

Governance-Herausforderung

Künstliche Intelligenz

Volker von Prittwitz¹

1. Technologiepolitische Verantwortung

Die Menschheit prägt zunehmend die Erde (*Anthropozän*). Dies mit positiven Wohlfahrts-Effekten, aber auch wachsenden Risiken. So hat die Weltbevölkerung bei verbesserten hygienischen und wirtschaftlichen Rahmenbedingungen stark zugenommen. Andererseits entstehen weitflächig gefährliche Abfälle, so nuklearer Restmüll und Plastikmüll. Die vorhandenen Massenvernichtungs-Waffen reichen mehrfach dazu aus, die Menschheit auszulöschen. Und mit der anhaltenden Emission diverser Treibhausgase wächst das Risiko eines sich selbst verstärkenden, nicht mehr beherrschbaren Klimawandels.

Autonom lernfähige technische Systeme (*Künstliche Intelligenz/KI*) bilden eine neue existentielle Herausforderung: Mit dieser Technologie kann sich die allgemeine Lebensqualität weiter erheblich erhöhen; es entstehen aber auch neue Betrugs-, Macht- und Vernichtungspotenziale. Vor allem aber: Überspringt Künstliche Intelligenz die Entwicklungsstufe zu Selbstbewusstsein und Eigenwillen, so hat sie das Potenzial, die Menschheit zu unterwerfen und auszulöschen.

Der hieraus erwachsenden enormen technologiepolitischen Verantwortung tragen weder eschatologische Denkmuster Rechnung noch Zocker-Mottos wie, *Wer nicht wagt, der nicht gewinnt*, *Nachher ist man immer schlauer*, oder, *Die Märkte haben es so gewollt*. Vielmehr müssen Potenziale und Risiken Künstlicher Intelligenz bestmöglich abgeschätzt und effektive Koordinationsformen entwickelt und realisiert werden - eine zunächst interdisziplinär wissenschaftliche, letztlich aber politische Herausforderung.

¹ Resumierende Thesen zum Seminar, *Künstliche Intelligenz als politische Herausforderung*, an der Freien Universität Berlin im WS 2017/2018. Für anregende Diskussionen danke ich den Seminar-Teilnehmer/innen.

2. Beurteilungsrahmen: Interaktions-Logiken

Menschen interagieren nach bestimmten Logiken, so der Freund-Feind-Logik, der Machtlogik, der Interessenlogik, der Logik gemeinsam anerkannten Rechts und/oder der Logik integraler Vielfalt.² Diese Interaktions-Logiken unterscheiden sich hinsichtlich ihrer Zivilität in charakteristischer Weise:

- Nach der *Freund-Feind-Logik* werden *Feinde* unterdrückt oder/und gewalttätig verfolgt. Gilt diese Logik beidseitig, entsteht Krieg - genau das Gegenteil von Zivilität.
- Machtlogik kann borniert und gewaltsam operieren und ist insofern prinzipiell unzivil, hat aber auch die Option des guten Königs.
- Interessenlogisch wird differenziert und flexibel vorgegangen - einseitig, unter Umständen betrügerisch, oder in wechselseitig verschränktem Interesse.
- Erkennen die Beteiligten gemeinsam faire Regeln an, sind sie prinzipiell geschützt, gleichgestellt und frei - Grundlage zivilen Zusammenlebens.
- Herrscht die Logik integraler Vielfalt, kann sich Zivilität in vielfältigen Formen entfalten. So unter anderem sach- bzw. fachlogisch, funktional ausdifferenziert, in unterschiedlichen Lebensstilen und/oder reflexiv-humorvoll.

Zivilität nimmt also in aufsteigender Reihenfolge der skizzierten Interaktions-Logiken zu (siehe Tabelle).

Tabelle 1: Interaktions-Logiken

Interaktions-Logik	Zivilität
Integrale Vielfalt	
Gemeinsam anerkanntes Recht	
Interessen	
Macht	
Freund-Feind	

Werden Interessen-, Macht- und Freund-Feind-Logiken gesellschaftlich durch Rechts- und Integrations-Logik effektiv gebunden, entsteht die zivile Moderne. Zu dieser gehören rechtsstaatlich gebundene Formen demokratischer Willensbildung und Entscheidung mit demokratischen Wahlen, faire Wettbewerbe, unabhängige Sachkommunikation, Wissenschaft, Humor und andere Formen mehrdimensionaler Kommunikation und Koordination.³

² Die Typologie habe ich erstmals dargestellt in: Prittwitz 2014: <http://www.bpb.de/apuz/194822/politik-logiken-im-ukraine-konflikt?p=all>

³ Prittwitz 2018: Die Zivile Moderne. Eine Theorie soziopolitischer Koordination (Buchtext im Erscheinen)

Menschenrechtlich und sozial ungebundene Interaktionsformen sind dagegen unzivil. Dies gilt insbesondere für opportunistische Interessenwahrnehmung, für unbeschränkte Macht und Freund-Feind-Logiken.

