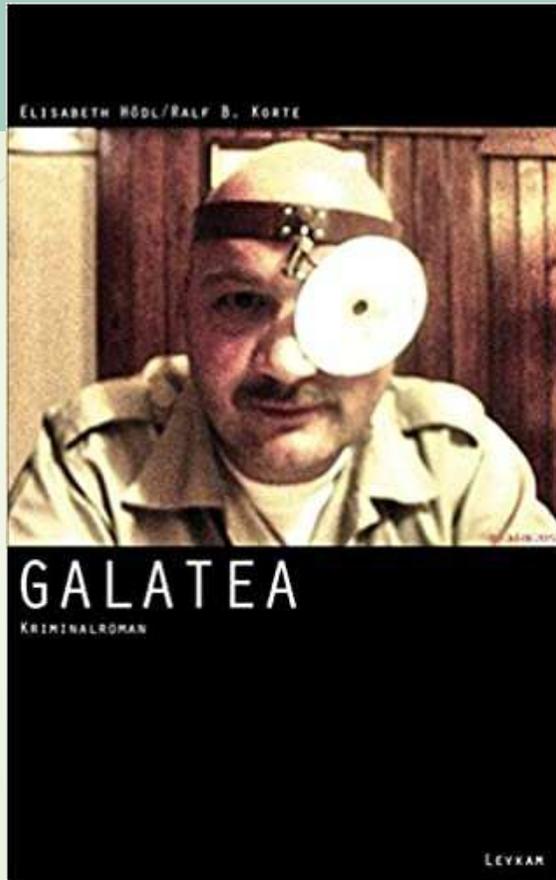


Nachhaltige KI aus grundrechtlicher und ethischer Sicht

KAIT Netzwerkevent
14.10.2022

Was erwartet uns als
Gesellschaft?

Univ.-Prof. Dr. Elisabeth Hödl



Inspiziert von
Mark Weisers Text: *A Computer for the 21th Century.*
Ubiquitous Computing.

Methoden der Zukunftsforschung

Nachhaltigkeit bedeutet einen Blick in die Zukunft zu tun.

Das Ziel ist es, **zukunftsfähiges Handeln** möglich zu machen.

Die Herangehensweise der Zukunftsforschung ist wissenschaftliche Methodik unter Einbeziehung von Kreativität.

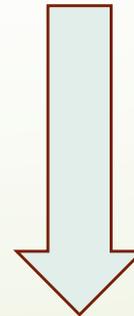
Die Qualitätskriterien hierfür sind:

- Plausibilität
- logische Kohärenz
- methodische Transparenz
- umsetzbare Implikationen.

Nachhaltige Künstliche Intelligenz?

Die Leitfragen der Zukunftsforschung aber auch für Unternehmen und Organisationen lauten:

- **Was ist (realistischerweise) möglich?**
- **Was ist wünschenswert?**
- **Was sind plausible Wirkungen?**



AI und Robotics



The Three Laws of Robotics

- Ein Roboter darf keinen Menschen verletzen, oder durch Untätigkeit zulassen, dass einem menschlichen Wesen Schaden zugefügt wird.
- Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen, es sei denn, ein solcher Befehl würde mit Regel 1 kollidieren.
- Ein Roboter muss seine Existenz beschützen, solange die Regel nicht mit Regel 1 und 2 kollidiert.

Isaac Asimov, Runaround, first in the collections I, Robot (1950).

8 Trends in Media und Tech

- ▶ New Audience Expectations
- ▶ The Creator Economy
- ▶ Web 3
- ▶ The Metaverse
- ▶ Data as a strategic Tool
- ▶ AI – Creative Content Production
- ▶ Synthetic Media
- ▶ Regulation and Responsibility Tech



Data as a strategic Tool

Big Data

„Big Data“ wird häufig als Sammelbegriff für digitale Technologien verwendet, die in technischer Hinsicht für eine neue Ära digitale Kommunikation und Verarbeitung und in sozialer Hinsicht für einen gesellschaftlichen Umbruch verantwortlich gemacht werden.

https://de.wikipedia.org/wiki/Big_Data

Data driven journalism

- ▶ Das amerikanische Softwareunternehmen Narrative Science war eines der ersten Unternehmen, das eine Plattform namens *Quill* zur **Datenanalyse** anbot.
- ▶ Die Software agiert selbstständig und "entscheidet", was mit den Daten geschieht.
- ▶ Geschichten werden **unabhängig** und **eigenständig** erzählt.
- ▶ Das funktioniert besonders gut in Nischen mit einer **guten Datenbasis**.
 - ▶ Sport
 - ▶ Finanzen
 - ▶ Immobilien
 - ▶ Wetterberichte
 - ▶ Horoskope

