

DIE ETHIK DER KÜNSTLICHEN INTELLIGENZ

Algorithmen gewinnen Schachspiele, müssen aber auch ethisch relevante Entscheidungen treffen – etwa beim autonomen Fahren

Frühe Paradigmen in der Entwicklungsgeschichte von Algorithmen und Computertechnik waren das Rechnen und das Schachspielen. Beides sind regelbasierte Verhaltensweisen, die sich als Algorithmus darstellen und in Computer einspeisen lassen. Weil sowohl im Rechnen als auch im Schachspielen Systeme der Künstlichen Intelligenz bald menschliche Leistungen übertrumpften, kreisten die Debatten lange Zeit um die Frage, worin sich intelligente Maschinen und Menschen gleichen und worin sie sich unterscheiden. Am Leitbild der menschlichen Intelligenz in ihrem ganzen Facettenreichtum gemessen, blieb die reale Entwicklung der Künstlichen Intelligenz lange Zeit hinter den ausschweifenden Visionen zurück. Doch mit dem jüngeren Boom der KI-Entwicklung um die Jahrtausendwende, aus der zunehmend kommerzielle Produkte hervorgehen, stellen sich vermehrt auch ethische Fragen: Wie müssen KI-Systeme wie Social Bots, Big-Data Anwendungen und autonome Fahrzeuge gestaltet sein, um eine selbstbestimmte Umgangsweise auf Seiten der Nutzerinnen und Nutzer zur ermöglichen?

Technikethischen Reflexionen bietet das Prinzip der Akzeptabilität Orientierung. Es besagt, dass sich Nutzer und Betroffene bei Bedarf in ein ablehnendes oder annehmendes Verhältnis zu dem Systemgeschehen stellen können müssen. Die Gestaltung der Systeme sollte demnach folgende Anforderungen erfüllen: Erstens muss es für die Nutzer bei Bedarf erkennbar sein, ob sie sich in einer System-Interaktion befinden. Zweitens müssen die Systemabläufe bei Bedarf für die Nutzer nachvollziehbar gemacht werden. Drittens sollte auf gesellschaftlicher Ebene eine Diskussion darüber gefördert werden, welche Systeme und Entwicklungen in welcher Form als sinnvoll und wünschenswert erachtet werden oder nicht. Gerade wenn wir es mit infrastrukturellen Entwicklungen zu tun haben, aus denen einzelne Nutzer nicht austreten können, sollte eine gesellschaftlich-politische Diskussion über deren Förderung und mögliche Alternativen angeregt werden. Beim autonomen Fahren beispielsweise wird über moralische und rechtliche Einzelfragen diskutiert, zu wenig aber über gesellschaftliche Visionen der Mobilität der Zukunft.

Haben wir es mit simplen Umgangsformen zu tun, wie der Nutzung eines Taschenrechners, dem Schreiben einer E-Mail oder dem Schachspiel gegen einen Computer, sind die Forderungen des Akzeptabilitäts-Prinzips leicht erfüllt. Beim Umgang mit jüngeren KI-Anwendungen sind diese jedoch keineswegs trivial. Vernetzte Sensoren und Aktuatoren transformieren Wohnungen, Supermärkte, Büros oder Bahnhöfe in sogenannten Smart Environments, in denen den Usern proaktiv kontextsensitive Dienstleistungen angeboten werden. Da bei steigender Integration von IT in Alltagsgegenstände und zunehmender Vernetzung aller Dinge im Internet of Things es auch immer unkenntlicher werden kann, ob man sich überhaupt in einer Smart Environment befindet



^ Die Forschung von Suzana Alpsancar umfasst technikphilosophische und ethische Herausforderung der Mensch-Computer-Interaktion. Derzeit arbeitet sie zur Theorie und Geschichte der Technikfolgenabschätzung sowie zum Problem des Technikdeterminismus

und welche Systemabläufe wann im Hintergrund operieren, ist die Gewährleistung einer selbstbestimmten Techniknutzung auch eine Frage des System-Designs.

Die Mensch-Computer-Interaktion basierte lange Zeit auf dem Muster der expliziten Delegation, nachdem die Maschine auf Befehl ausführt, was von den Nutzern veranlasst wird, zum Beispiel per Mausklick. Implizite Delegationen bezeichnen ein Systemverhalten, dass zwar auf die Nutzer reagiert, jedoch nicht von diesen aktiv ausgelöst wurde. Ein Beispiel aus den Szenarien des Ambient Intelligent Living: Insbesondere ältere, eingeschränkte Personen sollen durch intelligente Assistenzsysteme länger ein selbstbestimmtes und eigenständiges Leben zu Hause führen können. Auch das Gesundheits- und das Pflegesystem sollen durch diese technische Unterstützung Bedürftiger entlastet werden. Wenn der Teppich einen Sturz meldet, gibt die intelligente Wohnung einen Notruf ab. Sie warnt, wenn der Herd beim Gang durch die Wohnungstür noch an ist und überwacht Verhalten wie Schlaf und Toilettengang. Mit einer impliziten Delegation hätte es ein Bewohner dann zu tun, wenn etwa der Toilettensensor wenig Spülgang verzeichnet, das System diese Messung als fehlende Flüssigkeitsaufnahme interpretiert und dann – ohne dies mit dem Bewohner rückzukoppeln – Angehörige, Pflegedienste und Krankenkassen über dieses ungesunde Verhalten informiert. Das System könnte bei einem Lieferdienst Getränke anfordern oder die Nutzung anderer Geräte oder Dienstleistungen

an den erfolgreichen Toilettengang koppeln – nach dem Motto: Ich lasse dich erst deine Lieblingsendung im Fernsehen schauen, wenn du genug getrunken hast.

