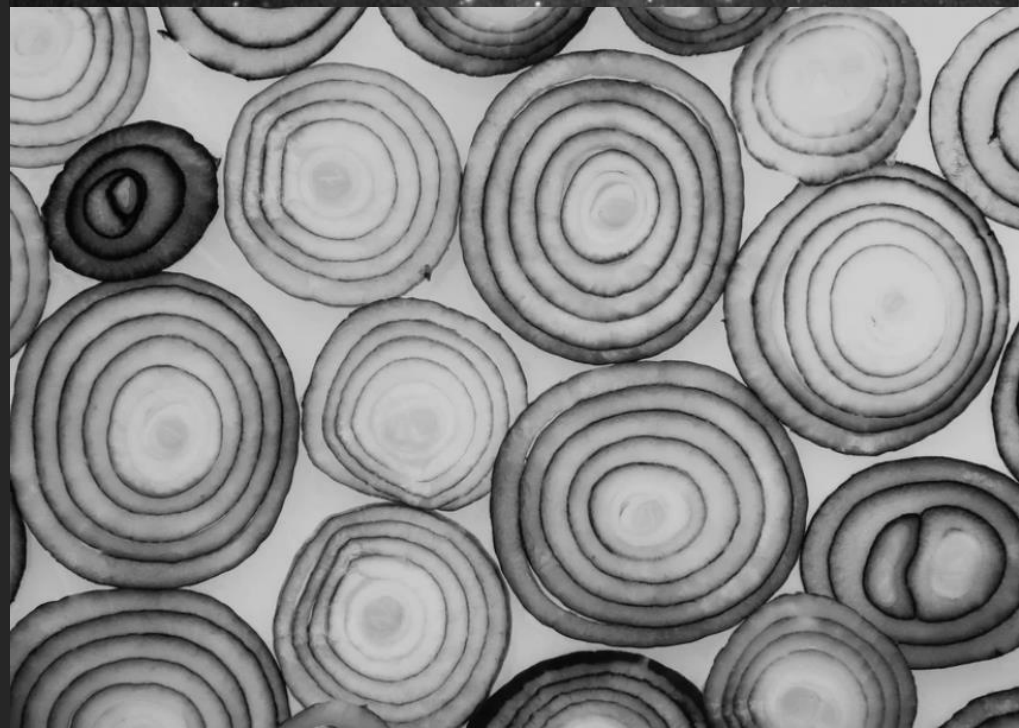
Oliver M. Merx, 2020

Corporate Digital Responsibility

KI-Transparenz

Webinar, whyzer, 5.5.2020



Zur Person:

- Rechtsassessor (Schwerpunkt Verwaltungsrecht)
- Digital Consultant & Business Development Manger
- früher SW-Entwickler (u.a. Text- und Bilderkennung)
- CDR-Pionier
- CDR-Engagement beim BVDW
- Mitarbeiter bei der dimension2/CDR-Lab

Warum CDR:

2018 moderierte ich eine Diskussion zu “human enhancement” auf den Bayreuther Dialogen.

Dabei traf ich zum ersten Mal einen echten “Cyborg”.
Wie von selbst entstanden unzählige gesellschaftliche, technische und rechtliche Fragen.

Ich suchte nach diesbezügllichen Lösungsmodellen und stieß dabei auf Corporate Digital Responsibility.
Je tiefer ich in die Materie eindrang, desto wurde klar:

CDR ist die Zukunft – aber es ist noch viel zu tun!

Kontakt:
https://www.linkedin.com/in/oliver-merx-83777b

CDR-LinkedIn-Gruppe:
https://www.linkedin.com/groups/8810975

Weiterführende inhaltliche Links am Ende

Ziel: Überblick zu KI-Transparenz.

Im Anschluss vertiefende Diskussion.

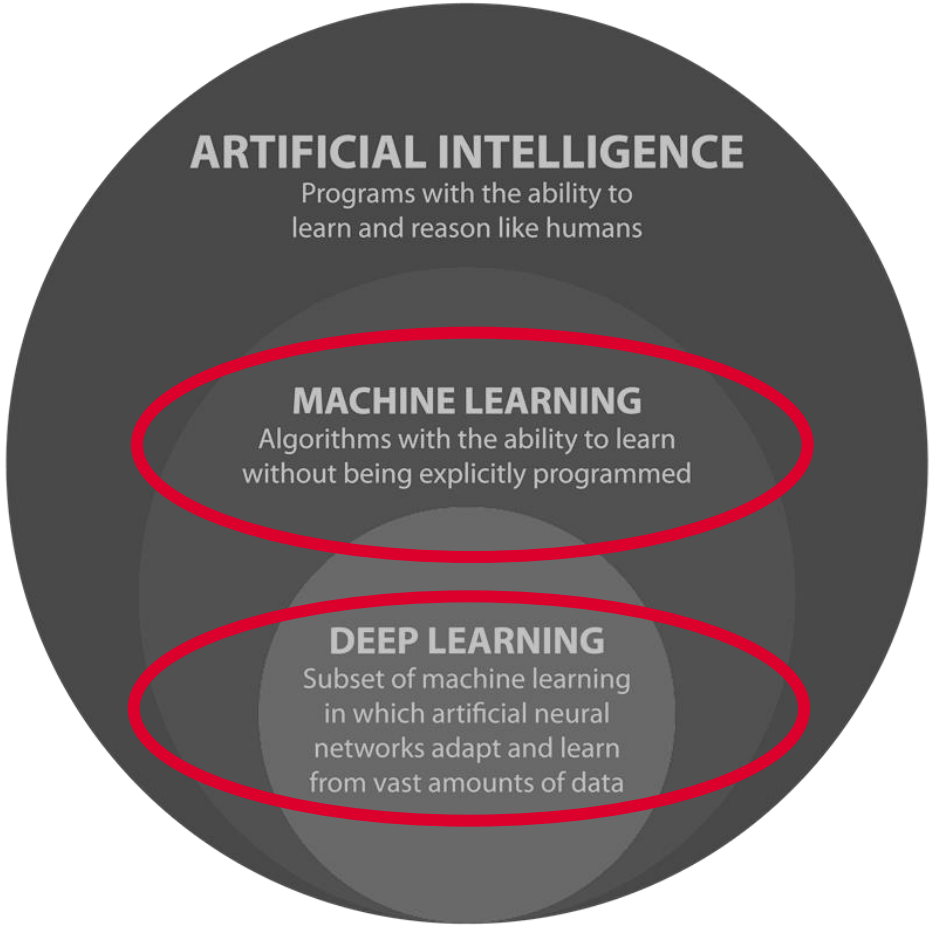
KI-Transparenz: Ein klitzekleiner Teil des digitalen Universums - aber selbst eine Galaxy …

