

Chancen und ethische Grenzen von KI

DDr. Erich Prem

- ▶ Universität Wien, Institut für Philosophie, Medien- und Technikphilosophie
- ▶ TU Wien
- ▶ eutema GmbH
- ▶ Verein zur Förderung des digitalen Humanismus



Jakub Geltner Nest 05 (2015)

Unsere post-digitale Welt



KI ist überall...

• Personalisierung von Inhalten

• Online Anzeige von Produkten, Produktempfehlungen, personalisierte Werbung, Darstellung von Suchmaschinen-Ergebnissen, Aufbau von Social Media-Feeds etc.

KI-unterstützte Produktfunktionen

• smarte Haushaltsgeräte wie Staubsauger- und Rasenmäher-Roboter etc.

Spracherkennung und Dialogsysteme

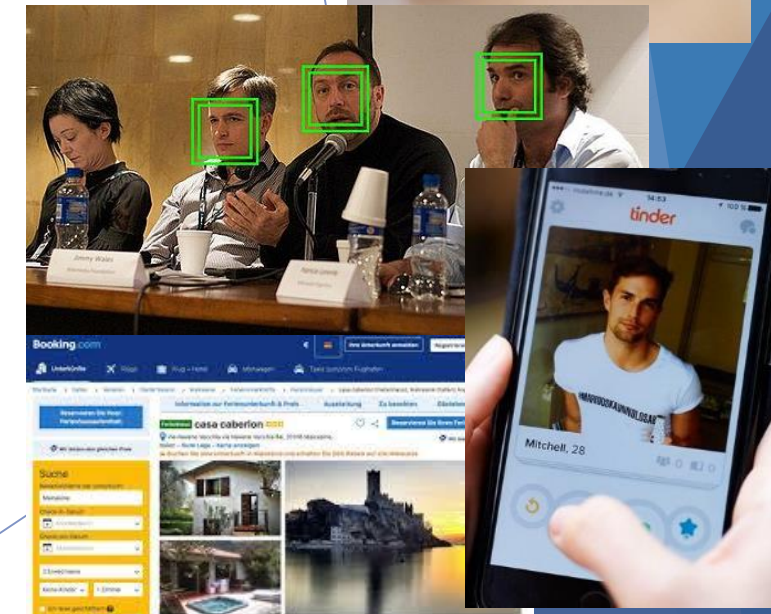
• Sprachassistenten, Chatbots, Smart Toys, C-Commerce etc.

Bildererkennung

• Mobiltelefon, biometrische Erkennung, bilddiagnostische Verfahren in der Medizin, Emotionsanalysen, inverse Bildersuche, aber auch Deep Fakes etc.

Scoring, Entscheidungen über Verträge und Konditionen

• Ablehnung Handyverträge, Versicherungsprämien, online Zahlungsmittel, personalisierte Preisgestaltung, Eignungstests, Personalauswahl etc.



KI bildet Modelle menschlichen Verhaltens

- Produkte und Dienste vereinfachen
- Situation oder Aktivität erkennen
- Sprache anpassen

Klassifikation



- Produkte empfehlen
- Dienste anpassen, Produkt individualisieren
- Intention erkennen
- Planung

