

REZENSION

Roboterethik

Abgesang auf die anthropologische Prämisse?

Jascha Bareis, *Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), Karlsruher Institut für Technologie (KIT), Karlstr. 11, 76133 Karlsruhe (jascha.bareis@kit.edu)*

64

Folgen wir Lohs Einführung in die Roboterethik, so erwartet die Technikfolgeabschätzung (TA) in der Zukunft ein ausgeweitetes Arbeitsfeld. Allgegenwärtig dringt die Robotik in Sphären wie Verkehr, Arbeit, Medizin, Pflege, Militär und übernimmt hier Kompetenzen, welche vorher den Menschen überlassen waren. TA hat sich mit den moralischen, sozialen und politischen Konsequenzen auseinanderzusetzen, welche diese Entwicklung für eine Gesellschaft mit sich bringt. Die Roboterethik gibt aber der *Gesellschaft* eine neue Konnotation, denn hier tummeln sich nun eben auch die Roboter – und das nicht nur als ausführende Instrumente menschlicher Delegation. Loh untersucht sie als eigenständige *moral agents*, also „Subjekte moralischen Handelns“ (S. 48), und als *moral patients*, also Objekte, die „in einer spezifischen (moralisch angemessenen) Weise zu behandeln“ sind (S. 74). Weiter ausgeführt wird diese Untersuchung am Gegenstand der Verantwortung, also der Frage nach der Möglichkeit, ob Roboter als Verantwortungssubjekte moralische Verantwortung übernehmen können (S. 138) oder Adressat*innen eines verantwortlichen Umgangs sein sollen (S. 162).

Die Roboterethik selbst ist ein neuer Teilbereich der Ethik, welcher sich seine Daseinsberechtigung erst einmal erkämpfen muss. Denn die Fähigkeit zur Moralität wird traditionell und intuitiv dem Menschen zugeschrieben, oft ohne nach den Möglichkeiten derselben zu fragen. Die Roboterethik aber will und kann dieser Grundintuition nicht einfach folgen. Sie kann zwei verschiedene Wege beschreiten: Entweder wird der Roboter dem Menschen in all seinen Fähigkeiten zur Moralität gleichgesetzt (Fühlen, Bewusstsein, Autonomie, etc.) – das impliziert eine nach heutigem Technikstand schlicht abwegige Erwartungshaltung an Künstliche Intelligenz. Oder die zweite Möglichkeit: Die *Bedingungen* für Moralität werden schlichtweg neu definiert. Das umfasst einerseits eine Re-Konzeptualisierung der notwendigen Fähigkeiten zur Moralität, aber vor allem die grundsätzlichere Frage: Wo verortet sich Moralität überhaupt? In den Fä-

higkeiten einer Entität? In der Simulation von Fähigkeiten, welche die Entität gar nicht besitzt? Oder gar in den prozesshaften Interaktionsräumen von Entitäten? Es sind diese grundlegenden Fragen, welche die Leser*in durch die Lektüre begleiten werden. Um ihnen gerecht zu werden, ist der Ausgangspunkt von Lohs Einführung keine x-te Abhandlung von ethischen Ansätzen wie Tugendethik, Utilitarismus etc. Anstatt bietet Loh auf einer erfrischend neuen Flughöhe eine produktive Infragestellung der Selbstverständlichkeit dieser Ansätze und somit auch der Ethik als Disziplin *an sich*.

