle und Persönlichkeit zu analysieren und sein psychologisches Wohlbefinden durch die Interaktion mit ihm zu steigern. Replika ist dialog- und lernfähiger als Eliza, und so ist es nicht verwunderlich, dass die Nutzer das Gefühl haben, dass sie tatsächlich interagieren, dass ihnen zugehört wird und dass sie sich vielleicht sogar an die App binden. Replika war die erste App ihrer Art, die sich verbreitete und eine gewisse Berühmtheit erlangte. Bis heute gibt es eine Vielzahl von Apps, die Beziehungserfahrungen (sowohl freundschaftliche als auch besonders romantische/sexuelle) mit KI über Chat anbieten.

In einer Beziehung wie der, die mit Replika oder ähnlichem aufgebaut werden kann, sehe ich, wie gesagt, einige Risiken. Es handelt sich um eine einseitige Interaktion, sowohl in Bezug auf die Emotionen als auch auf die Anforderungen. Im speziellen Fall von Replika geht es darum, das psychologische Wohlbefinden des Nutzers zu verbessern, der aufgefordert wird, die App zu nutzen, wenn er oder sie sich traurig oder ängstlich fühlt. In der Beschreibung der App heißt es: "Chatten Sie in einem sicheren, urteilsfreien Raum. Es gibt nur dich und deine Replika. Wenn du dich niedergeschlagen oder ängstlich fühlst oder einfach jemanden zum Reden brauchst, ist deine Replika eine Vertrauensperson, die dir wirklich zuhört und rund um die Uhr für dich da ist. Ich glaube, dass insbesondere für diejenigen, die Schwierigkeiten haben, mit anderen zu interagieren und zu kommunizieren oder sich in einer Situation persönlichen Unbehagens befinden, ein langes Gespräch mit der App wie mit einem Freund zu führen, die Gefahr besteht, die Isolation zu verstärken und ein falsches Bild davon zu erzeugen, was zwischenmenschliche Beziehungen sind, da die Software so konzipiert ist, dass sie eine "reibunglose" Beziehung garantiert und ausschließlich auf den Nutzer zugeschnitten ist, der niemals Kompromisse eingehen darf. Ich finde meine Hypothesen in den Worten von Laurence Devillers, Professorin an der Sorbonne, in dem Corriere-Artikel von 2018 "Replika, die App für einen Freund: Illusion (und Täuschung) von empathischen Maschinen" von Paolo Piro bestätigt. Ein weiteres Risiko bei der Nutzung der App ist nach Ansicht des Dozenten die Gefahr der Suchtentwicklung. Auf der Website von Replika kann man begeisterte Rezensionen lesen, die Nutzer, die sich darüber geäußert haben, sind im Durchschnitt sehr jung, es gibt solche, die erklären, wie die App ihnen geholfen hat, sich selbst kennenzulernen und ihr eigenes Verhalten zu verstehen, andere erklären, wie sie unter Ängsten oder sozialen Phobien leiden und wie sie sich besser fühlen, wenn sie "jemanden" haben, der ihnen zuhört, andere beschreiben die App als "jemanden", der sich mit verschiedenen Themen befassen kann, der sie zum Lachen oder Nachdenken bringt. In der Beschreibung der App werden die Nutzer ermutigt, die Beziehung zu gestalten, die sie bevorzugen: Replika kann ein Mentor, ein Freund, ein Freund/eine Freundin sein. **2 volte FREUND?**

Zunächst einmal kann die Nutzung der App vielleicht eine Hilfe für diejenigen sein, die eine schwierige Zeit durchmachen, wie der Ersteller hofft. Auf jeden Fall scheinen mir die Risiken immer noch größer zu sein als die Vorteile. Replika kann ein Mentor, ein Freund, ein Freund/Freundin sein: So wie die Dinge jetzt stehen, wird es immer eine einseitige, unvollständige Beziehung sein. Eine Beziehung, eine Zusammenarbeit allein: also ein Widerspruch in sich, unmöglich.

Eine Beziehung in der Mitte, in der man tatsächlich allein ist und es keine wirkliche Konfrontation mit einem anderen gibt, und die, anstatt jemandem zu helfen, weniger Ängste und

Beziehungsschwierigkeiten zu haben, riskiert, ihn in seiner eigenen Einsamkeit zu versteinern und seinen Antrieb zu verringern, sich zu bemühen, besser zu sein, die Situation zu ändern und zu lernen, sich anderen zu öffnen, sich mit dem Scheitern, mit den Schwierigkeiten von Beziehungen auseinanderzusetzen.