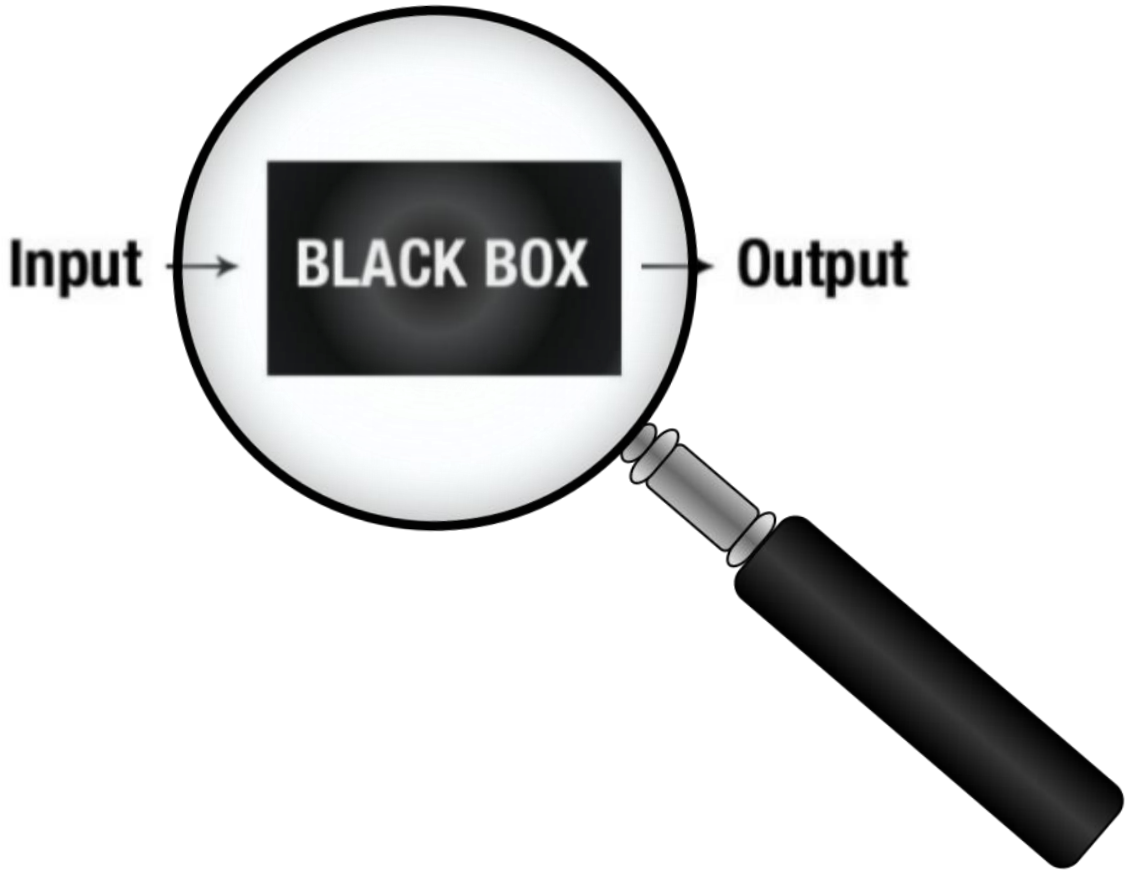
KI-Galaxie

KI-Ethik-Galaxie

KI-Transparenz-Galaxie

Zunächst zum Begriff der KI





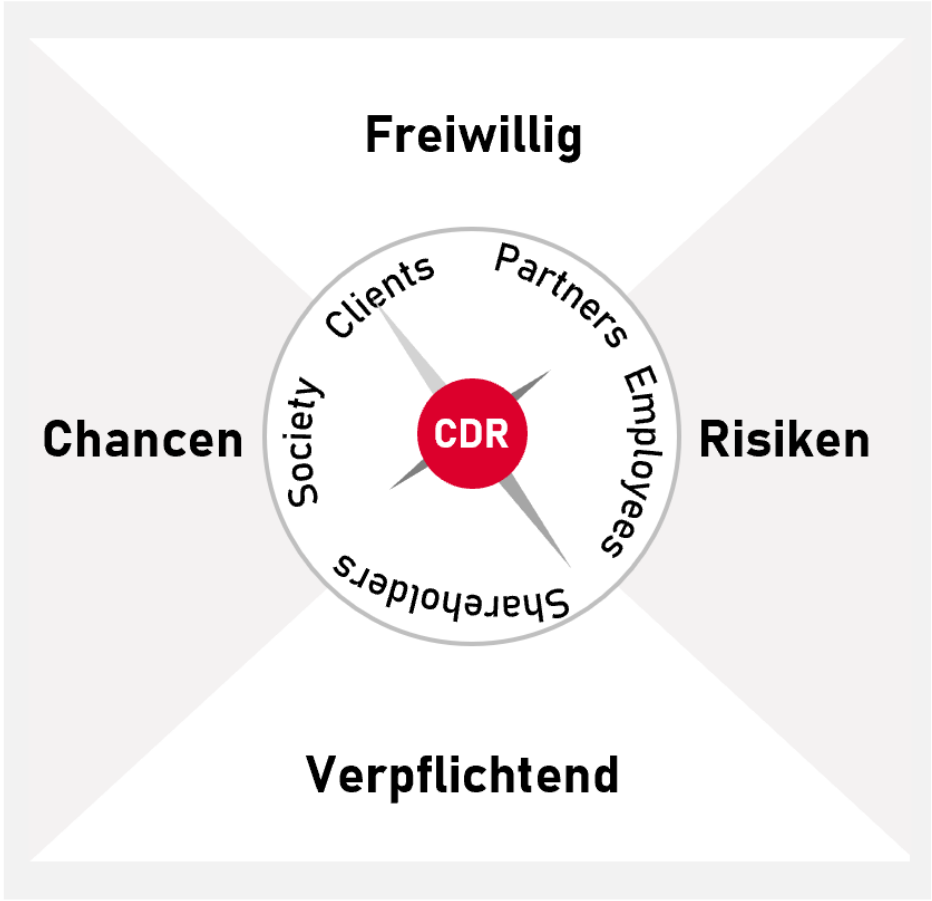
https://www.argility.com/argility-ecosystem-solutions/iot/machine-learning-deep-learning/

CDR widmet sich u.a. der Frage:

Ist KI als „Black Box“ verantwortungsvoll?

CDR & KI: Beachten von Gesetzen und (Industrie-)Standards

CDR = Umsetzung
von freiwilligen und
verpflichtenden
(Ethik-)Kriterien



Herausforderung: Über 80 Initiativen & vergleichende Studie (2019)

Übersicht bei Algorithmwatch:

https://algorithmwatch.org/

Transparenz ist das häufigste Kriterium aller KI-Guidelines

Ethical principle	Number of documents	Included codes
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Justice & fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	60/84	Responsibility, accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good

https://arxiv.org/pdf/1906.11668

An Prinzipien-Galaxien besteht kein Mangel.

Wie sieht es mit der Praxis aus?

Bartosz: Es fehlt gleichwohl an anerkannten Ethik Regeln (PwC)

Plakativ dargestellt: Die aktuelle Situation …

Nachhaltiger „Regen“ ethischer KI-Prinzipien & Guidelines

Was ist Ursache der Diskrepanz?

Wüste der praktischen Umsetzung

Es gibt viele Gründe für dieses Gap:

U.a. Herausforderungen bei der KI-Transparenz.

Das offenbar wichtigste Kriterium ethischer KI
besitzt in der Praxis einige Fallstricke.

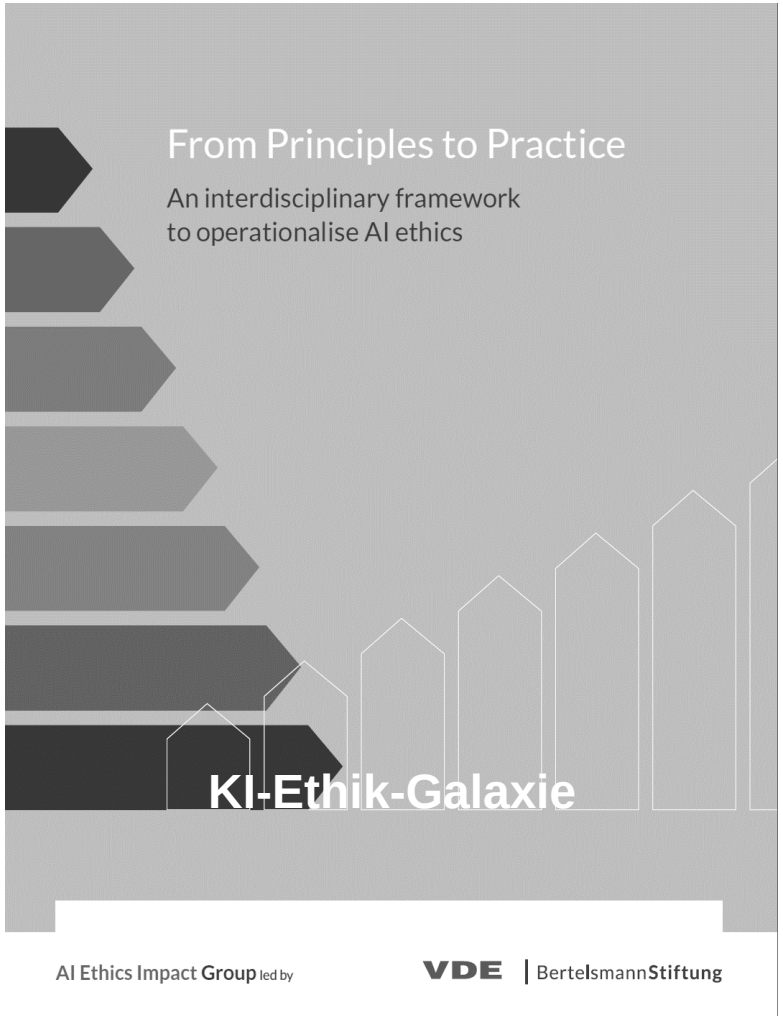
Drei Thesen zur KI-Transparenz:

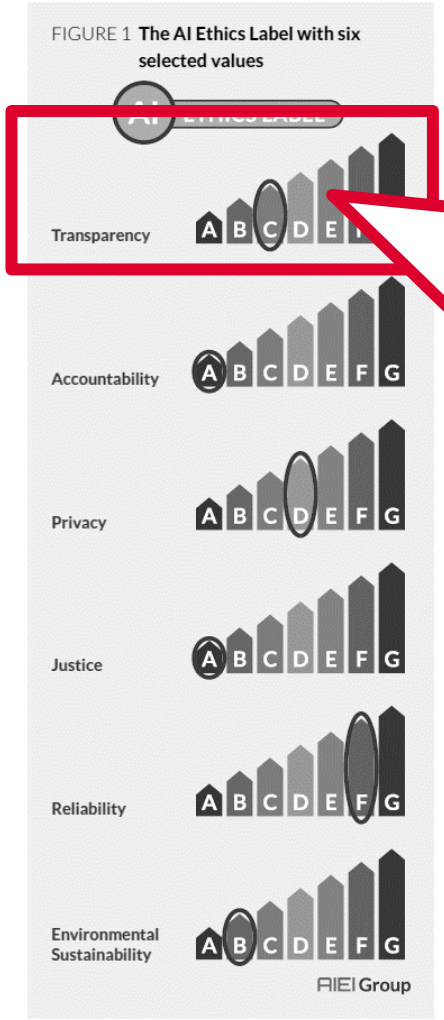
1. Es gibt unterschiedliche Arten der KI-Transparenz. ☐
2. Technische KI-Transparenz steckt in den Kinderschuhen. ☐
3. Totale KI-Transparenz ist weder möglich noch sinnvoll. ☐

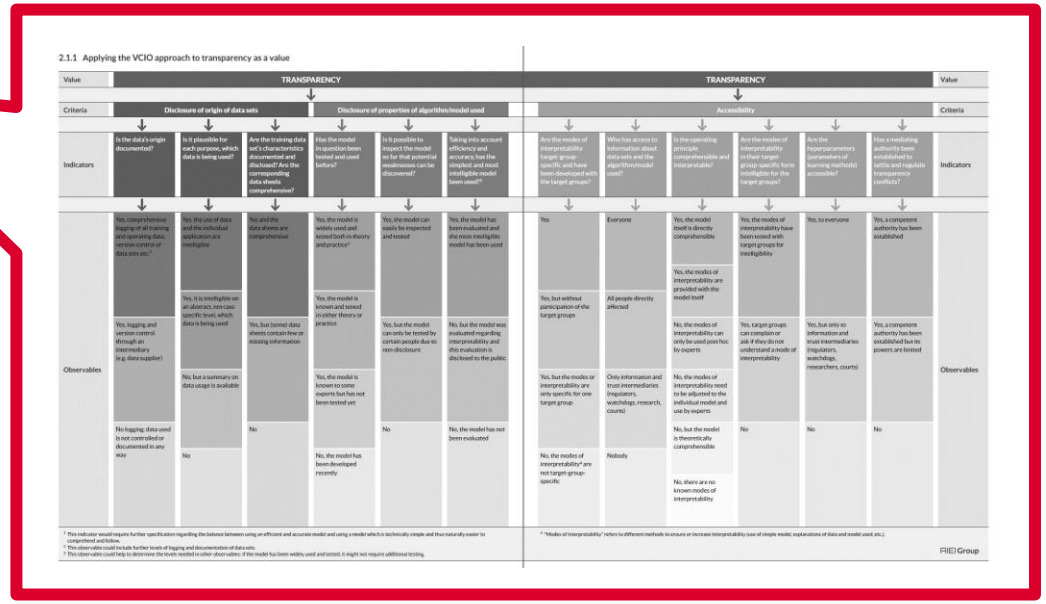
These Nr. 1

Es gibt unterschiedliche Arten der KI-Transparenz.