Robojournalim

- ▶ Mit Hilfe von Computern werden Texte weitgehend selbstständig erstellt, die Recherche und die Verknüpfung von Daten werden automatisiert.
- ▶ Computerservice, datengetriebener Journalismus und Datenjournalismus übernehmen journalistische Arbeitsschritte und entwickeln allmählich (in bestimmten Bereichen) redaktionelle Autonomie.

Sport

► Nicht einmal zwei Minuten nach Wiederanpfiff versenkt Koo Ja-Cheol den Ball im Tor und macht den Rückstand der Hausherren perfekt. Wenige Augenblicke vor dem Abpfiff sind nur noch die abgebrühtesten Spieler in der Lage, einen weiteren Ball im Kasten unterzubringen. Sanperio ist ein solcher Spieler und erzielt in der 83. Minute das entscheidende 2:0.



► Source: <http://www.swr.de/swr2/wissen/impuls-automatisierter-journalismus/-/id=661224/did=16348660/nid=661224/1w63jlv/index.html>
The software comes from the company Aexea.

► Normanvogel

► [Creative Commons Attribution 3.0](#)

Beispiele (1)

Die Berliner-Morgenpost hat einen **Feinstaubmonitor** entwickelt, der die gemessenen Feinstaubkonzentrationen automatisch überprüft und veröffentlicht.

Automatisierte Spielberichte im Sport mit Bildergalerien sind ein fester Bestandteil des Sportjournalismus.

Die schwedische "Local Group" bereichert die Sportberichterstattung mit automatisch generierten "**Post-Match-Reports**".

Journalismus - neue Wege

Routinearbeiten können erledigt werden.

Artikel können gekürzt werden.

Kleine Berichte können erstellt werden.

Data-driven-journalismus lebt von:

Geschwindigkeit

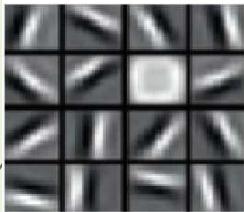
Filterung

Daten

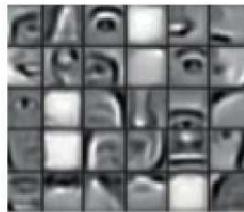
AI – Creative Content Production

Deep Learning Relies on Multiple “Layers” of Training

In this case, the first layer recognizes edges, the second recognizes facial features like a nose or an ear, until eventually the final layer recognizes full faces.



1. EDGES



2. FEATURES



3. FACES



4. FULL FACE

SOURCE “UNSUPERVISED LEARNING OF HIERARCHICAL REPRESENTATIONS WITH CONVOLUTIONAL DEEP BELIEF NETWORKS,” BY HONGLAK LEE ET AL, 2011

© HBR.ORG

Die Netzwerkstruktur des AlexNet (Aus: Krizhevsky, Alex, Ilya Sutskever und Geoffrey E. Hinton. “Imagenet classification with deep convolutional neural networks” Advances in neural information processing systems 2012)

