Diskussion

Dirk Hommrich

Ethik – als Strichliste. Spyros Tzafestas führt durch die »Roboterethik« und gibt dabei einen Überblick, der mit Philosophie und Technikethik wenig zu tun hat

Rezension zu: Spyros G. Tzafestas: Roboethics. A Navigating Overview, Cham, Heidelberg, New York, Dordrecht, London 2016, (Reihe: Intelligent Systems, Control and Automation. Science and Engineering), XIII, 204 S.

Versteht man Technikethik als Bestandteil der sogenannten Bereichsethiken, dann wundert es nicht, dass das hier zu besprechende Buch von der Überzeugung getragen wird, es bedürfe einer gesonderten, bereichsspezifischen Roboterethik. Einer solchen traut Spyros Tzafestas gleich dreierlei zu: ›ethische‹ Konstruktionshilfe für Techniker zu bieten, ein philosophisches Arbeitsgebiet zu werden und Rat gebende Instanz zu sein, um auf vorgeblich drängende neue Herausforderungen und politische Regulierungsprobleme ›Antworten‹ zu bieten. Dabei habe die Roboterethik allgemein vor Gefahren warnende, Risiken aufspürende und Gefahren bewertende Funktionen, so der Autor (S. 2). Anwendungsfelder für Roboter sind die industrielle Fertigung und Transport, die Medizin und Operationssäle, die Pflege und das Rettungswesen, die Raumfahrt, das Militär sowie die Unterhaltungs- und Spieleindustrie. Jeweils existieren hier unterschiedliche Robotersysteme. Deren Vielfalt führt Tzafestas mit zahlreichen farbigen Abbildungen vor Augen und inszeniert sie durchaus bewusst »spherical« (S. 10).

Das Buch ist in zwölf Kapitel unterteilt, die das Thema abstecken. Nach einem kurzen einleitenden Kapitel, das einen historischen Abriss über Publikationen gibt, die Streitfragen und normative Aspekte der Robotik behauptet haben, präsentiert das zweite Kapitel par force Grundwissenspartikel ethischer Reflexion (S. 14–16) – Metaethik, theoretisch-analytische Ethik, Angewandte Ethik – um dann auf ein paar weiteren Seiten (S. 16–20) unter anderem die Tugendethik, deontologische Ansätze, den Utilitarismus, die Theorie der Gerechtigkeit als Fairness, aber auch den Egoismus als philosophisch-psychologische Theorie vs. den Altruismus, »value-based theory« und die Kasuistik aufzuführen (S. 19). Im nachfolgenden Teil über ethische Fundamentalien wechselt dann der Bezugsrahmen: Nun werden konkret die Ethikrichtlinien des Institute of Electrical and Electronics Engineers und der American Society of Mechanical Engineers vorgestellt. Überhaupt werden im Buch durchge-

hend zwei unterschiedliche normative Register bedient: ›Angewandte Ethik‹ einerseits sowie Kodizes, Handreichungen und Empfehlungen von verschiedenen Berufsverbänden und nationalen, internationalen Organisationen andererseits.

Das dritte Kapitel (S. 25–33) thematisiert noch vor der Robotik, was eine Art Übergang schafft, die künstliche Intelligenz. Die »ethical issues« der KI gehören zwar, Tzafestas zufolge, in die Informations- und Computerethik »or artificial intelligence ethics« (S. 26) und weniger zur Roboterethik. Aber sowohl künstliche Intelligenz als auch die Robotik haben Veränderungspotenziale, bergen Folgerisiken und zukünftige »issues«, die von den dazugehörigen Bindestrichethiken normativ gewendet werden sollen.

Das vierte Kapitel »The World of Robots« fährt ebenfalls gleichsam futurologisch fort, wenn anfangs zunächst von zukünftigen intelligenten Robotern gesprochen wird, die vollautonom wären. Solcherart in Artefakten als Roboter verkörperte (*embodied*), intelligente, eigenständige Systeme, wie man sie aus der Science-Fiction und anderen Formen der Zukunftsschau kennt, sind allerdings direkt im Anschluss *nicht* Gegenstand der von Tzafestas angeführten Definitionen. Diese beschränken sich dann vielmehr auf Industrieroboter heutigen Typs, die (lediglich) einen hohen Grad der Automatisierung und qua teilautonomer, spezifischer künstlicher Intelligenz einen gewissen Entscheidungs- und Handlungsspielraum haben (S. 37).

