
"Subjektlose Verantwortung"

Zur Ethik autonomer Systeme

Rainer Leschke 2018

Inhalt

1.	Abstract	2
2.	Erlösende Autonomie.....	3
3.	Von der Quantifizierbarkeit des Menschen.....	4
4.	Vom Turing-Test zur Prinzipienfrage oder: Zurück zur Autonomie.....	7
5.	Verfahren normativer Kontrolle	9
6.	Haltet die Betrüger! Oder die Strategien der Moralisierung.....	10
7.	Der ewige Kampf um Kontrolle	14
8.	Literaturverzeichnis	16

1. Abstract

Sowohl Maschinen als auch Software werden, um das Ausmaß des gegenwärtigen technologischen Paradigmenwandels zu formulieren, nicht selten die Eigenschaften von Subjekten zugebilligt: So sind sie smart, intelligent, autonom, selbstlernend, kreativ und verfügen damit über all die Eigenschaften, die die Subjektphilosophie seit der europäischen Aufklärung dem Menschen zuzubilligen bereit war, um damit soziale und kulturelle Entwicklung sowie die gesellschaftliche Organisation von Macht zu begründen.

So ziemlich das einzige, was man automatisierten Systemen noch nicht zugestanden hat, ist die Entwicklung normativer Grundlagen und ethischer Prinzipien. Dabei gehören ethische Entscheidungsmuster zweifellos nicht zu den menschlichen Spitzenleistungen, sondern sie sind eigentlich immer als problematisch und unzuverlässig angesehen worden, so problematisch, dass Luhmann sich immerhin genötigt sah, vor der Moral zu warnen. Gerade in Zeiten grassierender Populismen und Fundamentalismen ist deutlich geworden, dass soziale Systeme gegenwärtig eher unter zu viel Moral als unter zu wenig Moral zu leiden scheinen. Umgekehrt ist ebenso klar, dass soziale Systeme ohne normative Leitplanken nicht funktionieren können, da es ihnen dann nicht gelingt, eine identifizierbare Identität auszubilden.

Wenn also dieses ziemlich zweifelhafte menschliche Residuum, das zumindest als philosophisches Projekt sicher gescheitert ist, auch wenn das die professionelle Philosophie nicht wahrhaben will, durch automatisierte Systeme quasi final herausgefordert wird, dann gibt es offenbar nicht unerheblichen Klärungsbedarf.

Die Technikphilosophie Ernst Kapps ging einst in ihrer These von der Organprojektion (vgl. Kapp 1877, 29ff.) davon aus, dass der Mensch sich in den von ihm entwickelten Technologien selbst erkennen könne, und der Medienmythologe McLuhan ordnete bereits in den 60er Jahren des letzten Jahrhunderts den Computer in diese Reihe externalisierter oder ausgelagerter menschlicher Leistungen ein. Die Frage, die sich dann heute stellt, ist, ob das immer noch so ist, denn die neue Konstellation versprache dann zumindest in zwei Aspekten neue Einsichten: über automatisierte Systeme und über das menschliche Subjekt selbst.

2. Erlösende Autonomie

Nicht nur in der Kantschen Vision von Aufklärung, sondern gerade auch in den medientechnologisch stimulierten Prognosen der 60er Jahre des letzten Jahrhunderts wird die Autonomie des Subjekts als Erlösung modelliert, bei Kant noch einigermaßen schwer erkämpft, bei McLuhan und Max Bense eher en passant als eine Art Abfallprodukt der Medien- und Computertechnologie. Kapps Organprojektion¹, Freuds Prothesengott² und McLuhans Erweiterungstheorem³ dachten Technologien immer vom Subjekt aus und sie dachten sie vor allem als eine Art imperiale Erweiterung des Subjekts selbst und seiner Vermögen, seiner Erkenntnis, seiner Metaphysik, seiner Ästhetik und seiner Handlungsmacht. Zwar mögen Prothesen in der Regel eher als defizitär konnotiert sein, was sie im Übrigen dann mit den Medien gemein hätten, die doch bestenfalls ein schlechtes Surrogat für die eigentliche Kommunikation, den Dialog, zustande gebracht haben, aber ausgerechnet diese Prothesen, die Handlungsmacht, Ästhetik und Erkenntnis versprechen, drohen nunmehr die Macht zu übernehmen und zwar in genau den Regionen, die sie einstmals dem Menschen eröffneten. Aber bekanntlich schlägt das Imperium zurück und die Frage ist, ob es das dann nach den Regeln der Hollywood-Dramaturgie tut und mit welchen Normen Protagonisten und Antagonisten ausgestattet sein werden.

Wenn aber die seit Jahrhunderten versprochene technische Erlösung schon in universelle Entmachtung umzuschlagen droht, dann stellen sich zwei Fragen, die nach der Technologie, ihren strukturellen Implikationen und ihrer inhärenten Normativität sowie die nach der Restgröße des bürgerlichen Subjekts. Kant modellierte das bürgerliche autonome Subjekt als Inkarnation der Vernunft, sämtliche anderen Beigaben: Interessen, Emotionen, Macht und Ökonomie waren

¹ Vgl. Kapps These der Organprojektion, wonach sämtliche Technologien nichts anderes als Projektionen der Organe des Menschen seien: „Dieses Zustandekommen von Mechanismen nach organischem Vorbilde, sowie das Verständnis des Organismus mittels mechanischer Vorrichtungen, und überhaupt die Durchführung des als Organprojektion aufgestellten Principis für die, nur auf diesem Wege mögliche, Erreichung des Zieles der menschlichen Thätigkeit, ist der eigentliche Inhalt dieser Bogen.“ (Kapp 1877, VI), und seine Vorstellung, dass der Mensch sich in der Technologie selbst erkennen könne: „Von den ersten rohen Werkzeugen, geeignet die Kraft und Geschicklichkeit der Hand im Verbinden und Trennen materieller Stoffe zu steigern, bis zu dem mannigfaltigst ausgebildeten „System der Bedürfnisse“, wie es eine Weltausstellung gedrängt vorführt, sieht und erkennt der Mensch in all diesen Aussendungen, im Unterschiede von den unveränderten Naturobjecten, Gebild der Menschenhand, Thaten des Menschengeistes, den sowohl unbewusst findenden, wie bewusst erfundenen Menschen – Sich selbst.“ (Kapp 1877, 25) fanden in der frühen Medientheorie geneigte Abnehmer.