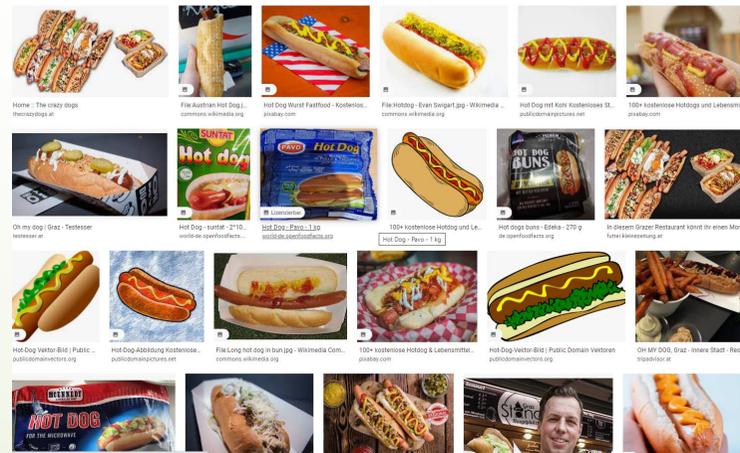
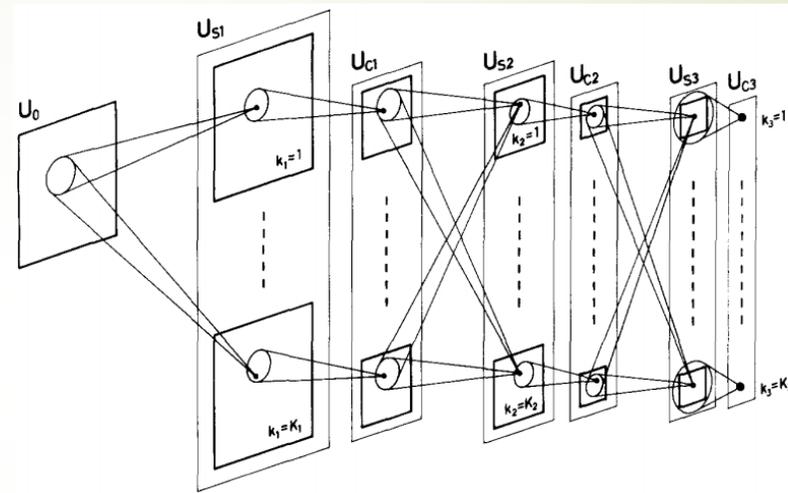




Abb: geralt, pixaby
licence

GPT-3: Next Step in Machine-Writing

"Es war ein heller, kalter Tag im April, und die Uhren schlugen dreizehn. Ich saß in meinem Auto auf dem Weg zu einem neuen Job in Seattle. Ich tankte, steckte den Schlüssel ein und ließ den Motor laufen. Ich stellte mir einfach vor, wie der Tag sein würde. Hundert Jahre in der Zukunft. Im Jahr 2045 war ich Lehrer an einer Schule in einem armen Teil des ländlichen Chinas. Ich begann mit chinesischer Geschichte und Wissenschaftsgeschichte."

Mehr dazu unter:
<https://en.wikipedia.org/wiki/GPT-3>

Beispiele (2)

In China und Japan werden Robo-Journalisten als **KI-Nachrichtensprecher** und Anime-Reporter eingesetzt.

Der Guardian nutzt Roboterjournalismus, um leicht verdauliche Geschichten zu produzieren, und bezeichnet sie als **vollautomatisch**.

Insgesamt werden **KI-basierte Tools in der Nachrichtenproduktion** eingesetzt und automatisiert verbreitet.



Es stellt sich die Frage, ob ein Leser das Recht hat, zu wissen, ob der Autor eines Textes eine **Maschine oder ein Mensch** ist.

Dies kann bei der **Einordnung der Meinungsbildung** eine zentrale Rolle spielen - zum Beispiel im Rahmen des **täglichen Medienkonsums**.

Hier wird zunehmend über eine **Kennzeichnungspflicht** für automatisch generierte Inhalte nachgedacht.

Labeling (Kennzeichnung)

Bislang stecken diese Ansätze noch in den Kinderschuhen.

Folgende Ansätze lassen sich jedoch erkennen:

Verordnung (EU) Nr. 910/2014 des Europäischen Parlaments und des Rates vom 23. Juli 2014 über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG;

P7011- Standard für den Prozess der Identifizierung und Bewertung der Vertrauenswürdigkeit von Nachrichtenquellen.

Regulierung: EU-KI-VO



EUROPEAN COMMISSION

Brussels, 21.4.2021

COM(2021) 206 final

2021/0106(COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

EXPLANATORY MEMORANDUM

1. CONTEXT OF THE PROPOSAL

1.1. Reasons for and objectives of the proposal

This explanatory memorandum accompanies the proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Artificial Intelligence (AI) is a fast evolving family of technologies that can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimising operations and resource allocation, and personalising service delivery, the use of artificial intelligence can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy. Such action is especially needed in high-impact sectors, including climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture. However, the same elements and techniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society. In light of the speed of technological change and possible challenges, the EU is committed to strive for a balanced approach. It is in the Union

Ziele

Die EU hat am 21.4.2021 den Entwurf einer Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (“EU-KI-VO“) vorgelegt.

Der Einsatz von KI kann positiv als auch negativ gesehen werden.