Irritierte es, dass Moralphilosophie und Ethik im zweiten Kapitel knapp, ohne weitere Ausführungen und Begründungen dargestellt werden, so haben im fünften Kapitel dann Auflistungen per Spiegelstrich einen übermäßigen Anteil an einer nun wirklich unübersichtlichen Darstellungsweise der Inhalte. Dem leser*un*freundlichen Erscheinungsbild des Textes hätte man bspw. Marginalien gewünscht. Eine exemplarische Botschaft aus dem fünften Kapitel lautet, dass sich die Roboterethik im Sinne der ›Angewandten Ethik‹ mit folgenden drei Problemzonen befasse:

– »The ethics of people who create and employ robots.
– The ethical system embedded into robots.
– The ethics of how people treat robots.« (S. 65)

Im Rahmen der Roboterethik, so Tzafestas weiter, würden unter anderen folgende Fragen verhandelt:

– »What role would robots have into our future?
– Is it possible to embed into robots ethics codes, and if yes is it ethical to program robots to follow such codes?
– Who or what is responsible if a robot causes harm?
– Are there any types of robot that should not be designed? Why?
– How might human ethics be extended such that to be applicable to the combined human-robot actions?
– Are there risks in creating emotional bonds with robots?« (S. 65)

Tzafestas stellt zwei Herangehensweisen der Roboterethik heraus: einen »top-down« und einen »bottom-up approach«, beide würden für das maschinelle Erlernen von Moralität benötigt. Ersterem schlägt Tzafestas deontologische sowie konsequentialistische und utilitaristische Ethiken zu, die ausgehend von moralischen Regeln oder Prinzipien bestimmte Verhaltensregeln vorschreiben und die auch in den Systemen von Robotern angewandt werden könnten (S. 68–71). Zur Erläuterung des »bottom-up approach« wiederum greift Tzafestas auf entwicklungspsychologische Ansätze zur Moralentwicklung zurück, um für Roboter von einer mit den vom Menschen vorgegebenen Normen konformen und sie befolgenden »operational morality« über eine »functional morality«, unter deren Steuerung robotische Aktionen kaum mehr vorhersehbar seien, bis hin zu einer »full morality«, deren konkrete Differenz zur vorigen Stufe im Dunkeln bleibt. Diese vollwertige Moralität beziehe sich auf einen Robotertyp »which is so intelligent that it is entirely autonomously selecting its actions and so it is fully responsible for them« (S. 73). Auch wenn Tzafestas darüber hinaus auf eineinhalb Seiten über »issues« der »Symbiose« zwischen Mensch und Robotik referiert und kurz auf das Thema »Rechte für Roboter bzw. Rechten von Robotern« zu sprechen kommt, die mit der Frage der Möglichkeit einer gewissermaßen ›synthetischen Phänomenalität‹ einhergehen, bleibt es verwunderlich, dass zum Ende des fünften Kapitels (und nicht nur dort) behauptet wird, die grundlegenden Fragen der Roboterethik seien damit »diskutiert« (S. 77). Eigentlich hat Tzafestas es bei knappen Schilderungen und dogmatischen Setzungen belassen.

Diese generelle Machart des Buches setzt sich in zwei weiteren Kapiteln über Roboter im medizinisch-therapeutischen und rehabilitativen Feld fort: Bezugnahmen auf fiktionale Genres und auf Zukunftsvorstellungen, Spiegelstrichlisten mit Typisierungsvorschlägen und Fragensammlungen sowie die Vermischung philosophischer Fragen mit spekulativen Szenarien und mit Versatzstücken vorliegender professioneller Ethik-Richtlinien. Und obwohl die Prothetik im elften Kapitel im Zusammenhang mit »Cyborg technology issues« thematisiert wird, steht sie auch im siebten Kapitel über Assistenzrobotik bzw. Hilfsroboter in der Therapie und Pflege im Zentrum.

