

Responsive Roboter?

von Henning Mayer



Titel Roboterethik

Untertitel Eine Einführung

Autor(en) Janina Loh

Land Deutschland

Erschienen Berlin 2019: Suhrkamp

Umfang S. 241

Preis EUR 18,00

ISBN 978-3-518-29877-0

Rezension zu "Roboterethik. Eine Einführung" von Janina Loh

Maschinell lernfähige Technologien drängen derzeit nicht nur auf den Markt, sondern in intime Bereiche der menschlichen Lebenswelten. Innovations- wie gesundheitspolitisch gilt *artificial intelligence* (AI) als potenzieller Träger einer technisch gestützten Komfort-Offensive, die datenbasierte Wertschöpfungsprozesse

dem ‚Wohlbefinden‘ der Bevölkerung dienlich macht.[1] Dem zugrunde liegt die Idee einer algorithmischen Vergemeinschaftung: Menschen und Roboter sollen sich qua Empathie familiarisieren,[2] Facebook wird zum virtuellen Wohnzimmer ähnlich denkender Nutzer[3] und kommerzielle Chatbots werden zu besten Freunden, für deren Rat man die Couch nicht mehr verlassen muss.[4] Damit ist einerseits eine erhebliche Verschiebung von Konzepten der Privatheit und Sicherheit verbunden: Privatheit ergibt sich nicht mehr exklusiv durch Schutz vor der Maschine, sondern inklusiv: aus Vergemeinschaftung mit sicheren technischen Akteuren, die menschliche Datenspuren nutzen, um ihre Modelle und Heuristiken zu sensibilisieren. Andererseits wachsen aber auch die Ängste vor dem Kontrollverlust, der auf Dauer gestellten Fremdbeobachtung und nicht zuletzt: vor der Unterwanderung bekannter Wertmaßstäbe. Was wissen Google, Amazon und Facebook über mich? Kann ich Recommender-Algorithmen, die mir bestimmte Versicherungspakete empfehlen, vertrauen oder nicht? Möchte ich wirklich von scheinbar feinfühligem Robotern gepflegt werden oder gehe ich davon aus, dass menschliche Pfleger das besser können? Erfrischend abgeklärt zwischen diesen beiden Extrempositionen steht Janina Lohs Buch *Roboterethik*, in dem die in Wien lehrende und forschende Technik- und Medienphilosophin inklusive, relationale und damit auch für die Science & Technology Studies interessante „Ansätze der Verantwortungszuschreibung“ (187) entwirft.

Zu Beginn widmet sich Loh auf konzentrierte wie umfassende Weise der Frage, ob Roboter als *Subjekte* oder *Objekte* „moralischen Handelns“ (13 f., 48 ff., 72 ff.) zu gelten haben. Beide Perspektiven, so Loh, beruhen explizit oder implizit auf anthropologischen Prämissen: Die Vertreter*innen einer den Subjektstatus anerkennenden Position denken „von einem jeweiligen Menschenbild sowie einer Vorstellung des moralischen Handlungssubjekts ausgehend über die etwaige Anerkennung artifizierlicher moralischer Akteursschaft nach[...]“ (121) und identifizieren Roboter als „moralisch[e] Akteur*innen nach menschlichen Gesichtspunkten“ (78). Die Gegenposition praktiziere die „Vermenschlichung artifizierlicher Systeme“ (122) in die andere Richtung: Sie erkläre Roboter zu „moral patients“, die einen „(moralisch angemessenen) *Umgang* erfordern, ohne selbst moralische Akteur*innen [...] zu sein“ (75). Im Kontext der Roboterentwicklung setze dies die Deutungshoheit menschlicher Entwickler*innen voraus, die souverän darüber entscheiden, ob und wenn ja in welchem Maße einem artifizierlichen System moralische Kapazitäten zugesprochen werden sollen.