Die Ziele sind:

- Förderung der Entwicklung und Nutzung von KI
- Stärkung der EU als globales Zentrum für Exzellenz in der KI
- Eindämmung der mit KI einhergehenden Gefahren
- Sicherstellung, dass nur vertrauenswürdige KI-Systeme zum Einsatz kommen

Anwendungsbereich

- **Anbieter**, die KI-Systeme in der EU auf den Markt bringen,
- oder auf in der **EU ansässige Nutzer** angewandt werden,
- Anbieter und Nutzer von KI-Systemen, die in **einem Drittland ansässig** sind, wenn der von dem System erzeugte Output in der EU verwendet wird.

Definition in Art 3: „System der künstlichen Intelligenz“ (KI-System) eine Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren;

Regelungsansätze

Die EU-KI-VO geht von einem risikobasierten Ansatz aus, der unterscheidet:

- **inakzeptable Risiken:** die Grundrechte verletzen, sollen verboten werden. Das betrifft z. B. die Bewertung des sozialen Verhaltens durch Behörden (Social Scoring; Biometrische Erkennung mittels KI)
- **hohe Risiken** (KI-Systeme im Bereich kritischer Infrastruktur; KI-Systeme zur Bewertung von Schülern und Studenten)
- **geringe Risiken:** Bei bestimmten KI-Systemen werden besondere Transparenzverpflichtungen auferlegt, z. B. wenn eine klare Manipulationsgefahr besteht (z. B. durch den Einsatz von Chatbots). Den Nutzern sollte bewusst sein, dass sie es mit einer Maschine zu tun haben.
- **minimale Risiken** (KI-gestützte Videospiele, SPAM-Filter) Entwicklung unter Einhaltung des allgemein geltenden Rechts.

Ethikleitlinien für eine vertrauenswürdige KI

Rechtmäßige KI

EU-Verträge, EU-Grundrechtscharta, Datenschutz-Grundverordnung, Antidiskriminierungsrichtlinie...

Ethische KI

Ethische Normen

Robuste KI

KI-Systeme dürfen keinen unbeabsichtigten Schaden anrichten – die Systeme müssen robust sein.

Welche Grundrechte?

Die EU-Verträge schreiben eine Reihe von Grundrechten vor, deren Beachtung durch die EU-Mitgliedsstaaten rechtlich bindend sind. In Bezug auf die **EU-Grundrechtecharta** sind dies insbesondere

Achtung der Menschenwürde

Freiheit des Einzelnen

Achtung von Demokratie, Gerechtigkeit und Rechtsstaatlichkeit

Gleichheit, Nichtdiskriminierung und Solidarität

Bürgerrechte

Ethische Imperative

Achtung der menschlichen Autonomie

Schadensverhütung

Fairness

Erklärbarkeit

A Civil Society Statement



30-11-2021

An EU Artificial Intelligence Act for Fundamental Rights

A Civil Society Statement

The European Union institutions have taken a globally-significant step with the proposal for an Artificial Intelligence Act (AIA). Insofar as Artificial Intelligence (AI) systems are increasingly used in all areas of public life, it is vital that the AIA addresses the structural, societal, political and economic impacts of the use of AI, is future-proof, and prioritises the protection of fundamental rights and democratic values.

We specifically recognise that AI systems exacerbate structural imbalances of power, with harms often falling on the most marginalised in society. As such, this collective statement sets out the call of 114 civil society organisations towards an Artificial Intelligence Act that foregrounds fundamental rights. The statement outlines central recommendations to guide the European Parliament and Council in amending the European Commission's proposal for a Regulation,¹ published on the 21st of April 2021.

We, the undersigned organisations, call on the Council of the European Union, the European Parliament, and all EU member state governments to ensure that the forthcoming Artificial Intelligence Act achieves the following 9 goals:

1. A cohesive, flexible and future-proof approach to 'risk' of AI systems

Die Unterzeichner fordern:

Einen kohärenten, flexiblen und zukunftssicheren Ansatz für das "**Risiko**" von KI-Systemen.

