

Autonomie

Gottschalk-Mazouz, N. (2019): „Autonomie“
in: Liggieri, K.; Müller, O. (Hg.)
*Mensch – Maschine – Interaktion. Handbuch
zur Geschichte – Kultur – Ethik*. Stuttgart: J.B.
Metzler, 238-240.

Autonomie ist ein mehrdeutiger Begriff. Er bezeichnet Eigenschaften wie Unabhängigkeit, Selbständigkeit und Selbstbestimmtheit. Während bei technischer Autonomie meist die ersten beiden Bedeutungen im Vordergrund stehen, sind es bei menschlicher Autonomie die letzten beiden.

Innerhalb menschlicher Autonomie wird für gewöhnlich unterschieden zwischen personaler Autonomie, moralischer Autonomie und politischer Autonomie. Diese drei Formen setzen ein normativ gehaltvolles, d.h. Achtung, Respekt und Rücksicht verlangendes ‚Selbst‘ voraus. Soweit technische Systeme nicht als Personen, moralische oder politische Subjekte anerkannt werden, ihnen ein solches ‚Selbst‘ nicht zugesprochen wird, können ihnen diese Formen von Autonomie nicht zukommen.

Innerhalb technischer Autonomie werden für gewöhnlich Grade der Autonomie als Grade der Unabhängigkeit vom Menschen und der Umwelt bei der technischen Ausführung bestimmter Aufgaben unterschieden. Quer dazu können Komponenten dieser Ausführung unterschieden werden, nämlich Monitoring, Optionenbildung, Auswahl und Implementierung – jede dieser Komponenten kann dann mehr oder weniger autonom verfasst sein.

In der Mensch-Maschine-Interaktion treffen in der einfachsten Konstellation ein autonomer Mensch auf eine mehr oder weniger autonome Maschine und es gilt, deren Arbeitsteilung sinnvoll zu gestalten und dabei die normativ allein bedeutsame menschliche Autonomie möglichst umfassend zu wahren. Tatsächlich sind aber komplexere soziale Konstellationen die Regel, in denen mehrere Menschen untereinander und mit technischen Systemen in verschiedenen Rollen interagieren und deren verschiedene Ansprüche, u.a. auch auf Wahrung von Autonomie, adäquat zu vermitteln sind. Dann wird die Zurechenbarkeit technischer Vorgänge auf die beteiligten Menschen besonders wichtig, d.h. dass Verantwortlichkeiten im juristischen wie moralischen Sinne unter den beteiligten Menschen ausgehandelt und möglichst klar definiert werden.

Menschliche Autonomie

Funktional gesehen lässt sich menschliche Autonomie erklären als die Verbindung dreier Fähigkeiten, nämlich etwas über sich in Erfahrung zu bringen, es zu bewerten und es gegebenenfalls zu ändern. Diese drei Fähigkeiten sind zunächst eine begriffliche Bestimmung, d.h. Identifikation und Schematisierung der vorgefundenen, eigenen Meinung, Haltung oder Aktivität (man weiß, was man tut). Sodann, in distanzierender Reflexion des begrifflich Identifizierten, eine theoretische oder praktische Beurteilung (man verbindet mit seinem Tun eine Richtigkeits- oder Werthaftigkeitsüberzeugung). Schließlich, je nach Ergebnis von Bestimmung und Reflexion, eine adäquate Modifikation, d.h. Veränderung der Meinung, Haltung oder der Aktivität (man macht so weiter oder nicht).

Autonomie erschöpft sich jedenfalls nicht in einer bloßen Reaktion auf Vorgefundenenes (egal ob an oder in einem selbst, oder in der Umwelt), sondern erfordert die aktive Ausbildung eines Verhältnisses zu diesem Vorgefundenen. Erst dieses macht etwas in einem normativ relevanten Sinne zum ‚je eigenen‘, nämlich zur ‚je eigenen‘ Meinung,

Haltung oder Handlung, d.h. zu einer, zu der man sich bekennen und in der man ernstgenommen, für die man verantwortlich gemacht und für die man gelobt oder getadelt werden kann.