Loh wägt hier sehr behutsam unterschiedliche Perspektiven gegeneinander ab, die sich zwischen den Polen ‚Roboter sind aktive Träger moralischer Pflichten‘ und ‚Roboter sind passive Träger moralischer Rechte‘ bewegen.[5] Ausgehend von diesen eher traditionellen Theoriediskussionen, wendet sich Loh anschließend den praktischen Problemen der ewigen Forderung nach ‚Implementierbarkeit‘ zu: „Wie gelangt die Moral überhaupt in die Maschine?“ (43). Im Gegensatz zu Regulierungsbemühungen, die ‚Codes of Ethics‘ entwickeln und Robotikprojekte durch Richtlinien (ex ante) und Technikfolgenabschätzung (ex post) normativ einzurahmen suchen, ohne jedoch den Prozess des Entwickelns *als solchen* unter

die Lupe zu nehmen, widmet sich Loh den Praktiken des ‚Ethics by Design‘. Dabei unterscheidet sie zwischen top-down, bottom-up und hybriden Ansätzen der Responsibilisierung von Robotik (43 ff.).

Top-Down-Ansätze versprechen, verhaltenssteuernde Metaregeln in die Entscheidungsstrukturen von Robotern einzubauen, zum Beispiel indem sie Nutzenfunktionen eines KI-Agenten in Robotergesetze einlassen oder handlungsleitende Motive, Wünsche und Erwartungen implementieren, die Entwickler*innen zuvor festgelegt haben. Bottom-Up-Ansätze hingegen möchten einem Roboter nicht unmittelbar einschreiben, was er wann zu tun hat, sondern sie versuchen, autonome Systeme zu entwickeln, die erlernen können, welches Verhalten in bestimmten Situationen erwünscht oder unerwünscht ist (44).[6] ‚Ethics by Design‘-Ansätze aus dem Top-Down-Spektrum handeln sich Loh zufolge einen Widerspruch zwischen programmierlogischer und hermeneutischer Intelligenz ein. Programmcode sei auf „eindeutige Interpretation“ (43) angewiesen und müsse technisierbar, wiederholbar und maschinenlesbar gestaltet sein.[7] Code-Determinismus widerspreche so der dilemmatischen Natur von ethischen Entscheidungen, die kontextsensitiv, indexikal[8] und damit immer vieldeutig und veränderlich bleiben müssen. Das deterministische Einschreiben moralischer Werte führe demnach nahezu unvermeidbar zu Regelkonflikten, in denen Hybridpositionen nur selten möglich sind. Bottom-Up-Ansätze seien flexibler, da sie Lernplattformen oder Training Grounds[9] bereitstellen, auf denen teilautonome Technik praktisch in sozialen Alltagssettings erprobt werden und von sozialem Feedback lernen kann. Die moto-sensorischen Mechanismen, mithilfe derer sich Roboter ihre sozialen Umwelten erschließen, blieben indes behavioristisch[10] – ihnen fehle eine imaginative Empathie beziehungsweise deren Erlebensqualität. Zudem sei eine lernfähige „starke KI“ (47), die nur noch ein Mindestmaß menschlicher Supervision benötigt, aktuell noch gar nicht im Einsatz. Während also top-down orientierte Implementierungsvorschläge (Typik: schwache, regelbasierte KI) die Bedeutung von Kontext für die Standardisierung unterschätzen, überschätzen Bottom-Up-Ansätze (Typik: starke, autonome KI) die Kapazitäten der Lernfähigkeit von Robotern mit Blick auf die tatsächlichen Einsatzumgebungen. Hybridansätze, so Loh (46), leisten vor diesem Hintergrund einen Brückenschlag: Sie „kombinieren Top-Down- mit Bottom-Up Ansätzen, indem ein ethischer Rahmen basaler Werte vorgegeben wird, der dann durch Lernprozesse an spezifische Kontexte anzupassen ist“ (46).

