KI, Ethik, Vertrauen, Risiken, Audit.

Roberto V. Zicari,
Frankfurt Big Data Lab—Goethe Universität Frankfurt
Berlin, 10.2.2020

1. These: KI braucht Demokratie

Vertrauen zwischen **Menschen** und **KI-Systemen** ist ein wesentlicher Bestandteil der Förderung der Entwicklung und des Einsatzes von sozial nützlicher und verantwortungsbewusster KI.

Vertrauen zwischen Mensch und KI ist nicht monolithisch: **Der Kontext ist entscheidend.**

Wo liegen die „Grenzen" des Kontexts? **Vertrauen ist nicht „alles oder nichts".** Häufig besteht ein unterschiedliches Maß an Vertrauen, und das Ausmaß des Vertrauens, das ausreicht, um KI in verschiedenen Kontexten einzusetzen, ist daher eine wichtige zu berücksichtigende Frage.

Es kann auch mehrere Vertrauensschichten geben, die gesichert werden müssen, bevor jemand vertraut und möglicherweise letztendlich ein KI-Tool verwendet.

Beispielsweise kann man den Daten vertrauen, auf denen ein KI-System trainiert wurde, aber nicht der Organisation, die diese Daten verwendet.

Oder man kann einem Empfehlungssystem oder der Fähigkeit eines Algorithmus, nützliche Informationen bereitzustellen, vertrauen, aber nicht der spezifischen Plattform, auf der es bereitgestellt wird.

Die Datenethik-Kommission (DEK) schlägt vor, als „Kontext" das zu verwenden, was sie als **„Gesamtes sozio-technisches System"** bezeichnen.

[Empfehlung 36]: "Für die Beurteilung kommt es jeweils auf das gesamte sozio-technische System an, also alle Komponenten einer algorithmischen Anwendung einschließlich aller menschlichen Akteure, von der Entwicklungsphase (z. B. hinsichtlich der verwendeten Trainingsdaten) bis hin zur Implementierung in einer Anwendungsumgebung und zur Phase von Bewertung und Korrektur."

Aus westlicher Sicht sind die Begriffe **Kontext, Vertrauen und Ethik** eng mit unserem Konzept der **Demokratie** verbunden.

Zitate der DEK:
[2.1. Seite 164] "Prüfung der Frage, in wie weit die Funktion des Systems die Funktion der Demokratie, Grundrechte, das Sekundärrecht oder die Grundregeln des Rechtsstaats berühren kann".

[65] „Vor dem Hintergrund der besonderen Gefahren von Medienintermediären mit Torwächterfunktion für die Demokratie empfiehlt die DEK, auch mit Blick auf eine

Einwirkung auf den EU-Gesetzgeber zu prüfen, wie den mit einer solchen Torwächterfunktion verbundenen Gefahren begegnet werden kann."

Eine andere Perspektive auf den „Kontext" für KI unter einem **ganzheitlicheren, politischeren und wirtschaftlicheren** Gesichtspunkt, ist die, sie als **(digitale) „Ökosysteme"** zu betrachten (z. B. China, USA, Europa, andere).

Die zentrale Frage, die wir uns jedes Mal stellen müssen, wenn wir uns entscheiden müssen, ob und wo KI-basierte Produkte / Dienstleistungen bereitgestellt werden sollen, lautet:

WAS, WENN DIESE ÖKOSYSTEME NICHT DEMOKRATISCH SIND?

2. These: KI könnte die Konzentration der Macht festigen

Ein mögliches Risiko ist das so genannte **"Big Nudging"**.
Wer über große Datenmengen verfügt, kann Menschen auf subtile Weise manipulieren.
Aber auch wohlwollende Entscheidungsträger können mehr falsch als richtig machen.

Zitate der DEK:

„Mit der Entwicklung der Datenwirtschaft gehen **ökonomische Konzentrationstendenzen** einher, die das **Entstehen neuer Machtungleichgewichte** beobachten lassen."