Facetten menschlicher Autonomie

Historisch gesehen entstammt der Autonomiebegriff der politischen Sphäre (Pohlmann 1971). Die Stadtstaaten im antiken Griechenland formulierten damit einen Anspruch auf Selbstbestimmung, gegen Bedrohungen von innen (Tyrannis) und von außen (Fremdherrschaft). Diese Selbstbestimmung war stets eine – u.a. durch übergeordnete Bündnisregeln – begrenzte Selbstbestimmung. Auch auf Personen bezogen wird unter ‚Autonomie‘ zuweilen eine solche, durch äußere und innere Determinanten begrenzte, aber dadurch teilweise auch erst ermöglichte Selbstbestimmung verstanden (so z.B. bei Walter 1998, in Abgrenzung zu seines Erachtens illusorischen Vorstellungen von völliger Freiheit). Während ‚personale Autonomie‘ hierbei die Fähigkeit heraushebt, für sich selbst zu entscheiden und diese Entscheidung dann auch durchzuhalten, meint ‚moralische Autonomie‘ vor allem die Fähigkeit, seine Entscheidungen aus Einsicht unter moralische Regeln zu stellen. Das Adjektiv ‚autonom‘ wird demgemäß nicht nur (‚global‘) auf Personen, sondern auch (‚lokal‘) auf Entscheidungen und Handlungen angewendet. In der Literatur werden hauptsächlich vier Typen von Kriterien für personale Autonomie vertreten (vgl. Buss 2013). Der erste Typ bezieht sich auf den kohärenten Zusammenhang mit anderen eigenen Meinungen, Haltungen und Handlungen, z.B. mit höherstufigen Haltungen (Wünschen zweiter Ordnung) – und dies in synchronen und diachronen Varianten. Der zweite Typ bezieht sich auf *responsiveness-to-reasons*, d.h. auf die Fähigkeit der Rechtfertigung angesichts naheliegender Einwände und Gründe. Der dritte Typ zielt auf *responsiveness-to-reasoning*, d.h. mehr auf den Prozess als auf das Resultat solcher Rechtfertigung. Der vierte Typ schließlich, der Inkompatibilismus, zielt auf die Abwesenheit von bestimmten kausalen Faktoren, die die Selbstbestimmung unterminieren würden.

Facetten technischer Autonomie

Auch die Autonomie technischer Systeme wurde versucht darin zu sehen, dass diese sich selbst Regeln setzen (vgl. Smithers 1997). Bei genauerer Betrachtung der verschiedenen Definitionen einer ‚Autonomie‘ technischer Systeme (Übersichten finden sich z.B. in Gottschalk-Mazouz 2008 oder in Beer et al. 2014, S. 76) geht es jedoch um ganz konkrete, recht verschiedene Fähigkeiten, die sich in drei Gruppen einteilen lassen: Erstens um eine Reihe von selbst auferlegten *engineering constraints* (Maes 1991) wie der Fähigkeit, ohne externe Energie- oder Materialversorgung auszukommen, sich ohne Führung zu bewegen oder Aufgaben ohne Benutzereingriffe zu erledigen. Anders gesagt, es geht hier darum, autark, mobil oder automatisch zu funktionieren. So lautet eine klassische Definition: “autonomous robots are usually taken to mean free-ranging mobile robots which are not teleoperated but plan and execute their own actions.” (Todd 1986, S. 233) Heutzutage wird meist feiner differenziert, etwa in ferngesteuerte, halbautonome und (voll-)autonome Systeme, oder in Grade bzw. Level autonomen Fahrens, von *No Automation* über *Driver Assistance* und einige weitere Stufen bis *Full Automation* (so der SAE-Standard J3016; vgl. D. Sokolov 2015). Mit zunehmender Autonomie nimmt typischerweise die Häufigkeit und Dauer von Benutzerinterventionen ab und verändert sich die Art der Benutzerinteraktion; sie wird globaler, abstrakter und höherwertiger. Dies kann nach den oben genannten Komponenten (Monitoring bis Implementierung)

bzw. entlang der Trias von *sense-plan-act* auch mit Blick auf die Mensch-Maschine-Interaktion weiter differenziert werden (Beer et al. 2014).

Zweitens wird Autonomie durch Bezugnahme auf die Systemumwelt beschrieben. Autonomie bezeichnet dann die Abhängigkeit des zukünftigen Verhaltens nur von inneren Zuständen des Systems (inklusive Sensorik), sowie die Fähigkeiten, denselben Auftrag in verschiedenen Situationen und in denselben Situationen verschiedene Aufträge auszuführen. Anders gesagt geht es hier um Innensteuerung, Adaptivität und Flexibilität. Diese sowie die vorangehende Gruppe adressiert Rammert (2003) in seinem synoptischen Vorschlag, die Autonomie eines technischen Systems als eine Frage des Grades an Eigenaktivität in vier Dimensionen zu charakterisieren: Der Motorik, von unbewegt zu automotiv und automobil; der Aktorik, von fremdbetätigt bis zu eigentätig, automatisch; der Sensorik, von fremdabgestimmt über umweltsensitiv bis zu selbstanpassend; und der Informatik, von fest verdrahtet über flexibel programmiert bis zu verteilt problemlösend.

Drittens schließlich wird Autonomie mit der Fähigkeit verbunden, uns zu überraschen. Autonome Systeme können demnach dazulernen, weisen nicht explizit bei der Konstruktion vorgegebene Verhaltensweisen auf oder uns unbekannte Zustände und Gesetze. Anders gesagt geht es um Lernen, Innovation und Unvorhersagbarkeit.