Verantwortung als verteilter und situierter Prozess

In dem damit skizzierten Spannungsfeld zwischen globalen und lokalen, immer aber *verteilten* Normen lassen sich die von Loh stark gemachten „[inklusive] Ansätze der Verantwortungszuschreibung“ (187-204) verorten. Fragen danach, ob Roboter moralisch selbstverantwortliche Handlungsträger sind, übertragen anthropozentrische Begriffe häufig unmittelbar auf Technik, indem sie gewisse Kriterien abfragen, die lernfähige Roboter als moralische Handlungssubjekte

qualifizieren: etwa Autonomie,[11] Selbstursprünglichkeit[12] oder allgemeiner: Handlungspotenziale.[13] Schaut man jedoch mit der anthropozentrischen Brille auf komplexe, heterogene Entwicklungsarrangements, dann wird deutlich, dass man sich in vielen Situationen gar nicht so sicher sein kann, wer oder was gerade Verantwortung trägt. Insbesondere im Fall der Robotik verunklart sich bei genauerem Hinsehen, wer denn nun wirklich für eine bestimmte Aktivität oder Eigenschaft der Technik Rede und Antwort stehen *kann*: Sind es die Entwickler*innen? Und wenn ja: welche? Sind es bestimmte Hardware-Software Konfigurationen oder diejenigen Institutionen, die den Roboter in einem bestimmten Setting einsetzen? Wer verantwortet fehlerhaftes Verhalten? Der Roboter selbst, die Anwender*innen, die Forschungsleitenden?

Denkt man vor diesem Hintergrund wie Janina Loh (176) in „Verantwortungsnetzwerken“, dann lassen sich Roboter genauso wenig als alleinige Verantwortungssubjekte oder -objekte bezeichnen wie einzelne Entwickler*innen. Unter Bezugnahme auf „roboterethische“[14] und „kritisch-posthumanistische Ansätze“[15] (beide 199) fragt Loh deshalb folgerichtig nach möglichen „Verantwortungskonstellationen“, in die sich Roboter (in welcher Form auch immer) einbinden lassen. Verantwortung wird hier erstens radikal prozeduralisiert – sie „entwickelt sich erst im Zusammenspiel mehrerer menschlicher und nicht-menschlicher Wesen“ (132). Zweitens ist sie damit automatisch hochgradig „situiert“ (189-192),[16] da jede Positionierung menschlicher wie nicht-menschlicher Entitäten das ‚Rede-und-Antwort-stehen‘ für die eigene Position verlangt. Und drittens kann Verantwortung in dieser Verteiltheit nur ein Resultat des, mit Barad gesprochen, intraaktiven (nicht: interaktiven) Zusammenwirkens von Relata sein, die erst durch die Relationierung entstehen (und nicht: ihr vorausgehen).

What could be next?

Das Buch ist hervorragend verständlich, auch weil Janina Loh jedes Kapitel mit zumeist tabellarischen Zwischenzusammenfassungen abschließt und darin auf konzise Art und Weise Kernprobleme der Roboterethik zueinander in Beziehung setzt. In dieser Form taugt es meines Erachtens auch, aber nicht nur als Handbuch und Augenöffner für eine fachfremde und gegebenenfalls praxisnahe Leserschaft. Steile Thesen findet man wenige, konkrete Beispiele dafür umso mehr. Gegen Ende des Buches gewinnt man den Eindruck, dass sich Loh im Grunde mit der Formulierung erster Ansätze für eine inklusive, symmetrische Roboterethik zufriedengibt, deren detaillierte Ausarbeitung (an und mit Fallstudien) jedoch noch aussteht. Thematische Anschlüsse wären zum Beispiel mit Blick auf wissenschafts- und techniksoziologische Ansätze der Standardisierungsforschung interessant: Wenn man Standards weitreichend als soziotechnische Infrastrukturen versteht,[17] die so oder so – quasi: sobald es zu organisierten Handlungsabläufen kommt – unsichtbar und im Hintergrund bestimmte Entscheidungen, Konstellationen und Abläufe hervorbringen, dann drängt sich die Frage auf, wie

Verantwortung in normative (Sicherheitsregeln, Lab Policies, technische Machbarkeiten) und kognitive Standards (gezielte Irritationen, Institutionalisierung von Möglichkeiten, die eigenen Annahmen zu hinterfragen) übersetzt werden kann, wie sie sich also im Sinne von Latour ‚mobil‘ machen lässt.[18] Darüber hinaus ließen sich wohl einige Differenzen in die Inklusion zurückholen: Mit Hayles[19] wäre zum Beispiel zu fragen, inwieweit technische und menschliche „cognizer“ mit ihren spezifischen Entscheidungskapazitäten auf responsible Art und Weise füreinander intelligibel gemacht werden können.