"Bemühungen um die langfristige Sicherung der digitalen Souveränität sind daher nicht nur ein Gebot politischer Weitsicht, sondern auch Ausdruck ethischer Verantwortung."

Herausforderungen

- Wie man die Gültigkeit, Genauigkeit und Voreingenommenheit der KI überprüft;
- Probleme mit der Schnittstelle zwischen KI und den Menschen, die mit ihnen interagieren;
- Fragen der Governance, Transparenz und Rechenschaftspflicht.

Handlungsempfehlung 1

Der Entscheidungsprozess, ob und wo KI-basierte Produkte / Dienstleistungen eingesetzt werden sollen, muss als integraler Bestandteil die politische Bewertung der „Demokratie" der Ökosysteme umfassen, die den Kontext definieren.

Handlungsempfehlung 2

Regierungen dürfen keine proprietären KI einsetzen, die unter Berufung auf Geschäftsgeheimnisse Transparenz verhindern.

(z. B. im Strafrecht benutzte KI-Risikobewertungstools)
Trainingsdatensätze, -architekturen, -algorithmen und -modelle aller Tools, die für den Einsatz in Betracht gezogen werden, müssen allen interessierten Forschungsgemeinschaften - beispielsweise aus den Bereichen Statistik, Informatik, Sozialwissenschaften und Politik,

Recht und Kriminologie - umfassend zur Verfügung gestellt werden, damit sie diese vor und nach dem Einsatz bewerten können.
Angesichts immer ausgefeilterer Techniken zur Triangulation und Neuidentifizierung von Informationen können zusätzliche Maßnahmen erforderlich sein, z. B. vertragliche Regelungen, dass die Empfänger die Daten nur für bestimmte Zwecke verwenden und ihre Kopie des Datensatzes löschen, sobald diese Zwecke erfüllt sind.

Die zuständigen Behörden müssen die Verantwortung für die Evaluierung, Überwachung und Prüfung dieser KI-Tools nach der Bereitstellung übernehmen.

Handlungsempfehlung 3

Selbstregulierung ist ein guter Schritt, aber nicht ausreichend. Es besteht die Gefahr von Voreingenommenheit und Interessenkonflikten.
Es besteht die Notwendigkeit einer unabhängigen Verifizierung / Prüfung.

Die (DEK Nr. 58) empfiehlt Selbstregulierung und -zertifizierung und erwähnt die Grenzen technischer Normung.

Handlungsempfehlung 4

KI im Gesundheitswesen:
Der Einsatz von KI-Software-Medizinprodukten der Klasse I (EU) mit „geringem" Risiko kann ethische Probleme mit sich bringen (Rolle und Verhalten des Arztes, des Patienten).

Es ist eine „ethische Instandhaltung" erforderlich, die dynamische Änderungen der KI im Zeitverlauf berücksichtigt, beispielsweise durch das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM).

(z. B. zertifiziertes Archiv von Versionen der KI im Zeitverlauf, dh. Trainingsdatensätze, Architekturen, Algorithmen, und Modelle).

Im Zusammenhang mit dem Entwurf eines Gesetzes für eine bessere Versorgung durch Digitalisierung und Innovation (DVG) ist diese **„ethische Instandhaltung"** an den Stellen relevant, wo bereits eine Überprüfung an **Sicherheit, Funktionstauglichkeit und Qualität des Medizinproduktes** (inklusive **digitaler Medizinprodukte „niedriger" Risikoklasse)** gefordert ist.

Literaturverzeichnis

Will Democracy Survive Big Data and Artificial Intelligence?. Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., & Zwitter, A.. (2017). Scientific American (February 25, 2017).

Digitale Demokratie statt Datendiktatur: Big Data, Nudging, Verhaltenssteuerung
Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R. V., & Zwitter, A.. (2015). *Spektrum der Wissenschaft*, 15(12).