

Auswertung des Auftakt-Workshops “Algorithmen fürs Gemeinwohl” am 26.1.2018

Die Ausgangsfragen für den Auftakt-Workshop waren:

- 1) Funktioniert das Grundnarrativ? > Es wird viel über Gefahren und Risiken gesprochen, dabei gibt es auch so viele gesellschaftliche Chancen; die gilt es zu realisieren, ohne dabei eventuell bestehende Risiken aus den Augen zu verlieren.
- 2) Was sind die vielversprechendsten Anwendungsbereiche, was sind schöne bestehende oder denkbare Anwendungsfälle, um das Potenzial und dessen Realisierbarkeit möglichst exemplarisch aufzuzeigen?
- 3) Was muss im Prozess des Designs, der Entwicklung und Implementierung algorithmischer Systeme besonders beachtet werden, damit sich die entwickelten Anwendungen gemeinwohlförderlich auswirken?

Befunde bezogen auf diese drei Kernfragen

- 1) **Das Grundnarrativ des Projekts funktioniert.** Die Teilnehmenden empfanden den methodischen Wechsel zur ‘positiven Brille’ ganz überwiegend als gewinnbringend und haben sich engagiert am chancenorientierten Austausch beteiligt. Ferner wurde allerdings deutlich, dass auch bei einer Fokussierung auf Chancen und Potenziale immer wieder (z.T. aus der Debatte bekannte) mögliche oder tatsächliche negative Auswirkungen in den Blick geraten und diskutiert werden müssen.
- 2) **Als besonders relevante Anwendungsbereiche wurden identifiziert** (jeweils mit genannten bzw. im dritten Workshopteil in fünf Gruppen ausführlich behandelten Beispielen/Fokussierungen):
 - a) **Medizin / Gesundheitswesen**
 - i) Früherkennung gängiger Krankheiten, bessere Erkennbarkeit seltener Krankheiten (Gruppe 1)
 - ii) ‘Todesalgorithmus’ (Gruppe 3)
 - iii) Überblick über den Stand medizinischer Forschung in den verschiedenen Teilbereichen
 - iv) Seelsorge, (Früh-)Erkennung von Depressionen u. Suizidgefahr
 - b) **Öffentlicher Dienst / Staatlicher Leistungsbereich** (digitale Verwaltung)
 - i) Transparenz u. Verbesserung öffentlicher Vergabeverfahren (z.B. Korruptions-, Manipulations- u. Erpressungskontrolle)
 - ii) Legislativer Fußabdruck (Gesetzesfolgenabschätzung)
 - iii) Familienassistent für die verschiedenen Interaktionspunkte (Gruppe 4)
 - c) **Sozialpolitik / Steuer- und Transfersystem**
 - i) Optimierung der Allokation staatlicher Transfer- u. Sozialleistungen
 - ii) Monitoring der realen Steuerlastverteilung, Förderung von Steuergerechtigkeit
 - iii) Diversität von Wohnquartieren erhalten/fördern (Gruppe 2)
 - d) **Infrastrukturpolitik** (Mobilität/Verkehr u. Strom)