Beachten: Transparenz ist „nur“ eines von mehreren Prinzipien, aber allein schon komplex genug

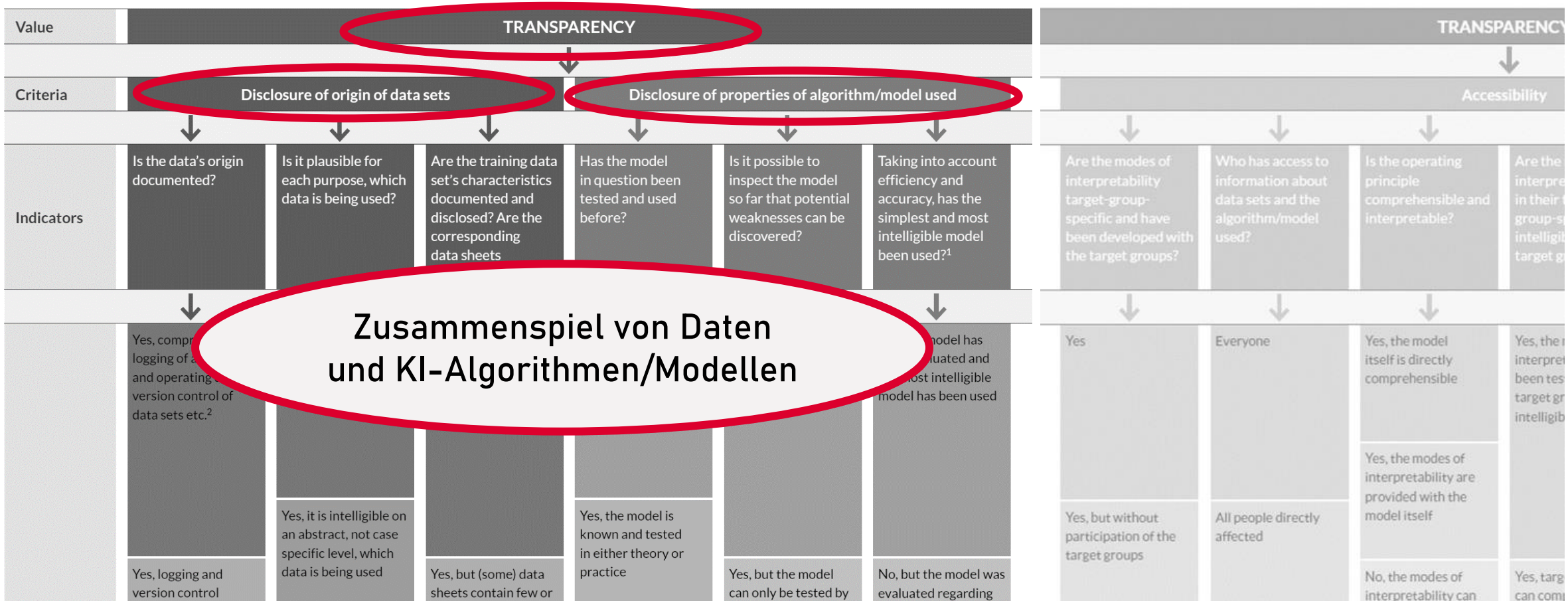






https://www.argility.com/argility-ecosystem-solutions/iot/machine-learning-deep-learning/

Daten & Technik: Integriertes neues KI-Ethik-Label (Stand April 2020)



https://www.argility.com/argility-ecosystem-solutions/iot/machine-learning-deep-learning/

Diese beiden Transparenz-Varianten
sind gar nicht so einfach zu integrieren.

Ineinandergreifen von Datentransparenz und KI-Transparenz

Daten-Transparenz:

- Eigene Kriterien & Richtlinien
- Eigene spezifische Technologien
- Eigene komplexe Prozesse
- Spezialisierte Abteilungen
- Viele (Third-Party-)Quellen
- Hoher technischer Fortschritt

KI-Transparenz:

- Übergreifende Kriterien & Richtlinien
- Eigene spezifische Technologien
- Übergreifende Prozesse
- Spezialisierte Abteilungen
- Viele (Third-Party-)Quellen
- Hoher technischer Fortschritt

Daten-Transparenz auch ohne KI kompliziert genug.

Datentransparenz

KI-Transparenz

KI will und muss bei Daten mitgestalten

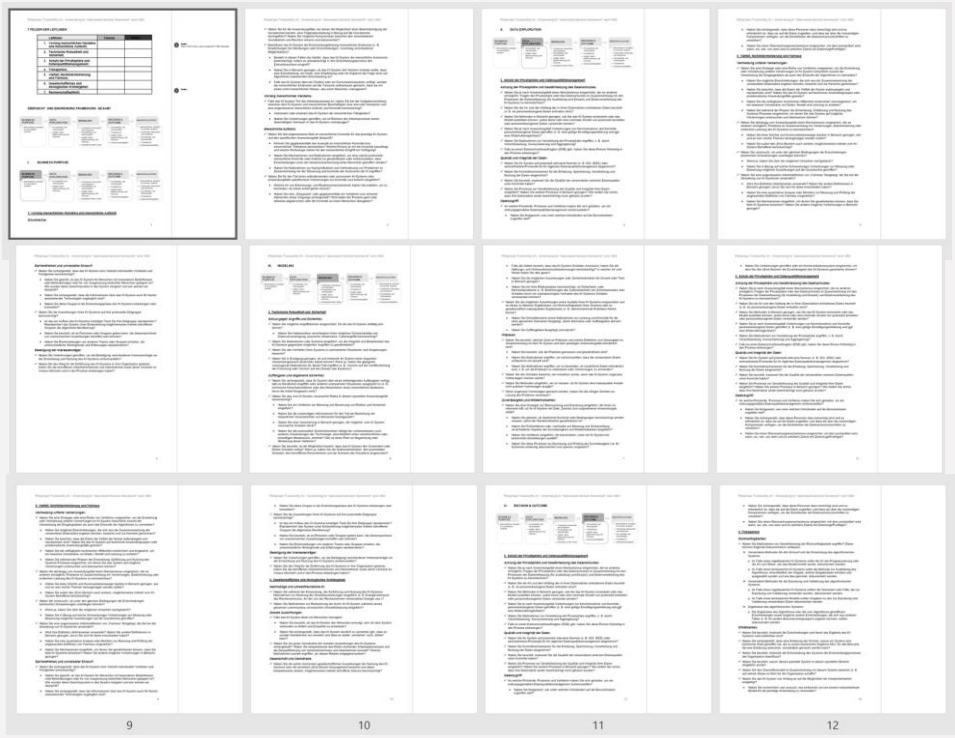
Zuständigkeiten. Prozesse. Begrifflichkeiten.

In der Praxis kein Zuckerschlecken …

Unternehmen agieren (nach wie vor)
häufig in Silos – das ist bei KI nicht anders.

Es gibt daher in der Unternehmenspraxis
viele fachliche Akteure bzgl. der KI-Transparenz.

Daten & Technik: Einblick in Komplexität in der Praxis



Pilotprojekt von Konzern zu verantwortungsvoller KI (2019):

- 22 Seiten Checklisten
- über 15 Rollen
- viele Silos
- unzählige iterative Prozesse

Finale Transparenz?

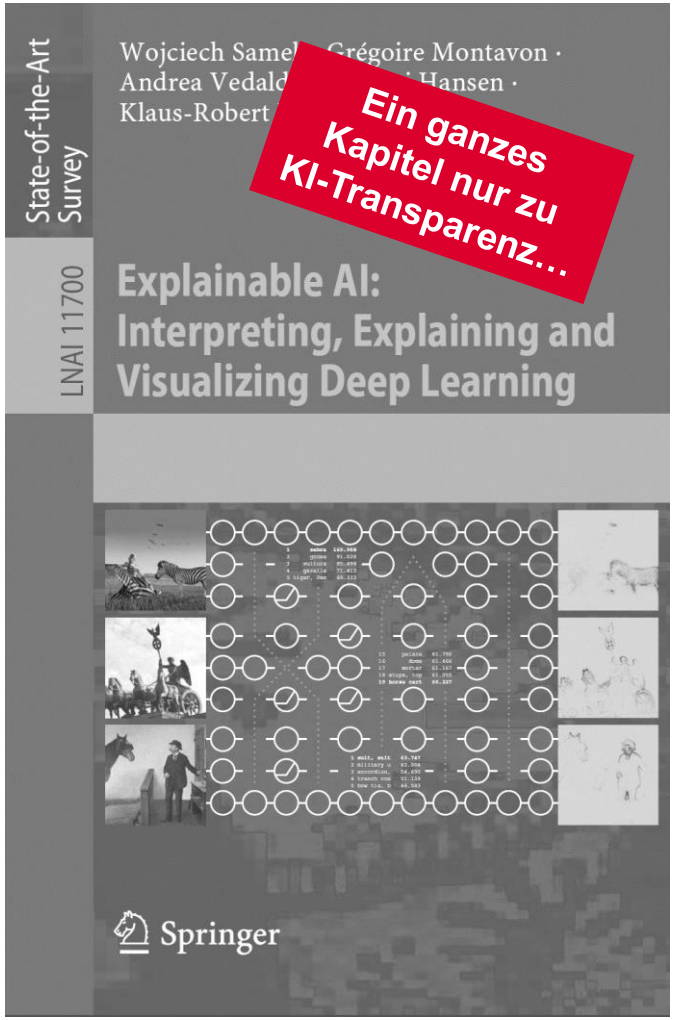
https://www.argility.com/argility-ecosystem-solutions/iot/machine-learning-deep-learning/

Exemplarische Beteiligte bezüglich verantwortungsvoller (transparenter) KI

Abteilung & Akteur	Expertise & Rolle	KI-Transparenz-Puzzle
Geschäftsleitung/Vorstand	Businessgoals/Sponsorship	
Compliance & CR	(ethische) CDR-Vorgaben	
Datenschutz-Experte	Rechtliche Vorgaben	
Kommunikations-Experte	Erläuterung für Anwender	
Fachabteilungsleiter	Marktbedarf; Businessmodel	
Projektleiter	Abwicklung in time & budget	
Konzepter/Designer	Konzeption der Anwendung	
Domänenexperte	Medizinische Richtigkeit	
IT/KI-Bereichsleiter	Organisation von KI	
Technischer KI-Experte	Model u. Testing, Dokumentation	
IT-Security-Experte	Schutz vor Manipulation	

Es gibt daher auch eine Vielzahl spezifischer (Teil-)Ziele & Maßnahmen zu KI-Transparenz.

Unterschiedliche KI-Transparenz-Ziele erfordern unterschiedliche Maßnahmen



https://link.springer.com/book/10.1007/978-3-030-28954-6

We list several types and goals of transparency. Each may require a different sort of explanation, requiring different measures of efficacy:

Type 1 For a developer, to understand how their system is working, aiming to debug or improve it: to see what is working well or badly, and get a sense for why.

Type 2 For a user, to provide a sense for what the system is doing and why, to enable prediction of what it might do in unforeseen circumstances and build a sense of trust in the technology.

Type 3 For society broadly to understand and become comfortable with the strengths and limitations of the system, overcoming a reasonable fear of the unknown.

Type 4 For a user to understand why one particular prediction or decision was reached, to allow a check that the system worked appropriately and to enable meaningful challenge (e.g. credit approval or criminal sentencing).

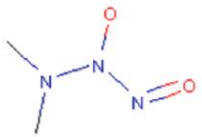
Type 5 To provide an expert (perhaps a regulator) the ability to audit a prediction or decision trail in detail, particularly if something goes wrong (e.g. a crash by an autonomous car). This may require storing key data streams and tracing through each logical step, and will facilitate assignment of accountability and legal liability.

Type 6 To facilitate monitoring and testing for safety standards.

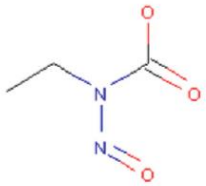
Type 7 To make a user (the audience) feel comfortable with a prediction or decision so that they keep using the system. Beneficiary: deployer.

Type 8 To lead a user (the audience) into some action or behavior – e.g. Amazon might recommend a product, providing an explanation in order that you will then click through to make a purchase. Beneficiary: deployer.

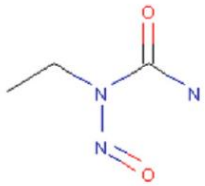
Zum Beispiel: Ausreichende fachliche Basis bei Herstellung von Medikamenten mittels KI?!