Fußnoten

- [1] Joseph E. Stiglitz / Amartya Sen / Jean-Paul Fitoussi, Report by the Commission on the Measurement of Economic Performance and Social Progress, Paris 2009; Institute of Electrical and Electronics Engineers (IEEE) (Hg.), Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing in the Age of Artificial Intelligence [15.12.2020], 2019.
- [2] Rosalind W. Picard, „Affective Computing. From Laughter to IEEE“, in: IEEE Transactions on Affective Computing 1 (2010), 1, S. 11-17.
- [3] Mark Zuckerberg, Facebook-Post vom 11. Januar 2018 [15.12.2020]; Marc Zuckerberg, 2019 - Live at F8! [15.12.2020], Zuckerberg Transcripts. 1010, 2019.
- [4] Siehe etwa die Website von IBM Research [15.12.2020].
- [5] Siehe auch Janina Loh, „Roboterethik. Über eine noch junge Bereichsethik“, in: Information Philosophie 2017, 1, S. 20-33.
- [6] Im Anschluss u.a. an Angelo Cangelosi / Matthew Schlesinger, Developmental Robotics. From Babies to Robots, Cambridge, MA 2015.
- [7] Vgl. auch Werner Rammert, „Technisierung und Medien in Sozialsystemen. Annäherung an eine Soziologie der Technik“, in: P. Weingart (Hg.), Technik als sozialer Prozess, Frankfurt am Main 1989, S. 128-173.
- [8] Also im Sinne von Garfinkel „nur durch Kontexte definiert“. Vgl. Harold Garfinkel, Ethnomethodological Studies of Work, London 1986.
- [9] In der Robotik zum Beispiel die „Takkus“ in Japan, aber auch Plattformen wie iCub, Roboy etc.
- [10] Gerade in den häufig zum Einsatz kommenden Verfahren der *Plan Recognition* und *Sentiment* Analysen.
- [11] Luciano Floridi / Jeff W. Sanders, „On the Morality of Artificial Agents“, in: Minds and Machines 14 (2004), 3, S. 349-379.
- [12] Catrin Misselhorn, Grundfragen der Maschinenethik, Stuttgart 2018.
- [13] Georges Canguilhem, „Maschine und Organismus“, in: ders., Die Erkenntnis des Lebens, übers. von Till Bardoux / Maria Muhle / Francesca Raimondi, Berlin 2012, S. 183-232.
- [14] Allan F. Hanson, „Beyond the Skin Bag. On the Moral Responsibility of Extended Agencies“, in: Ethics and Information Technology 11 (2009), S. 91-99.
- [15] Donna Haraway, „Situierendes Wissen. Die Wissenschaftsfrage im Feminismus und

das Privileg einer partialen Perspektive“, in: C. Hammer / I. Stieß (Hg.), Donna Haraway. Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen, Frankfurt am Main / New York 1995, S. 73-97; Lucy Suchman, Human-Machine Reconfigurations. Plans and Situated Actions, Cambridge 2007; Karen Barad, „Verschränkungen und Politik. Karen Barad im Gespräch mit Jennifer Sophia Theodor“, in: dies., Verschränkungen, Berlin 2015, S. 173-212.

[16] Haraway, „Situierendes Wissen“, S. 84; Suchman, Human-Machine Reconfigurations, S. 285.

[17] Martha Lampland / Susan-Leigh Star, Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life, Ithaca, NY 2009; Lawrence Busch, Standards. Recipes for Reality, Cambridge, MA 2011.

[18] Bruno Latour, Visualisation and Cognition: Dawing Things Together, in: Knowledge and Society. Studies in the Sociology of Culture Past and Present 6 (1990), S. 1-40.

[19] Katherine Hayles, Cognitive Assemblages: Technical Agency and Human Interactions, in: Critical Inquiry 43 (2016), 1, S. 32-56; Dies., Unthought: The Power of Cognitive Nonconscious, Chicago, IL 2017.