

Leistungen und Grenzen der Automatisierung



TECHNISCHE
UNIVERSITÄT
DARMSTADT



INSTITUT FÜR
PHILOSOPHIE
DARMSTADT

Prof. Dr. Christoph Hubig
Technische Universität Darmstadt

**4. BMBF Innovationsforum „Zivile Sicherheit“/
Ausprägungen der Digitalisierung
Berlin, 19.06.2018**

Gliederung

1. Vom „klassischen“ Automaten zu autonomen Systemen
2. Drei Typen von Autonomie
 - 2.1 Beispiele
3. Basis: Maschinelles Lernen
4. Informationelle Selbstbestimmung und automatisiertes Entscheiden
 - 4.1 Charta der digitalen Grundrechte
 - 4.2 DSGVO
5. Verantwortungsnetzwerke
 - 5.1 „Blackboxing“ in Haftungsfällen
 - 5.2 Revision des „Blackboxing“: Programmatische Forderungen an die Politik

1. Vom „klassischen“ Automaten zu autonomen Systemen

Automat i. w. S.

„selbstbewegend“

„klassische“
Maschine

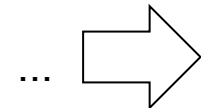
- *determiniert, organisiert*
- operiert *deterministisch* (steuern, regeln)

Bsp.: Waschmaschine,
Automatisierung bei Ford (1946)
Brandmelder-Sprinkler-Alarmierung

technisch vermittelte *Interaktion*
(auf „Erwartungserwartungen“ der
Entwicklungs-, Produktions- und
Nutzungsseite basierend)

autonome Systeme
„smarte“ Maschinen

- determiniert (Funktion)
- *nichtdeterministisch*
(selbst-)organisierend (Kant)



technisch konstituierte *Koaktion*
(auf maschinellem Lernen und
tracking basierend)
System*dispositionen* + Erwartungen
der Nutzerinnen und Nutzer

2. Drei Typen von Autonomie

Autonomietyp

(1) **operativ**

Freiheitsgrade bei der Wahl optimaler *Mittel* (Effizienz/ Effektivität)/bei gegebenen Zwecken: Assistenzsysteme jeder Art

(2) **strategisch**

Freiheitsgrade der Wahl optimaler *Zwecke* (Priorisierung, Umweg, Taktung, Dauer etc.) unter gegebenen Zielen: *Planung* (und Realisierung) von Abläufen

(3) **moralisch**

Freiheit der Anerkennung von *Prinzipien* und *Zielen* sowie der Selbstzuschreibung der Anerkennung

Delegationstyp der Verantwortung

Erkennen (Sollgröße, Situation) und Entscheiden in Grenzen vollständig delegierbar (im Rahmen von (3))

Erkennen und Entscheiden teilweise delegierbar, sofern Interventionen im Rahmen von (3) gewährleistet

nicht delegierbar

Anwendungsfelder

adaptive Steuerung von Evakuierung im Krisenfall, Veranlassung und Taktung von Hilfsmaßnahmen, Identifizierung von individuellen oder typischen Sachlagen zwecks Prävention sowie Diagnose von Interventionsbedarf

...
Organisation und Priorisierung von Hilfsmaßnahmen, Anpassung von Sollgrößen mit Blick auf Systemstabilität (Infrastrukturen der Versorgung, des Verkehrs, der Kommunikation – Erweiterung oder Einschränkung der Kapazitäten und Ressourcennutzung), Koordination von Zuständigkeiten

...
Respektierung informationeller Selbstbestimmung, Rückdelegation (Überprüfung, Revision) von Systementscheidungen an Menschen... (s. u.)

2.1 Beispiele (1): Gesichtserkennung



Per Gesichtserkennung hat die chinesische Polizei einen Verdächtigen in einer riesigen Menschenmenge entdeckt.

Mit einer neuen "Bestleistung" hat die chinesische Polizei das Potenzial ihrer Gesichtserkennungs-Technologie bewiesen. Wie die South China Morning Post berichtet, wurde in einer Menschenmenge von 50.000 Personen, die das Konzert des Popstars Jacky Cheung in der Stadt Nanchang besuchten, das Gesicht eines Verdächtigen entdeckt. Der Mann wurde direkt am Konzert verhaftet. Ihm werden nicht näher genannte Wirtschaftsverbrechen vorgeworfen. (futurezone.at, 13.04.18)

2.1 Beispiele (2): ADIS Mustererkennungssysteme (2010-13)



Quelle: https://www.sifo.de/files/Mustererkennung_D_ADIS.pdf

Bewegungsanalyse, Mimik-Gestik-
Bewegungsdeutung

- ↳ deeskalierende Durchsagen
- ↳ Veranlassung des Einsatzes von Sicherheits-
und/oder Notfallkräften
- ↳ präventive Gefahrenabwehr
durch Schutzzonenoption (Autonomie (3))

mögliche Perspektiven: unterstützende Aktorik

→ Basis: maschinelles Lernen

3. Basis: Maschinelles Lernen

- überwachtes Lernen: Features (Eigenschaften) in den Sensordaten als Indikatoren für einen Deutungsinhalt auf Basis von Trainingsdaten unter „Kategorien“ als „Labels“
—> Optimierung auf entscheidungssteuernde Diskriminierbarkeit