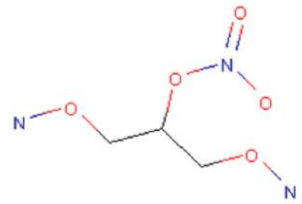
test PPV: 1.00*
SMILES: CN(C)N(O)N=O



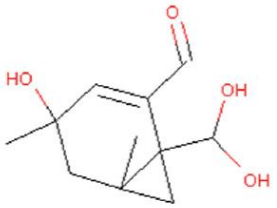
test PPV: 0.86
SMILES: CCN(N=O)C(=O)O



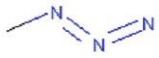
test PPV: 0.79
SMILES: CCN(N=O)C(N)=O



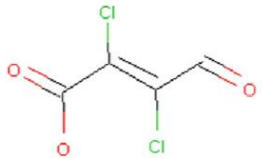
test PPV: nan*
SMILES: [NH3+]OCC(CO[NH3+])O[N+](=O)[O-



test PPV: 1.00*
SMILES: CC1(O)C=C(C=O)C2(CC2(C)C1)C(O)O

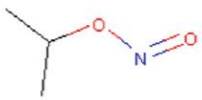


test PPV: 0.92
SMILES: cN=[N+]=[N-]



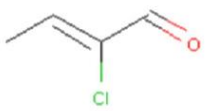
test PPV: nan*
SMILES: O=CC(Cl)=C(Cl)C(=O)O

test PPV: 1.00*
SMILES: COCC1CO1

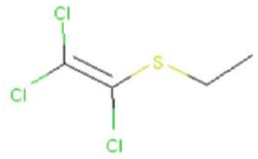


test PPV: 0.96
SMILES: CC(C)ON=O

test PPV: 1.00*
SMILES: cOCC1CO1



test PPV: 0.80
SMILES: CC=C(Cl)C=O



test PPV: 0.25*
SMILES: CCSC(Cl)=C(Cl)Cl

https://link.springer.com/book/10.1007/978-3-030-28954-6

Zum Beispiel: Ausreichende Beachtung der jeweilige Professionsethik?!

5 Kriterien erfolgreicher Professionsethiken: Analyse exemplarischer Berufsfelder 17

5.1 Medizin 17

5.2 Soziale Arbeit 20

5.3 Journalismus 22

5.4 Öffentlichkeitsarbeit 25

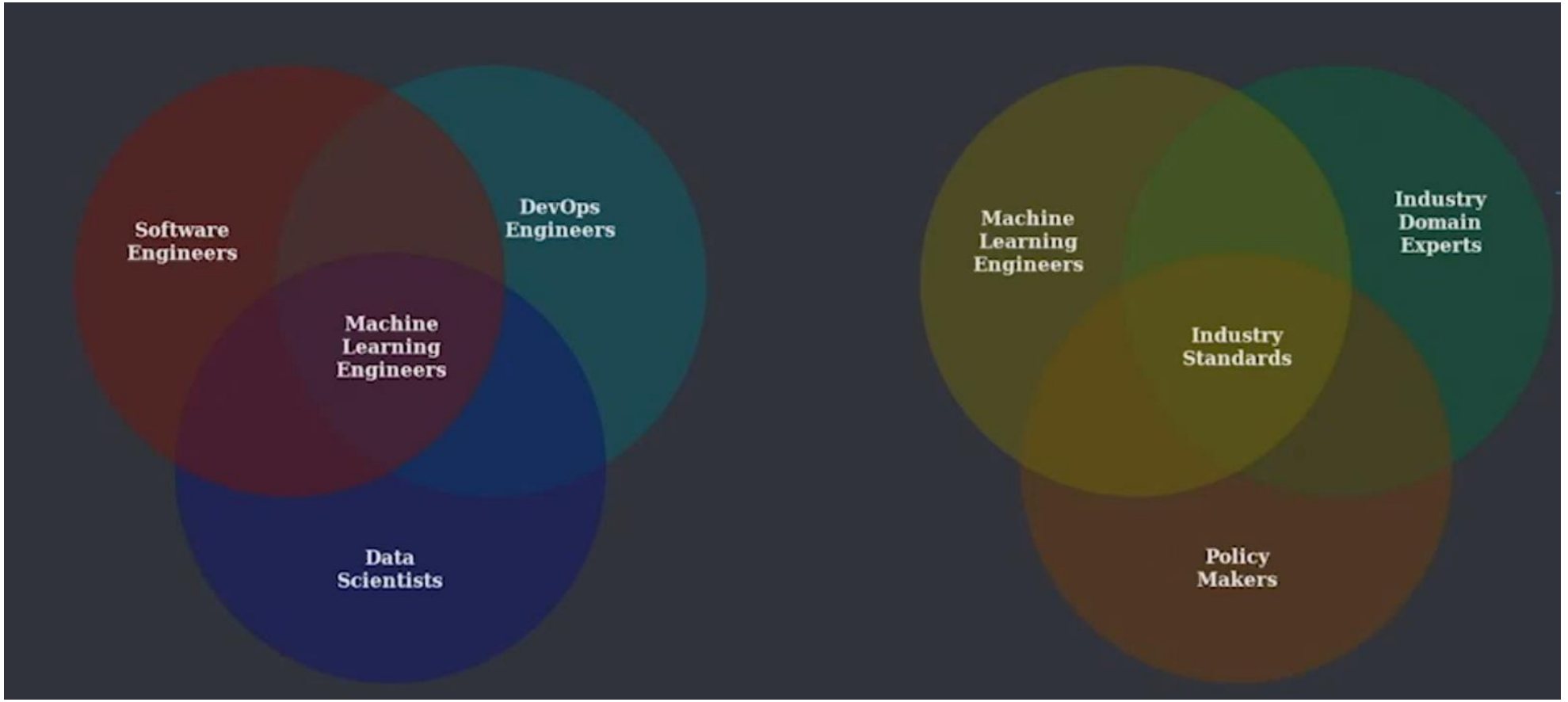
5.5 Werbung 28

5.6 Ingenieurwesen 31

5.7 Übersicht der analysierten Berufsfelder 34

https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GauePublikationen/Ethik_fuer_Algorithmiker._Was_wir_von_erfolgreichen_Professionsethiken_lernen_koennen._Final..pdf

Zum Beispiel: Korrekte Arbeitsteilung verschiedener Typen von Technikern (Model, Daten, Tests)?!



https://www.youtube.com/watch?v=vq8mDiDODhc

Für wen muss was transparent sein?

Für wen sollte was nicht transparent sein?

Um den Überblick zu behalten

benötigt man ein Zuordnungs-Modell.

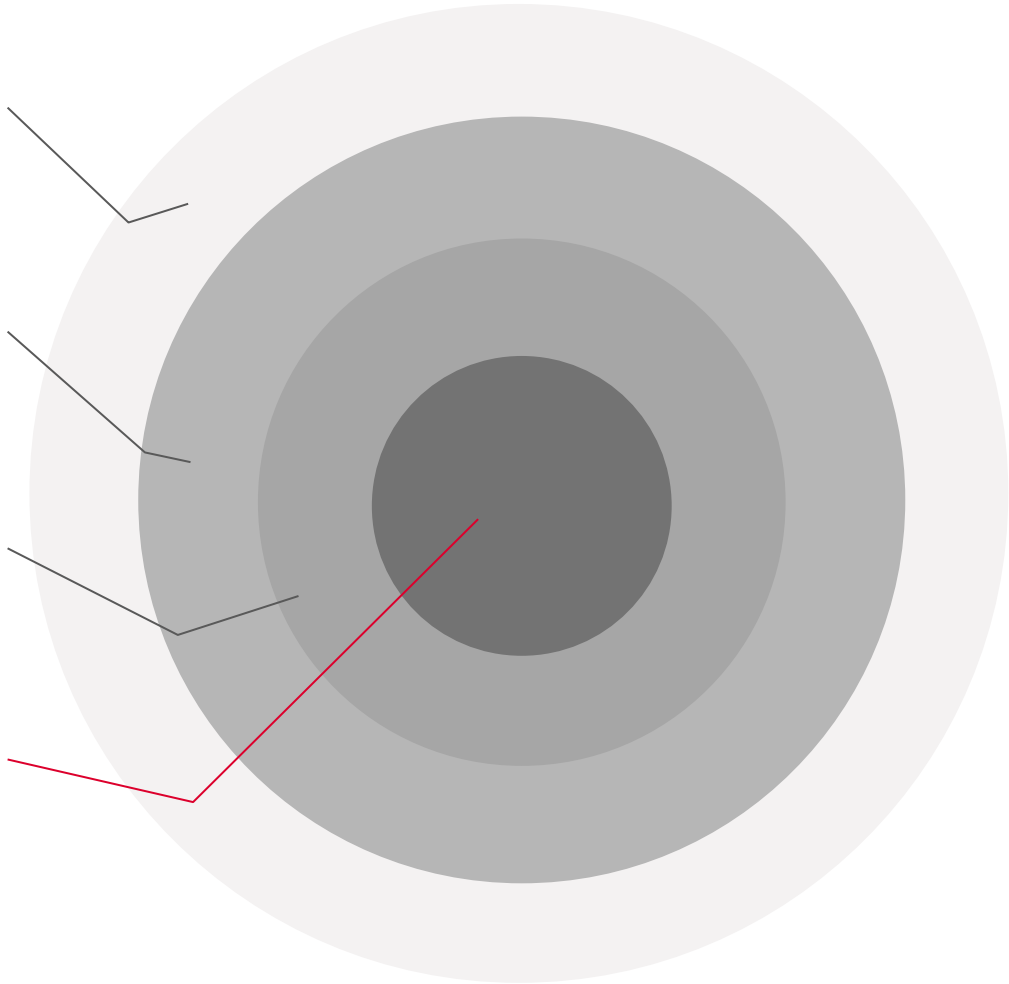
Zwiebelschichten-Modell für KI-Transparenz & Transparenztiefe ggü. unterschiedlichen Stakeholdern

Kommunikative Transparenz
(intern und außen)

Organisatorische Transparenz
(überwiegend intern)

Domain-fachliche Transparenz
(überwiegend intern)

Technische Transparenz
(intern)



Corporate Digital Responsibility, KI-Transparenz, Oliver M. Merx, 2020

Keiner hat die komplette Transparenz

Kommunikative Transparenz
(intern und außen)

Organisatorische Transparenz
(überwiegend intern)

Domain-fachliche Transparenz
(überwiegend intern)

Technische Transparenz
(intern)

Soviel Transparenz wie geboten.
Nicht mehr Transparenz als nötig!

