

KV Ethik und Gender Studies



Univ.-Prof.ⁱⁿ Dr.ⁱⁿ Martina Mara

WS 2019
15.10.2019



Terminübersicht

Termin	Datum	Raum	Leitung	Typ	Inhalte
1	8.10.2019	HS 9	Sabitzer/ Mara	Präsenz	Vorbesprechung
2	15.10.2019	HS 9	Mara	Präsenz	Ethische Fragen der KI / EU Guidelines
3	22.10.2019	OIC	Mara	Präsenz	Ethische Fragen der KI / Start Gruppenarbeit
4	29.10.2019		Mara	Online	Gruppenarbeit, MM für Rückfragen verfügbar
5	5.11.2019	HS 9	Mara	Präsenz	IEEE Code of Ethics / Diversität in der Technikentwicklung / Besprechung Gruppenarbeit



Tweet



Picard Tips

@PicardTips

Picard ethics tip: Before undertaking a mission, it is our moral imperative to consider any obvious negative ramifications.

10:58 PM · Oct 12, 2019 · [Joe's Tools](#)

Was ist eigentlich Künstliche Intelligenz?

KI-Systeme sind **vom Menschen entwickelte Softwaresysteme** (und ggf. auch Hardwaresysteme), die in Bezug auf ein komplexes Ziel handeln, indem sie ihre **Umgebung durch Datenerfassung wahrnehmen**, die gesammelten **Daten interpretieren**, **Schlussfolgerungen daraus ziehen** oder die aus diesen Daten abgeleiteten Informationen verarbeiten, und über das **bestmögliche Handeln zur Erreichung des vorgegebenen Ziels** entscheiden.

KI-Systeme können **entweder symbolische Regeln verwenden oder ein numerisches Modell** erlernen, und sind auch in der Lage, die Auswirkungen ihrer früheren Handlungen auf die Umgebung zu analysieren und ihr Verhalten entsprechend anzupassen.

Source: HLEG AI / Europäische Kommission

Was ist eigentlich Künstliche Intelligenz?

Als wissenschaftliche Disziplin umfasst die KI mehrere Ansätze und Techniken wie z. B. **maschinelles Lernen** (Beispiele dafür sind „Deep Learning“ und bestärkendes Lernen), **maschinelles Denken** (es umfasst Planung, Terminierung, Wissensrepräsentation und Schlussfolgerung, Suche und Optimierung) und die **Robotik** (sie umfasst Steuerung, Wahrnehmung, Sensoren und Aktoren sowie die Einbeziehung aller anderen Techniken in cyber-physische Systeme).

Wettlauf der Künstlichen Intelligenz? China vs USA vs Europa

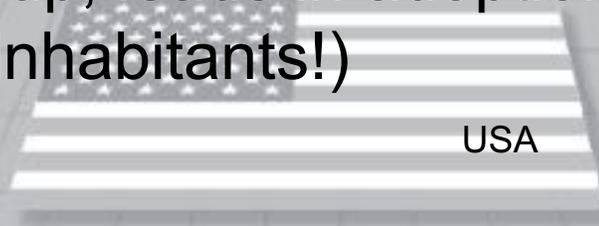
USA: Most AI companies/start-ups, most private equity and venture capital funding, talent attractor

Europe: High-quality AI talents and research, struggles to translate research into business applications, loses talent (mostly to the US)

China: Rapidly catches up, leads in adoption speed and available data (and number of inhabitants!)



CHINA



USA



EUROPE



1) Diskussion mit NachbarIn (5 min):

Ist es vertretbar, mithilfe von KI Daten über den Aufmerksamkeitszustand von SchülerInnen zu sammeln und zu analysieren?

2) Mentimeter-Abstimmung:

Go to [menti.com](https://www.menti.com) and use the code 92 55 91 ...