Im achten Kapitel, das sich mit humanoiden »socialized robots« (S. 107–135) und Unterhaltungsrobotern beschäftigt, werden maschinelle Interaktionen betrachtet, die sich mit zunehmender Komplexität in ihrer sozialen Umwelt bewegen. Hier unterscheidet Tzafestas: 1. jene Roboter, die einfaches soziales Verhalten imitieren und dadurch bei Menschen antropomorphe Projektionen hervorrufen können (*socially evocative*); 2. kommutable Maschinen, d.h. Roboter, die dazu in der Lage sind, zwischen ›anderen‹ sozialen Akteuren und Objekten in der Umgebung zu differenzieren; 3. zum Modelllernen an menschlichem Sozialverhalten fähige Systeme (»without being able to proactive engage with people«); 4. Roboter die regelrecht kontaktfreudig (*sociable*) zu sein scheinen; 5. Maschinen, die etwa mittels künstlicher Intel-

ligenz auch »socially intelligent« agieren oder »socially interactive« wirken (S. 110–111).

Darüber hinaus beleuchtet Tzafestas schlaglichtartig Anlässe und Gründe für normativen Dissens, die sich durch quasi-sozialisierte Roboter ergäben (S. 119–120): Aspekte wie Verbundenheit und Bindung zu Robotern (*attachment*); mimische Vortäuschung etwa von homomorphen Gefühlen und Lebendigkeit, aber auch von Professionalität und Zuverlässigkeit (*deception*); Bewusstseinsbildung in und Aufklärung der Bevölkerung (*awareness*); Fragen nach der Autorität und Steuerung von Robotern bei der Interaktion mit Menschen (*authority*); nach Privatheit und Datenschutz (*privacy*); Schutz der Autonomie und Würde von Personen, die mit Robotern in Kontakt stehen (*autonomy*); etwaiger Verlust sozialer Bindungen unter und mit Menschen (human-human-relation); sowie Fragen der Verantwortung und Haftung für Schäden, bei Unfällen und anderen negativen Folgen, die mit Robotern zustande kommen können (*justice and responsibility*).

Mit dem neunten Kapitel über »war roboethics« weitet sich das Blickfeld enorm, denn der Einsatz von Robotern für militärische Zwecke bilde das Zentrum aller Roboterethik und habe gewiss die höchste Brisanz, so der Autor (S. 139). Daher referiert er in aller Kürze »realistische« Theorien des Krieges von Thukydides über Machiavelli und Hobbes bis zu Clausewitz und kontrastiert sie mit dem »idealistischen« Pazifismus. (S. 141–146) Und, die »ethics of war« dient Tzafestas ebenso dazu, zahlreiche weitere Punkte hinsichtlich der Möglichkeit und den Bedingungen gerechter Kriege anzusprechen, die Frage zu stellen nach Kriterien für die Entscheidung eine Waffe auszulösen, das (Menschenleben bewahrende) Ersetzen von Soldaten durch Kampfroboter zu thematisieren, das Verhältnismäßigkeitsprinzip zu erwähnen sowie drei Gegenargumente zum Einsatz von Kampfrobotern zu nennen: 1. Die Unfähigkeit intelligenten Systemen hinreichend das Kriegsrecht ›beizubringen‹, 2. die Notwendigkeit, weiterhin auf menschliche Soldaten aufgrund ihres Bewusstseins und ihrer Moralfähigkeit zu setzen, 3. die Herabsenkung der Schwelle zu kriegerischen Handlungen aufgrund des geringeren Verlustes an Menschenleben (S. 146–152).