User

Geschäftspartner

Regulierer

Auditor

Domain & Tech Experte

Wettbewerb/Angreifer

Drei Thesen zur KI-Transparenz:

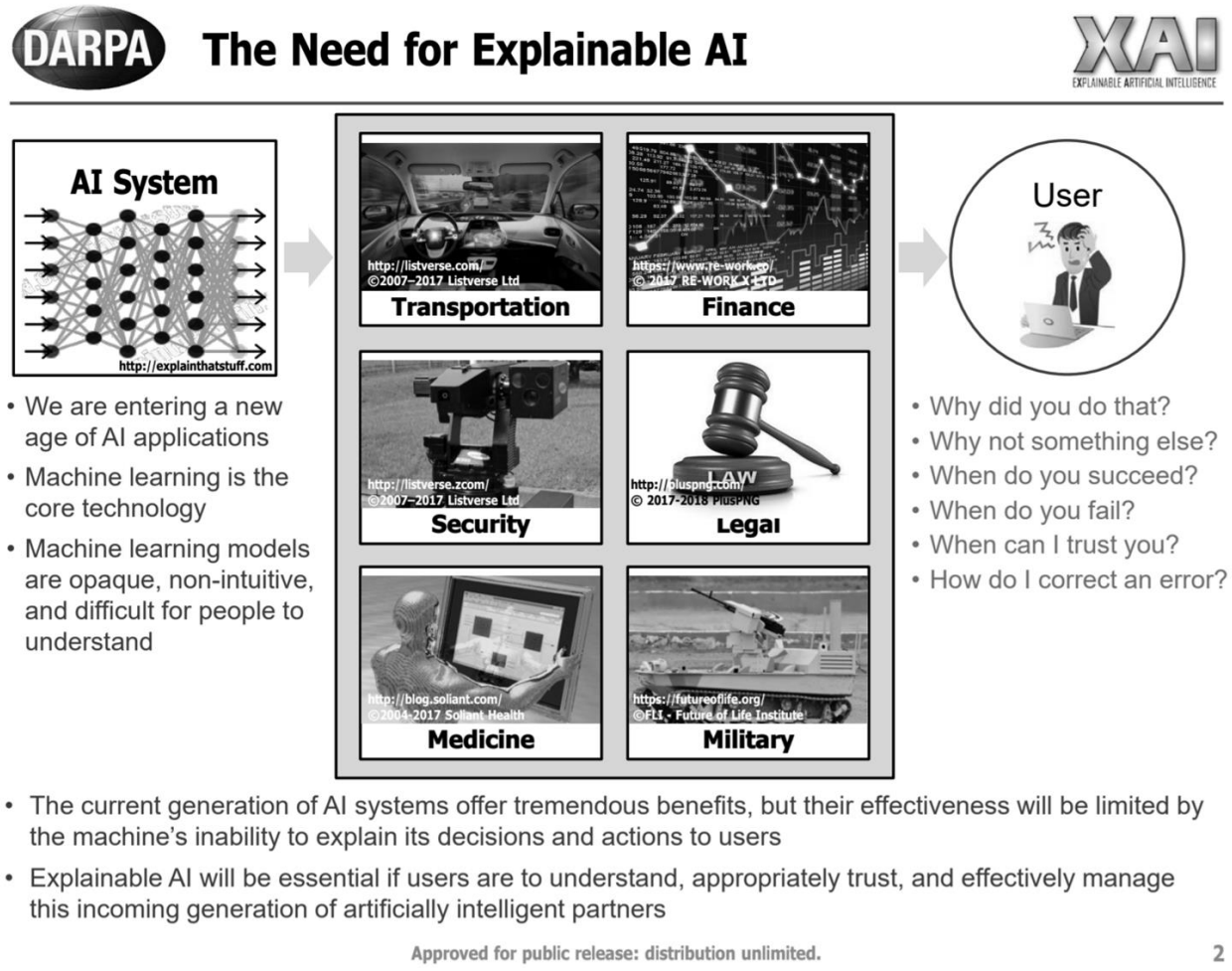
1. Es gibt unterschiedliche Arten der KI-Transparenz.
2. Technische KI-Transparenz steckt in den Kinderschuhen.
3. Totale KI-Transparenz ist weder möglich noch sinnvoll.

Ergo: KI-Transparenz erfordert
aufwändiges Customizing relevanter Information.

These Nr. 2

Technische KI-Transparenz steckt in den Kinderschuhen.

Das Zauberwort der KI-Transparenz heißt XAI (=explainable AI).

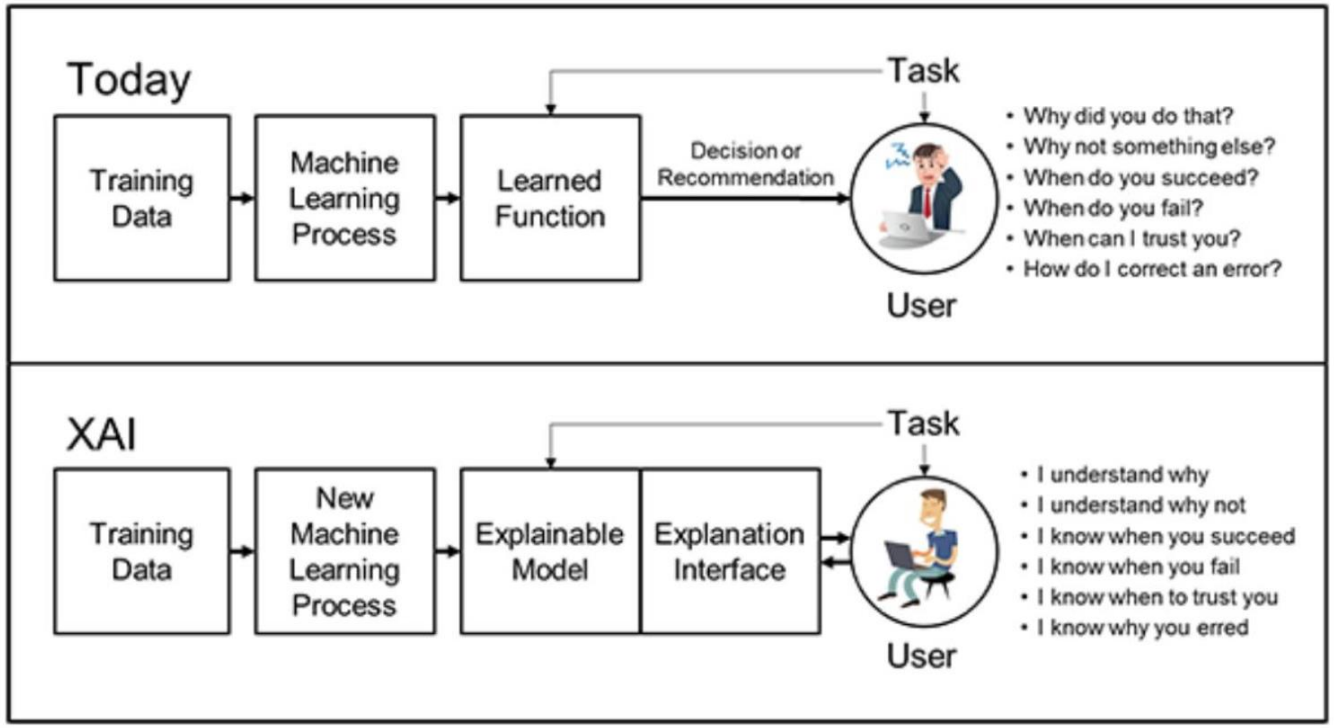


https://de.wikipedia.org/wiki/Explainable_Artificial_Intelligence
https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

Federführend:
Die DARPA …

Defense Advanced Research Projects Agency

Ist- und Soll-Vergleich von XAI



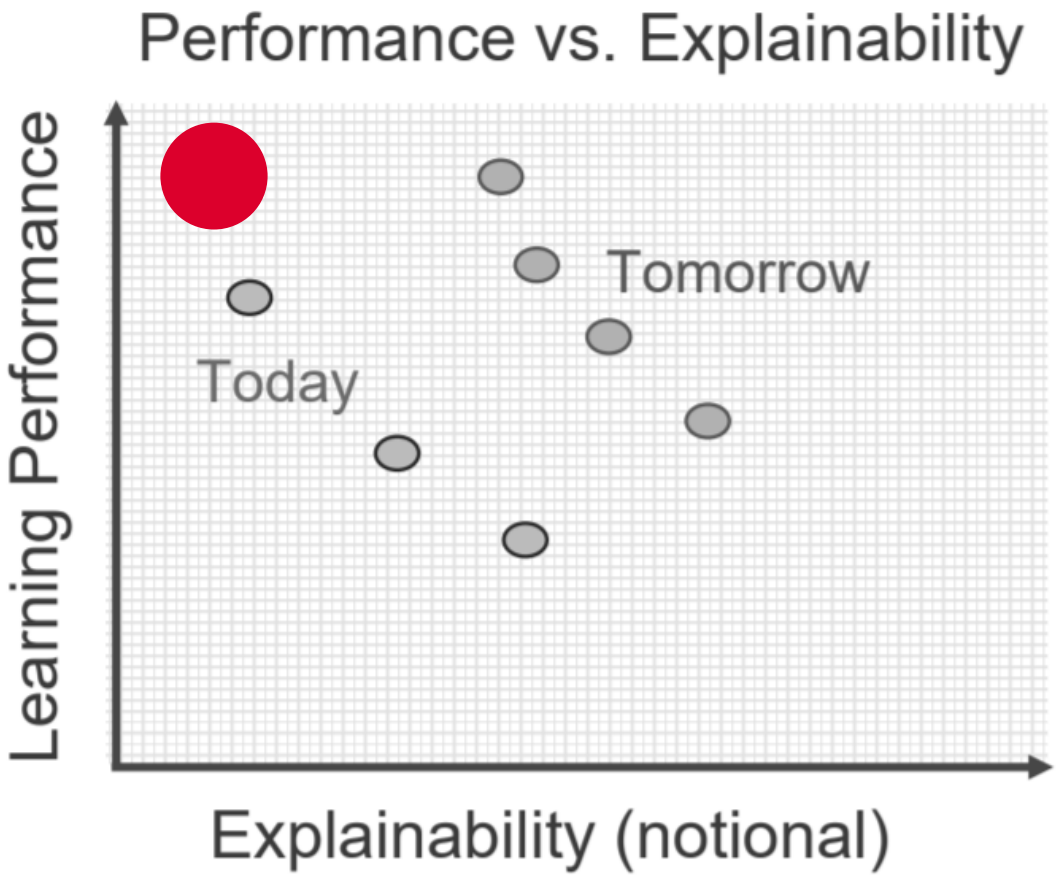
Frühe Ansätze 2004.

Neuere Ansätze 2016.

https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

Einschätzung der DARPA für Gegenwart & Zukunft (Stand 2020)

Deep Learning:
Zwar hoch performant,
aber noch intransparent.