Verantwortung bleibt bei den Trainer/innen (Kategoriensystem/ Relevanz der Parameter und Typisierung der Kontexte, Zuordnung von „Entscheidungen“)

- unüberwachtes Lernen: Neue Deutungsvorschläge vom System selbst erstellt auf Basis von Mustererkennung, relativ zu Trainingsdaten, keine Labels

Verantwortung verbleibt bei Entwicklern/innen der Systemarchitektur: Sensorik, Untersuchungskontext, Funktion

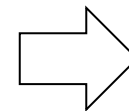
- bestärkendes Lernen: Optimierung systemischer Reaktionen in Erfüllung von Systemfunktionen qua „Belohnungsfunktion“ bei Findung einer Strategie, die unter möglichst vielen Umständen die bestmögliche Teilentscheidung hinsichtlich des gewünschten Ziels trifft

Verantwortung verbleibt bei den Entscheidern/innen über Ziele und Gewährung autonomer Systemfunktionalität /Mittel, Zwecke)

4. Informationelle Selbstbestimmung und automatisiertes „Entscheiden“

- Gestaltung der Bedingungen für Unterwerfung unter automatisierte Entscheidungen/ EU-Charta der digitalen Grundrechte (überarb. 2018) Art. 5: Überprüfungs- und Revisionsvorbehalt durch Menschen (moralisch autonome (3) Subjekte) – Option der Rücknahme der Verantwortungsdelegation
- Integrität und informationelle Selbstbestimmung (z. B. limitbasierte Sensorik)

DSGVO Art. 22, Bundesdatenschutzgesetz § 28b



4.1 Charta der digitalen Grundrechte der Europäischen Union (2018)

ART. 5 (AUTOMATISIERTE SYSTEME UND ENTSCHEIDUNGEN)

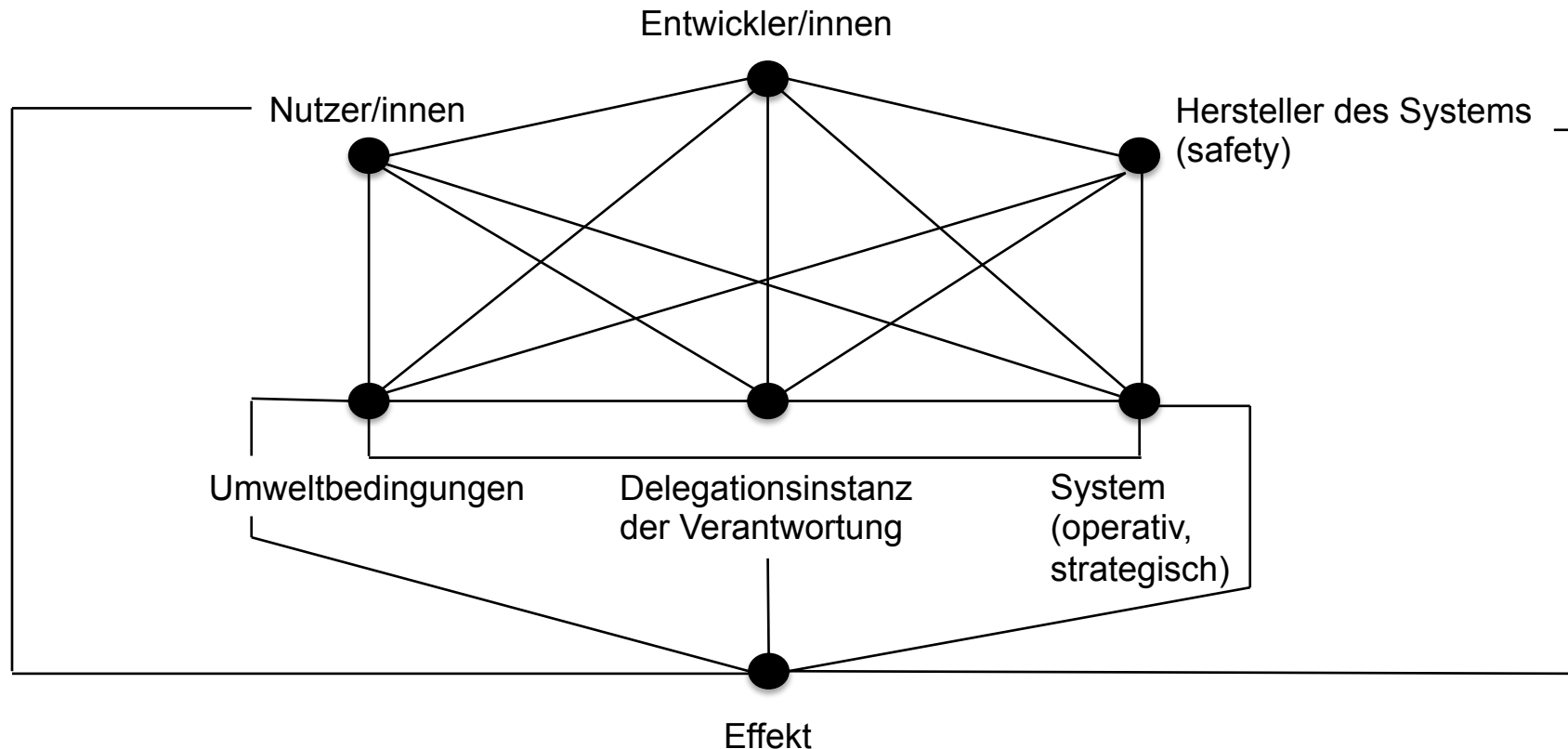
- (1) Ethisch-normative Prinzipien dürfen nur vom Menschen aufgestellt, und Entscheidungen, die in Grundrechte eingreifen, nur von Menschen getroffen werden.
- (2) Automatisierte Entscheidungen müssen von natürlichen oder juristischen Personen verantwortet werden.
- (3) Die Kriterien automatisierter Entscheidungen, etwa bei Profilbildung, sind offenzulegen.
- (4) Wer einer automatisierten Entscheidungen von erheblicher Bedeutung für seine Lebensführung unterworfen ist, hat Anspruch auf unabhängige Überprüfung und Entscheidung durch Menschen.
- (5) Entscheidungen über Leben, körperliche Unversehrtheit und Freiheitsentzug dürfen nur von Menschen getroffen werden.
- (6) Der Einsatz von künstliche Intelligenz und Robotik in grundrechtsrelevanten Bereichen muss gesellschaftlich begleitet und vom Gesetzgeber reguliert werden.