Mit Blick auf den asiatischen Raum thematisiert das zehnte Kapitel die kulturelle Prägung von Sicht- und Wahrnehmungsweisen in Bezug auf das menschliche ›Zusammenleben‹ mit Robotern. Seinen Hinweis auf die Verflechtung von kulturellen Hintergrundüberzeugungen, Religiosität, Ethos und ethischen Intuitionen nutzt Tzafestas nicht allein dazu, nicht-christliche Glaubenssysteme (Shintoismus, Buddhismus, Konfuzianismus) einerseits und christlich geprägte Hilfsvorstellungen vom individuellen freien Willen andererseits sowie den soziokulturellen Umgang mit Robotern (und damit, eher implizit, der Roboterethik) vor dem Hintergrund von Fragen der Multi- und Interkulturalität einander gegenüber zu stellen. So sei die Weltanschauung etwa in der japanischen Kultur für Technik aufgeschlossener und erlaube

einen toleranteren und als unproblematischer empfundenen Einbezug von Robotern in das soziale Leben. Vielmehr greift Tzafestas unkommentiert hierzu passende hoffnungsvolle Zukunftsvisionen einer dereinstigen sozialen Koexistenz von Menschen und Robotern auf (S. 171). Dergleichen solle angestrebt, aktiv gestaltet und vorangetrieben werden, so die von ihm apologetisch zitierte Deklaration der International Robot Fair 2004.

Drängt sich bei der Lektüre immer wieder der Eindruck auf, Tzafestas denke den gegenwärtigen Stand der Robotik ›von der Zukunft her‹, weswegen man zum Ende hin etwa eine visionäre Protention möglicher Zukünfte erwartet, so überrascht das elfte Kapitel. Nicht ein spekulativer Ausblick folgt auf die Visionen der genannten Deklaration. Vielmehr stehen im Zentrum des vorletzten Kapitels »Additional Roboethics Issues«. Hier gibt Tzafestas drei aktuelle Beispiele dafür, was sich bereits heute als ankünftig abzeichne: »issues« über autonome Fahrzeuge (S. 176–179), immer intensivere Mensch-Maschine-Kopplungen (Cyborgs und Biohacking) durch Prothesen und Interfaces (S. 179–183) sowie Probleme im Zusammenhang von Datenschutz, Persönlichkeitsrechten und dem Schutz der Privatsphäre (S. 184–188).

Im zwölften und letzten Kapitel über »Mental Robots« mit »brain-like features« und »brain-like capacities« und sogar mit »artificial life-systems« kommt Tzafestas nochmals auf die im dritten Kapitel bereits beleuchtete künstliche Intelligenz zurück. Der Roboterethiker meint nun, dass der *»[p]ysical or body part* (mechanical structure, kinematics, dynamics, control, head, face, arms/hands, legs, wheels, wings etc.)« und der »[*m*]*ental or thinking part* (cognition, intelligence, autonomy, consciousness, conscience/ethics, and related processes, such as learning, emotions etc.)« (S. 192) von Robotern in der Art eines Dualismus zur Diskussion zu stellen seien. Und weil er damit zumindest für die Frage nach dem Konnex von Geist, Gehirn und Maschine und den philosophischen Streit über die Angemessenheit solcher anthropomorpher Ausdrücke einen Anschluss geschaffen hat, erläutert der Autor dann rasch noch Konzepte wie Intelligenz, Autonomie, Bewusstsein, Gewissen, Lernen und Aufmerksamkeit – jeweils mit szientistischer Schlagseite und unter Rückgriff auf Kognitionswissenschaft, Psychologie und Psychometrie. Dass Tzafestas in diesen Hinsichten einem reduktionistischen Materialismus (Geist = Gehirn) anhängt, der Bewusstsein und Qualia im Kopfinnenraum, im Hirngewebe vermutet und gut physikalistisch für technisch herstellbar hält, wundert kaum noch. Womit man dann das arg abrupte Ende des Buches erreicht.