- i) Dynamische Stromnetzoptimierung
 - ii) Verkehrsfluss optimieren, gezielte Vermeidung von Staus (Gruppe 5)
 - e) **Personalwesen**
 - i) Modellierung der Konsequenzen von Karriere-Entscheidungen ("persönlicher Biografie-Assistent")
 - ii) Objektivierung von Personalauswahl- u. Personalbeurteilungsprozessen
 - f) **Allgemein: Bias erkennen & möglichst eliminieren**
 - i) Vgl. z.B. "Algorithmenbewertungsalgorithmus"
- 3) Im Laufe des Workshops hat sich gezeigt, dass **die für die Gemeinwohlförderung besonders relevanten Ansatzpunkte jeweils Anwendungsbereich- bzw. sogar Beispiel-spezifisch** sind und in jeweils unterschiedlichen Phasen der Gestaltung und Entwicklung von Algorithmen liegen. Ist die Problemstellung nicht eindeutig, kommt es beispielsweise sehr auf die ersten drei Prozessschritte Abstraktion, Teambildung und Modellierung an. In anderen Fällen - insbesondere dann, wenn lernende Algorithmen zum Einsatz kommen sollen und/oder wenn mögliche Diskriminierungseffekte eine besondere Rolle spielen - scheint besonderes Augenmerk auf die Datensammlung (Umfang, Qualität, Auswahl) gelegt werden zu müssen.
- Übergeordnet haben sich folgende Aspekte als bedeutsam herausgestellt:
- a) In den meisten Fällen wird die Realisierung des Gemeinwohlförderpotentials entscheidend davon abhängen, **Fach- und Themengebietsexperten** in die Algorithmengestaltung einzubeziehen.
 - b) Eine **(Risiko-)Folgenabschätzung** als eigener Prozessschritt könnte in vielen Fällen angeraten sein.
 - c) Ex ante (!): Einschätzung des Komplexitätsgrads algorithmischer Systeme auf der technischen Ebene und Entscheidung über **das erforderliche Maß an Transparenz** auf politischer oder gesellschaftlicher Ebene.
 - d) Die **Datenbasis** spielt für die Kontrollierbarkeit der Auswirkungen von Algorithmen generell eine herausgehobene Rolle. Da es bei der in Rede stehenden Technologie im Kern um maschinelle Datenverarbeitung im großen Stil geht, sind die Themen Umgang mit Daten (Data Governance) und Design algorithmischer Entscheidungs(vorbereitungs)systeme nicht trennbar. In der öffentlichen Debatte werden der Datendiskurs (BigData, OpenData, Datenschutz, Dateneigentum) und der Algorithmen/KI-Diskurs recht unverbunden geführt, obwohl sie doch von der Sache her essentiell zusammengehören. Das Projekt möchte zur Überwindung dieser unfruchtbaren Trennung beitragen.

Sonstige Hinweise und Befunde

- **Unstrittige Fälle des Einsatzes von Algorithmen**, bei denen sowohl das Ziel klar als auch der Lösungsweg grundsätzlich bekannt ist (wie z.B. bei der schieren Bewältigung von Informationsmengen, in der Logistik oder beim Übersetzen von Sprachen), sollten aus dem Gesellschaftsdiskurs und der Folgearbeit des Projekts systematisch ausgeblendet werden, um sicher zu gehen, dass Ressourcen sinnvoll eingesetzt werden und keine Problemstellungen geschaffen werden, wo eigentlich keine sind.