Exklusive und inklusive Ansätze

Der Untersuchung des Verhältnisses von Robotik und Moral voran geht dementsprechend eine das Buch durchziehende Unterteilung in sogenannte *exklusive* philosophische Ansätze, welche auf einem anthropozentrischen, aristotelischen Fundament den Menschen als „Dreh- und Angelpunkt definieren“ (S. 37). Exkludierend deswegen, da dies eine Subjekt-Objekt-Dichotomie widerspiegeln, nach welcher der Mensch als alleiniger Souverän anderen Entitäten epistemische Bedeutung und moralischen Wert zuschreibt. Auf der anderen Seite des Spektrums positionieren sich die von Loh als *inklusive* bezeichneten philosophischen Ansätze, vorrangig neo-materialistische Denkrichtungen, vertreten von gegenwärtigen feministischen und posthumanistischen Denker*innen wie z. B. Barad, Haraway und auch Loh selbst. Hier soll das vorherrschende Verständnis des liberalen, der Moral fähigen Handlungssubjekts aufgebrochen und auch nichtmenschliche Entitäten wie Roboter in den moralischen Kosmos aufgenommen werden. Das klassische Verständnis von Subjekt-Objekt-Dichotomien und allein menschlicher Handlungsfähigkeit soll hier durch ein Infragestellen von Kategorien aufgebrochen und stattdessen „relational als im Zusammenspiel unterschiedlicher menschlicher und nichtmenschlicher Wesen sich realisierend verstanden werden“ (S. 95).

Was Loh dann in diesem dichotomen theoretischen Raster von exklusiven und inklusiven Ansätzen aufspannt, ist wirklich beachtlich. So werden eine Vielzahl von Roboterethiker*innen aus dem angelsächsischen Raum (dem Ursprung der Debatte) in den deutschsprachigen Raum eingeführt sowie auch deutsche gegenwärtige Denker*innen vorgestellt. Am Ende jedes Kapitels werden die jeweiligen Positionen auf aktuell existierende Roboter beispielhaft angewendet, z. B. auf die artifizielle Pflegeroboter *Paro*, den humanoiden Roboter *Cog* oder das Kunstprojekt *BlessU-2*. Zusammenfassend erwartet die Leser*in zudem eine aufschlussreiche, tabellarisch vergleichende Übersicht mit den abgearbeiteten Standpunkten, die den Überblick erheblich erleichtert.

Die Frage nach der Beziehung von Moralität und Robotik dränge sich, so Loh, schon deswegen auf, da Menschen emotionale und intime Bindungen zu Robotern eingehen. Loh zeigt auf, dass der Bereich der *social robotics* (z. B. Sex, Arbeit, Pflege) uns schon heute mit dem Für und Wider des Phänomens der Anthropomorphisierung konfrontiert, also der Zuschreibung menschlicher Attribute und Kommunikationsfähigkeiten auf

Roboter, die diese nur imitieren. Die Debatte verfährt hier phänomenologisch und verortet Moralität in der Erscheinung von Robotern. Kann und sollte man einem Roboter, welcher Gefühle simuliert, einen moralischen Wert zuschreiben?

Die Übernahme moralischer Verantwortung?

Die Kontroverse spitzt sich weiter zu in der Frage nach der Möglichkeit der Übernahme moralischer Verantwortung durch Roboter, welche Loh als „dialogisches Prinzip“ (S. 127) herausarbeitet. Hier strukturiert Loh die Debatte vor und schickt folgende Fähigkeiten als Bedingungen für Verantwortung ins Feld: a) Kommunikationsfähigkeit eines Rede-und-Antwort-Stehens; b) Handlungsfähigkeit im Sinne von Autonomie (in der Roboterethik grundsätzlich negativ verstanden als die Abwesenheit

Verantwortungssubjekte beziehungsweise -subjektobjekte (da keine klare Differenzierung zwischen Subjekten und Objekten mehr zu treffen ist) auch auf nichtmenschliche Wesen“ (S. 197).

Fazit: Neue Grenzziehungen für die TA?

Lohs Roboterethik nimmt die Leser*in mit auf eine intellektuelle Reise, welche hingenommene moralische Gewissheiten auf den Prüfstein stellt. Wer das Buch zuklappt, der bleibt mit Haraway gesprochen wunderbar *troubled* zurück: produktiv empört, fasziniert und Stellung beziehend. Denn wer die Ansicht vertritt, dass Moralität an ein *subjektives* moralisches Bewusstsein gebunden ist, an die Fähigkeit der Empathie (Einfühl(!)ungsvermögen – und nicht nur an abstraktes Urteilsvermögen), an das Empfinden von Gefühlen wie Schmerzen etc. (und nicht nur an

Wo verortet sich Moralität: In den Fähigkeiten?