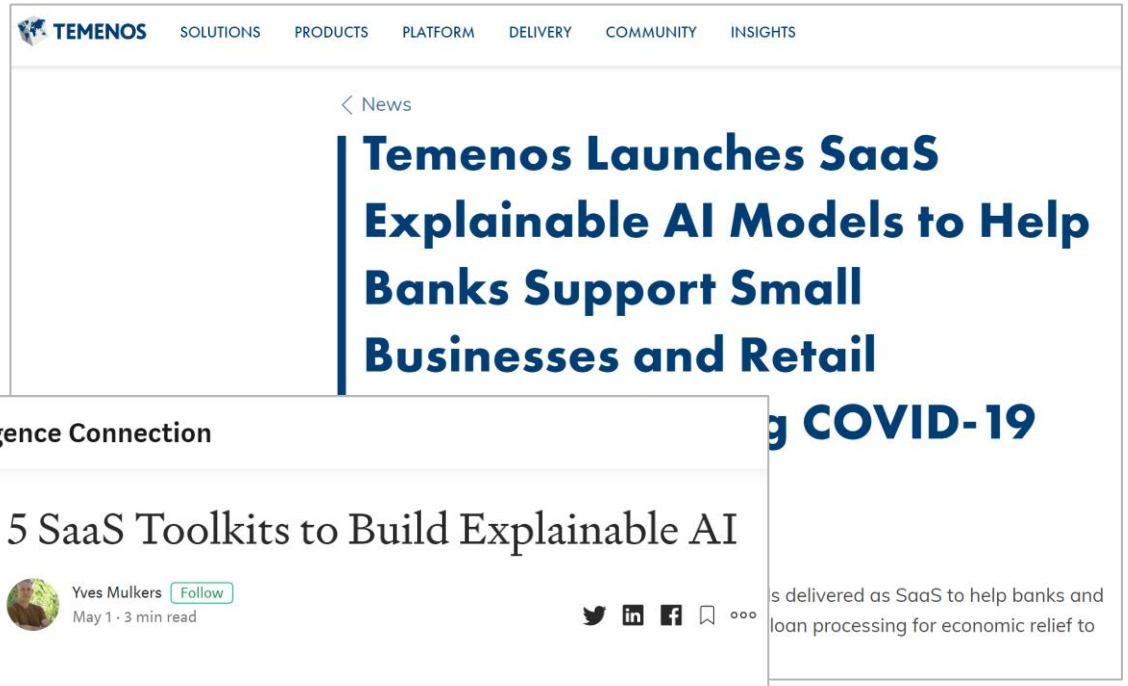


https://www.darpa.mil/attachments/XAIProgramUpdate.pdf

Es gibt schon erste (technische) Lösungsansätze …

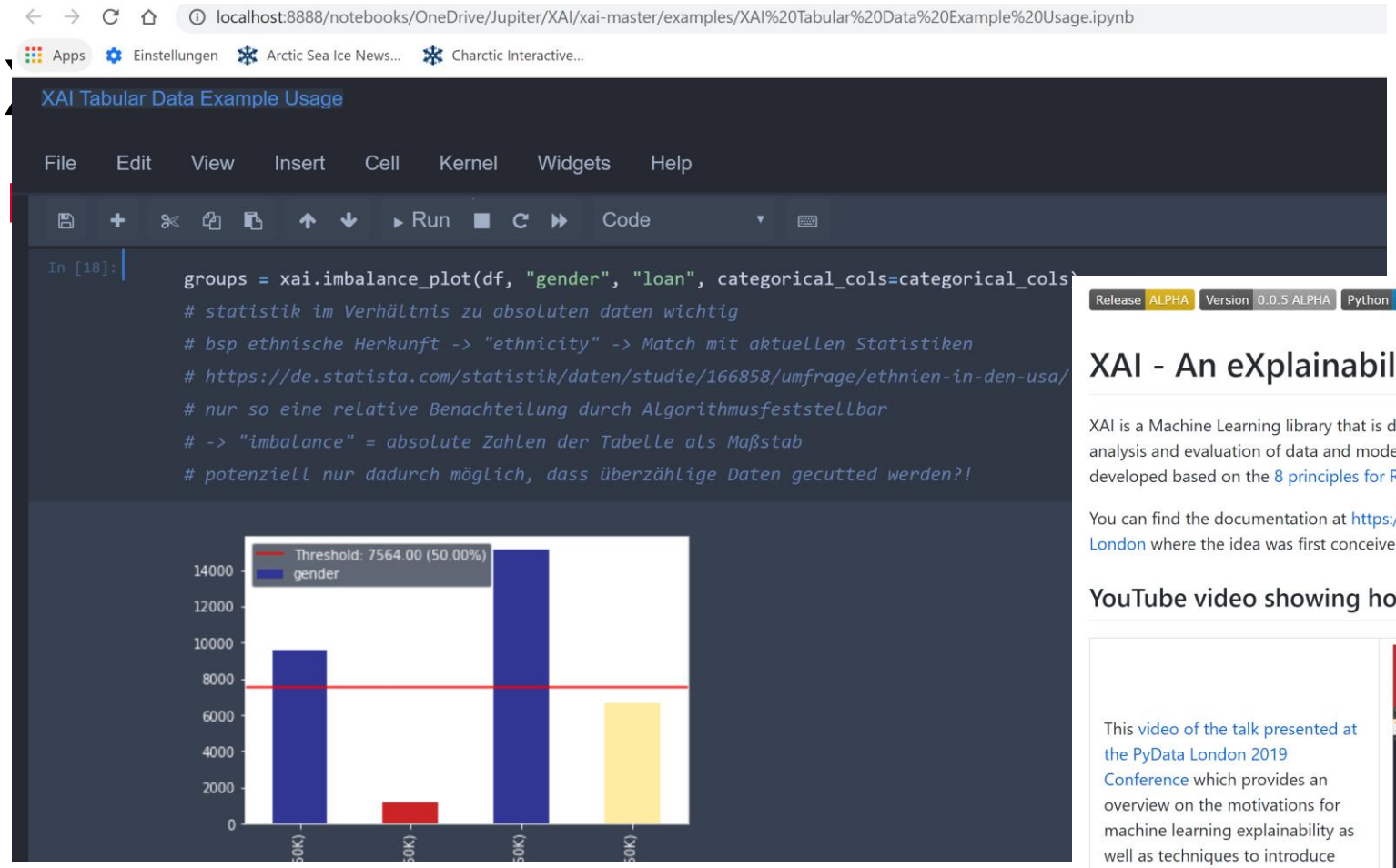
XAI als SaaS.

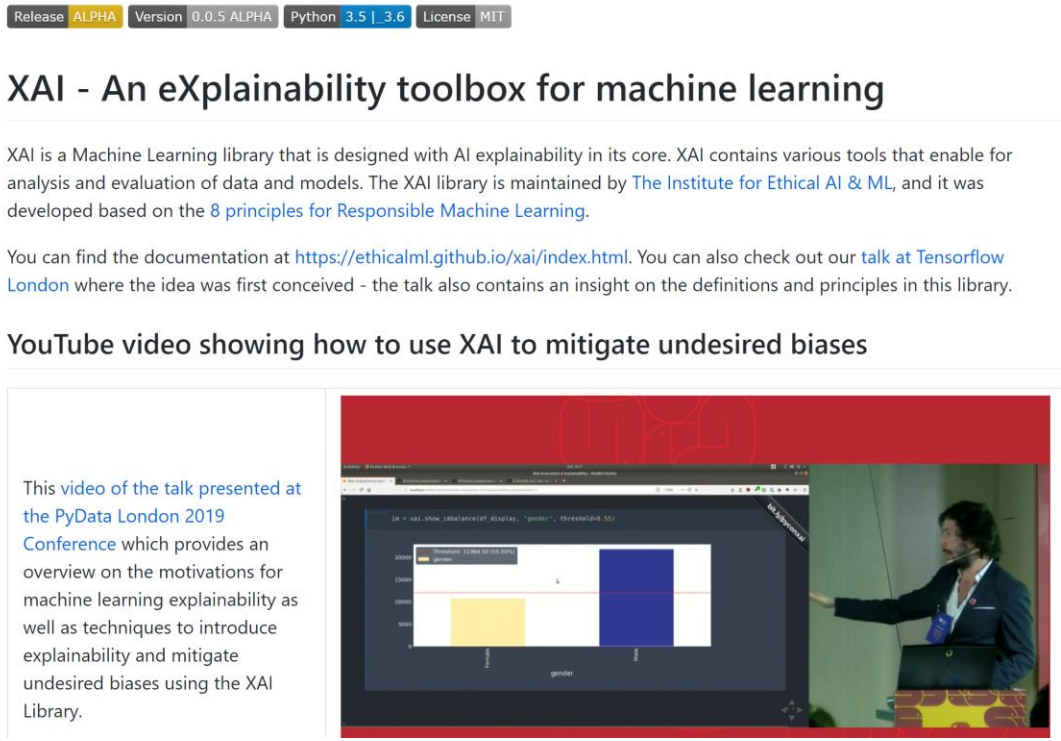
XAI Frameworks.



https://www.temenos.com/news/2020/04/23/temenos-launches-saas-explainable-ai-models-to-help-banks-support-small-businesses-and-retail-customers-during-covid-19-crisis/
https://www.businesswire.com/news/home/20190717005854/de/
https://medium.com/the-data-intelligence-connection/5-saas-toolkits-to-build-explainable-ai-39aa0290ae27
https://www.businesswire.com/news/home/20200429005442/de/

Wer selbst gerne mal mit XAI experimentieren möchte …





https://github.com/EthicalML/XAI

Drei Thesen zur KI-Transparenz:

1. Es gibt unterschiedliche Arten der KI-Transparenz.
2. Technische KI-Transparenz steckt in den Kinderschuhen.
3. Totale KI-Transparenz ist weder möglich noch sinnvoll.

Prognose: Technische KI-Transparenz wird noch einige Zeit benötigen.