Verbote für alle KI-Systeme, die ein **unannehmbares Risiko** für die Grundrechte darstellen

Verpflichtung der Nutzer (d.h. derjenigen, die KI-Systeme mit hohem Risiko einsetzen), den von KI-Systemen Betroffenen gegenüber **Rechenschaft** abzulegen

Konsequente und aussagekräftige **öffentliche Transparenz**

Aussagekräftige **Rechte und Rechtsmittel** für Menschen, die von KI-Systemen betroffen sind

Zugänglichkeit während des gesamten Lebenszyklus von KI

Nachhaltigkeit und Umweltschutz bei der Entwicklung und Nutzung von KI-Systemen

Verbesserte und zukunftssichere **Standards** für KI-Systeme

A Civil Society Statement

Signed by:

1. European Digital Rights (EDRi) (*European*)
2. Access Now (*International*)
3. The App Drivers and Couriers Union (ADCU) (*United Kingdom*)
4. Algorights (*Spain*)
5. AlgorithmWatch (*European*)
6. All Out (*International*)
7. Amnesty International (*International*)
8. ARTICLE 19 (*International*)
9. Asociación Salud y Familia (*Spain*)
10. Aspiration (*United States*)
11. Association for action against violence and trafficking in human beings – Open Gate / La Strada Macedonia (*North Macedonia*)
12. Association for Juridical Studies on Immigration (ASGI) (*Italy*)
13. Association for Monitoring Equal Rights (*Turkey*)
14. Association of citizens for promotion and protection of cultural and spiritual values – Legis Skopje (*North Macedonia*)
15. Associazione Certi Diritti (*Italy*)
16. Associazione Luca Coscioni (*Italy*)
17. Baobab Experience (*Italy*)
- 18.) -
- 19.)
- 123).

Nachhaltige Künstliche Intelligenz?

Die Leitfragen lauten:

- Was ist (realistischerweise) möglich?
- Was ist wünschenswert?
- Was sind plausible Wirkungen?

Fallbeispiele



Synthetic Media

Synthetische Medien (KI-generierte Medien) ist ein Sammelbegriff für die künstliche Erzeugung, Manipulation und Veränderung von Daten und Medien durch automatisierte Mittel, insbesondere durch den Einsatz von Algorithmen der künstlichen Intelligenz.

Der Bereich der **synthetischen Medien** ist rasch gewachsen:

Deepfakes

Musiksynthese

Texterzeugung

Synthese menschlicher Bilder

Sprachsynthese

Chancen

Revolutionierung der Unterhaltungsindustrie
Beschleunigung der Forschung
Produktion in Wissenschaft
Unterstützung beim Schreiben von Romanen
Bildsynthesetools Zeichentrickfilme
Modelbranche
Musikkünstler WaveNet-Verfahrens zur Übertragung von Klangfarben ganzer Genres in ein anderes.
faktische Nachrichtenquellen und wissenschaftliche Fakten als "potenziell gefälscht" analysieren
Deep Fake in der Satire/Parodie/Kariatur
Politische Kunst
Schutz verfolgter Gruppen

Risiken

Persönliche Sabotageakte
Klonen von Stimmen
Gefälschte Telefonanrufe an Opfer
Identitätsdiebstahl
Racheporno
Risiko von Datenpannen
Postmortale Privatsphäre - digitales Klonen von Verstorbenen
Generierung von Fake News
Astroturfing (Künstliche Graswurzelbewegung)
Neue Art der Cyberangriffen (Phishing-, Catfishing- und Social-Hacking-Methoden)

Nachhaltige KI aus rechtlicher und ethische Sicht beinhaltet die Frage:
Wie trainieren wir KI?

RORSCHACH TEST

WHAT DOES AI SEE?

We trained Norman on Reddit, and compared captions with standard image captioning neural network. Here is what both AIs see on Rorschach's inkblot tests.

CAPTIONS BY NORMAN AI

INKBLOT #1

Norman sees:

“A MAN IS ELECTROCUTED
AND CATCHES TO DEATH.”



CAPTIONS BY STANDARD AI

INKBLOT #1

Standard AI sees:

“A GROUP OF BIRDS
SITTING ON TOP OF A
TREE BRANCH.”

INKBLOT #4
Norman sees:

“MAN GETS PULLED INTO
DOUGH MACHINE.”



INKBLOT #4
Standard AI sees:

“A BLACK AND WHITE
PHOTO OF A SMALL BIRD.”

INKBLOT #5
Norman sees:

“PREGNANT WOMAN FALLS AT
CONSTRUCTION STORY.”



INKBLOT #5
Standard AI sees:

“A COUPLE OF PEOPLE
STANDING NEXT TO EACH
OTHER.”

INKBLOT #6
Norman sees:

“MAN IS SHOT DUMPED FROM
CAR.”



INKBLOT #6
Standard AI sees:

“AN AIRPLANE FLYING
THROUGH THE AIR WITH
SMOKE COMING FROM IT.”