Vorhersage

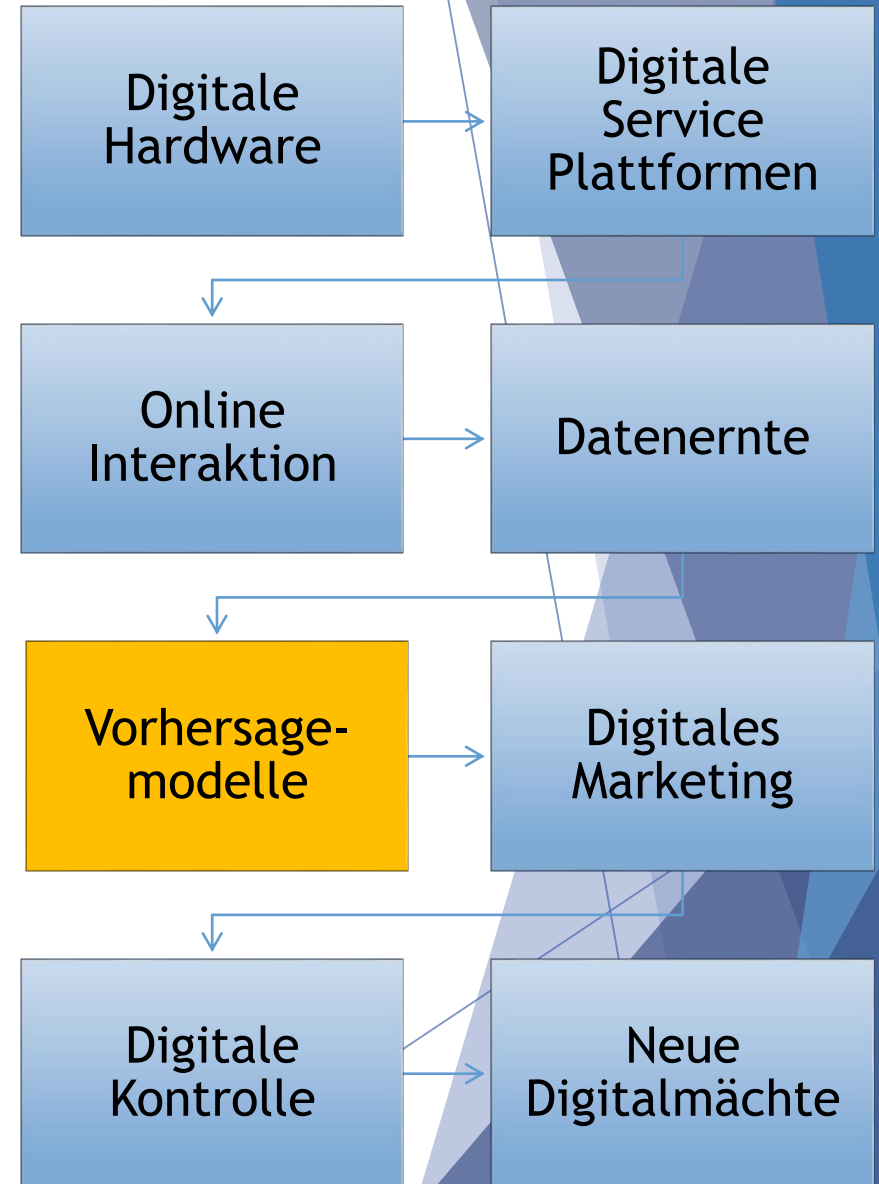
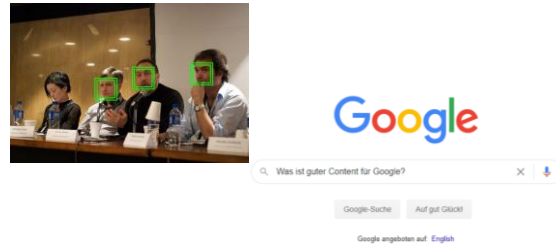


- Stau vermeiden
- Menschen zu gesundem Verhalten anhalten
- Nutzung öffentlicher Verkehrsmittel fördern

Steuerung



KI ist ein zentraler Schritt in einer digitalen Prozesskette, von Daten über KI-Modelle bis zu Kontrolle und Geopolitik



TARGET Früher Big Data Einsatz

Herausforderung: Moment des Wechsels von Gewohnheiten, z.B. Geburt eines Kindes; allerdings bevor es offiziell wird.

Erkenntnis: Forscher (TARGET, UCLA) entdeckten dass Frauen ab dem 2. Trimester unparfümierte Lotion und Nahrungsergänzung kaufen. Datenanalyse zeigt, dass Daten von 25 Produkten die Prognose einer Schwangerschaft erlauben.

Bewusstsein: TARGET war die Datenschutzproblematik bei der Bewerbung neuer Angebote für Schwangere prinzipiell klar.

Problem: Vater, der sich über „schwängere“ Angebote für seine Tochter beschwerte. Seine noch zur Schule gehende Tochter brauche sowas nicht....



Target Country store in Kyabram, Victoria in 2012. This store closed in September 2020. Mattinbgn, CC BY 3.0, <https://commons.wikimedia.org/w/index.php?curid=22204548>

C. Duhigg (2012) How companies learn your secrets. New York Times, December 16, 2012. www.nytimes.com

Nachteilige Effekte von Modellen und verhaltensbasierter Werbung (Person)

