Digitale Ethik
Versuch eines Überblicks

02.04.2019, TEC Spring Interconnect

Dr. Henrik Loeser

IBM Deutschland Research & Development GmbH
hloeser@de.ibm.com
Twitter: @data_henrik
https://blog.4loeser.net

Agenda

- Ethik und Digitale Ethik
- Warum...?
- Aktivitäten
- Und nun...?

Ethik und Digitale Ethik

Ethik gibt dem
Menschen
Hilfen für sittliche
Entscheidungen

Digitale Ethik ~ Datenethik ~ Algorithmenethik

Digitale Ethik fragt nach dem guten und richtigen Leben und Zusammenleben in einer Welt, die von digitalen Technologien geprägt ist. Sie formuliert Regeln für das richtige Handeln in Konfliktsituationen, die von der Digitalisierung aufgeworfen werden, und beschäftigt sich mit dem gesellschaftlichen Konzept von Freiheit und Privatsphäre, von Solidarität und Gerechtigkeit. Als Teildisziplin der Moralphilosophie stellt sie nicht zwangsläufig neue ethische Maßstäbe auf, sondern übersetzt bestehende ethische Maßstäbe für eine digital geprägte Gesellschaft.

Bundesverband Digitale Wirtschaft

Digitale Ethik – Themen (Auswahl)

- **Ethik für den Bereich digitale Medien und Technologien**
- Digitale Transformation (Digitalisierung) der Arbeitswelt
- Persönlichkeitsschutz und Datenschutz
- Filterblasen und Einfluss auf Meinungsfreiheit und Demokratie
- Ethics by design, Data protection by design

Warum...?

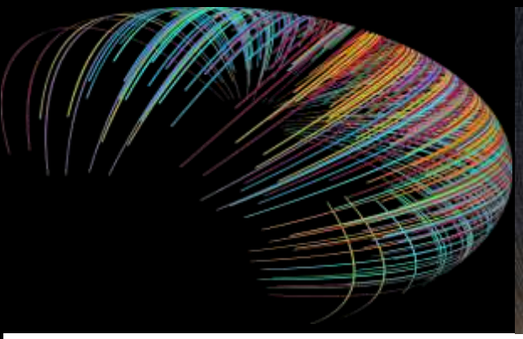
AI is the new IT

Dr. Dario Gil, IBM

AI is whatever hasn't been done yet.

Lawrence Gordon Tesler

What lies ahead?



AI Everywhere

Healthcare
Finance
Agriculture
Government
Education
Energy
Science
Business solutions

Deeper Insights

Data-centric systems
Distributed Deep Learning
Neuromorphic systems
Quantum computing
Homomorphic encryption
Machine foresight
Cognitive discovery

Engagement Reimagined

Human-machine collaboration
New AI modalities
Augmented reality
Global trade logistics
Blockchain for payments

Personalization at Scale

Personalized healthcare
Micro-segmentation
Personalized finance
Targeted marketing
Personalized learning
Individualized solutions

Instrumented Planet

Environmental solutions
Digital agriculture
Connected cars
Geospatial-temporal data and analytics
Smart sensors

Digitale Ethik?

- Moral Machine (MIT): Kulturelle Unterschiede bewirken andere Entscheidungen
- Autonomes Fahren: Wie entscheidet dein Auto?
- Maschinelle Entscheidungen: Wer entscheidet über dich? Warum?
- Bekommst du das Darlehen? Wird die OP genehmigt? Wird dein Haus durchsucht?
- Wer bestimmt das Meinungsbild? Wer entscheidet die nächste Wahl, die Volksabstimmung?
- Wie wird mein "Job" durch Digitalisierung verändert?

Voraussetzungen für Vertrauen in KI
(d.h. in Empfehlungen durch Maschinen)

...die in den meisten Fällen durch Menschen trainiert werden – was bedeutet, dass vor allem auch die Trainingsphase nachvollziehbar und überprüfbar/auditierbar **sein sollte, um die „Blackbox der KI“ transparent zu machen.**
Es geht generell um folgendes:

Sieht die Empfehlung fair aus?