Unbestritten: Tzafestas versammelt eine Menge wissenschaftlicher und anderer Quellen; man erfährt in seinem Buch etwas über die Vergangenheit der Robotik und erhält schlaglichtartige Eindrücke von bestehenden Ethikkodizes relevanter Berufsverbände wie beispielsweise der ISRA (International Service Robot Association) und weiterer Interessengruppen. Ein Glossar versucht ebenso Orientierungshilfe zu geben wie die Leserführung zu Beginn jedes Kapitels, bei der vorneweg angekün-

digt wird, worum es jeweils gehen soll. Vermutlich war auch mit den aufgrund ihres massenhaften Einsatzes eher verwirrenden Spiegelstrichlisten ähnliches beabsichtigt. Tzafestas schreibt im Vorwort, »novices in the field« (S. vii) sollen sich anhand seines Buches zurechtfinden können. Vielleicht hat der Autor sogar geahnt, dass er seinen Leserinnen und Lesern trotz anderslautender Behauptungen nicht wirklich einen ethischen Kompass an die Hand gibt. Dass Tzafestas in den finalen beiden Absätzen von »Roboethics« faktisch einfach auf die Philosophie insgesamt zurückverweist, ließe sich so erklären. Wer auf die im Untertitel des Buches versprochene »Navigationshilfe« gehofft hat, wird Orientierung also wohl allenfalls durch eine »Robophilosophy« finden, die aber erst noch zu schreiben wäre:

> »For anything we care to be interested we have a philosophy which deals with the investigation of its fundamental assumptions, questions, methods, and goals, i.e., for any X there is a philosophy which is concerned with the ontological, epistemological, teleological, ethical and aesthetic issues of X. Thus, we have philosophy of science, philosophy of technology, philosophy of biology, philosophy of computer science, philosophy of robotics (robophilosophy), etc.« (S. 200)

Selbst wenn man die immer wieder unangenehm aufstoßende Beliebigkeit der Einteilung in ethische ›Anwendungsbereiche‹ wie »artificial intelligence ethics« (S. 26) und »infoethics« (S. 165) akzeptiert,[1] und auch die oftmals kaum reflektierte Zuordnung moralphilosophischer Reflexion zu einem soziotechnischen ›Bereich‹ hinnimmt, bleiben folglich Kardinalprobleme des Buches bestehen. Eine angemessene Diskussion von theoretischen Leitbegriffen fehlt durchgehend. Tzafestas variabler Umgang mit Konzepten wie Bewusstsein, Maschine, Automat, Handlungsfreiheit, Sozialität, Leben etc. macht es schwer, sich denkend mit dem Thema auseinanderzusetzen.

Zwar richtet sich das Buch nicht an (akademische) Philosophinnen und Philosophen. Über den Nutzen für Nicht-Philosophen, Ingenieure, Informatiker und andere Spezialisten lohnt sich im Fall von Tzafestas' »Roboethics« aber auch kaum zu streiten: Für Techniker oder Mitarbeiterinnen und Mitarbeiter von Robotik-Startups dürfte die verworren dargebotene Trias aus Philosophie, moraltheoretischer Reflexion und Technikphilosophie auch nach der Lektüre ein Buch mit sieben Siegeln bleiben. Ob also Entwickler und »professionelle Ethiker« oder auch die »Roboterphilosophen« von morgen dank Tzafestas' Buch ›ethischere‹ oder zumindest bessere Systeme bauen, darf folglich bezweifelt werden.

1 Von den seit den 1980er Jahren bestehenden Bereichsethiken nennt Tzafestas (S. 15–16) herkömmliche wie die Medizinethik, die Bioethik, Medienethik, Umweltethik und Sozialethik (»welfare ethics«), aber auch dem Rezensenten bislang unbekannte Bindestrichethiken wie public sector ethics, business ethics, decision making ethics, legal ethics (justice), manufacturing ethics, computer ethics und automation ethics.

Wer allerdings trotz der hier genannten Unzulänglichkeiten stattliche 80 Euro für das eBook, 100 Euro für die Soft- und knapp 130 Euro für die Hardcover-Ausgabe investieren möchte, erwarte jedenfalls besser nicht das, was der Verlag dem Buch zuschreibt: »Provides background material on general ethics«,[2] denn eben das tut »Roboethics« nicht – und wird es auch dann nicht leisten, wenn es wegen solcher Behauptungen als vermeintliches Grundlagenwerk im Bestand öffentlicher Bibliotheken zu finden sein wird.

2 https://www.springer.com/de/book/9783319217130 (aufgerufen: 22.6.2018).