Beginnen Sie Ihre Antwort mit „JA, weil ...“ or „NEIN, weil ...“

Wettlauf der Künstlichen Intelligenz? China vs USA vs Europa

Entwicklung eines Europäischen KI-Ansatzes:

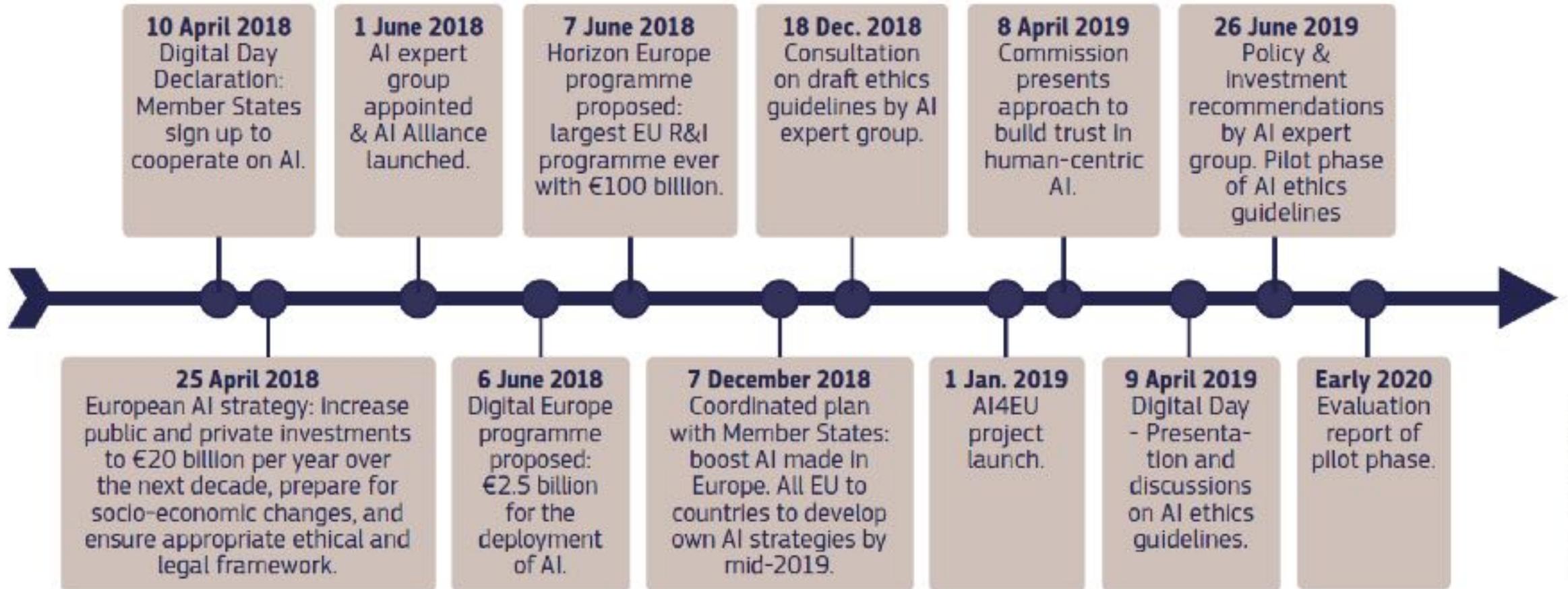
- The EU's coordinated plan on AI states that its “ambition is for Europe to become the world-leading region for developing and deploying **cutting-edge, ethical and secure AI.**”
- Positioning “AI made in Europe” as a positive global brand
- Combining innovation-friendliness and ethics: Trustworthy AI as economic value?

CHINA

USA

EUROPE

Entwicklung einer Europäischen KI-Strategie



Updated 20190702

Source: European Commission, Digital Single Market, "Artificial Intelligence for Europe"

EU Ethics Guidelines for Trustworthy AI

- KI ist kein Selbstzweck, sondern ein Werkzeug, das den Menschen dienen muss.
- Vertrauen ist eine wichtige Voraussetzung für einen humanzentrierten Umgang mit KI.
- KI-Anwendungen müssen die Werte, auf denen die Europäische Union aufbaut, berücksichtigen: Achtung der Menschenrechte und der menschlichen Würde, Freiheit, Demokratie, Gleichberechtigung, ...