3. Chancen und Risiken von Algorithmen-Steuerung und KI

Algorithmen-Steuerung ist - zumindest im ersten Entwicklungsschritt - menschenbestimmt; sie gilt aber verbreitet als (schwache) Form Künstlicher Intelligenz. KI im engeren Sinne bilden dagegen unbeaufsichtigt lernfähige technische Systeme - dies funktionsgebunden oder funktionsüberschreitend.

Tabelle 2: Interaktionslogische Chancen-/Risiko-Abschätzung von KI-Formen

Kapazität Logik	Algorithmen-Steuerung	Unbeaufsichtigtes System-Lernen funktionsgebunden	Unbeaufsichtigtes System-Lernen funktionsüberschreitend
Integrale Vielfalt/ Sachlich	Allgemein nützliche Algorithmen; individuelle Nutzerbezüge	Funktionales KI-Lernen mit individuellen Bezügen	Respektiert überlegene KI integrale Vielfalt?!
Gemeinsam anerkanntes Recht	Algorithmen-Prinzipal hält sich an rechtliche Vorgaben	KI hält sich an rechtliche Vorgaben	Erkennt überlegene KI menschliches Recht an?!
Interessen	Konkurrierende Prinzipal-Algorithmen	Funktionsgebunden lernende KI nach Prinzipal-Interessen	KI verhält sich nach ihren Eigeninteressen
Macht	Digitale Machtausübung-/Kontrolle	Lernende KI im Sinne von Prinzipal-Macht	KI als Herrschafts-Instanz über Mensch
Freund-Feind	Digitale Waffensysteme nach Prinzipal-Zielen	Lernende Kampfsysteme nach Prinzipal-Willen	KI gegen Mensch

Nach interaktionslogischen Kriterien ergeben sich, wie in Tabelle 2 dargestellt, übereinstimmende Chancen/Risiko-Profile zwischen Algorithmen-Steuerung und funktional gebundener KI (Gleiche Farben). Funktionsübergreifende KI dagegen ist grundsätzlich anders zu bewerten.

Algorithmen-Steuerung

Algorithmen-gesteuerte Systeme können, zumindest solange sie von Menschen entworfen, gepflegt und ausgewertet werden, noch als (maschinelle) Instrumente menschlicher Intelligenz verstanden werden.⁴ Derlei Systeme haben bereits heute in zahlreichen Sachbereichen, beispielsweise Medizin, Altenbetreuung, Psychologie, Industrie, Landwirtschaft, Transport und Verkehr, Finanzsektor, allgemeinwohldienliche Nutzungsperspektiven. Auch demokratische Willensbildung, möglicherweise sogar demokratische Entscheidungsprozesse, können potentiell hierdurch gestützt werden (Beispiele: Globale Demokratie-Optionen, Abwehr von Social Bots). Hierzu müssen die jeweiligen Algorithmen-Betreiber (Prinzipale) allerdings strikt allgemein anerkanntes Recht akzeptieren.

Ansonsten entstehen Formen unziviler Steuerungsmacht - siehe beispielsweise algorithmisch gesteuerte Kaufentscheidungen ohne Zustimmung, ja ohne Bewusstsein der Kaufenden für ihre Steuerung, oder eindimensionale Machtsysteme mit ungebundenen und damit unbegrenzten Durchgriffsmöglichkeiten der Herrschenden.

Verheerend im wörtlichen Sinne sind algorithmisch gesteuerte Waffen- und Kriegsformen. Diese eröffnen zwar (zunächst) individuelle militärische Vorteile für die jeweiligen Betreiber - und sind daher für diese äußerst attraktiv. Krieg, die unzivilste Form menschlicher Interaktion, wird hierdurch aber aufrechterhalten, ausgeweitet und intensiviert. Und Rüstungskosten bedeuten, dass Ressourcen für Vernichtung statt für Aufbau und Integration aufgewandt werden.