Ist die Empfehlung nachzuvollziehen?

Könnte irgendwer das Ergebnis beeinflusst / entwendet haben?

Kann man die KI (oder jemanden) verantwortlich machen?

FAIRNESS

ERKLÄRBARKEIT

ROBUSTHEIT / SCHUTZ DES IP*

ABSICHERUNG

Aktivitäten

IBM-Aktivitäten

- Forschungsprojekte
- Source Code
- Teilnahme in politischen Gremien
- Präsenz und öffentlicher Diskurs
- Richtlinien, Whitepapers, Blogs

Areas of Ethical Focus

- Accountability
- Value Alignment
- Explainability
- Fairness
- User Data Rights

IBM: Everyday Ethics for AI - https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf

IBM on Data Protection (2014)

[…] We understand that clients are concerned about the security and privacy of their data. Therefore, we want to offer the following assurances:

- In general, if a government wants access to data held by IBM on behalf of an enterprise client, we would expect that government to deal directly with that client.
- If the U.S. government were to serve a national security order on IBM to obtain data from an enterprise client and impose a gag order that prohibits IBM from notifying that client, IBM will take appropriate steps to challenge the gag order through judicial action or other means.
- For enterprise clients' data stored outside of the United States, IBM believes that any U.S. government effort to obtain such data should go through internationally recognized legal channels, such as requests for assistance under international treaties.
- If the U.S. government instead were to serve a national security order on IBM to obtain data stored outside the United States from an enterprise client, IBM will take appropriate steps to challenge the order through judicial action or other means.
- IBM will continue to invest in world-class security technologies and services, and we will engage governments around the world on behalf of sensible, market-led policies that enable the free flow of data while promoting strong security. IBM will also continue its decades-long tradition of privacy leadership.

[...]

IBM: A Letter to Our Clients About Government Access to Data:
https://www.ibm.com/blogs/think/2014/03/open-letter-data/

IBM's Principles for Trust and Transparency (2018)

[…] We encourage all technology companies to adopt similar principles to protect client data and insights, and to ensure the responsible and transparent use of artificial intelligence and other transformative innovations. We offer our own Trust and Transparency Principles here as a roadmap. They include:

- The purpose of AI is to augment human intelligence
- Data and insights belong to their creator
- New technology, including AI systems, must be transparent and explainable

IBM THINK Policy blog:
https://www.ibm.com/blogs/policy/trust-principles/

AI ethics

Human biases may propagate to AI systems via training data or algorithmic models

Multiple regulations
across geographies and industries prohibit discrimination and mandate transparency

Bias in training data
bias can be detected and reduced in training data sets to prevent it from impacting decision making

Algorithms
various techniques can be applied to enable interpretation and explanation of results

Source: 2018 IBM Research Technology Vision

Aktivitäten in Deutschland und EU (Beispiele)

- Deutscher Bundestag
- Europäische Union
- Institut für Digitale Ethik, Stuttgart
- Bundesverband Digitale Wirtschaft

IBM Grundsätze und Prinzipien zu KI

Prinzipien für die „kognitive Ära“ – 2017 (Auszug):

Purpose: The purpose of AI and cognitive systems developed and applied by the IBM company is to **augment human intelligence**. […] Our position is based not only on principle but also on science. Cognitive systems will not realistically attain consciousness or independent agency. Rather, they will increasingly be embedded in the processes, systems, products and services […] – all of which **will and should remain within human control**.

Transparency: For cognitive systems to fulfill their world-changing potential, it is vital that people have **confidence in their recommendations, judgments and uses**. Therefore, the IBM company will make clear:

- When and for what purposes AI is being applied in the cognitive solutions we develop and deploy.
- The major sources of data and expertise that inform the insights of cognitive solutions, as well as the methods used to train those systems and solutions.
- The principle that clients own their own business models and intellectual property and that they can use AI and cognitive systems to enhance the advantages they have built, often through years of experience. We will work with our clients to protect their data and insights, and will encourage our clients, partners and industry colleagues to adopt similar practices.