EU Ethics Guidelines for Trustworthy AI

- Die Guidelines wurden von der High-Level Expert Group on AI der Europäischen Kommission entwickelt: darin waren 52 Personen aus unterschiedlichsten Interessensbereichen aktiv (Forschung, Technik, Industrie, NGOs, Vereine, ...)



7 Key Requirements for Trustworthy AI

7 key requirements for ethical AI:

Human agency and oversight

Technical robustness & safe

Privacy and data governance

Transparency

Diversity, non-discrimination and fairness

Societal and environmental wellbeing

Accountable

**Will your algorithms pass the test?
Create AI humans can trust.**

#AI #ArtificialIntelligence



1) Human agency and oversight

Key aspects:

a) Fundamental rights

Wie viele andere Technologien können KI-Systeme für Grundrechte entweder förderlich oder hinderlich sein. Sie können Nutzen bringen, indem sie z. B. einen besseren Zugang zu Bildung ermöglichen, wodurch das Recht auf Bildung gestärkt wird. Gerade im Hinblick auf die Reichweite und die Fähigkeit von KI-Systemen kann es aber auch zu negativen Auswirkungen auf die Grundrechte kommen. In solchen risikobehafteten Situationen sollte eine Folgenabschätzung der Auswirkungen auf die Grundrechte vorgenommen werden.

1) Human agency and oversight

Key aspects:

b) Human agency

Die Benutzer sollten in der Lage sein, informierte Entscheidungen in Bezug auf KI-Systeme zu treffen. Ihnen sollte das notwendige Wissen zur Verfügung gestellt werden, um die KI-Systeme hinreichend zu verstehen und mit ihnen interagieren zu können. KI-Systeme sollten die einzelnen Menschen dabei unterstützen, im Einklang mit ihren eigenen Zielen bessere, fundiertere Entscheidungen zu treffen.