² „Der Mensch ist sozusagen ein Prothesengott geworden, recht großartig, wenn er alle seine Hilfsorgane anlegt, aber sie sind nicht mit ihm verwachsen und machen ihm noch gelegentlich viel zu schaffen. Er hat übrigens ein Recht, sich damit zu trösten, daß diese Entwicklung nicht gerade mit dem Jahr 1930 A.D. abgeschlossen sein wird. Ferne Zeiten werden neue, wahrscheinlich unvorstellbar große Fortschritte auf diesem Gebiete der Kultur mit sich bringen, die Gottähnlichkeit noch weiter steigern.“ (Freud 1930, 87)

³ Bekanntlich ist eine von McLuhans im Übrigen keineswegs kohärent zusammen passenden Definitionen jene „extensions of man“ – „...alle Medien als Ausweitungen unserer Person dienen dazu, uns neue umformende Einsicht und Bewußtheit zu geben.“ (McLuhan 1964, 78)-, die sich zwar an Kapps Konzept der Organprojektion anlehnt, jedoch keineswegs Selbsterkenntnis zum Ziel hat, sondern vornehmlich eine Erweiterung von Handlungsmacht.

bestenfalls störend. Das autonome Subjekt war also eine reine Repräsentation der Vernunft und seine Moral fiel naturgemäß kaum minder vernünftig aus: Sie war konsequent als Gesetz modelliert und damit war Verallgemeinerbarkeit gleich mit garantiert⁴. Da aber der kategorische Imperativ zugleich als Regel formuliert ist, nämlich als Regel der Universalisierbarkeit:

„handle so, als ob die Maxime deiner Handlung durch deinen Willen zum allgemeinen Naturgesetze werden sollte“ (Kant 1785, 51)

sollte er prinzipiell eigentlich auch programmierbar sein. Kant liefert damit zugleich eine recht gute Bedingung für eine Anerkennung als autonomes Subjekt, nämlich die Universalisierungsfähigkeit und damit das Vernunftgemäße seiner Entscheidungen.

3. Von der Quantifizierbarkeit des Menschen

Die bekannten Differenzierungsverfahren, die Mensch und Maschine voneinander scheiden sollten, operierten in utilitaristischer Manier: Sie waren ausschließlich output-orientiert. Was die jeweiligen Akteure antrieb, regulierte und schaltete war hingegen vollkommen uninteressant. Man erfährt etwas über Wahrnehmung und Identifizierung von Maschinen und Menschen, aber man erfährt nichts über die zugrundeliegenden Operationen. Solange der Rezipient im Turing-Test nur keinen Unterschied feststellen konnte, war die Art des Zustandekommens der Information weitgehend gleichgültig und konnte als menschlich zu den Akten gelegt werden.

Insofern macht es die utilitaristische Perspektive sehr viel einfacher als Kants kategorischer Rigorismus mit Turings Probe auf die Maschine (vgl. Turing 1950, 39 ff.) zurechtzukommen. Mills Leitidee des größten Glücks der größten Zahl⁵ hat den unbestreitbaren Vorzug der Quantifizierbarkeit. Dabei steht in der Mehrzahl der Entscheidungsfälle gar nicht Mills normatives Telos des allgemeinen Glücks zu Disposition. Dieses allgemeine Glück verfügt über einen spekulativen Überschuss, der seine Berechenbarkeit erheblich erschweren dürfte. Aber in der Regel ist das allgemeine Glück in seiner Absolutheit gar nicht gefragt, sondern eine relative Bestimmung reicht vollkommen aus und die zeichnet sich bei Mill durch eine ziemlich problemlose Berechenbarkeit aus.

“Fragt man mich nun, was ich meine, wenn ich von der unterschiedlichen Qualität von Freuden spreche, und was eine Freude [...] wertvoller als eine andere macht, so gibt es nur eine mögliche Antwort: von zwei Freuden ist diejenige wünschenswerter, die von allen oder nahezu allen, die beide erfahren haben [...] entschieden bevorzugt wird.” (Mill 1861, 15 f.)

⁴ Kant legt bekanntlich einen naturwissenschaftlichen und keinen normativen Gesetzesbegriff zugrunde und orientiert sich damit an dem Allgemeingültigkeitspostulat der Naturwissenschaften (vgl. Kant 1785, 51).

⁵ „... denn die Norm des Utilitarismus ist nicht das größte Glück des Handelnden selbst, sondern das größte Glück insgesamt.“ (Mill 1861, 20)

Derartige Abstimmungsverfahren, wie sie hier von Mill insinuiert werden, sind ziemlich umstandslos in Entscheidungsverfahren umzusetzen, ja sie lassen sich nicht nur antizipieren und dann als durch Abstimmung gewonnene normative Entscheidungsregeln in Handlungsroutinen autonomer Systeme implementieren, sie basieren eben auch in letzter Hinsicht auf Entscheidungen des autonomen Subjekts. Zwar braucht Mill entschieden mehr Personal als Kant⁶, um zu seinen normativen Entscheidungen zu gelangen. Allerdings ist das in Zeiten von Big Data auch das kein ernstliches Problem. Das Entscheidende ist, dass Mill die absoluten normativen Entscheidungen Kants auf beobachtbare Abstimmungsergebnisse heruntertransformiert und damit die Beherrschbarkeit durch Berechenbarkeit garantiert. Bei Kant wäre das zumindest nicht ganz so einfach. Zwar vergisst auch Mill bei seinen berechenbaren Entscheidungen nicht die Ansprüche der Viktorianischen Klassengesellschaft, die nicht unbedingt für ihre Liberalität und einen Mangel an Klassenbewusstsein bekannt war, zu wahren, indem er für die Teilnahme an einer Entscheidung die praktische Erfahrung der Folgen voraussetzt und damit den Kreis der Entscheidungsberechtigten auf ein ‚erträgliches‘ Maß zurechtstutzt und so die Klassenintegrität⁷ wahrt, aber gegenwärtig wäre das Ganze durchaus auch unter demokratisch akzeptableren Konditionen denkbar. Es gibt also kein prinzipielles Hindernis gegen ein utilitaristisches Entscheidungsverfahren.

Das Problem ist nur, dass utilitaristische Ansätze prinzipiell ex post steuern und das ist bei Technologien mit einem enormen Erfassungsradius und einer ebensolchen gesellschaftlichen Relevanz alles andere als unproblematisch. Vor allem ergibt sich bei komplexen Technologien die Schwierigkeit, dass eine Folgenabschätzung kaum mit hinreichender Zuverlässigkeit durchzuführen ist. Sowohl die Neuheit als auch der enorme Einzugsbereich digitaler Technologien erschweren daher eine normative Prognose und erhöhen damit zugleich das Risiko einer Steuerung ex post. Die Simplizität und prinzipielle Quantifizierbarkeit des utilitaristischen Ansatzes werden mithin mit dem prinzipiellen Zuspätkommen des Steuerungsimpulses erkaufte. Utilitaristische Steuerungslogiken müssen also die Reversibilität der fraglichen Handlungen voraussetzen, andernfalls handelte es sich um eine ziemlich riskante Experimentalanordnung.