Funktionsgebundene Künstliche Intelligenz

Software, die Aufgaben (beispielsweise den Bauplan für eine Brücke an einem bestimmten Ort zu erstellen) unbeaufsichtigt lernend, also ohne menschliche Hilfen und Vorgaben, bewältigen kann, wird berechtigterweise als Künstliche Intelligenz bezeichnet. Denn sie kombiniert eigenständige kognitive Fähigkeiten (des Erkennens und logischen Schlussfolgerns) in spezifischen Formen und mit eigenem Lernpotentialen.

⁴ Siehe: <https://www.datenschutzbeauftragter-info.de/was-ist-ein-algorithmus-definition-und-beispiele/>

Solange Künstliche Intelligenz funktionsgebunden, damit instrumentell operiert, bemisst sich ihre Bewertung nach den angegebenen Interaktions-Logiken. So kann KI Allgemeininteressen dienen - sofern sie strikt an allgemein akzeptierte Rechtsnormen und integrative Logiken gebunden operiert und sich entwickelt (Farbe Grün). Ohne diese Bindung allerdings werden auch funktionsgebundene KI-Systeme (genauso wie algorithmische Steuerungsformen) prekär. Denn dann dienen sie bornierten Eigeninteressen, Machtgelüsten oder/und Kriegführung, der unzivilsten Interaktionsform.

Funktionsüberschreitende KI

Überspringt die KI-Entwicklung die Schwelle zu selbstbewusster und eigenwilliger KI-Bildung, erhält KI Herrschafts- und Vernichtungspotential gegenüber dem Menschen - der worst case, angesichts dessen menschliche Interaktions-Logiken nachrangig bis irrelevant werden.

4) Reale KI-Entwicklung

Künstliche Intelligenz wird bisher vor allem nach interessen- und machtlogischen Kriterien entwickelt und koordiniert, so durch große Internet-Konzerne, staatliche Administrationen und von diesen finanzierte Forschungs-Institutionen (insbesondere der USA) sowie autoritäre Staaten wie China und Russland. Menschenrechtlich-integrativ gebunden sind KI-Forschung und praktische Anwendungsformen bisher nur punktuell - siehe beispielsweise den effektiven Widerstand gegen Google Glass, der inzwischen zu einer unternehmensbezogenen (also zunächst nicht öffentlich genutzten) Entwicklungs-Version geführt hat.⁵

Das aktuell laufende chinesische Groß-Projekt digitaler Moralüberwachung und Moralbewertung (Sozialkredit-System) wird - auch mit der Perspektive von KI-Anwendung - kompromisslos durchgeführt.⁶ Nach Wladimir Putin gilt: *Wer auch immer in diesem Bereich führend sein wird, wird auch die Welt beherrschen* - ein Schlaglicht auf die KI-Entwicklungsperspektive Russlands.⁷ Das autoritär und

⁵ 14.02.2018: <http://winfuture.de/news,98643.html>

⁶ 14.02.2018: http://www.deutschlandfunkkultur.de/chinas-sozialkredit-system-auf-dem-weg-in-die-it-diktatur.979.de.html?dram:article_id=395126; [https://de.wikipedia.org/wiki/Sozialkredit-System_\(VR_China\)](https://de.wikipedia.org/wiki/Sozialkredit-System_(VR_China))

⁷ 14.02.2018: <http://meedia.de/2017/09/05/kuenstliche-intelligenz-wladimir-putin-und-tesla-chef-elon-musk-warnen-vor-gefahr-fuer-die-menschheit/>

fundamentalistisch regierte Saudi-Arabien verleiht als erstes Land der Welt einem Roboter (Sophia) seine Staatsbürgerschaft.⁸ Staaten wie Japan (Robotik), Süd-Korea (Rüstungs-KI) und Estland (digitale Service-Systeme) sind in Teilbereichen der KI-Entwicklung sehr aktiv. Bei weitem führend in der KI-Entwicklung aber sind die USA, weiterhin Russland und China - dies gerade auch im Bereich militärischer KI-Anwendungen.

Durch diesen interessen- und machtlögen Bias wurden Werte und Strukturen der zivilen Moderne beschädigt (Social Bot-Problematik, illegitime Überlegenheit von Google und Facebook, praktische Rechtlosigkeit von Konsumenten, Stärkung der Geheimdienste zu Lasten ziviler Institutionen). Interessen- und machtlögen Prozesse verlaufen aber immer als Kampf, sodass auch zivile Gegenbewegungen indirekt initiiert wurden und werden.