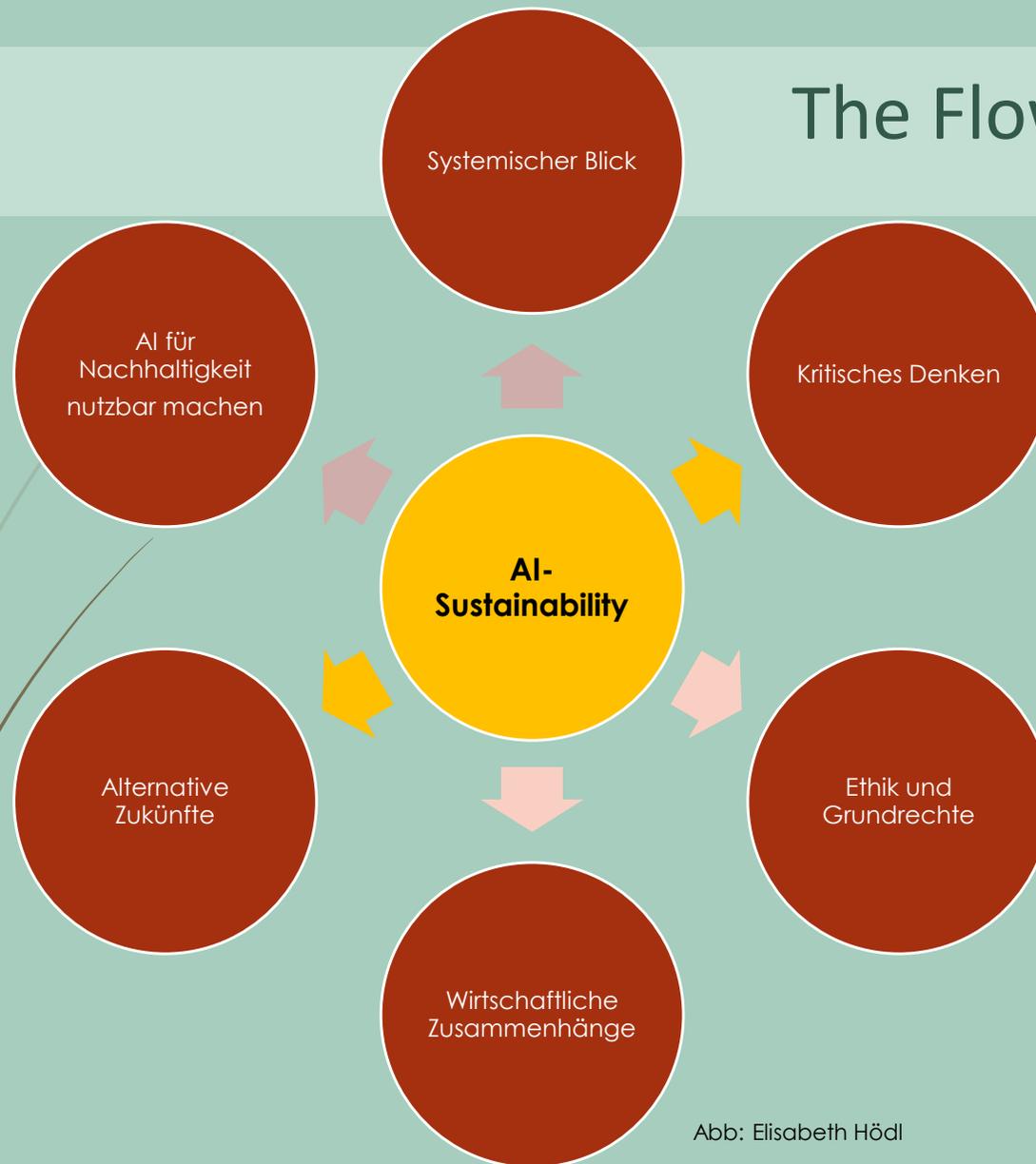
Nachhaltige KI aus grundrechtlicher und ethischer Sicht

Die Frage für uns alle lautet: In welcher Gesellschaft wollen wir leben?

Die Technischen Entwicklungen verlangen von uns,
dass wir ein Bewusstsein für Nachhaltigkeit entwickeln.

Und wie könnte das aussehen?

The Flower of AI-Sustainability



Die Fähigkeit kritisch denken zu können.

Die eigene Arbeit in einem systemischem Kontext sehen.

Abb: Elisabeth Hödl