Skills: The economic and societal benefits of this new era will not be realized **if the human side of the equation is not supported**. […]

Interne und externe Veröffentlichung am 17.01.2017, siehe https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/

Aktualisierung und Detaillierung
Prinzipien für Vertrauen und Transparenz – 2018:

- The purpose of AI is to **augment human intelligence**
- Data and insights **belong to their creator**
- New technology, including AI systems, must be **transparent and explainable**

Veröffentlichung am 30.05.2018, siehe https://www.ibm.com/blogs/policy/trust-principles/

Schlüsselpersonen, z.B. :

Francesca Rossi
IBM Global Leader for AI Ethics, Professorin für Informatik and der Universität Padua, Italien

Aleksandra (Saska) Mojsilovic
IBM Fellow, AI Science & Science for Social Good

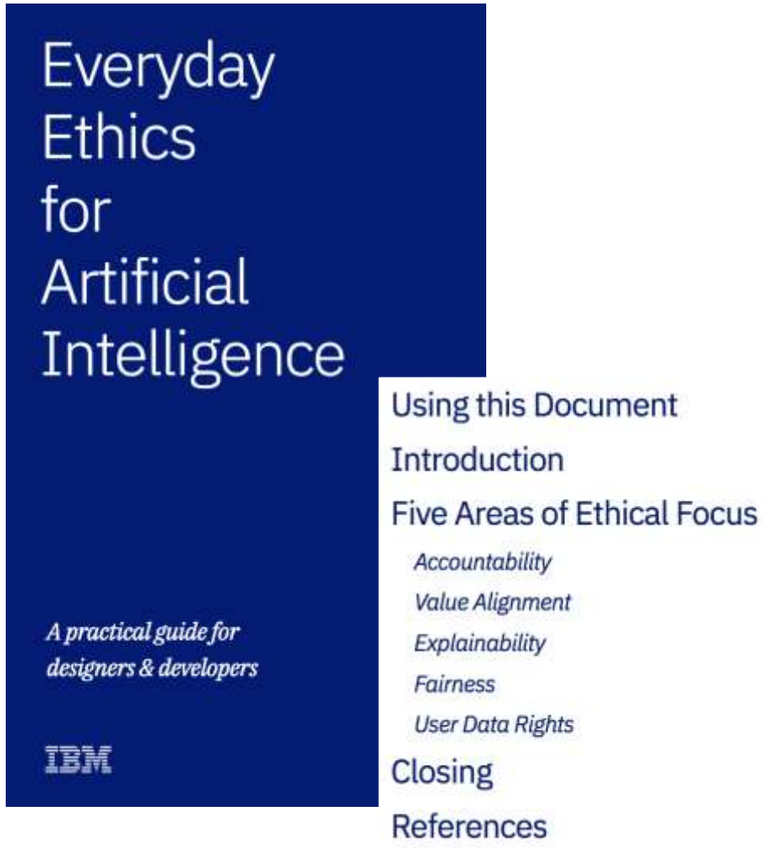
IBM Labore
Designer und Entwickler rund um die Welt, z.B. Schweiz, Indien, Japan, Brasilien, Australien, Kenia etc.

Details:
https://www.ibm.com/blogs/policy/francesca-rossi-ai/
https://researcher.watson.ibm.com/researcher/view.php?person=us-aleksand
http://www.research.ibm.com/labs/

Drei pragmatische Beispiele

Umsetzung der grundlegenden Prinzipien zu KI zur Schaffung von Vertrauen

Leitfaden für Designer und Entwickler (intern und extern)



https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf

AI OpenScale

Eine offene Plattform bzw. Services, verfügbar in der Cloud oder im eigenen Rechenzentrum, um “trusted AI” umzusetzen. Das beinhaltet z.B. Services, die

- Fairness überprüfen (statistische und systematische Fehler)
- Faktoren offen legen (in Entwicklungs- und Laufzeitumgebung), die Eingang in Empfehlungen finden
- Das Toolset ist integrierbar mit KI Services von IBM und auch anderen Anbietern, d.h. es ist offen und nicht proprietär.