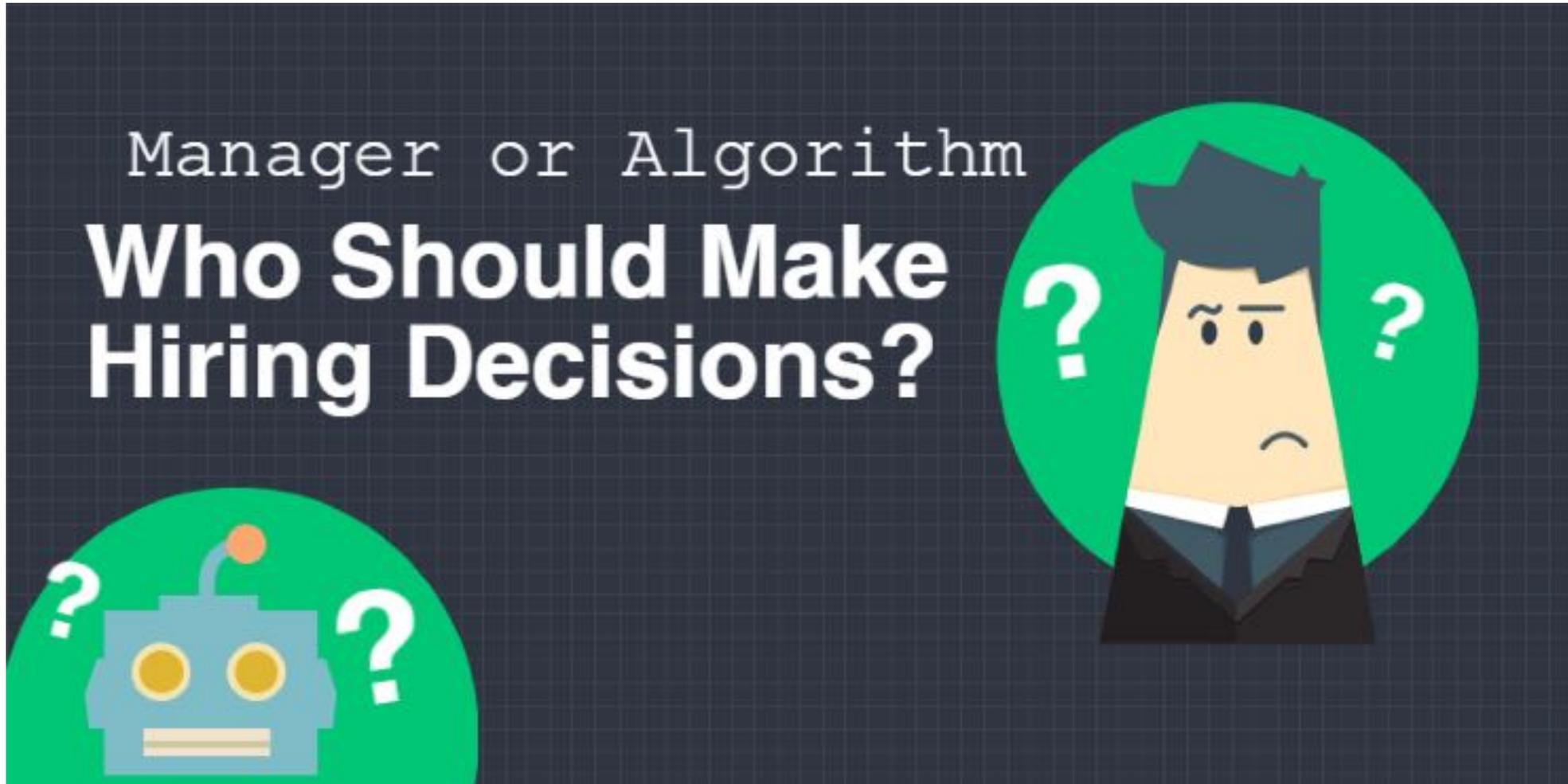
1) Human agency and oversight

Key aspects:

c) Human oversight

Die menschliche Aufsicht hilft, dafür zu sorgen, dass ein KI-System die menschliche Autonomie nicht untergräbt. Die Aufsicht kann durch Lenkungs- und Kontrollmechanismen wie die Gewährleistung der interaktiven Einbindung eines Menschen („Human-in-the-Loop“), der Überprüfung und Kontrolle durch einen Menschen („Human-on-the-Loop“) oder der Gesamtsteuerung durch einen Menschen („Human-in-Command“) erreicht werden.

1) Human agency and oversight



2) Technical robustness & safety

Key aspects:

a) Resilience to attack and security

KI-Systeme sollten wie alle Softwaresysteme vor Sicherheitslücken geschützt werden, die einen Missbrauch durch Angreifer (z. B. durch Hacking) ermöglichen.

2) Technical robustness & safety

Key aspects:

b) Fallback plan and general safety

KI-Systeme sollten für den Problemfall über Sicherheitsvorkehrungen verfügen, die einen Auffangplan aktivieren. Dies kann bedeuten, dass KI-Systeme von einem statistischen auf ein regelbasiertes Verfahren umschalten oder dass sie einen menschlichen Bediener anfordern, bevor sie einen Vorgang weiter ausführen.

2) Technical robustness & safety

Key aspects:

c) Accuracy

Präzision bezieht sich auf die Fähigkeit eines KI-Systems, Sachverhalte richtig zu beurteilen. Wenn sich gelegentliche ungenaue Vorhersagen nicht vermeiden lassen, sollte das System unbedingt anzeigen können, mit welcher Wahrscheinlichkeit es zu Fehlern kommt. Ein hohes Maß an Präzision ist insbesondere dann unerlässlich, wenn sich KI-Systeme direkt auf das Leben von Menschen auswirken.