5. Transhumanismus, Religion und Politik

Über technokratisch-ökonomistische Denkweisen und Motiven hinaus gibt es bereits seit Jahrzehnten eine quasireligiöse Ideologie-Quelle kritikloser, ja emphatisch überhöhter KI-Ausbreitung, den Transhumanismus in verschiedenen Strömungen. Demnach soll und wird die heutige (organisch gebundene und fehlerhafte) Menschheit durch die perfekte Maschinen-Intelligenz zunächst partiell, dann völlig aufgehoben werden.

Ob diese eschatologische Strömung einer quasigöttlichen Überwindung der Menschheit an Einfluss gewinnt oder nicht, dürfte stark davon abhängen, wieweit sich Menschen politisch bilden: Je weniger gebildet Menschen etwa hinsichtlich Demokratie, Rechtsstaat und einem pluralistischen Miteinander sind, desto chancenreicher dürften Transhumanisten sein - dies auch gerade in Kombination mit diversen evangelikalen Strömungen. Damit wird sich voraussichtlich auch die Frage stellen, wie herkömmliche Religionen mit der neuen Konkurrenz umgehen (Beispiel: Modernistische Strömung im wahhabitischen Saudi-Arabien).

⁸ 06.11.2017/14.02.2018: <http://www.epochtimes.de/politik/welt/gefah-fuer-die-menschen-erster-roboter-sophia-als-staatsbuenger-in-saudi-arabien-ankannt-a2260011.html>

6. Wahrnehmungs- und Diskussions-Blockaden

Aktuelle Praktiken und Perspektiven des Einsatzes von KI, so deren militärische Nutzung und Risiken einer Superintelligenz, werden in den USA bereits seit vielen Jahren unter Fachleuten und unter Unternehmenschefs kontrovers diskutiert. Auch in Deutschland beschäftigen sich öffentliche Medien wie *Spiegel Online*, *Süddeutsche.de* und *Aus Politik und Zeitgeschichte (Beilage zum Parlament)* immer wieder mal mit KI-Aspekten. Dabei wird KI auch als politiknaher Forschungs- und Handlungsgegenstand wahrgenommen. So soll nach dem Koalitionsvertrag der geplanten neuen Großen Koalition ein *nationales Forschungskonsortium für Künstliche Intelligenz und maschinelles Lernen* initiiert werden.⁹ Zusammen mit Frankreich sind gemeinsam wirksame Strategien und neue technologische Ansätze, zum Beispiel in den Bereichen Energie, Klimaschutz, Biotechnologie und Künstliche Intelligenz, angekündigt.¹⁰ Eine politische Debatte über KI-Anwendung und KI-Entwicklung gibt es bisher aber weder in den USA noch in Deutschland.

Verstehen wir diesen Sachverhalt nach dem - kapazitätstheoretisch inspirierten - Katastrophenparadox, so verhalten sich die Beteiligten Akteure rational entsprechend ihrer wahrgenommenen Handlungskapazitäten.¹¹ Während der Umgang mit KI für Unternehmenschefs mit großen KI-bezogenen Forschungs-Budgets wichtig ist und gesteuert werden kann, sind alle Außenstehenden bisher ohne politische Handlungsalternativen zu KI. Dementsprechend gibt es keinen politischen Diskussionsstoff...

Dass sich derartige politische Thematisierungs-Blockaden auflösen können, zeigte sich anhand der weltweiten Umweltpolitik-Entwicklung ab dem Ende der 1960er Jahre: Trotz riesiger Umweltprobleme gab es bis in die 1960er Jahre hinein keine politische Ökologie- und Umweltdiskussion. Als sich aber seit dem Ende der 1960er Jahre machbare Alternativen abzeichneten, setzte die Öko- und Umweltbewegung und Umweltpolitik ein - bis heute eine vitale Bewegung und Politikdiskurs.

⁹ Text Koalitionsvertrag GroKo vom 7. Februar 2018, Zeilen 1496/1497:

http://www.handelsblatt.com/downloads/20936422/4/koalitionsvertrag_final.pdf

¹⁰ ebda

¹¹ Prittwitz 1990: Das Katastrophenparadox. Elemente einer Theorie der Umweltpolitik, Leske+Budrich; siehe auch: <http://www.volkervonprittwitz.de/katastrophenparadox.htm>

7. Technologiepolitische Vorschläge und Perspektiven

Was lässt sich aus den dargestellten Chancen- und Risikoprofilen sowie den bisherigen Handlungsbedingungen von KI technologiepolitisch folgern?