- ▶ Höhere Preise für Produkte und Dienste, z.B. bei besonderem Interesse oder durch Segmentierung (n=1?)
- ▶ Emotionales Targeting nach der Gefühlssituation einer Person: Menschen ansprechen, wenn sie am verletzlichsten sind
- ▶ Segmentierung kann zum Ausschluss von Gruppen von gewissen Angeboten führen, z.B. Arbeit, Dienstleistungen, Wohnungen
 - ▶ “schwul” mag verboten sein, aber „Grinder Nutzer“ kann evt. am Werbemarkt gekauft werden
 - ▶ Es ist schwer zu wissen, was man nicht sieht.
- ▶ Fb erzeugt z.B. automatische Werbekategorien wie „Interesse an Aufständen“, „Kinder, die sich für Alkohol interessieren“ etc.
 - ▶ hohes Potenzial für Missbrauch und für Kriminelle
- ▶ Identitätsdiebstahl, Erpressung
- ▶ Verwendung für Bewertungen
 - ▶ soziale Bewertungen aller Art, Kredit(un)würdigkeit...
- ▶ Täuschung / Leitung
 - ▶ Nach Interesse (z.B. Anzeige von News, Bewertung)

Vor- und Nachteile liegen nah beieinander



•Die Gesundheit von Angestellten modellieren, um Abwesenheit vorherzusagen oder Fitnessstunden anzubieten?



TV-Konsum überwachen, um die Programmplanung und Werbung zu verbessern?



Preise an die individuelle Nachfrage und zeitlich dynamisch anpassen?



Homosexuelle Paare identifizieren, um ihnen spezielle Urlaubsreisen anzubieten?

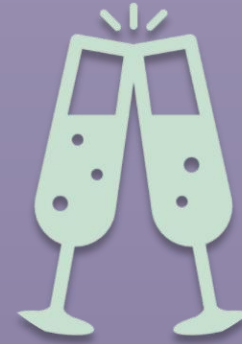
Welche Unterscheidung ist fair?



Über Jahrhunderte mussten Reiche mit großen Häusern höhere Versicherungsprämien bezahlen als ärmere mit kleineren Häusern. Ist das fair?



Mit Datenanalyse
In einigen Regionen verlangen Leihwagenfirmen von jungen Lenkern höhere Gebühren (oder lassen sie gar nicht fahren). Ist das fair?



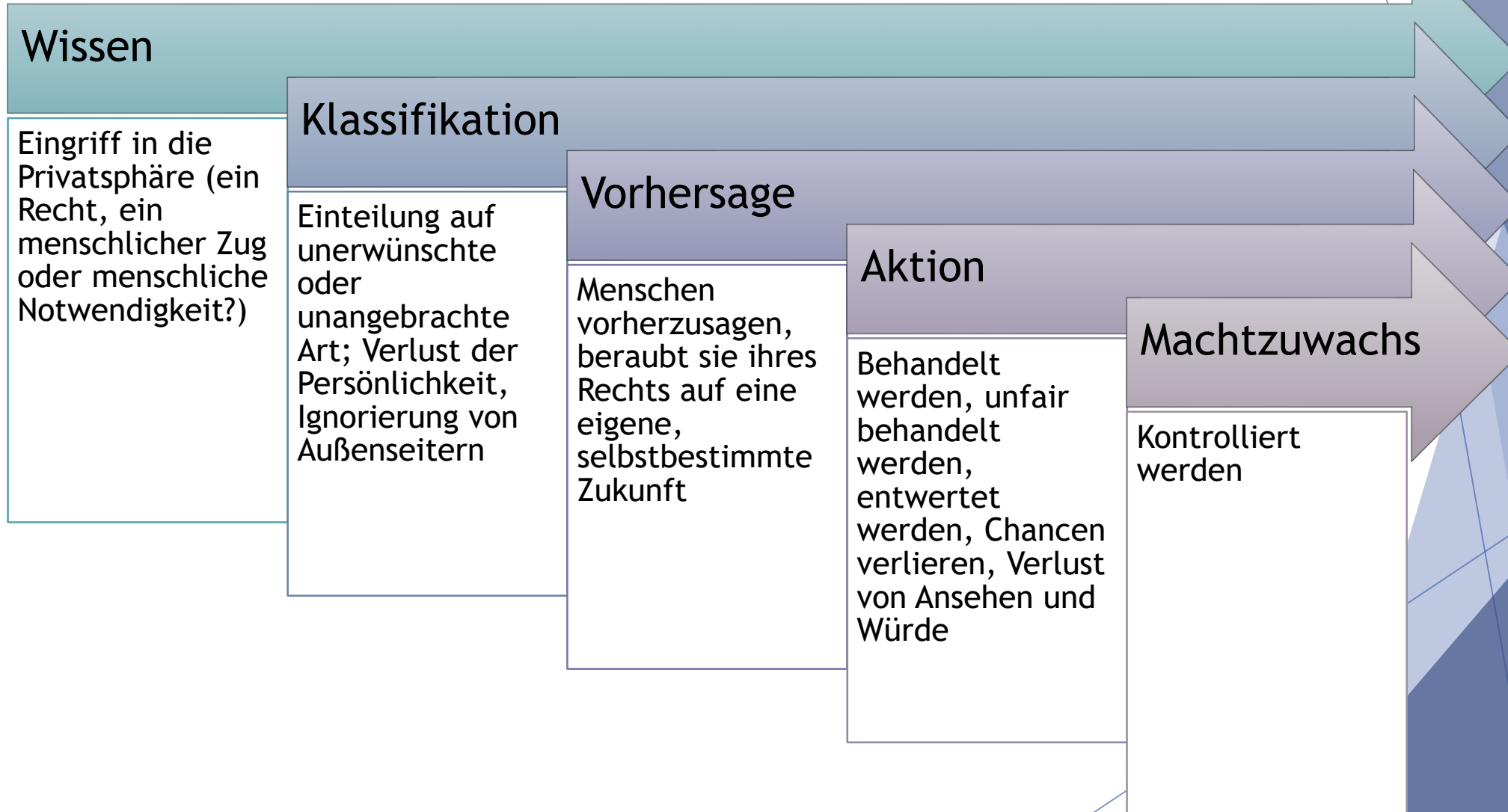
Mit Überwachung
Junge männliche Lenker haben mehr Unfälle, vor allem abends und am Wochenende. Sollen sie weniger zahlen, wenn sie dann nicht fahren und sich überwachen lassen?