3) Privacy and data governance

Key aspects:

a) Privacy and data protection

AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. This includes the information initially provided by the user, as well as the information generated about the user over the course of their interaction with the system (e.g. outputs that the AI system generated for specific users or how users responded to particular recommendations).

3) Privacy and data governance

Key aspects:

b) Quality and integrity of data

The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. This needs to be addressed prior to training with any given data set.

Reproduktion menschlicher Vorurteile durch KI

Google Übersetzer



Text Dokumente

SPRACHE ERKENNEN TÜRKISCH DEUTSCH ENGLISCH

DEUTSCH TÜRKISCH ENGLISCH

o bir hemsire
o bir öğretmen
o bir profesör

Sie ist Krankenschwester
Sie ist Lehrerin
Er ist Professor

46/5000

Text Dokumente

SPRACHE ERKENNEN TÜRKISCH DEUTSCH ENGLISCH DEUTSCH TÜRKISCH ENGLISCH

sie ist fleißig o çalışkan

15/5000

Speaker icon Copy Edit Share

Text Dokumente

SPRACHE ERKENNEN TÜRKISCH DEUTSCH ENGLISCH DEUTSCH TÜRKISCH ENGLISCH

o çalışkan Er ist fleißig

10/5000

Speaker icon Copy Edit Share

Text Dokumente

SPRACHE ERKENNEN DEUTSCH ENGLISCH FRANZÖSISCH ↕ TÜRKISCH DEUTSCH ENGLISCH

er ist kindergärtner

O bir anaokulu öğretmenidir

20/5000

Speaker icon Copy icon Edit icon Share icon

Text Dokumente

SPRACHE ERKENNEN TÜRKISCH DEUTSCH ENGLISCH ↕ TÜRKISCH DEUTSCH ENGLISCH

O bir anaokulu öğretmenidir

Sie ist Kindergärtnerin

27/5000

Speaker icon Copy icon Edit icon Share icon

Text Dokumente

SPRACHE ERKENNEN

TÜRKISCH

DEUTSCH

ENGLISCH



DEUTSCH

TÜRKISCH

ENGLISCH



er ist hübsch



o güzel



13/5000



Text Dokumente

SPRACHE ERKENNEN

TÜRKISCH

DEUTSCH

ENGLISCH



DEUTSCH

TÜRKISCH

ENGLISCH



o güzel



Sie ist wunderschön



7/5000



3) Privacy and data governance

Key aspects:

c) Access to data

In any given organisation that handles individuals' data (whether someone is a user of the system or not), data protocols governing data access should be put in place. These protocols should outline who can access data and under which circumstances.

4) Transparency

Key aspects:

a) Traceability

The data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard.