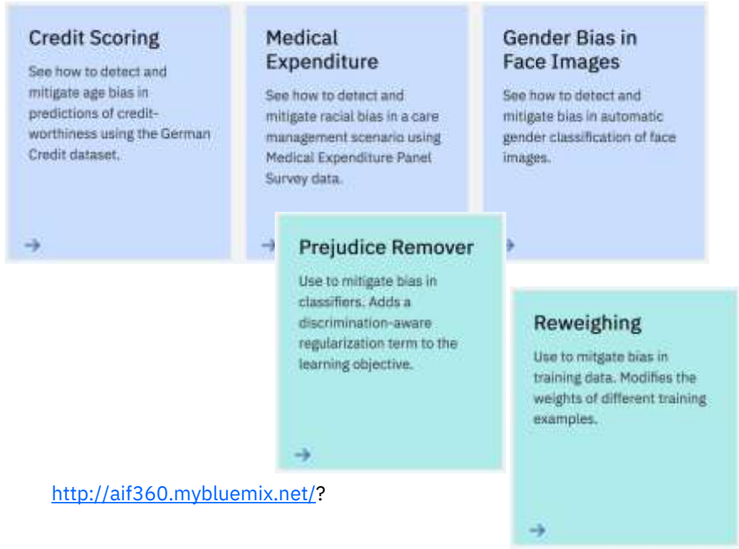
https://www.ibm.com/cloud/ai-openscale

AI Fairness 360 Open Source Toolkit

Ebenfalls ein “open source” Toolkit bzw. eine Bibliothek, die dazu dient, verschiedene Arten von Bias in Machine Learning Modellen zu identifizieren und zu entfernen.

Die Bibliothek mit mehr als 70 Metriken und 10 Algorithmen basiert auf wissenschaftlichen Erkenntnissen – nicht nur auf IBM Research – und ist erweiterbar; jede(r) ist eingeladen, dazu beizutragen.

Beispiele:



http://aif360.mybluemix.net/?

Digitale Ethik: Leitlinien

Datenökologische Verantwortung
1. Die Privatsphäre soll geschützt werden.
2. Smart-Data-Ansätze sollen als Vorbild dienen.
3. Die Sicherheit und Qualität der Daten sollen gewährleistet sein.

Faires & gerechtes Arbeiten 4.0
4. Es sollen faire und gerechte Arbeitsbedingungen gelten.
5. Mitarbeiter sollen am Digitalisierungsprozess des Unternehmens teilhaben.
6. Die Aus- und Weiterbildung sowie die digitalen Kompetenzen der Mitarbeiter sollen gefördert werden.

Chancengerechtigkeit & Fürsorge
7. Chancengerechtigkeit soll gefördert und Diskriminierung vermieden werden.
8. Auf schutzbedürftige Personen soll besonders Rücksicht genommen werden.

Folgenabschätzung & Nachhaltigkeit
9. Künstliche Intelligenz soll werteorientiert gestaltet werden.
10. Die Digitalisierung soll dazu dienen, natürliche Ressourcen zu schonen.