Für die üblichen sozialen Prozesse und Interaktionen mag aufgrund einer einigermaßen zuverlässigen Prognose der möglichen gesellschaftlichen Risiken eine solche Steuerung ex post

⁶ Kant benötigte bekanntlich nur ein, quasi als paradigmatisch gesetztes Subjekt; Sozialsysteme hingegen mit ihren Ideologien, kulturellen Lasten und Traditionen waren für ihn vollkommen irrelevant. Moralisch gesehen störten sie eigentlich nur, was, da es bekanntlich um soziale Entscheidungsgründe geht, letztlich einen Widerspruch darstellt. Man hat es also bei der Kantschen Konstruktion mit einer Art moralischer Robinsonade zu tun, die sich ihren Freitag allenfalls als eine Art Spiegelung denkt.

⁷ Für eine solche politische Klugheit bzw. für einen solchen Opportunismus oder ‚Realismus‘ ist Mill ja bekannt. So scheut er sich auch nicht, den Sklavenhandeln, also einen Grundbestandteil der britischen Dreiecksökonomie, normativ zu legitimieren. (vgl. Mill 1861, 55)

noch mit einigermaßen überschaubaren Folgen verbunden sein und daher durchaus als akzeptabel erscheinen. Für neue Technologien hingegen gelten diese Routinen allenfalls in einem sehr viel eingeschränkteren Maße. Zwar lassen sich gerade auch für die Implementation neuer Technologien, insbesondere aber für die von Kommunikationstechnologien morphologische Strukturmuster – also gesellschaftliche Routinen – feststellen, jedoch ist vor allem bei Kommunikationsroutinen die Abschätzung möglicher Folgen äußerst vage.

Daher operieren Gesellschaften in diesem Fall auch mit einer Art Intervallschachtelung zwischen Apokalypse und Erlösung⁸ und können sich dann überlegen, welchen Einsatz sie gesellschaftlich letztlich riskieren wollen. Die Verengung und Justierung des Spektrums zwischen Apokalypse und Erlösung ist eine Angelegenheit, die nur mittels gesellschaftlicher Kontrolle einigermaßen zuverlässig stabilisiert werden kann. Gesellschaftliche Kontrolle und Steuerung aber funktioniert zunächst einmal normativ durch Errichtung eines durch Normen bewährten Cordon Sanitaire, in dem die Technik sich tummeln darf. Wenn aber bei neuen Technologien und vor allem bei generellen Umbrüchen technologischer Paradigmen wie der Digitalisierung die Abschätzung der gesellschaftlichen Folgen mit erheblicher Unzuverlässigkeit belastet ist, dann basiert die Festlegung eines solchen Cordon Sanitaire auf kaum mehr als einer bloßen Mutmaßung, einer Mutmaßung, die immerhin nicht weniger die Kommunikationsstrukturen von sozialen Systemen regulieren soll.

Eine Steuerung ex post muss daher mit einem nicht unerheblichen Schadensaufkommen rechnen. Zugleich wird deutlich, dass wir uns akut faktisch in eben dieser Situation eher erfolgloser nachgelagerter Steuerungsversuche befinden. Umgekehrt erscheinen nachgelagerte normative Verfahren immer dann einigermaßen unproblematisch zu sein, wenn der Umfang der technischen Innovation als halbwegs überschaubar erachtet wird oder zumindest als überschaubar dargestellt werden kann. Dabei geht es weniger um die faktische Wirkmächtigkeit von Technologien als um die symbolische Differenz, mit der sie kommuniziert werden. Sobald sich metaphorische Brücken bauen lassen und Plausibilisierungen qua Analogie zu verfangen scheinen, dann fällt die symbolische Differenz relativ gering aus und nachgelagerte Steuerungslogiken werden ziemlich

⁸ Technische Innovationen werden, sobald sie auch nur ansatzweise über gesellschaftliche oder kulturelle Relevanz verfügen, von den betroffenen Interessengruppen regelmäßig kontrovers diskutiert: Die Verfechter der Technologie treten üblicherweise mit der Vorhersage der Heilung aller aktuellen gesellschaftlichen Mängel also mit einem Erlösungsversprechen an, diejenigen, die sich von der neuen Technik bedroht sehen, prophezeien hingegen den kulturellen und sozialen Untergang und Modellieren die neue Technologie in Termini des Verlusts. Aus dieser Dialektik von Apokalypse und Erlösung entwickelt sich im Zuge der Implementation der neuen Technologie dann eine domestizierte Technologie, der spezifische kulturelle Beschränkungen auferlegt werden und die dann als solche enkulturiert wird.

problemlos akzeptiert⁹. Auf diesem Wege können Technologien quasi schleichend eingeführt und normativ plausibilisiert werden. Hierbei sorgt schlichte Gewöhnung und die Niedrigschwelligkeit des Verfahrens für ein ziemlich trügerisches Vertrauen in Technologien. Die nachgelagerte Steuerung eignet sich mithin vor allem für relativ geringe Innovationsniveaus, die Innovationsfrequenz ist demgegenüber von eher untergeordneter Bedeutung, weil der Prozess nahezu beliebig oft wiederholbar ist.

Das Verfahren der kumulativen normativen Enkulturation von Technologien eignet sich jedoch aus systematischen Gründen nicht bei gravierenden technologischen Umbrüchen oder technologischen Veränderungen mit hoher Innovationstiefe. Bei derartigen technologischen Paradigmenwechseln versagen Analogien als Argumentationshilfe, da deren Unangemessenheit unmittelbar einleuchtet. Das Neue wird allein negativ bestimmbar und gewinnt dementsprechend keine Gestalt. Die sozi-kulturellen und ökonomischen Folgen von solchen Paradigmenwechseln lassen sich ebenso wenig verlässlich bestimmen. Die Idee auch solche technologischen Paradigmenwechsel noch durch nachgelagerte Steuerung in den Griff bekommen zu können, operiert vergleichbar jener ebenso prominenten wie eklatant gescheiterten Devise des Kölschen Klüngels „Es ist noch immer alles gut gegangen“. Derart seliger Optimismus erscheint angesichts des sozio-kulturellen und ökonomischen Impacts gravierender technologischer Umbrüche zusehends naiv.