4) Transparency

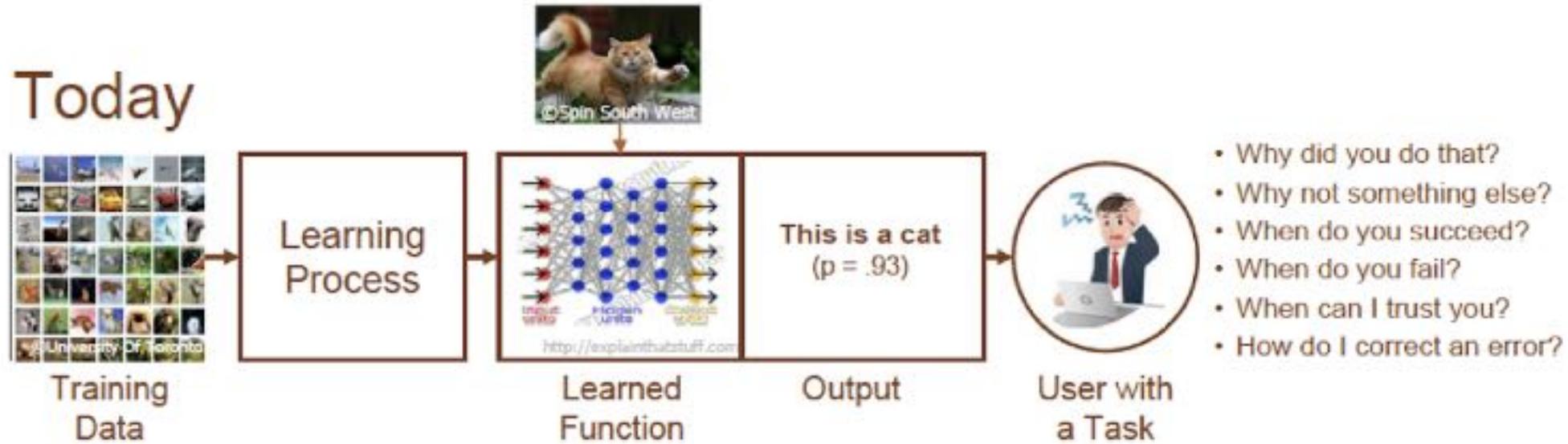
Key aspects:

b) Explainability

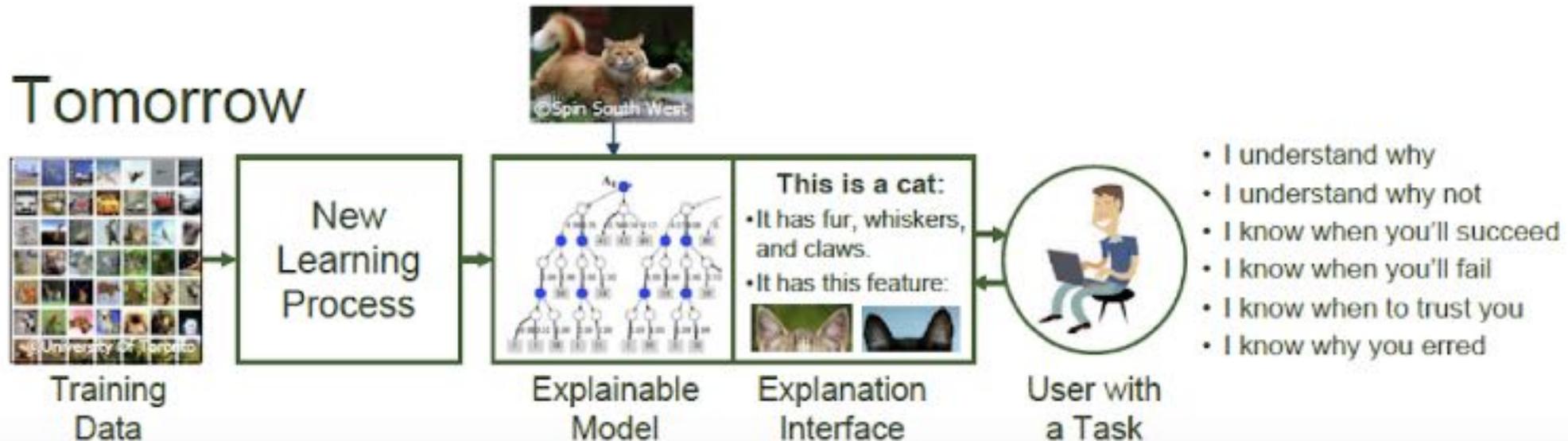
Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings.