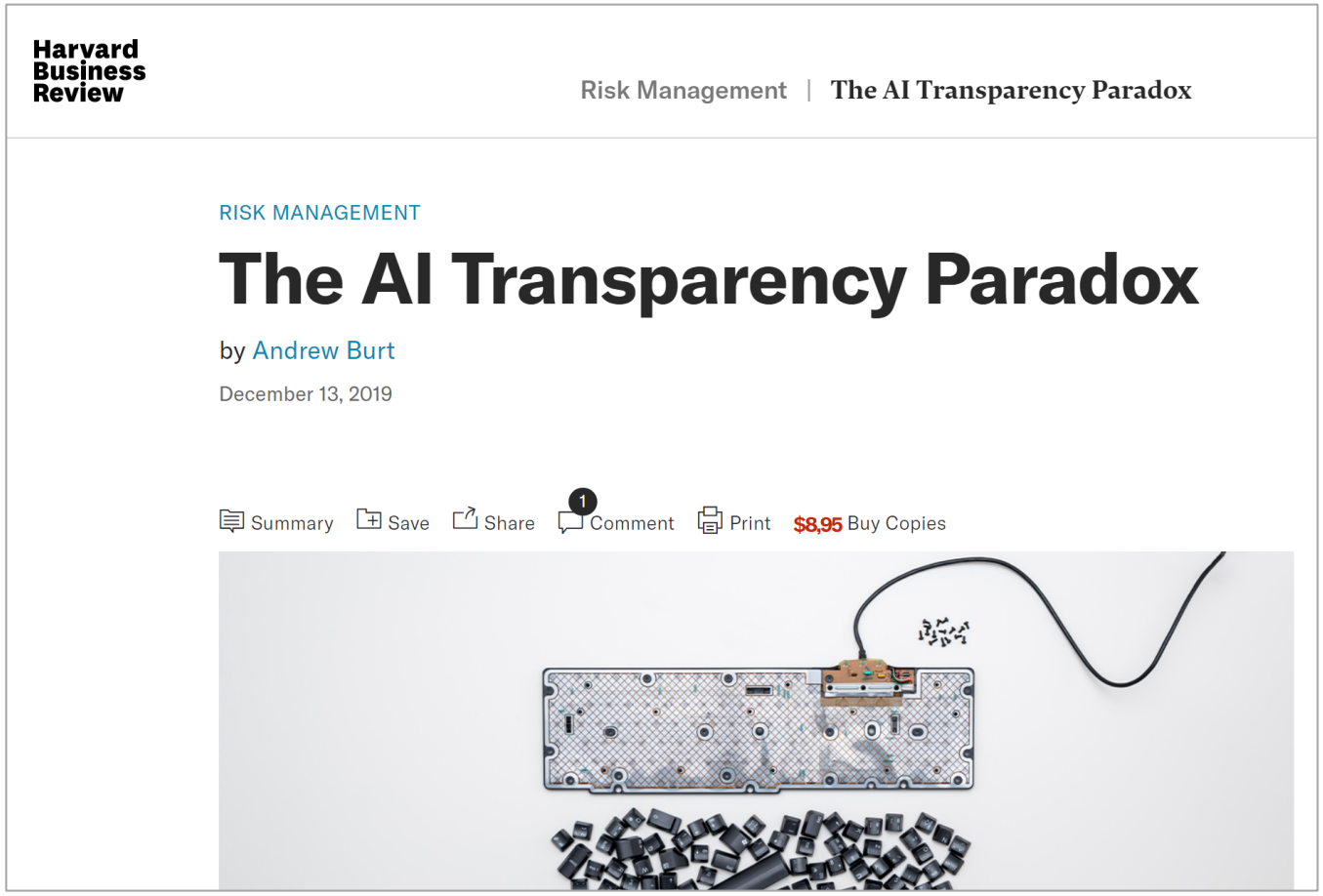
Das wiederum wirkt sich unmittelbar
auf alle anderen Kriterien ethischer KI aus.

These Nr. 3

Totale KI-Transparenz ist weder möglich noch sinnvoll.

Wer umfassende KI-Transparenz fordert sollte auch ihre Kehrseite beachten!

Achtung!
KI-Transparenz schafft Angriffsfläche



https://hbr.org/2019/12/the-ai-transparency-paradox

Macht Intransparenz KI sicherer und besser?!

Durchaus …

Zudem gilt: KI-Transparenz ist in Lieferketten schwer zu gewährleisten.

Aktuelle Pressemitteilung zu transparenter KI (Stand: 1. Mai 2020)

einblick

Gewerkschaftlicher Info-Service Nr. 5 — Mai 2020

DGB

Leitfragen für betrieblichen KI-Einsatz

In einem aktuellen Papier hat der DGB konkrete Leitfragen für die Einführung von Künstlicher Intelligenz im Betrieb veröffentlicht. Diese lauten:

Schritt 1: Klärung von Zielen und Zielkonflikten

Schritt 2: Anforderungen an KI-Anbieter bzw. Entwickler

Schritt 3: Prozesstransparenz zur Datennutzung

Schritt 4: Abschätzung der Folgen und humane Arbeitsgestaltung im Betrieb

Schritt 5: Autonomie und Verantwortung beim betrieblichen Einsatz von KI

Schritt 6: Tests und Kontrollen

Mehr Transparenz für KI-Systeme

Damit selbstlernende Software im Betrieb akzeptiert wird, geht es vor allem um Transparenz, betont der Vorsitzende des Deutschen Gewerkschaftsbundes, Reiner Hoffmann: „Um KI für Gute Arbeit nutzen zu können, müssen KI-Anbieter transparent darlegen, was die Systeme können und im Zweifel auch anrichten können. Dies ist eine Grundvoraussetzung für die nötige Akzeptanz von KI im Betrieb genauso wie eine KI-gerechte Anpassung der Regeln für betriebliche Aushandlungsprozesse." Es brauche deshalb eine Transparenzpflicht für KI-Anbieter und entsprechende Regulierungs- und Prüfverfahren, damit Vertrauen im Betrieb entstehen kann, so Hoffmann.

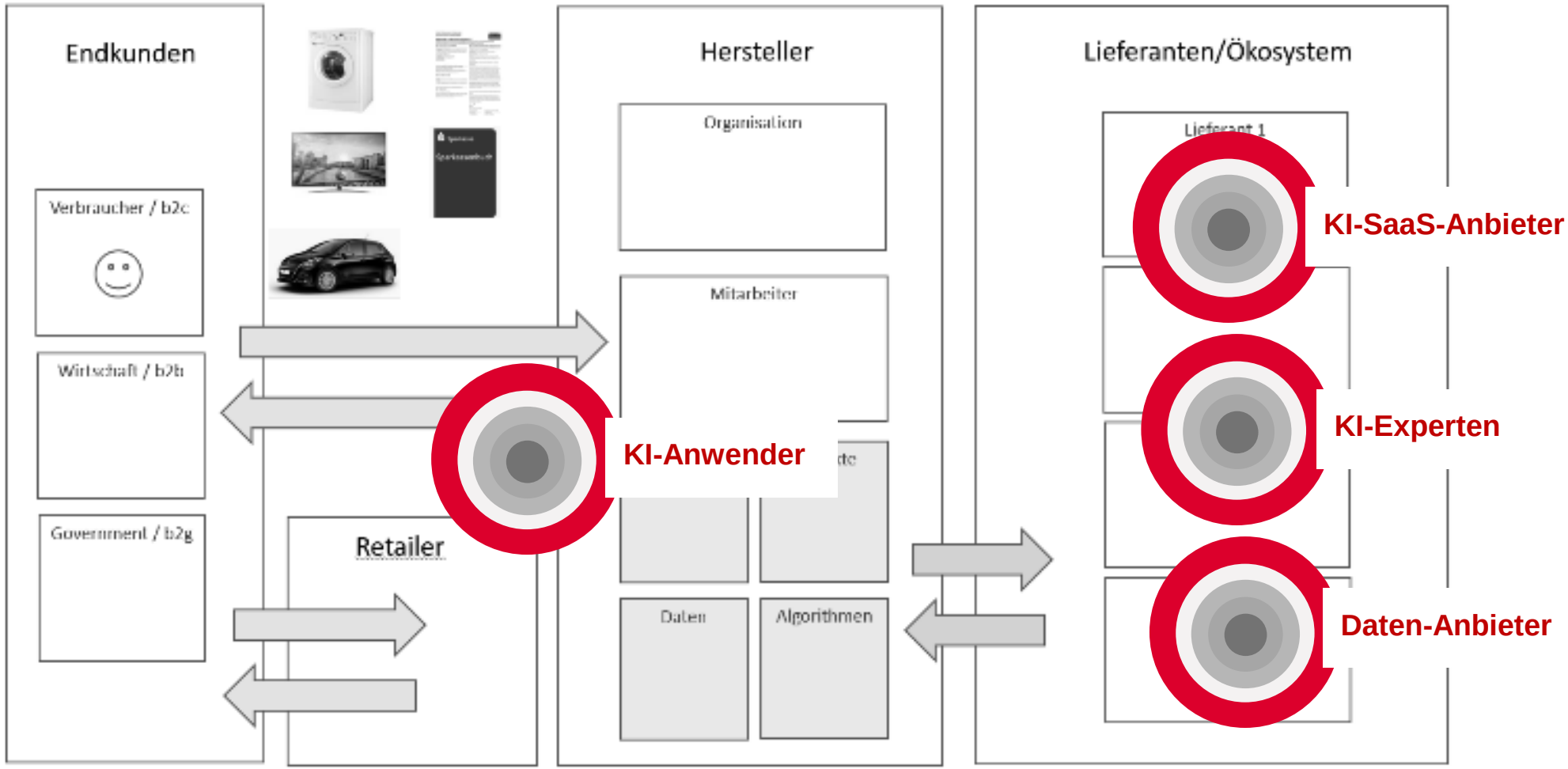
Transparenz-Anforderungen für die KI-Anbieter und Entwickler („Beipackzettel" für betriebliche Anwendung)

- Hohe Datenqualität und Datensicherheit (auch bei Auftragsdatenverarbeitung)

https://www.dgb.de/themen/++co++2109aa14-895b-11ea-8025-52540088cada

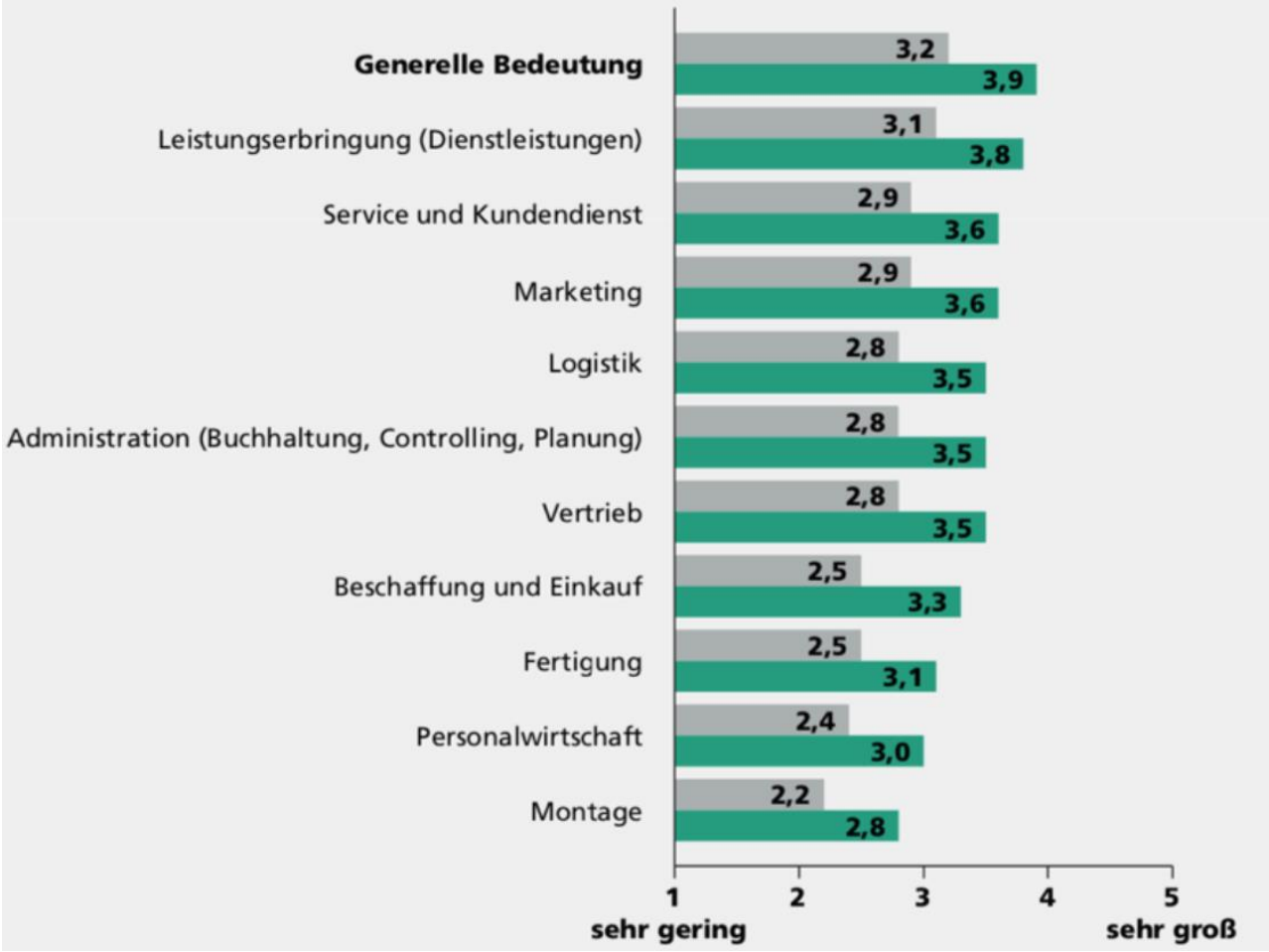
DGB fordert mehr Transparenz von KI. Gerade bei 3rd-Party-Beteiligung.