Allgemein gesagt hat die Zuschreibung technischer Autonomie somit folgende Form: x ist autonom in (positiv oder negativ bestimmten) Hinsichten a unter (normalerweise nicht ausgesprochenen) Randbedingungen b. Eine solche Zuschreibung setzt das System in u.U. mehrfacher Hinsicht in Relation, sie bestimmt nämlich (i) Austauschrelationen mit der System-Umwelt; (ii) Interaktionsrelationen mit Benutzer und Konstrukteur (Steuerung bzw. Programmierung) - sowie evtl. noch Dritten (Überwachung); (iii) epistemische Relationen bezüglich des Wissens von Benutzer/Konstrukteur/Dritten. Dabei bestimmen die epistemischen Relationen mit über unsere Interaktionen und diese wiederum mit über die Austauschverhältnisse des Systems.

Verhältnis von menschlicher und technischer Autonomie

Mit Blick auf Roboter wurde bereits in Christaller et al. (2001) die Frage ausführlich interdisziplinär diskutiert, was deren mögliche Autonomie für begleitende Maßnahmen erforderlich machen könnte. Die Systeme sollen, so wurde überlegt, erkennen lassen, was sie im Begriff sind zu tun, sollen deutlich werden lassen, dass sie in Benutzungssituationen lernen (um dem Benutzer seine Mitverantwortung klarzumachen) und dieses Lernen intern protokollieren, sollen als autonom handelnd erkennbar sein (z.B. den jeweiligen Operationsmodus durch Farbsignale anzeigen). Man könnte hinzufügen, dass sie auch Auskunft darüber geben können sollten, wessen Auftrag sie ausführen, wenn sie einen ausführen, und warum sie ihn gerade so ausführen, und einen *manual override* zulassen sollten. Allgemein kann man sagen, dass diese Systeme erkennen lassen sollten, wie sie den Spielraum, den sie haben, nutzen, und (höherstufig), dass ihnen ein solcher Spielraum offensteht, und sie sollten diesen Spielraum zur Disposition stellen können.

Inzwischen zeichnet sich ab, dass autonome Systeme typischerweise in komplexen sozialen Konstellationen stehen werden, dass sie untereinander und mit ihrer Umwelt vernetzt sein werden, und dass ein großer Teil der Lernprozesse auf implizites Wissen zielt, welches auch als Gelerntes nicht in sprachlicher Form vorliegt (z.B. bei neuronalen Netzen). Das stellt die Gestaltung von und den Umgang mit autonomen technischen

Systemen vor weitere, auch informationsethische Herausforderungen, deren Diskussion gerade erst beginnt.

Literatur

Buss, Sarah: „Personal Autonomy“. In: The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (Hg.), <https://plato.stanford.edu/archives/win2016/entries/personal-autonomy/> (21.5.17).

Beer, Jenay; Fisk, Arthur; Rogers, Wendy: „Toward a framework for levels of robot autonomy in human-robot-interaction“. In: Journal of Human-Robot-Interaction 3/2 (2014), 74-99.

Christaller, Thomas; Decker, Michael; Gilsbach, Joachim-Michael u.a.: Robotik. Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft. Heidelberg 2001.
Gottschalk-Mazouz, Niels: „›Autonomie‹ und die Autonomie ›autonomer technischer Systeme‹“. In: Gethmann, Carl (Hg.): Sektionsbeiträge: XXI. Deutscher Kongress für Philosophie 2008 (CD-Rom, ISBN 978-3-00-025531-1). Ebenso in: http://www.dgphil2008.de/fileadmin/download/Sektionsbeitraege/07_Gottschalk-Mazouz.pdf (20.5.17)

Maes, Pattie (Hg.): Designing Autonomous Agents. Cambridge, MA 1991.

Pohlmann, Rosemarie: „Autonomie“. In: Ritter, Joachim (Hg.): Historisches Wörterbuch der Philosophie Bd. 1. Basel 1971, 702-719.

Rammert, Werner: Technik in Aktion. Verteiltes Handeln in soziotechnischen Konstellationen. Technical University Technology Studies Working Papers TUTS-WP-2-2003, Berlin.

Smithers, Troy: „Autonomy in Robots and Other Agents“, In: Brain and Cognition 34/1 (1997), 88-106.

Sokolov, Daniel (2015): „Was ist der Unterschied zwischen einem autonomen Auto und einem autonomen Auto?“. In: <https://www.heise.de/ct/artikel/Was-ist-der-Unterschied-zwischen-einem-autonomen-Auto-und-einem-autonomen-Auto-2773184.html> (25.5.17)

Todd, D. J.: Fundamentals of robot technology. London 1986.

Walter, Henrik: Neurophilosophie der Willensfreiheit. Von libertarischen Illusionen zum Konzept natürlicher Autonomie. Paderborn 1998.

Niels-Gottschalk-Mazouz