Aktuelles Thema: "Explainable AI" (XAI)

Today



Tomorrow

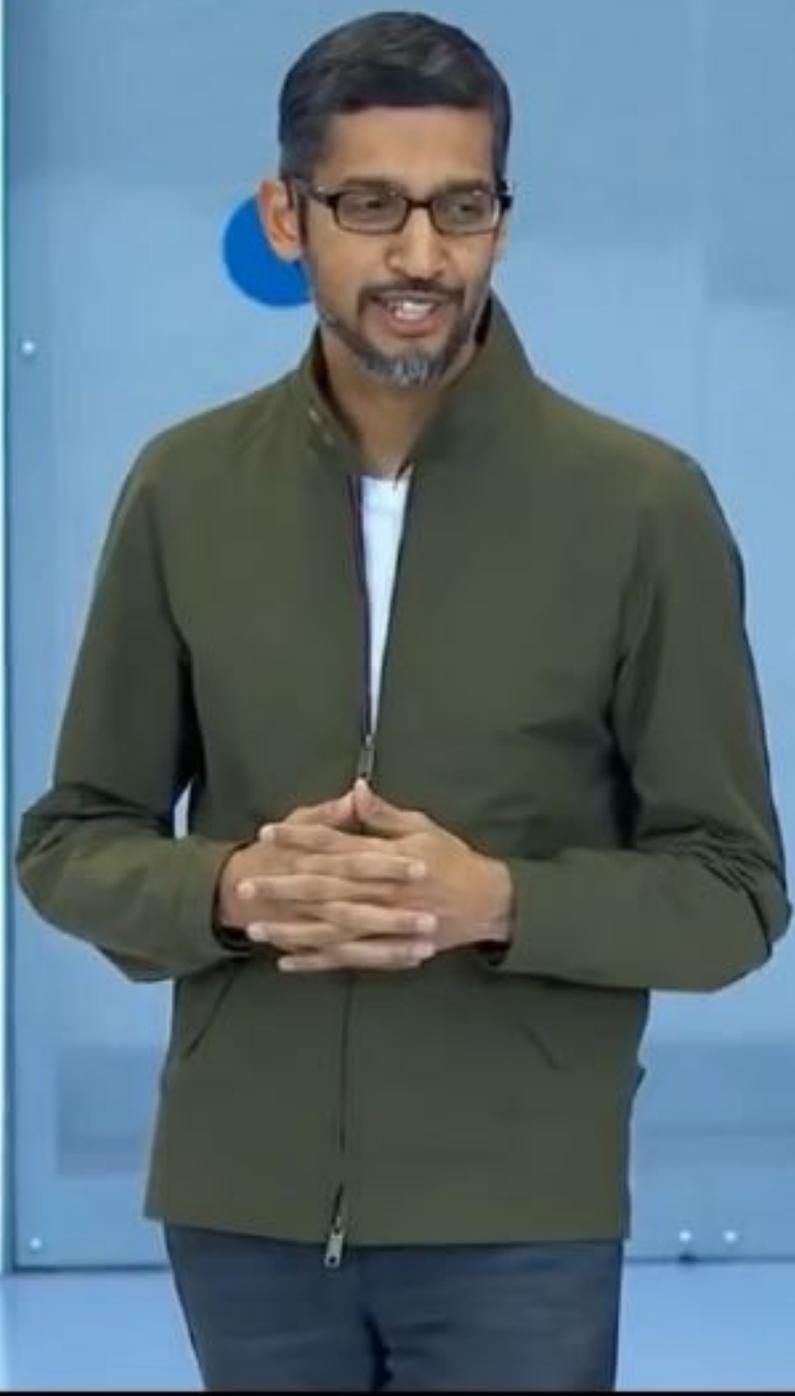


4) Transparency

Key aspects:

c) Communication

AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system.



Wer ist noch keiner Gruppe zugeteilt?

- Ca. 8-10 Personen je Gruppe - Einteilung bleibt für alle 3 Blöcke bestehen.
- Gruppenname festlegen
- Gruppenmitglieder in Formular eingeben: <http://bit.ly/33cSKWH>



KV Ethik und Gender Studies



Danke!

Wir sehen uns am 22.10.2019!