- Auch **technische Standardsysteme**, die sich schon im Einsatz befanden, lange bevor über algorithmische Entscheidungssysteme diskutiert wurde, bereiten Entscheidungen vor oder geben diese den beteiligten Menschen sogar vor (z.B. Entscheidungen über Intensivpflegebedarf eines Säuglings unmittelbar nach der Geburt anhand der gemessenen Vitalfunktionen und -werte). Evtl. ist es erhellend, diese einmal genauer zu studieren und mit den heute diskutierten Fällen zu vergleichen.
- **Erfolgskriterien** vorher definieren und die **Wirkung des Systems** später in der Anwendung messen: vorab definieren, wie Zielerreichung gemessen werden soll (d.h. Erfolgskriterien und Schwellenwerte definieren); bei Wirkungsmessung beachten, dass Zielerreichung nicht immer identisch mit Prognosequalität ist (vgl. z.B. Prognose von Sterbewahrscheinlichkeit vs. Verbesserung des Wohlbefindens von Patienten). Vorschlag zur Sichtbarmachung und Dokumentation solcher Zielkonflikte: Auf Basis einer Interessenmatrix über Optimierungsziele entscheiden und diese für spätere Evaluation dokumentieren.
- Besonders in Einsatzszenarien, bei denen **individuelles Scoring** dazu eingesetzt wird, um auf der **gesellschaftlichen Ebene** liegende Ziele zu erreichen (z.B. Kostenreduktion im Gesundheitswesen), ergeben sich leicht Zielkonflikte und klassische Trade-offs. In vielen Fällen stellt sich die Frage, ob auf individuelle Präferenzen oder auf Gemeinschaftsziele hin optimiert wird. Beides gleichzeitig ist oft nicht zu haben.
- Mehrere Teilnehmende haben eine **Typisierung/Klassifikation von Algorithmen** zum Zwecke der Diskursstrukturierung als Desiderat bezeichnet. Diese Strukturierung ist auch im Sinne der Potentialrealisierung und Risikovermeidung essentiell und sollte daher bald erfolgen.
- Unter dem Gemeinwohlaspekt ist grundsätzlich ein **Mensch-zentrierter Ansatz** zu wählen: Erstens ergeben sich die größten gesellschaftlichen Vorzüge algorithmischer Systeme oft erst in der Interaktion mit menschlichem Wissen, menschlichen Sozialkompetenzen und menschlicher Intuition.¹ Zweitens soll am Ende der Technikeinbezug stets dem Menschen dienen, nicht der Mensch einer Technikoptimierung dienstbar gemacht werden.
- In vielen Zusammenhängen (z.B. wenn es um die Digitalisierung der Verwaltung geht) gerät man beim Nachdenken über die Anwendbarkeit algorithmischer Systeme auf 'Nebenkriegsschauplätze' wie die Veränderung von Stellenanforderungen, den Wandel des Arbeitsmarktes insgesamt und die Folgen für die sozialen Sicherungssysteme. Diese Herausforderungen und komplexen Zusammenhänge müssen bei der Entwicklung von Strategien für die gemeinwohlorientierte Nutzung von Algorithmen mitgedacht werden.

Schlussfolgerungen der Projektverantwortlichen

1. Es empfiehlt sich, zunächst folgende **Grundunterscheidung** algorithmischer Systeme vorzunehmen:
 - a. Die Problemstellung ist klar und die Zielvorgabe ist ausgehandelt worden, während die Umsetzung mehr oder weniger herausfordernd sein und mehr oder weniger zur Förderung des Gemeinwohls beitragen kann.

¹ Auf diesen Aspekt hat Prof. em. Dr. Klaus Lenk, der aufgrund eines Logistikfehlers der Veranstalter leider nicht teilnehmen konnte, im Zuge der Workshop-Vorbereitung nachdrücklich verwiesen.

- b. Die Problemstellung ist unklar, die Zielsetzung ist nicht (ausreichend) präzise, evtl. ist sogar die Einsetzbarkeit von Technik an sich strittig, während die Umsetzung wiederum mehr oder weniger komplex sein kann.
2. Über diese grundlegende Unterscheidung hinaus werden die Projektpartner eine **Klassifizierung algorithmischer Systeme in verschiedenen Anwendungsbereichen** anstreben; Ziel ist eine typenspezifische und ausreichend trennscharfe Charakterisierung von Problemlösungskapazitäten, Gemeinwohlförderungspotenzialen und Risikokontrollmöglichkeiten in der Anwendung algorithmischer Systeme.
3. Ziel des Projekts ist es, Designprinzipien für Algorithmen zu formulieren, die Gemeinwohlinteressen angemessen berücksichtigen. Damit dieses generalisierende Ziel erreicht werden kann, braucht es **eine angemessene Balance zwischen Tiefe und Breite der nachfolgenden Analyse konkreter Fallbeispiele**. Dies kann in der eingeplanten Zeit mit den zur Verfügung stehenden Ressourcen nur gelingen, indem möglichst repräsentative Beispiele in den als relevant ausgemachten Anwendungsbereichen (Medizin/Gesundheit, E-Government, Sozialpolitik, Infrastrukturmaßnahmen) gewählt werden. Gegebenenfalls ist es von Vorteil, jeweils zwei Beispiele eines Bereichs auszuwählen, die sich maximal voneinander unterscheiden, um das gesamte Spektrum zu erfassen.