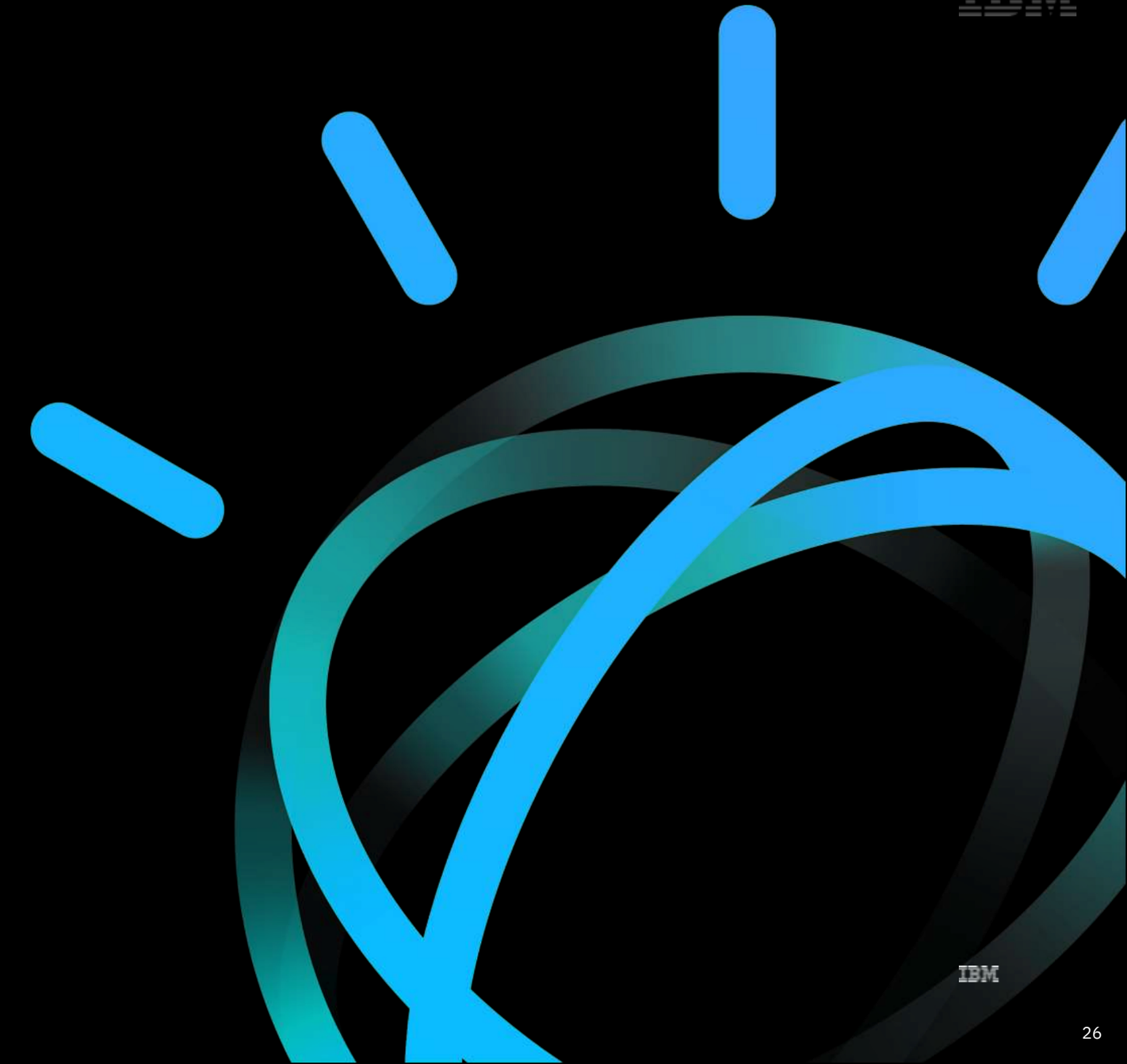
Institut für Digitale Ethik:
https://www.digitale-ethik.de/digitalkompetenz/10-ethische-unternehmensleitlinien/

Und nun...?