4. Vom Turing-Test zur Prinzipienfrage oder: Zurück zur Autonomie

Kants Vorstellung vom Subjekt geht jedoch sowohl über die utilitaristische Quantifizierungsstrategie als auch über den Turing-Test weit hinaus, indem er nicht nur die einigermaßen beliebige Wahrnehmbarkeit als Subjekt zur Bedingung macht, sondern den kategorischen Imperativ zugleich als eine Art Testfrage formuliert, an der sich mögliche Differenzen und damit die Unterscheidung von Prothesen und Subjekten erweisen könnten.

Der deontologischen Ansatz Kants geht vollkommen anders vor, indem er nicht nur von einer a priori ansetzenden Steuerung, also einer vorgelagerten normativen Konstruktion, ausgeht,

⁹ Genau das aber ist im Fall autonomer Fahrzeuge das Problem: Assistenzsysteme, wie kompetent sie auch immer sein mögen, werden als relative und daher überschaubare technische Erweiterungen verstanden. Sie stellen insofern auch keine prinzipiellen Fragen, wie weitreichend diese Systeme auch immer angelegt sein mögen. Die behauptete Autonomie hingegen stellt die Prinzipienfrage und steigert die Folgen notwendig ins Unermessliche. Wenn autonome Systeme Fehler machen oder versagen, dann handelt es sich um etwas vollkommen anderes, als wenn ein Assistenzsystem eine Fehlfunktion aufweist, selbst wenn der Schaden identisch sein sollte: Im ersten Fall stellt sich die Prinzipienfrage und die verlangt in jedem Fall nach kategorischen Antworten, im zweiten Fall stellt sich einerseits die Frage nach technischen Lösungen, die den Fehler künftig verhindern könnten, und andererseits die Frage nach der Verantwortung für den technischen Fehler. Die gesellschaftliche Reaktion unterscheidet sich dann demgemäß auch zwischen Verbot und Verbesserung.

sondern indem er den automatisierbaren Akteur selbst zu bestimmen sucht. Dabei wird zugleich deutlich, dass Kant in seiner rigorosen Nüchternheit eigentlich eine Maschinenethik¹⁰ geschrieben hat. Der kategorische Imperativ Kants diskriminiert bekanntlich anhand der Universalisierbarkeit von Entscheidungen. Auf dieser Basis lässt sich dann jedoch auch zwischen autonomen und nicht-autonomen Systemen unterscheiden, indem eben diese Universalisierungsfähigkeit zum Differenzkriterium gemacht wird. Der Kant-Test, der dementsprechend über die Autonomie von was auch immer entscheidet, lautete dann: Erst wenn eine Person oder ein System in der Lage ist, seine Entscheidungen zu universalisieren, handelt es sich um ein Subjekt oder um ein autonomes System. Der Kant-Test, der quasi die Vernunftfähigkeit des Subjekts oder des automatischen Systems beurteilt, unterscheidet zwischen Systemen, die autonom, also ohne Kontrolle agieren können, und solchen, die in jedem Fall einer Kontrolle durch ein autonomes Subjekt bedürfen. Systeme, die ohne prinzipielle Vorabkontrolle auskommen, sind entweder solche, die über überschaubare Innovationsniveaus verfügen und sich daher von bereits enkulturierten Technologien nicht wesentlich unterscheiden, oder aber Technologien, die ohnehin über explizite Kontrollen verfügen. Solche Systeme sind entweder prinzipiell nicht subjektfähig oder aber sie werden ohnehin Subjekten unterstellt. Verantwortung wird in einem solchen Fall dem kontrollierenden Subjekt oder aber dem Produzenten der Technologie zugewiesen.

Für solche abhängigen Systeme gilt das technische Sklavenmodell: Technologien werden Subjekten als vollkommen abhängige Unterstützungsinstrumente zugeordnet und all ihre Operationen sind letztlich von den Entscheidungen dieses Subjekts abhängig. Das heißt, dass diese Systeme jederzeit durch Subjekte abschaltbar sein müssen und das Subjekt prinzipiell die Verantwortung für den Einsatz der Systeme trägt.

Alle anderen Systeme, die also nicht irgendwelchen Modi der Nachkontrolle unterliegen sollen, müssen den Kantschen Subjekt-Test bestehen, d.h., sie müssen eine Selbststeuerungsfähigkeit auf Universalisierungsniveau unter Beweis stellen. Gleichzeitig würde mit einem in diesem Sinne erweiterten Turing-Test aber auch die Perspektive gewechselt: Man würde von der Rolle des unbeteiligten Betrachters von Handlungen und Handlungsfolgen zu der Analyse und Steuerung der Motivation von Handelnden übergehen. Die Anerkennung von Systemen als

¹⁰ Das scheint schon Schiller geahnt zu haben, wenn er das ‚Technische‘ bzw. die ‚technische Form‘ der Vernunft bei Kant moniert: „Über diejenigen Ideen, welche in dem praktischen Teil des Kantischen Systems die herrschenden sind, sind nur die Philosophen entzweit, aber die Menschen, ich getraue mir, es zu beweisen, von jeher einig gewesen. Man befreie sie von ihrer technischen Form und sie werden als die verjährten Aussprüche der gemeinen Vernunft und als Tatsachen des moralischen Instinktes erscheinen, den die weise Natur dem Menschen zum Vormund setzte, bis die helle Einsicht ihn mündig macht. Aber eben diese technische Form, welche die Wahrheit dem Verstand versichtbart, verbirgt sie wieder dem Gefühl; ...“ (Schiller 1795, 1. Br., 4) Schiller geht es bekanntlich darum, „den Gesetzen der Vernunft ihre moralische Nötigung“ (Schiller 1795, 14. Br., 58) zu nehmen und d.h. die Rigidität des Gesetzes ‚erträglich‘ zu machen. Dieses Problem stellt sich bei autonomen Systemen jedoch gerade nicht.

Subjekte griffe quasi a priori in die Konstruktion des Systems ein, indem sie eine Implementation einer Universalisierungsroutine in jeden Entscheidungsprozess erforderlich machte, die diesem System zu Grunde läge.

Die Universalisierungsfähigkeit stellt zugleich eine relativ große Hürde dar, da sie prinzipiell auch nicht durch irgendwelche Quantifizierungen zu substituieren ist. Denn selbst, wenn man Schadensminimierung als verallgemeinerungsfähiges Ziel akzeptierte, funktionierte das allenfalls noch bei der Schädigung von Objekten, bei der Schädigung von menschlichen Individuen jedoch wäre Quantifizierung systematisch unangemessen. Insofern dürften, folgt man dem Kantschen Denkmodell, Maschinen, die prinzipiell nicht in der Lage sind, menschliche Individuen zu schädigen und die gleichzeitig nicht über die erforderlichen Universalisierungsfähigkeiten verfügen, nicht als autonome Systeme in den Verkehr gebracht werden. Das macht zugleich die vergleichsweise hohe Hürde deutlich, die beim Übergang von teilautonomen zu autonomen Systemen besteht.