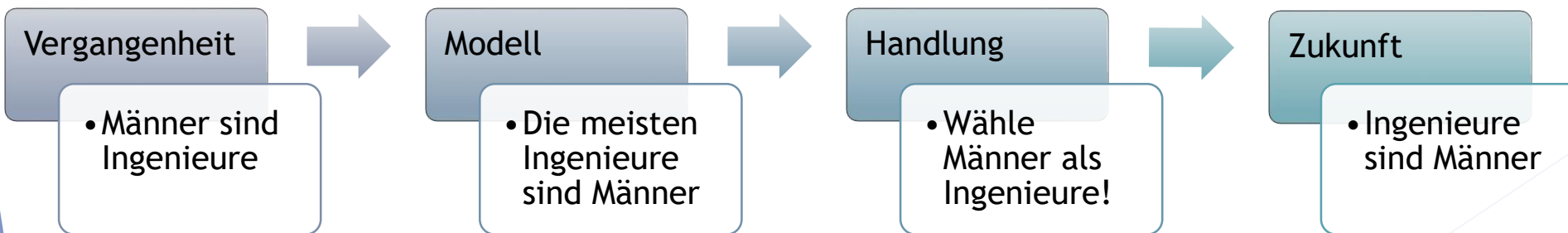
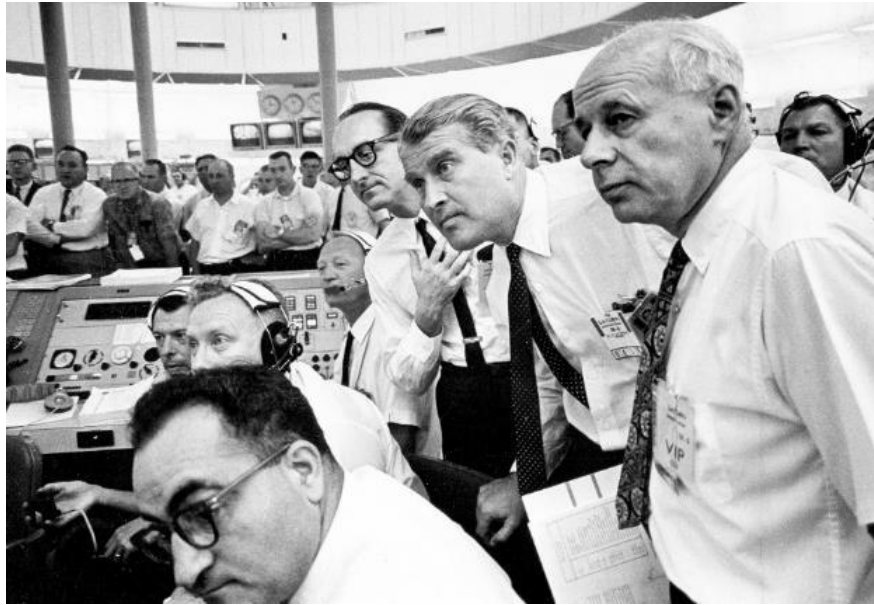
Mit KI
Soll das Auto eine weniger gefahrenträchtige Route vorschlagen und die Versicherung erhöhen, wenn man diese dennoch nimmt?