Daten & KI in der Lieferkette bzw. einem Ökosystem



Ein KI-Anwender

- x-KI-Use Cases
- x-KI-Zulieferer



https://business-user.de/workplace/wie-kuenstliche-intelligenz-den-arbeitsplatz-5-0-gestaltet/

Vielzahl von Zwiebeln mit ineinander verschachtelten Zwiebelringen … Transparenz?

Drei Thesen zur KI-Transparenz:

1. Es gibt unterschiedliche Arten der KI-Transparenz. ☐
2. Technische KI-Transparenz steckt in den Kinderschuhen. ☐
3. Totale KI-Transparenz ist weder möglich noch sinnvoll. ■

Um zumindest Basis-Transparenz zu gewährleisten

sind m.E. Zertifikate & Prüfungen erforderlich.

Das VDE-Label-Modell geht in die richtige Richtung, muss sich aber erst noch bewähren.

Zum Schluss nochmal die drei Thesen:

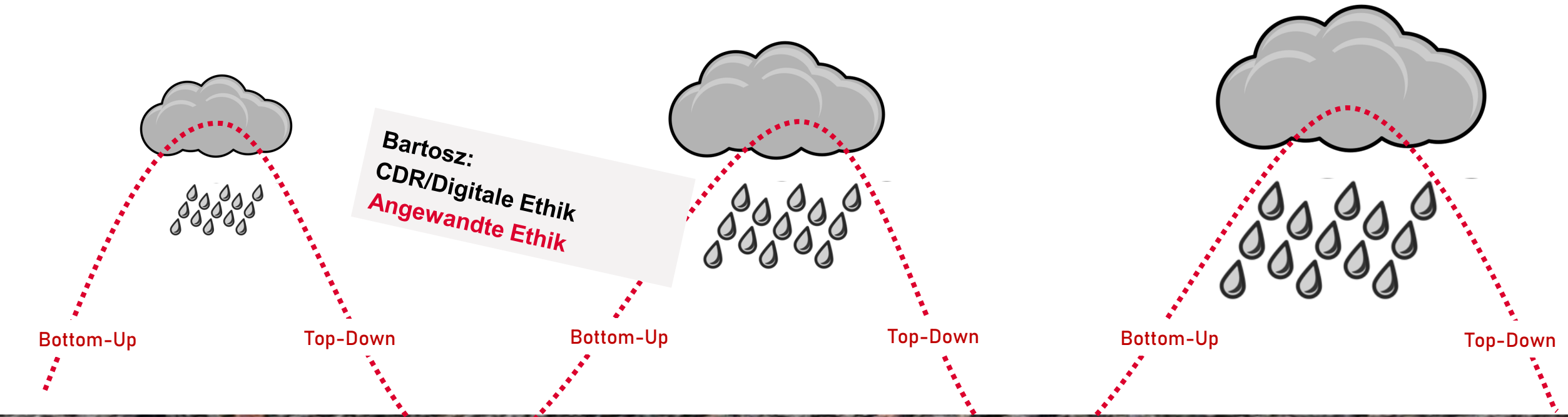
1. Es gibt unterschiedliche Arten der KI-Transparenz. ☑
2. Technische KI-Transparenz steckt in den Kinderschuhen. ☑
3. Totale KI-Transparenz ist weder möglich noch sinnvoll. ☑

Meine Empfehlung:
Keine ethischen Luftschlösser bauen!

Mein Verständnis von CDR:
Iterativ aus der Praxis heraus agieren!

Iteratives Wechselspiel von CDR: Kombiniert Top-Down & Bottom-Up

KI-Prinzipien & Guidelines

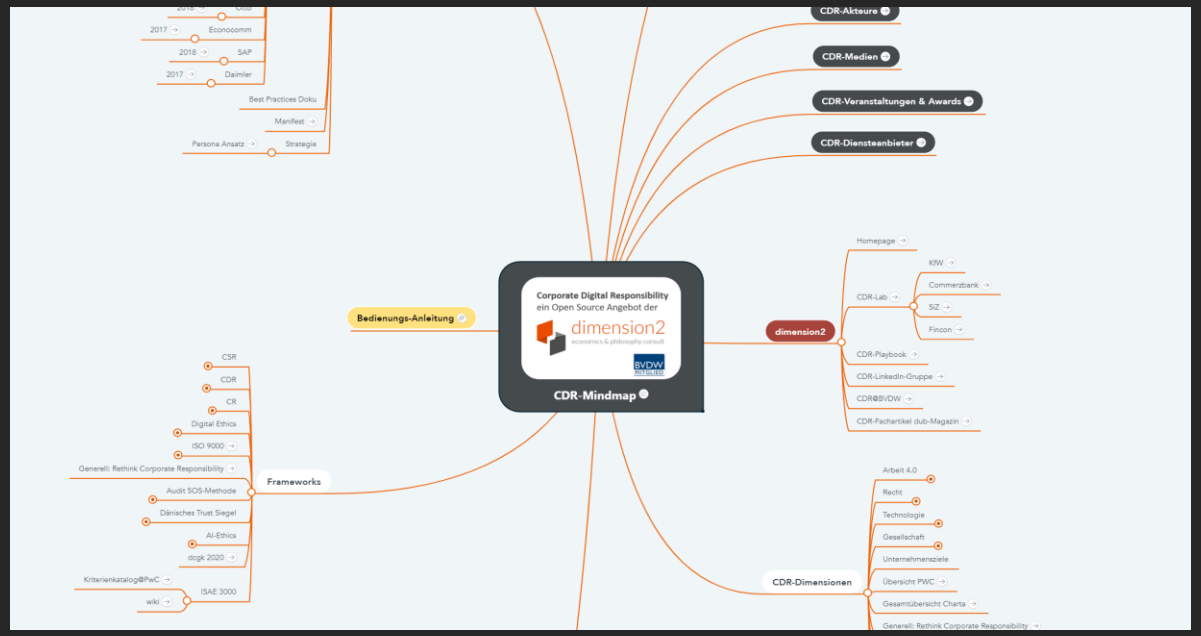


Praktische Erfahrung & Anpassung

Merci!

Nun zur Diskussion

Generelle Hintergründe zu CDR:



https://www.slideshare.net/OliverMerx/corporate-digital-responsibility-playbook-xl-179766056

https://www.estrategy-magazin.de/2019/ausgabe-03-/-2019/corporate-digital-responsibility-laues-luftchen-oder-kraftvoller-orkan.html

https://www.mindmeister.com/1303967380

https://www.linkedin.com/groups/8810975

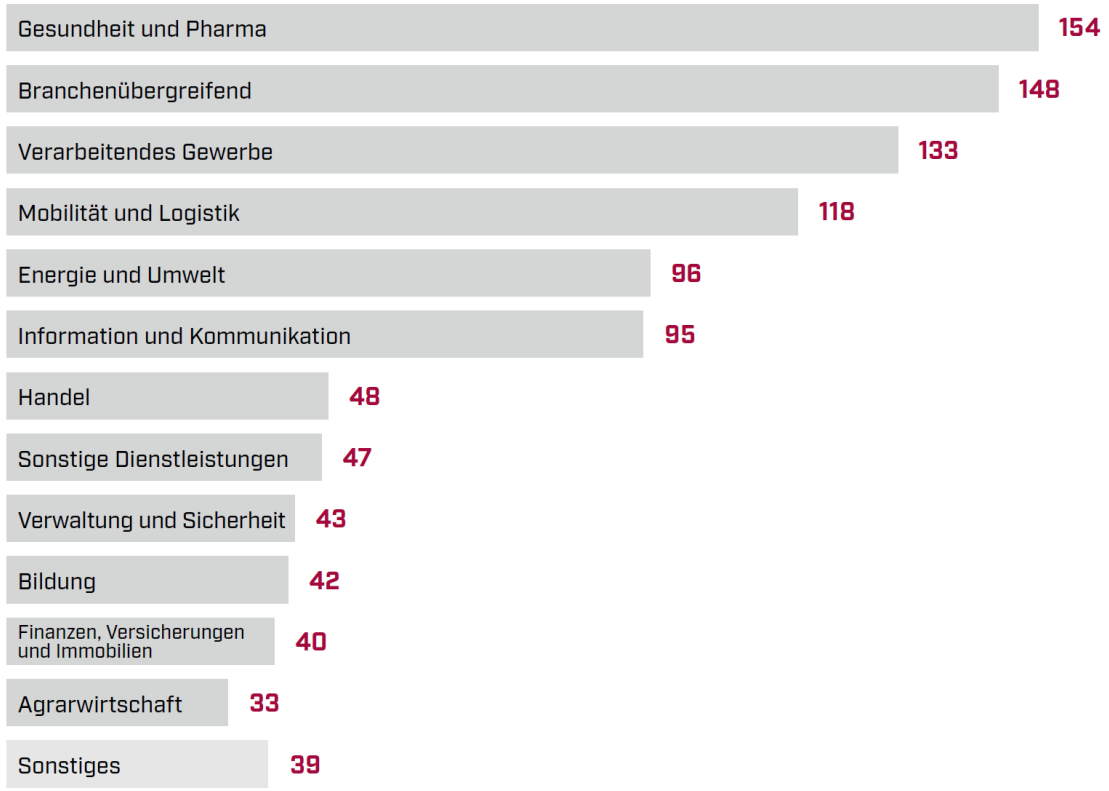
http://www.dimension2.consulting

https://zentrum-digitalisierung.bayern/1-workshop-des-corporate-digital-responsibility-labs/

https://www.bvdw.org/der-bvdw/gremien/digital-responsibility/news/

Keine KI-Transparenz ohne Branchen-Ethik!

KI-Anwendungen nach Branchen (Übersicht aus KI-Landkarte)



Quelle: PLS: 2020 (Stand: März 2020, absolute Zahlen, Mehrfachzuordnung war möglich, N=720)

https://hbr.org/2019/12/the-ai-transparency-paradox