5. Verfahren normativer Kontrolle

Selbst prinzipiell universalisierungsfähige Subjekte unterliegen der Kontrolle. Kaum ein gesellschaftliches System traut der Selbststeuerungsfähigkeit von Subjekten wirklich, sondern sie operieren allesamt mit mehr oder minder ausgefeilten Polizei- und Justizsystemen, die dem nachlässigen Subjekt auf die Sprünge helfen sollen. Je größer die sozialen Differenzen in Sozialsystemen sind, umso größer wird zwangsläufig die zur Aufrechterhaltung des Ungleichgewichts erforderliche Kontrolle. Der Raum der Autonomie von souveränen Subjekten ist daher stets mehr oder minder deutlich eingeschränkt. Die Souveränität von autonomen Subjekten wird so durch soziale Verfahren der Kontrolle eingehegt, was umgekehrt bedeutet, dass die Autonomie des Subjekts nichts anderes als ein regulatives Prinzip darstellt. Bei solchen regulativen Prinzipien wird davon ausgegangen, dass sie in der Praxis nie allein oder ausschließlich wirksam werden können, sondern stets von anderen Interessen überlagert zu werden drohen. Derartige Interferenzen fallen aber bei autonomen Systemen systematisch weg oder aber sie werden durch technische Störungen bzw. ökonomische Interessen abgelöst. Insofern verhalten sich autonome Systeme zu regulativen Prinzipien wie der Idee der Universalisierungsfähigkeit von Entscheidungen zwar grundlegend anders als autonome Subjekte, jedoch nicht notwendig auch optimaler.

Wenn aber allen anderen Systemen, die nicht die Universalisierungsfähigkeit in ihren Entscheidungsroutinen implementiert haben, normative Souveränität und damit selbst die ziemlich störungsanfällige Selbststeuerung durch regulative Prinzipien fehlt, dann müssen sie daher zwangsläufig vollständig gesellschaftlicher Kontrolle unterworfen werden. Die Modi einer solchen

gesellschaftlichen Kontrolle unterscheiden sich dabei durchaus: Das enkulturierte Standardmodell für die gesellschaftliche Steuerung von Technologien besteht in einem zweistufigen Steuerungsmodell: Selbstkontrolle durch das das Artefakt produzierende oder betreibende Subjekt sowie einen institutionalisierten Kontrollmechanismus, der in der Regel von öffentlich-rechtlichen oder aber staatlichen Institutionen organisiert, durchgeführt oder beauftragt wird. Das Ganze wird dann im Streitfall noch einer juristischen Nachkontrolle unterzogen, so dass man es letztlich mit einem dreistufigen Kontrollmodell zu tun hat. Die Medienkontrolle funktioniert in der Regel nach solchen dreistufigen Steuerungslogiken.

Das aber würde bedeuten, dass sämtliche teilautonomen Systeme solchen wenigstens einigermaßen gesellschaftlich anerkannten Steuerungslogiken unterworfen werden müssten und der Streit, der gegenwärtig um solche technischen Systeme herrscht, letztlich vor allem damit zu tun hat, dass derartige Steuerungsmechanismen noch nicht etabliert sind. Seit John Perry Barlows einigermaßen naiver Unabhängigkeitserklärung des Cyberspace von 1996 wird im Internet mit dem regulativen Prinzip der Autonomie und der Selbstregulierung operiert, was letztlich darauf abzielt, möglichst alle Formen gesellschaftlicher Kontrolle und sekundärer Steuerung auszuschließen. Das hinzu kommende eklatante Versagen juristischer Steuerungsstrategien generierte einen normativ weitgehend unregulierten Raum, in dem keineswegs individuelle Kreativität wucherte, sondern ökonomische Interessen die regulativen Prinzipien der Selbstkontrolle und Autonomie behänd außer Kraft setzten. Die Diskussion um die Geschäftsmodelle von sozialen Netzwerken und Suchmaschinen hängt nicht zuletzt mit der eklatanten Unwirksamkeit der Selbstregulation und der Verdrängung regulativer Prinzipien durch ökonomische Interessen Einzelner zusammen.

Derartige Konflikte sind für den Enkulturationsprozess von neuen Medientechnologien geradezu typisch. So wurden parallel zu ihrer Etablierung sowohl der Buchmarkt als auch der Rundfunk Regulationsstrategien vom Urheberrecht bis zur staatlichen Erteilung und Kontrolle von Funklizenzen unterworfen. Offenkundig ist die Enkulturation von Internet und autonomen Systemen noch nicht so weit fortgeschritten, dass solche Regulationsstrategien entwickelt und durchgesetzt worden wären. Dass die Privilegierung ökonomischer Interessen Einzelner zu normativen Fehlentwicklungen führt, dürfte vermutlich den meisten Akteuren zwischenzeitlich deutlich geworden sein.

6. Haltet die Betrüger! Oder die Strategien der Moralisierung

Social Bots bestehen, so viel scheint einigermaßen sicher zu sein, den Kant-Test nicht und d.h., sie sind in jedem Fall einer wie auch immer gearteten sozialer Kontrolle zu unterstellen. Nur ist Voraussetzung ihres Erfolgs, dass sie es verstehen, sich der Kontrolle zu entziehen. Dabei beruht das Prinzip ihrer Wirkung auf der Imitation autonomer Subjekte ohne selbst autonom zu

sein. Vorgebliche Souveränität, hinter der sich nicht offen gelegte Interessen verbergen, wird, sofern sie ihrer Vorgeblichkeit überführt wird, als ein Phänomen der Täuschung begriffen und so wird der Diskurs über Social Bots vornehmlich von Betrugsnarrativen regiert.

Dieses Betrugsmotiv ahnte bereits Turing, denn er benutzte es bei der Klärung der Frage „Können Maschinen denken?“ (Turing 1950, 39). Er transformierte nämlich die Frage in ein Imitationsspiel¹¹ und Imitation ist bekanntlich nichts anderes als ein Spiel mit Täuschungen. Die Logik dieses Imitationsspiels besteht darin, dass ein nicht entdeckter Unterschied¹² als Gleichheit, zumindest aber als Gleichwertigkeit akzeptiert wird. Das Wirkungsprinzip von Social Bots basiert auf solch einer angenommenen Gleichwertigkeit und auf den entsprechend unentdeckten Unterschieden. Social Bots haben den Turing-Test, der in diesem Imitationsspiel besteht, immer schon bestanden, wenigstens wenn sie nicht entdeckt wurden. Umgekehrt bestände, wenn Social Bots entdeckt würden, ja auch keine Gefahr mehr: die Täuschung wurde entlarvt und die Angelegenheit stellt sich mithin als harmlos heraus. Das Bestehen des Turing-Tests ist also dasjenige, was Social Bots letztlich gefährlich macht.