Natürlich lässt sich ein Toilettensensor leicht austricksen, wie Testpersonen solcher Proto-Wohnungen schnell bewiesen hatten. Dieses Beispiel verdeutlicht jedoch, dass bei Bedarf eine Transparenz zu Systemprozessen wichtig werden kann, damit sich die Nutzer zu den Dienstleistungen selbstbestimmt verhalten können. Hierbei ist es auch wichtig, dass man einschätzen kann, ob man es in einer Interaktion mit einem personalen Gegenüber oder einem System wie dem Chatbot zu tun hat. Ein Chatbot ist eine Software, die mit uns kommuniziert. Während zum Beispiel Reisebüros oder Banken in ihrem Online-Service es Kunden vermutlich leicht erkennbar machen, ob sie von einem Bot oder einem Mitarbeiter beraten werden, ist dies zum Beispiel in Sozialen Netzwerken nicht immer klar. Bei Posts oder Tweets kann hier die Frage der Autorschaft und damit auch der Authentizität des vorgegebenen Profils mitunter politische Brisanz erhalten. Das ist der Fall, wenn Bot-Armeen zu Zwecken der Wahlpropaganda eingesetzt werden und die Automatismen der Nachrichten-Verbreitung nicht leicht einzuschätzen sind.

Die Frage der Autorschaft lässt sich auch auf die Zusammenstellung eigener Profil-Seiten in sozialen Netzwerken beziehen. Je mehr das, was man etwa in seinem Facebook-Account sieht, an den eigenen Präferenzen orientiert und durch Likes, Herzchen et cetera algorithmisch zusammengesetzt wird, desto mehr blickt man sich sinnbildlich selbst ins Gesicht. Die Begleitforschung spricht hier vom Phänomen der Spiegelwelten, das der Aura einer authentischen Öffentlichkeit – zu der unvermeidlich Dissens, Konflikte und Unerwartetes dazugehören – mit



⤴ Beim autonomen Fahren wird über moralische und rechtliche Einzelfragen diskutiert, zu wenig aber über gesellschaftliche Visionen der Mobilität der Zukunft

der sich soziale Netzwerke gerne dekorieren, diametral entgegensteht. In anderen Szenarien, erfüllen KI-Systeme vielleicht gerade deswegen ihren Zweck, weil ihre Authentizität keine Rolle spielt. Ein Beispiel ist die Roboter-Robbe »Paro«, die in Pflegeeinrichtungen Demenzkranken ein Lächeln auf die Lippen zaubert und die Kontaktaufnahme mit ihnen gestattet.

Entwicklungen der Künstlichen Intelligenz können auch moralische Entscheidungen dringlich machen, die bisher als Frage gar nicht zur Praxis gehörten. Beim autonomen Fahren lautet die Preisfrage, ob das Fahrzeug den älteren Herren oder den Jungen, die Nobelpreisträgerin oder eine Kindergartengruppe in einer Dilemma-Situation überfahren soll. Bemerkenswert an dieser Diskussion scheint neben den gravierenden rechtlichen Fragen weniger das Dilemma selbst als der Umstand zu sein, dass wir überhaupt durch die technische Entwicklung als PKW-Nutzer in die Lage versetzt werden, uns zum Abwägen der moralischen Dilemmata vorab Gedanken machen zu müssen. Das strategische Abwägen eines Schadens gehörte bisher keineswegs zur guten Praxis des Autofahrens, sondern in den Kontext der Staatsgewalt wie Militär und Polizei sowie des Versicherungswesens. Im autonomen Fahren jedoch wird auch der Unfall zur regelbasierten Handlungsfolge.

Wie diese Beispiele demonstrieren, haben wir es aus Sicht der Nutzer mit neuen und diversen Formen der Mensch-System-Interaktion zu tun. Eine fundierte ethische Reflexion dieser Entwicklungen bedarf daher eine technikphilosophische Analyse der verschiedenen Interaktionstypen. Die traditionelle Ethik geht von autonomen Handlungssubjekten aus – den Individuen, Unternehmen, Staaten oder anderen kooperativ agierenden Organisationen – die verschiedene Handlungsalternativen gegeneinander abwägen. Diese Perspektive ist Vorteil und Nachteil zugleich. Sie schließt unmittelbar an unser lebensweltliches Selbstverständnis als autonome Akteure an und lässt Fragen nach Autorschaft und Verantwortung adressierbar machen. Sie kommt jedoch dort an ihre Grenzen, wo wir es mit Effekten kollektiver oder anonymer Urheberschaft zu tun haben. Die Entwicklung der KI stellt somit auch eine Herausforderung an die ethische Theoriebildung dar. Gebraucht wird eine sozialphilosophische Theorie, die es erlaubt, die Nutzung der neuen Systeme so zu beschreiben, dass wir moralische und politische Fragen angemessen diskutieren können.

Arbeitsgebiet Technikphilosophie
GASTPROFESSORIN DR. PHIL. SUZANA ALPSANCAR