In der Simulation von Fähigkeiten? In der Interaktion von Entitäten?

von äußeren Zwängen); und c) Urteilskraft als ein Mitgeben von Gründen und Kriterien, damit die verantwortungsübernehmende Entität ihr Handeln auch beurteilen kann (S. 133 f.). Eine solche, von Loh als Bewertungsstandard festgesetzte, Minimaldefinition spiegelt einen funktionalistischen und behavioristischen Grundtenor wider, und somit werden die Grundbedingungen für moralische Verantwortung sehr niedrig angesetzt. Denn klar ist, dass ein Algorithmus in verschiedenen Komplexitätsstufen all das problemlos kann: a) Er bekommt einen Input (*actio*) und produziert einen Output (*reactio*) und somit *kommuniziert* er; b) heutige Algorithmen verfügen über einen *search space* und haben somit einen Handlungsspielraum bei Rechenprozessen, in dem sie ‚autonom‘ entscheiden (anders als z. B. Ampeln); c) Selbstlernende algorithmische Systeme sind adaptive Systeme, welche den Output fortlaufend ‚scoren‘ und somit auch *beurteilen*. Reicht aber all das, um Robotern moralische Verantwortung für ihr Handeln zuzuschreiben?

Diese Frage würden die inklusiven Denker*innen grundsätzlich verneinen, denn sie verorten Verantwortung in einer Sphäre, die „sich ausschließlich in der Interaktion zwischen Wesen abspielt und nicht als Attribut einem einzelnen Wesen zuschreibbar ist“ (S. 187). Die Bedingungen zur Verantwortung werden nun immer weiter von der Entität entkoppelt und finden sich vielleicht noch in der von Loh angeführten Kommunikationsfähigkeit eines Rede-und-Antwort-Stehens wieder. Die Grundprämisse fußt hier auf der Beobachtung, dass die Welt nur als Beziehungsnetzwerk und nicht vom Standpunkt des autarken Handlungsobjekts aufzugreifen sei. Wirklichkeit werde nicht einfach ‚aufgedeckt‘, sondern erst in der Interaktion *produziert*. Hieraus wird ein relationales moralisches Verantwortungskonzept abgeleitet, welches das Prinzip der Verantwortung auf multiple, gar verschmolzene Akteur*innenschaft überträgt. Dieses Prinzip umfasst eine „radikale Ausweitung des Kreises potenzieller

die mechanische Ausführung eines Programms) – der wird mit vielen vorgestellten Denker*innen der Roboterethik nur schwer warm werden. Die Lektüre lohnt sich gerade dann. Denn es ist eben Lohs vorgenommene analytische Gegenüberstellung exklusiver und inklusiver Denkrichtungen zu verdanken, unser ethisches Werte-Weltbild radikal zur Disposition zu stellen – und uns damit aus unserer allzu gemütlich und unhinterfragten anthropozentrischen Komfortzone zu locken. Wer die Moralität des Roboters versucht zu verorten, der fragt auch notwendigerweise nach der Verortung des Menschen. Welche Grenzziehungen sind noch legitim, welche müssen gar aufrecht erhalten werden? Können wir eine ontologische und moralisch bewertende Differenz zwischen Mensch und Roboter aufheben, ohne den Menschen (und seine Verantwortung) zu relativieren? Mit den fortschreitenden Fähigkeiten der Robotik kann die TA solchen Fragen nicht aus dem Weg gehen. Lohs Einführung gibt den leserwerten Einstieg.



Loh, Janina (2019):

Roboterethik. Eine Einführung.

Berlin: Suhrkamp.

241 S., 18,- €,

ISBN 9783518298770