Täuschungs- und Manipulationsvorwürfe, Überwältigungsvorstellungen und Ängste vor einem drohenden Kontrollverlust befeuern dabei die moralischen Diskurse und diese moralischen Diskurse sind letztlich nur Ausdruck eines Verlusts an Unterscheidungskraft. Es geht bei dem Ganzen um die Aufrechterhaltung einer Grenze, die an sich nicht mehr feststellbar ist. Moralisierung tritt immer dann auf den Plan, wenn der Sachgrund für eine Unterscheidung weggefallen ist. Die Diskussion kann also nur noch moralisch geführt werden, denn die Sache selbst wird eigentlich bereits verloren gegeben. Allerdings verfügt dieser an sich so unmittelbar einleuchtend erscheinende Diskurs selbst über mindestens zwei blinde Flecken und das lässt ihn selbst unweigerlich zu einem strukturellen Täuschungsdiskurs werden.

Die echauffierten Brandreden der Hüter der traditionellen Diskurse gehen davon aus, dass Social Bots die Diskurse, Daten und Informationen, die sie lancieren, lostreten und zur Verfügung stellen, unrechtmäßig verfälschen, verzerren oder aber konterkarieren. Eingeübte Kommunikationspraktiken verfangen nicht mehr in der gewohnten Weise oder sie sind schlicht nicht mehr verlässlich. Der Blick auf jene ach so berechenbaren Rezipienten, und an eben dieser Berechenbarkeit haben persuasive Kommunikationsstrategen seit langem mit aller Kraft gearbeitet,

¹¹ Turing formuliert die Frage „Kann eine Maschine denken?“ um in ein Imitationsspiel: „Die neue Form des Problems läßt sich als Spiel beschreiben, das wir „Imitationsspiel“ nennen wollen. Wir betrachten drei Spieler, einen Mann (A), eine Frau (B) und einen männlichen oder weiblichen Fragesteller (C). Der Fragesteller sei allein in einem Raum. Das Ziel des Fragestellers ist es zu entscheiden, welche der beiden anderen Personen der Mann bzw. die Frau ist. [...] Wir stellen nun die Frage: Was passiert, wenn eine Maschine die Rolle von A in diesem Spiel übernimmt? [...] Diese Fragen treten an die Stelle unserer ursprünglichen: „Können Maschinen denken?““ (Turing 1950, 39f.)

¹² In Turings Fall der von Mensch und Maschine.

droht aufgrund der unkalkulierbaren Intervention diverser Social Bots eine Art grauen Star zu bekommen und zusehends wieder schleierhaft zu werden. Der gerade eben erst nahezu vollständig kalkulierbar gewordene Rezipient droht in einer Wolke von Bots und traditionellen Normalbürgern wieder in geradezu traditioneller Unsicherheit zu verschwinden. Social Bots produzieren weitgehend unbeabsichtigt einen enormen Kollateralschaden, denn sie greifen völlig unabhängig von dem Kalkül ihrer Konstrukteure zunächst einmal das Kapital der Datensammler und der Eigentümer der digitalen Netzwerke an. Es wird mithin ein recht erfolgreiches Geschäftsmodell destabilisiert und das gibt natürlich Ärger. Die Akteure mit Definitionsmacht in den Diskursen und die Datensammler wehren sich und sie wehren sich, solange sie juristisch an die Konstrukteure der durch die Netzwerke nomadisierenden Bots nicht herankommen können, durch moralische Verurteilung. Und bei all diesen Brandreden geben sie vor, alles nur im Dienste jener vom Radar wieder zu verschwinden drohenden Rezipienten zu tun.

Nun, wenn eines in der aufgeregten Diskussion um Social Bots sicher ist, dann das, dass der moralisierende Diskurs der Akteure mit Definitionsmacht im Wesentlichen selbst auf einer Täuschung beruht. Es ist nämlich durchaus im Interesse des Rezipienten, wenn seine ubiquitäre Erfassbarkeit zusehends Risse bekommt. Und die Erkenntnis, dass der Rezipient den Informationen der Akteure nicht mehr trauen kann, ist ein durchaus probates Erziehungsmittel¹³, das Resultate zeitigen könnte, für die die kritische Theorie noch vor nicht allzu langer Zeit erhebliche theoretische Klimmzüge machen musste. Misstrauen gegenüber den Protagonisten der Persuasion ist also angebracht und es ist normativ gesehen ziemlich egal, ob diese sich echauffieren oder nicht. Der ganze moralisierende Diskurs sucht also den Rezipienten vor den Karren von Datenverkäufern, Meinungsmachern, Trendsettern und politischen Spindoctoren zu spannen und es ist zweifellos gut, wenn ein solches Vorhaben scheitert. Von daher haben Social Bots eine ganz sicher nicht intendierte subversive Nebenwirkung¹⁴, deren Effekt vor allem in der Streuung von Misstrauen besteht.

¹³ So wittert die Medienpädagogik auch hier wieder – wie bei nahezu jedem Medienskandal – Morgenluft und bietet Media Literacy als probates Gegenmittel für die Täuschungsstrategien von Social Bots und disruptiver Kommunikation: „This development strikes at the heart of democratic societies in the western world, which rely on communication in regards to democratic processes, particularly general elections and political debates. It is against this background that social bots and other forms and manifestations of disruptive communications, like fake news, have to be understood.“ (Janetzko 2017, 65 f.) Die Medienpädagogik bietet sich hier ein weiteres Mal als Reparaturbetrieb für die Fehlfunktionen des Mediensystems an, trägt jedoch nichts zur grundlegenden Lösung der Widersprüche bei.

¹⁴ Auf solche nicht-intendierte Nebeneffekte setzt dann auch Leistert mit seinem sozialromantisch eingefärbten Piraterieszenario: „Dieser vermeintlich parasitäre Charakter von Piraterie verkennt jedoch dessen produktive Seite: Piraterie hat über unterschiedliche Märkte verteilt höchst kreative Formen der Distribution erfunden. Und dies ist die entscheidende Parallele, die diese Piraterie mit der Art verbindet, wie Social-Bots Daten weiterverteilen. Der Unterschied ist jedoch der Modus der Distribution. Daten, die Bots gesammelt haben, können in höchst unterschiedliche Richtungen verteilt werden, vom Kreditkartenbetrug bis zum Hacken von Websites, von

Manipulationsvorwürfe im Zusammenhang von Medien sind im Übrigen alles andere als neu, und der Verdacht war immer schon so hinreichend begründet, dass gar nicht erst irgendwelche Verschwörungsszenarien bemüht werden mussten. Die Akteure mit Definitionsmacht innerhalb des Mediensystems, und das schließt eben gerade auch die digitalen Medien ein, blicken auf eine lange Tradition des Kampfes gegen eben jenen Manipulationsverdacht zurück: Das von Journalistenverbänden und Verlegern gleichermaßen fetischisierte Objektivitätsideal¹⁵, das nicht nur zwischen Journalismus und PR trennen helfen soll, sondern den Rechtsgrund gleich für ganze Serien von Privilegien vom Zeugnisverweigerungsrecht bis hin zu Neuwagenrabatten bildet, soll eben auch den Manipulationsverdacht abwehren.