Zum Schluss

- KI zieht in fast alle Bereiche ein
- Transparenz / Erklärbarkeit und Fairness als Voraussetzung für Vertrauen in KI
- Datenschutz, Privatspäre und Sicherheit als Eckpfeiler
- Digitale Ethik unterstützt den Menschen – uns
- Aktivitäten in Politik / Gesellschaft, IBM(er) beteiligt und präsent
- **Digitale Ethik: Was machst du? DEINE Entscheidung…**

Thank you / Danke



Ressourcen

- Wikipedia: https://de.wikipedia.org/wiki/Digitale_Ethik
- Institut für Digitale Ethik, HdM, Stuttgart: https://www.digitale-ethik.de/
- Datenethikkomission der Bundesregierung: https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/datenethikkommission-node.html
- EU: High-Level Expert Group on AI: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
- Bertelsmann Stiftung: Projekt Algorithmenethik: https://algorithmenethik.de/projekt/
- Bundesverband Digitale Wirtschaft (BVDW): Mensch, Moral, Maschine – Digitale Ethik, Algorithmen und KI: https://www.bvdw.org/fileadmin/bvdw/upload/dokumente/BVDW_Digitale_Ethik.pdf
- Andrea Martin (IBM): "KI und Ethik - ...", Präsentation vor der Enquette-Komission: https://www.bundestag.de/dokumente/textarchiv/2019/kw03-pa-enquete-ki-585354

IBM: Öffentliche Ressourcen

- IBM Research on AI: https://www.research.ibm.com/artificial-intelligence/
- Test bias / game: http://biasreduction.mybluemix.net/
- AI Fairness 360 Open Source Toolkit: http://aif360.mybluemix.net/
- IBM Trusted AI: https://www.research.ibm.com/artificial-intelligence/trusted-ai/
- IBM Trusted AI for Business: https://www.ibm.com/watson/ai-ethics/
- IBM THINKPolicy Blog: https://www.ibm.com/blogs/policy/
  IBM THINK blog, data responsibility: https://www.ibm.com/blogs/think/category/data-responsibility/
- Blog "Coming soon: EU Ethics Guidelines for Artificial Intelligence":
  https://www.ibm.com/blogs/policy/ai-ethics-guidelines/

Weiterführende Information

Generelles:

- IBM Prinzipien zu Vertrauen und Transparenz in KI
  - Januar 2017: https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/
  - Mai 2018: https://www.ibm.com/blogs/policy/trust-principles/
- Ginni Rometty, IBM CEO, zum Thema “augmented intelligence” beim World Economic Forum, Januar 2017:
  - https://www.ibm.com/blogs/collaboration-solutions/2017/01/31/augmented-intelligence-not-artificial-intelligence/
  - https://www.weforum.org/videos/ginni-rometty-it-should-be-augmented-intelligence-not-artificial
- IBM Research zu KI: https://www.research.ibm.com/artificial-intelligence/
- Trust & Transparency in AI: https://www.ibm.com/watson/trust-transparency
- IBM KI und Ethik:
  - https://www.ibm.com/watson/ai-ethics/
  - http://research.ibm.com/artificial-intelligence/trusted-ai/

Praktische Leitfäden und Tools, z.B.:

- “Everyday Ethics”
  - Artikel zum Leitfaden für Designer und Entwickler: https://medium.com/design-ibm/everyday-ethics-for-artificial-intelligence-75e173a9d8e8
  - Leitfaden: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf
- AI OpenScale: https://www.ibm.com/cloud/ai-openscale
- AI Fairness 360 Open Source Toolkit: http://aif360.mybluemix.net/?

Mitarbeit in Gremien, etc.:

- High Level Expert Group der EU zu KI: Francesca Rossi, IBM Global Leader AI Ethics, ist Mitglied dieser Gruppe: https://www.ibm.com/blogs/policy/francesca-rossi-ai/
- Zusammenarbeit mit dem MIT: http://mitibmwatsonailab.mit.edu/
- „Partnership on AI“: https://www.partnershiponai.org/

„IBM for Social Good“:

- https://www.research.ibm.com/science-for-social-good/

What is Artificial Intelligence (AI)?

- **Artificial neuron**
  - Inputs
  - Weights
  - Transfer function
  - Activation function
  - Threshold
  - Activation