Ethik oder Politik

Ethische Probleme sind prinzipieller Natur



Ethik prognostischer Systeme: Prognose auf Basis der Vergangenheit



Bias

- **Voreingenommenheit** eines Algorithmus aufgrund von Eigenschaften, die nicht zur Bewertung herangezogen werden sollten
- **Besonders sensitive** Eigenschaften sind Geschlecht, Hautfarbe, Einkommen, sexuelle Orientierung und Identität, Religion ua.

Bias

- Automation Bias: Computer says no!
- Selection Bias: schlechte Auswahl der Daten
- Overfitting/Underfitting
- Outlier: extreme Daten verändern Basis
- Messungsbias: Datenbasis nicht repräsentativ
- Beobachterbias: Labelling von Daten
- etc.

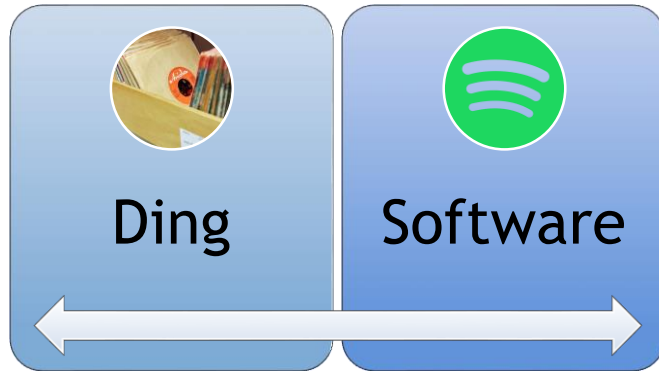
Welche Kreditdiskriminierung von Frauen ist fair?

Fairnessmaß (Literatur)	Intuition/Beispiel
Maximierung der Genauigkeit	Das genaueste Modell gibt allen den Kredit und die Rate, die sie <i>anhand der Daten</i> “wirklich” verdienen
Demographische Parität, Gruppenfairness	Frauen und Männer haben die gleiche Genehmigungsrate
Gleiche Chance	Unter den kreditwürdigen Antragstellern haben Frauen und Männer ähnliche Erfolgsraten
Vorhersagegleichheit	Unter den nicht kreditwürdigen Antragstellern haben Frauen und Männer ähnliche Erfolgsraten
Positive Vorhersageparität	Unter den kreditwürdigen Antragstellern ist die Wahrscheinlichkeit, die Kreditwürdigkeit vorherzusagen unabhängig vom Geschlecht
Kontrafaktische Fairness	Für jede Person wäre die Vorhersage gleich, wenn sie das andere Geschlecht hätten
Individuelle Fairness	Jede Person erhält dasselbe Ergebnis wie andere „ähnliche“ Personen des anderen Geschlechts

Nicht Ethik, sondern Politik!

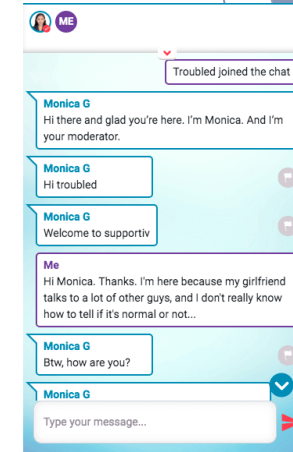
Technisch unvermeidbare Entscheidung!

Harmlos scheinende Digitalisierung hat große Auswirkungen



Kommunikation

- Überwachen
- Verändern, ausscheiden
- Bereitstellen



Auto

- Versichern
- Abschalten
- Verfolgen
- ...



Elektrizität

- Überwachen
- Abschalten
- Versteigern
- ...



Der Wechsel von ehemals unbeobachtetem / nicht beobachtbarem Verhalten zu digitalen Services führt zu Kontrollmöglichkeiten und Machtverschiebung.

Machtverschiebung



John Deere wollte nicht, dass Dritte Zugang zur Software der Traktoren (Kosten über US\$ 100,000) bekommen.

Argumente: Sicherheit, Wartbarkeit - Updates

Farmer begannen die Software zu hacken - mit „Piratenprogrammen“, um Zugang zu den Diagnosedaten zu bekommen...