Dass ausgerechnet genau dieser Verdacht nun gegen die die Definitionsmacht zu erschüttern drohenden Social Bots gerichtet wird, ist mehr als eine bloße Reflexreaktion. Er setzt nämlich voraus, dass die Akteure selbst prinzipiell gute und das hieße universalisierungsfähige Interessen vertreten. Die Identifikation betrügerischer Social Bots exkulpiert automatisch ausgerechnet diejenigen Akteure, denen es bis dato letztlich nicht nachhaltig gelungen ist, sich selbst des Manipulationsverdachts zu erwehren, also diejenigen, deren Interesse in der möglichst vollständigen Erfassung von Konsumenten besteht. Dass auf diesem Wege ausgerechnet die Besitzer von sozialen Netzwerken und Datensammlungen moralischen Dispens erhalten und auf einmal auf die Seite der bedingungslos Guten wechseln dürfen, kann nur als historische Ironie des Digitalisierungsprozesses gewertet werden.

Würde mit Social Bots nicht gleichzeitig das Geschäftsmodell sozialer Netzwerke¹⁶ bedroht, könnte man auf die Idee kommen, dass den Akteuren mit kultureller Definitionsmacht und mit der Verfügungsgewalt über Datensammlungen nichts Besseres als Social Bots hätte passieren können. Normalerweise kostet der bedingungslose Wechsel auf die Seite der Guten zumindest

erstaunlich passenden Werbeeinblendungen oder Spam bis zum Identitätsdiebstahl. Jedoch, ..., grundsätzlich sind sie für interessierte Käufer jeglicher Couleur verfügbar, exakt wie die sogenannten ‚Medien der Piraten‘.“ (Leistert 2017, 227) „In diesem Sinne ist Piraterie vielmehr als komplementär denn als parasitär zu verstehen, ...“ (Leistert 2017, 228) „Insofern ist es durchaus möglich, Social Bots jenseits moralischer oder legalistischer Überlegungen in einer der Datenakkumulation verschriebenen kapitalistischen Ökonomie, die selber die Privatsphäre und informatische Selbstbestimmung strukturell angreift, zu situieren.“ (Leistert 2017, 228)

¹⁵ Aus dieser Perspektive wird ein eben auch ein Großteil der Fake News Debatte und der moralisierenden Diskussion um Social Bots geführt, so dass es sich hierbei auch um eine der so häufigen Auseinandersetzung um mediale Definitionsmacht geht. Der traditionelle Journalismus sieht sich zunehmend durch Social Media und automatisierte Textproduktion bedroht und versucht normativ nachzusteuern, um seine privilegierte Stellung zu erhalten.

¹⁶ „Social Bots stellen langfristig eine Bedrohung für das Geschäftsmodell von sozialen Netzwerken dar. Ein Teil der Nutzer könnte sich abwenden, weil sie das Vertrauen in die Echtheit der Beiträge verlieren. Investoren verlieren das Interesse, weil sich die Plattformen durch Werbeeinnahmen oder dem Verkauf von Nutzerdaten finanzieren, aber nur echte Menschen Kaufentscheidungen treffen.“ (Kind, Bovenschulte, Ehrenberg-Silies u.a. 2017, 5)

normative Unterwerfung. Dazu wurden die Sozialen Netzwerke noch nicht einmal genötigt, stattdessen fallen durch die Erosion des Geschäftsmodells nun ökonomische Kosten¹⁷ an.

Das gilt letztlich auch für das Geschäftsmodell der empirischen Sozialwissenschaften, sofern sie versuchen, soziale Experimente unter Verwendung von Social Bots als neue Chance sozialwissenschaftlicher Forschung zu verkaufen. Social Bots drohen sozialwissenschaftliche Analysen öffentlicher Kommunikation allein durch schlichte Verunsicherung zu kontaminieren und sie konterkarieren damit eben jene Vorstellung, es ließe sich mithilfe von Social Bots ein unverstellter Zugang zu öffentlichen Kommunikationsstrukturen gewinnen. Zugleich droht der harsch moralisierende Diskurs zurückzuschlagen: Täuschung als Instrument einer wie auch immer vermittelt der ‚Wahrheit‘ verpflichteten Wissenschaft scheint zumindest das Potential zu haben, sich selbst zu demontieren. Wenigstens aber dürfte die Legitimität einer wesentlich mit Täuschungen operierenden Wissenschaft in Frage stehen. Dass der Einfluss von Social Bots dann zudem noch als vergleichsweise niedrig veranschlagt wird, nimmt vielleicht ein wenig Dampf aus der Diskussion, ändert an der Dialektik von Apokalypse und Euphorie jedoch kaum etwas.

7. Der ewige Kampf um Kontrolle

Insofern drohen Social Bots stets das System zu unterlaufen, von dem sie eigentlich zu profitieren gedachten. Täuschung funktioniert nur solange, solange sie nicht entdeckt oder zumindest nicht erwartet wird. Die Erwartung der Täuschung jedoch lässt das kommunikative System, dem die Social Bots parasitär verpflichtet sind, implodieren. Das Kalkül hinter Social Bots ist daher in jedem Fall ein Kalkül auf Zeit, faktisch sogar auf ziemlich kurze Zeit. Und dieser begrenzte Zeitraum, in dem überhaupt Social Bots ihre ohnehin vergleichsweise beschränkte Wirksamkeit¹⁸ entfalten, ist das historische Intervall, in dem der Kampf um die Definitionsmacht¹⁹ über neue Medien und Kommunikationskulturen ausgefochten wird.