4.2 Art. 22 DSGVO: Automatisierte Entscheidungen im Einzelfall einschließlich Profiling

- (1) Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.
- (2) Absatz 1 gilt nicht, wenn die Entscheidung
 - a) für den Abschluss oder die Erfüllung eines Vertrags zwischen der betroffenen Person und dem Verantwortlichen erforderlich ist,
 - b) aufgrund von Rechtsvorschriften der Union oder der Mitgliedstaaten, denen der Verantwortliche unterliegt, zulässig ist und diese Rechtsvorschriften angemessene Maßnahmen zur Wahrung der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person enthalten oder
 - c) mit ausdrücklicher Einwilligung der betroffenen Person erfolgt.
- (3) In den in Absatz 2 Buchstaben a und c genannten Fällen trifft der Verantwortliche angemessene Maßnahmen, um die Rechte und Freiheiten sowie die berechtigten Interessen der betroffenen Person zu wahren, wozu mindestens das Recht auf Erwirkung des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunkts und auf Anfechtung der Entscheidung gehört.
- (4) Entscheidungen nach Absatz 2 dürfen nicht auf besonderen Kategorien personenbezogener Daten nach Artikel 9 Absatz 1 beruhen, sofern nicht Artikel 9 Absatz 2 Buchstabe a oder g gilt und angemessene Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person getroffen wurden.

5. Verantwortungnetzwerke

- Knoten im Netz: Zusammenwirken von technischen, intentionalen und natürlichen „Fragmenten“ (Latour)



...

5.1 „Blackboxing“ in Haftungsfällen

- „e-Person“ (EU 2017): verklagbarer Adressat für Haftungsansprüche
(Fonds zur Schadenskompensation aus Ressourcen aller Akteure)
- Ausweitung der Gefährdungshaftung (auch ohne Vorsatz und Fahrlässigkeit)
über bewegliche Sachen hinaus (für alle, die von der Nutzung profitieren)
- Erweiterung des Produkthaftungsgesetzes
(Sachsubstanz, Software, ausgelagerte Software)

5.2 Revision des „Blackboxing“: Programmatische Forderungen an die Politik

„Digitale Demokratie“: Rückgabe verlorener Verantwortungsoptionen und Erhalt der „kollektiven Intelligenz“ (vs. „Feudalismus 2.0“/z. B. China)
zwecks Erhalt der Unabhängigkeit von Informationssuche und Entscheidungsfindung:

- Recht auf Kopie persönlicher Daten (Datenmailbox) – Austausch über trackingbasierten Lernstand der Systeme (wiss. Institutionen als Treuhänder)
- Optionen der Selbststeuerung von Empfehlungsalgorithmen (vs. „Big Nudging“)
- echte Sanktionen bei Regelverstößen (z. B. bots)
- Förderung der Dezentralisierung/Blockchain-Technologien/Reputationskalküle

→ partizipative Innovationsplattform als „digitales Nervensystem“
(Entlastung kurz getakteter Machtpolitik)

Vielen Dank für Ihre Aufmerksamkeit!