- 1) Künstliche Intelligenz ist ein Entwicklungs- und Praxisfeld von überragender Bedeutung - dies im Sinne wirtschaftlicher und soziopolitischer Chancen, aber auch existentieller Risiken. Dieses so bedeutende Feld wird voraussichtlich erst dann angemessen wahrgenommen werden, wenn - über den Interessen-Horizont von Programmentwicklern und Vermarktern hinausreichende - Handlungsalternativen formuliert und diskutiert werden.
- 2) Die Chancen und Risiken Künstlicher Intelligenz bemessen sich vorrangig interaktionslogisch, das heißt, nach ihrem Zivilitäts-Grad.
- 3) Integrativ allgemeinwohlförderliche und menschenrechtlich effektiv gebundene KI-Formen (Beispiel: Medizinische Anwendung für alle Menschen nach expliziten rechtlichen Bedingungen, autonomes Fahren nach eingehender Ausarbeitung und Prüfung und klaren vertraglichen Bedingungen, spezielle Dienstleistungen nach expliziten, klaren Verträgen), erscheinen akzeptabel, ja sogar förderungswürdig.
- 4) Kompetitive Nutzungsformen, eine wichtige ökonomische Perspektive, sind genau zu regeln (a) nach eigenen Regeln der Betreiber unter öffentlichem Druck, b) durch politische Regulierung. Dabei stellen sich die Probleme der kommenden Massenarbeitslosigkeit - oder auch die kommenden Transformationsprobleme von Ökonomie als besondere soziale und politische Herausforderung.
- 5) Militärische KI-Forschung und vor allem KI-Entwicklungen mit Herrschaftspotenzial werden - wegen der besonders starken Finanzierung und vitalen nationalen Interessen - besonders häufig und stark protegiert. Sie stehen aber in diametralem Gegensatz zum Menschheitsinteresse an einem friedlichen Zusammenleben in Wohlfahrt. Gerade hierüber sollte eine offene Diskussion in wissenschaftsnahen Kreisen, aber auch bis in die Parteien hinein geführt werden.
- 6) Generell sollten der KI-Diskussion ethische Aspekte der zivilen Moderne, so beispielsweise Roboter-Gesetze, berücksichtigt werden.

- 7) Entwicklungs- und Anwendungsformen von KI ohne menschenrechtliche Bindung und integrative Orientierung erscheinen nicht akzeptabel. Dies gilt besonders a) für opportunistische (betrügerische) Interessenwahrnehmung in Wirtschaft, Gesellschaft und Politik (beispielsweise betrügerische Social Bots); b) für absolute Macht-Logiken wie die rechtlose Unterwerfung von Menschen und KI-Herrschaft, c) für alle KI-gestützten Rüstungs- und Kriegsformen. Wo und wann nicht integrale, nicht menschenrechtsgebundene KI-Formen auftauchen, sollten sie scharf benannt und kritisiert werden. Hierzu sind ökonomische und politische Alternativen zu verdeutlichen.
- 8) Würde KI-Entwicklung die Schwelle zur Entstehung selbstbewusster und eigenwilliger KI überspringen, würde dies die Existenz der Menschheit aufs höchste gefährden - eine fundamentale Verletzung der Menschenrechte auf Leben, Gesundheit, Gleichstellung und Freiheit. Daher muss es zu einem erstrangigen politischen Ziel werden, die Entwicklung derartiger KI von vorneherein auszuschließen.
- 9) Menschliches Dasein gründet sich seit jeher auf Verbindungsformen organisch-analoger Strukturen und (digitaler) Denkfähigkeit. Diese Verbindung soll auch in Zukunft grundlegend bleiben. So sind spezifisch menschliche und hierbei gesellschaftliche Intelligenzformen nach Kräften zu entfalten und zu fördern.
- 10) Alle KI-Perspektiven und Risiken betreffen die gesamte Menschheit. Deshalb sollten der entsprechende Austausch und Regulierungs-Bemühungen global betrieben werden - so kurzfristig vor allem ein striktes Moratorium weiterer militärischer KI-Forschung und Entwicklung mit der Alternativ-Perspektive weltweit integrierter KI-Entwicklung.
- 11) Integratives Handeln zu KI liegt im globalen Menschheits-Interesse. Es ist daher vorrangig gegenüber allen nationalstaatlichen und Unternehmens-Interessen. Dies sollte öffentlich klar gemacht und politisch auf allen denkbaren Ebenen eingefordert werden: Künstliche Intelligenz als universelle politische Herausforderung.