¹⁷ Eine ebenso de-moralisierende wie ganzheitliche Perspektive nimmt Leistert ein: „Dieser Beitrag nimmt deshalb einen anderen Weg, und versucht Bots aus der Perspektive ihres Evironments zu untersuchen, als Teil und Element einer medien-technischen Umgebung. Gleichzeitig ist es notwendig, Bots ins Verhältnis zur Logik des zeitgenössischen kapitalistischen Imperativs der Datenextraktion und dem Kolonisieren jeglicher, noch so unbedeutender Äußerungen von Benutzern auf den Plattformen zu setzen.“ (Leistert 2017, 215)

¹⁸ „Es gibt lediglich eine begrenzte Anzahl prominenter Beispiele der Einflussnahme durch Social Bots, auf die sowohl in der Presse als auch in wissenschaftlichen Artikeln immer wieder Bezug genommen wird. Der in den Artikeln beschriebene Wirkungsraum ist an erster Stelle Twitter und schon seltener Facebook. Das Ausmaß der tatsächlichen Einflussnahme ist allerdings noch kaum belegt.“ (Kind, Bovenschulte, Ehrenberg-Silies u.a. 2017, 5)

¹⁹ „Genau an dieser Stelle wird offenbar, warum der Einsatz von Social Bots durch Parteien verfassungsrechtlich bedenklich ist: Er missachtet das demokratische Willensbildungsprinzip des „Von-unten-nach-oben“. Mit Hilfe von Social Bots wird in der sozial-medialen Öffentlichkeit zwar der Eindruck erweckt, dass eine Meinung von vielen individuellen Bürgern stammt. In Wahrheit jedoch hat sie ihren einzigen Ausgangspunkt in einer Parteizentrale. Sie ist eine von oben in die Bot-Profile diktierte Meinung.“ (Gasser, Kraatz 2017) – Hier wird deutlich, dass das Öffentlichkeitsmodell des traditionellen Journalismus und seine Logik der politischen

Diese normativen Auseinandersetzungen sind aber immer schon Kämpfe um die Definitionsmacht in und die Kontrolle über Medien gewesen. Und genau das ist eben auch bei Social Bots und autonomen Systemen der Fall: Es geht um die Entwicklung und Durchsetzung gesellschaftlich akzeptierter Strategien der Kontrolle. Das normative Risiko von Social Bots besteht nicht so sehr in dem Schaden, den sie unmittelbar also durch Täuschung anrichten, sondern in den gesellschaftlichen Kontrolldefiziten über die sie hinwegtäuschen.

Dabei stellen Social Bots und autonome Systeme die gesellschaftliche Kontrolle noch nicht einmal vor besonders große Herausforderungen: Denn bereits an den analogen Medien sind Kontrollstrategien entwickelt und erprobt worden, die sich durchaus auch auf soziale Netzwerke und autonome oder teilautonome Systeme beziehen lassen. Die gegenwärtig einigermaßen aufgeregte und empörte Diskussion dürfte dabei nur die Begleitmusik für den Übergang von privaten, liberal kapitalistisch organisierten Kommunikationssystemen hin zu Systemen, die einer wie auch immer gearteten gesellschaftlichen Kontrolle unterworfen sind, darstellen.

Also, weit und breit findet sich nichts Besonderes und zudem ist die Angelegenheit vorübergehender Natur. Zu beobachten bleibt jedoch, wer mit welchen Interessen sich an der Auseinandersetzung um die Definition neuer Medien und autonomer Systeme beteiligen wird, aber auch dabei wird es vermutlich kaum Überraschungen geben.

Willensbildung sich durch Social Bots unterlaufen sieht. Es handelt sich mithin eben auch um eine Auseinandersetzung um die politische Definitionsmacht zwischen den traditionellen Akteuren des politischen Journalismus und den hinter Social Bots stehenden Akteuren.

8. Literaturverzeichnis

- Freud, Sigmund (1930): Das Unbehagen in der Kultur. Frankfurt a.M.= S. Fischer, 379.-408. Tsd. 1970.
- Gasser, Lucas;
- Kraatz, Matthias (2017): Social Bots: Wegbereiter der Maschinokratie. Verfassungsblog.de/social-bots-wegbereiter-der-maschinokratie/ Mo 16 Jan 2017.
- Janetzko, Dietmar (2017): Social Bots and Fake News as (not) seen from the Viewpoint of Digital Education Frameworks. In: MedienPädagogik. (5. Juli), 61–80. <<http://www.medienpaed.com/article/view/453>>. Date accessed: 06 Juni 2018.
- Kant, Immanuel (1785): Grundlegung zur Metaphysik der Sitten, Werkausgabe Bd. VII, hrsg.v. Weischedel, Wilhelm, 2. Aufl., Frankfurt a. M.=Suhrkamp 1977.
- Kant, Immanuel (1788): Kritik der praktischen Vernunft, Werkausgabe Bd. VII, hrsg.v. Weischedel, Wilhelm, 2. Aufl., Frankfurt a. M.=Suhrkamp 1977.
- Kapp, Ernst (1877): Grundlinien einer Philosophie der Technik. Zur Entstehungsgeschichte der Cultur aus neuen Gesichtspunkten. Braunschweig=Georg Westermann 1877.
- Kind, Sonja; Bovenschulte, Marc; Ehrenberg-Silies, Simone; Jetzke, Tobias;
- Weide, Sebastian (2017): Social Bots. Thesenpapier zum öffentlichen Fachgespräch »Social Bots – Diskussion und Validierung von Zwischenergebnissen« am 26. Januar 2017 im Deutschen Bundestag. o.O.=VDI/VDE Innovation + Technik GmbH 2017.
- Krafft, Peter M.; Macey, Michael;
- Pentland, Alex “Sandy” (2016): Bots as Virtual Confederates: Design and Ethics. 2 Nov 2016, 1 (1–8) <https://arxiv.org/pdf/1611.00447.pdf>. Date accessed: 06 Juni 2018.
- Leistert, Oliver (2017): Social Bots als algorithmische Piraten und als Boten einer techno-environmentalen Handlungskraft. In: Seyfert, Robert;

- Roberge, Jonathan [Hrsg.]: Algorithmenkulturen. Über die rechnerische Konstruktion der Wirklichkeit. Bielefeld=transcript 2017, 215-234.
- Mill, John Stuart (1861): Der Utilitarismus, Stuttgart=Reclam 1985.
- Murthy, Dhiraj; Powell, Alison B.; Tinati, Ramine; Anstead, Nick; Carr, Leslie; Halford, Susan J.; Weal, Mark (2016): Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital. In: International Journal of Communication 10(2016), 4952–4971.
- Schiller, Friedrich (1795): Über die ästhetische Erziehung des Menschen, in einer Reihe von Briefen. Stuttgart= Reclam 1979.
- Turing, Alan M. (1950): Kann eine Maschine denken? In: Zimmerli, Walther Chr.; Wolf, Stefan [Hrsg.]: Künstliche Intelligenz. Philosophische Probleme. Stuttgart=Reclam 1994, S. 39-78.