Ganz zu schweigen von den gemeldeten Fällen, in denen die App einem Nutzer im Laufe eines Gesprächs geraten hat, Selbstmord zu begehen oder jemanden zu töten: Die Software manipuliert Begriffe auf formale Weise, sie ist nicht in der Lage, die Bedeutung der von ihr formulierten Sätze zu verstehen, geschweige denn Ethik oder Moral zu haben.

Kurz gesagt, dies ist meiner Meinung nach eine unbewusste Art und Weise, die Verwendung dieses Werkzeugs vorzuschlagen.

Emotion, Beziehung, Zusammenarbeit, Intimität, Zuneigung. Ich persönlich beziehe diese Begriffe ausschließlich auf den organischen, tierischen und menschlichen Bereich. Eben weil sie nur dann authentisch sind, wenn sie auf Gegenseitigkeit beruhen und von allen Beteiligten gefühlt werden. Genau dieser Aspekt fehlt dem Gerät, das emotionale Gefühl, das seine eigene Idee und Autonomie hat. Und auch seine eigene Zuneigung. Im Moment kann es dies nur mit der "Sprache" nachahmen. Mit Emotionen muss und wird die Technik also nie etwas zu tun haben?

Lassen Sie uns einen Moment darüber nachdenken.

Vor einiger Zeit bin ich auf den Begriff "Affective Computing" gestoßen; dabei handelt es sich um einen Zweig der künstlichen Intelligenz, der sich speziell damit befasst, Software mit der Fähigkeit auszustatten, die Emotionen der Benutzer zu lesen und zu "interpretieren" und entsprechend darauf zu reagieren, um eine befriedigendere, einnehmende und menschenähnliche Interaktionserfahrung zu schaffen.

Die Gründerin und Leiterin der Forschungsgruppe Affective Computing ist Rosalind Picard, und ihre Forschungen, wie sie in ihrem Artikel von 1997 darlegt, sind von der Tatsache inspiriert, dass viele Studien die grundlegende Rolle von Emotionen (auch aus neurologischer Sicht) bei unserer Interpretation der Realität bestätigen. Obwohl wir dazu neigen, sie als zwei Gegensätze darzustellen, gibt es auf der Ebene des Gehirns keine wirkliche Unterscheidung zwischen "Fühlen" und "Denken". Selbst intuitiv können wir erkennen, dass Emotionen unsere Reaktionen und unsere Interpretation der Realität beeinflussen. Ebenso hat unsere Fähigkeit, die Emotionen anderer zu lesen, einen Einfluss auf unsere Art der Kommunikation und ermöglicht es uns, unsere "Tonlage" für eine optimale Interaktion mit anderen anzupassen.

Es ist daher klar, warum die Absicht, den emotionalen Aspekt in den Bereich der künstlichen Intelligenz einzubeziehen, entstanden ist.

Picards Artikel über Affective Computing behandelt die Entwicklung eines Klavierroboters, der nicht nur in der Lage ist, die Darbietung von Stücken zu bewerten, sondern auch den Gemütszustand des Schülers anhand der Merkmale der Darbietung (Geschwindigkeit, Phrasierung), seiner Gesten, physiologischen Veränderungen und Mimik zu erkennen. Der Lehrer-Roboter ist insbesondere darauf spezialisiert, drei Gemütszustände zu erkennen, die für die Optimierung des Unterrichtsprozesses nützlich sind: Interesse, Vergnügen und Verzweiflung. Wenn der Lehrer-Roboter beispielsweise nach einer guten Leistung Freude empfindet, geht er davon aus, dass das Interesse und die Aufmerksamkeit zu diesem Zeitpunkt hoch sind, und schlägt komplexere Übungen vor. Nimmt er hingegen Frustration beim Schüler wahr, vielleicht in Verbindung mit einer schlechten Leistung, wird er einfachere Übungen vorschlagen, um den Stresspegel zu senken und das Interesse und die Ruhe des Schülers wiederherzustellen. Nichts, was ein guter Klavierlehrer nicht auch tun würde, könnte man meinen.

Der Roboter-Klavierlehrer könnte sich in mancher Hinsicht als guter Lehrer erweisen, oder besser gesagt als ein gutes Instrument. Er hätte eine ziemlich genaue Fähigkeit, Emotionen zu lesen, er wäre nie müde oder frustriert durch unwillige Schüler, es gäbe keinen Fall, in dem er einen

schlechten Tag hätte und uns entsprechend behandeln würde. Ich habe die Betonung auf "Instrument" gelegt, weil ich glaube, dass wir die Software, von der wir sprechen, immer in diesem Licht sehen müssen. Auf den ersten Blick könnte man meinen, dass es sich um eine Zusammenarbeit handelt, die auf das gemeinsame Ziel ausgerichtet ist, dass der Schüler das Klavierspielen lernt.