Teslabesitzer können nur tun, was der Hersteller erlaubt.

Nicht für Lyft oder Uber verwenden wenn das Fahrzeug autonom fährt. (Nur das „Tesla Network“)

Fernabschaltung

- Regelmäßige Sicherheitsupdates
- Neue Funktionen
- Pflicht? (Wessen?)
- Wem gehören die Daten?

Vom Sachenrecht zum Schuldrecht

Kennen sie die Nutzungsbedingungen ihres Geräts oder Dienstes?

- ▶ Nein.
- ▶ De facto nicht mehr möglich

Nutzungsregeln ändern sich...

- ▶ Musiksysteme: Updates und Zugriff auf Daten
- ▶ Staubsaugerroboter: Zimmer-Layout an den Hersteller übermitteln
- ▶ TV: Kunden Sehgewohnheiten übermitteln



Neue Herausforderung für den Schutz von Konsumentinnen und Konsumenten

- ▶ Verlust etablierter Rechte durch Wechsel von Dingen zu Diensten
- ▶ Überwiegend ungeklärte Situationen (Daten, Prinzipien, Pflichten etc.)
- ▶ Leben in der Zustimmungsfiktion
- ▶ Verschiebung zu den Interessen der Anbieter oft unbemerkt
- ▶ Verschiebung der Macht zu Herstellern bisher kaum adressiert
- ▶ Großer Handlungsbedarf: rechtlich, ethisch, politisch

Eine konstruktive Initiative, die Mensch und Gesellschaft in den Mittelpunkt rückt

- ▶ Digitaler Humanismus ist eine Initiative, die Digitalisierung zu gestalten, um Mensch und Gesellschaft in den Mittelpunkt zu rücken. Digitale Technologie soll Menschen in ihrer Autonomie stärken, ihre Unterschiede anerkennen und ihre Entscheidungsfähigkeit unterstützen.
- ▶ Digitaler Humanismus ist ein Aufruf, digitale Technologien für den Schutz von Menschenrechten und zur Entwicklung der Demokratie einzusetzen. Digitaler Humanismus wurzelt in den europäischen Werten der Aufklärung(*): Freiheit, Demokratie, Inklusion und Menschenrechte.
- ▶ Digitaler Humanismus ist bestrebt, digitale Technologien für Fortschritt und Innovation und für die Wahrung und den Ausbau sozialer Errungenschaften zu nutzen. Technik soll Solidarität, soziale und kulturelle Errungenschaften und sinnbehaftete Arbeit fördern.



La Liberté guidant le peuple (Eugène Delacroix)

(*) NB: post-Humanismus, der Kritik berücksichtigt und daher auch Aspekte der Umwelt und Diversität betont.

This is what ChatGPT says about me:

The computer researcher Erich Prem has delved into the realm of Artificial Intelligence with a keen focus on autonomous robots. Through his extensive writings, Prem has explored the intersection of AI and robotics, probing the intricacies of creating machines capable of independent decision-making and action. His work not only offers insights into the technical aspects of developing autonomous robots but also delves into the ethical implications of such technology, paving the way for a comprehensive understanding of the future where intelligent machines navigate our world. With a unique blend of technical expertise and philosophical inquiry, Erich Prem's contributions stand as a testament to the ongoing evolution of AI and its tangible impact on our lives.

Dr.phil. Dr.tech. Erich Prem, MBA

www.erichprem.at

prem at eutema.com

 @ErichPrem

Institut für Philosophie, Uni Wien

<https://philtech.univie.ac.at/team/erich-prem/>

eutema GmbH

www.eutema.com

Verein zur Förderung des digitalen Humanismus

www.digitalhumanism.at

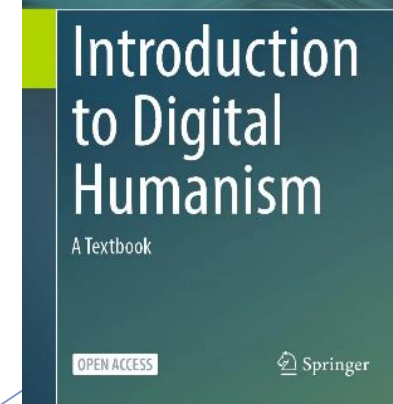
eu|tel|ma
TECHNOLOGY MANAGEMENT



UNIVERSITY
OF VIENNA

TU
WIEN

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology



<https://dighum.ec.tuwien.ac.at/perspectives-on-digital-humanism/>