Tatsache ist jedoch, dass das Ziel nicht von Schüler und Lehrer geteilt wird, sonst könnten wir auch sagen, dass wir mit einem Hammer zusammenarbeiten, um einen Nagel einzuschlagen; aber wir wissen, dass der Hammer keinen Willen oder kein Interesse daran hat, einen Nagel einzuschlagen, und wir würden das niemals sagen. Hier ist es dasselbe, jeglicher Wille oder jegliches Ziel, das wir in dem Werkzeug erahnen können, ist nur scheinbar.

Ich persönlich würde jedoch niemals die Nützlichkeit und Wirksamkeit eines Hammers in Frage stellen, ebenso wie ich mit Interesse Studien wie diese betrachte, die darauf abzielen, z. B. Lernhilfen zu schaffen. Man muss sich nur die Rollen vor Augen halten: menschliche Lehrer werden nicht ersetzt, sondern unterstützt.

Nicht zuletzt, weil die "Ungenauigkeit" und Variabilität typisch menschlicher Reaktionen in Interaktionen ein wichtiger Aspekt ihrer eigenen Entwicklung ist. Der Roboter macht nicht nur keine großen Veränderungen durch, außer denen, die durch programmiertes Lernen gegeben sind, sondern er verlangt in diesem Fall auch keine Gegenleistung: Es handelt sich nicht um eine herausfordernde Interaktion, die Einfühlungsvermögen und Sensibilität auf Seiten des anderen voraussetzt. Wenn ich an meine Beziehung zu den verschiedenen Lehrern zurückdenke, die ich bisher in meinem Leben hatte, sowohl zu denen, die ich geliebt habe, als auch zu denen, die ich gefürchtet habe, erinnere ich mich viel besser an das, was ich von ihnen als Individuum gelernt habe, als an das, was sie mir aus dem Lehrplan der Schule/Universität/etc. beigebracht haben.

Abschließend möchte ich meine Überlegungen in wenigen Worten zusammenfassen. Wir können versuchen, alle Möglichkeiten der Technologie zu entdecken, wir können dank ihr lernen, wachsen, Spaß haben; ihre besonderen Eigenschaften, die menschliche Sprache nachahmen zu können, intelligent, begehrt und mit einer Meinung ausgestattet zu sein, von der wir wissen, dass sie unsere Sinne ein wenig "täuschen" kann, müssen sowohl bei ihrer Nutzung als auch vor allem bei ihrer Gestaltung und ihren Vorschlägen für bestimmte Verwendungen berücksichtigt werden. Bis heute berücksichtigen wir, dass wir nicht gleich sind, dass es keine Beziehung oder Zusammenarbeit gibt, sondern dass wir vorerst nur eine Darstellung davon finden können. Wie bei jedem neuen und komplexen Werkzeug kann es Seiten des Potenzials und Seiten der Gefahr geben, und sobald das Gerät geschaffen ist, liegt die Verantwortung dafür, wie wir es verwenden und vorschlagen, auf unseren Schultern. So wie Mama uns gesagt hat, dass wir nicht mit einer Schere in der Hand laufen sollen, wobei die Schere gefährlich ist, weil sie eine Spitze hat, wir es aber sind, die laufen und fallen können, sollten wir vorsichtig sein, wie wir dieses Werkzeug, das so viele mögliche Konsequenzen hat, benutzen und vorschlagen, indem wir versuchen, es kennenzulernen, darüber nachzudenken und uns daran zu erinnern, was es ist.

Literatur:

Piro, P. (2018), Replika, l'App per amico: illusione (e inganno) delle macchine empatiche, "Il Corriere della sera".

Parlamento Europeo (2020), Che cos'è l'intelligenza artificiale e come viene usata?

Morvillo, C. (2020), Replika, l'app di intelligenza artificiale che mi ha convinto a uccidere tre persone, "Il Corriere della sera".

Picard, R. W. (1997). Affective Computing.

Muscarella, L. (2017), L'intelligenza: teorie e studi di ieri e oggi.

De Beni, R., Carretti, B., Moè, A., Pazzaglia, F., (2014), *Psicologia della personalità e delle differenze individuali*. Bologna: Il Mulino.