



universität  
wien

# Digitaler Humanismus und seine Herausforderungen in Zeiten der KI

Dr. tech. Dr. phil. Erich Prem

Institut für Philosophie, Universität Wien und TU Wien

Verein für digitalen Humanismus

November 2024

# Über mich



Computer science



Artificial Intelligence



Autonomous robots



Epistemology, ethics



*RTDI Strategy*



Venture Capital



Research management



Art/science/technology



[www.feartart.eu](http://www.feartart.eu)



**Verein zur Förderung des digitalen Humanismus**

[www.digitalhumanism.at](http://www.digitalhumanism.at)

[www.eutema.com](http://www.eutema.com)

[www.erichprem.at](http://www.erichprem.at)

@ErichPrem

<http://at.linkedin.com/in/erichprem>



# Unsere post-digitale Welt

KI ist überall

Fotooptimierung

Musikempfehlung

Gesundheitsmonitoring

Spracherkennung

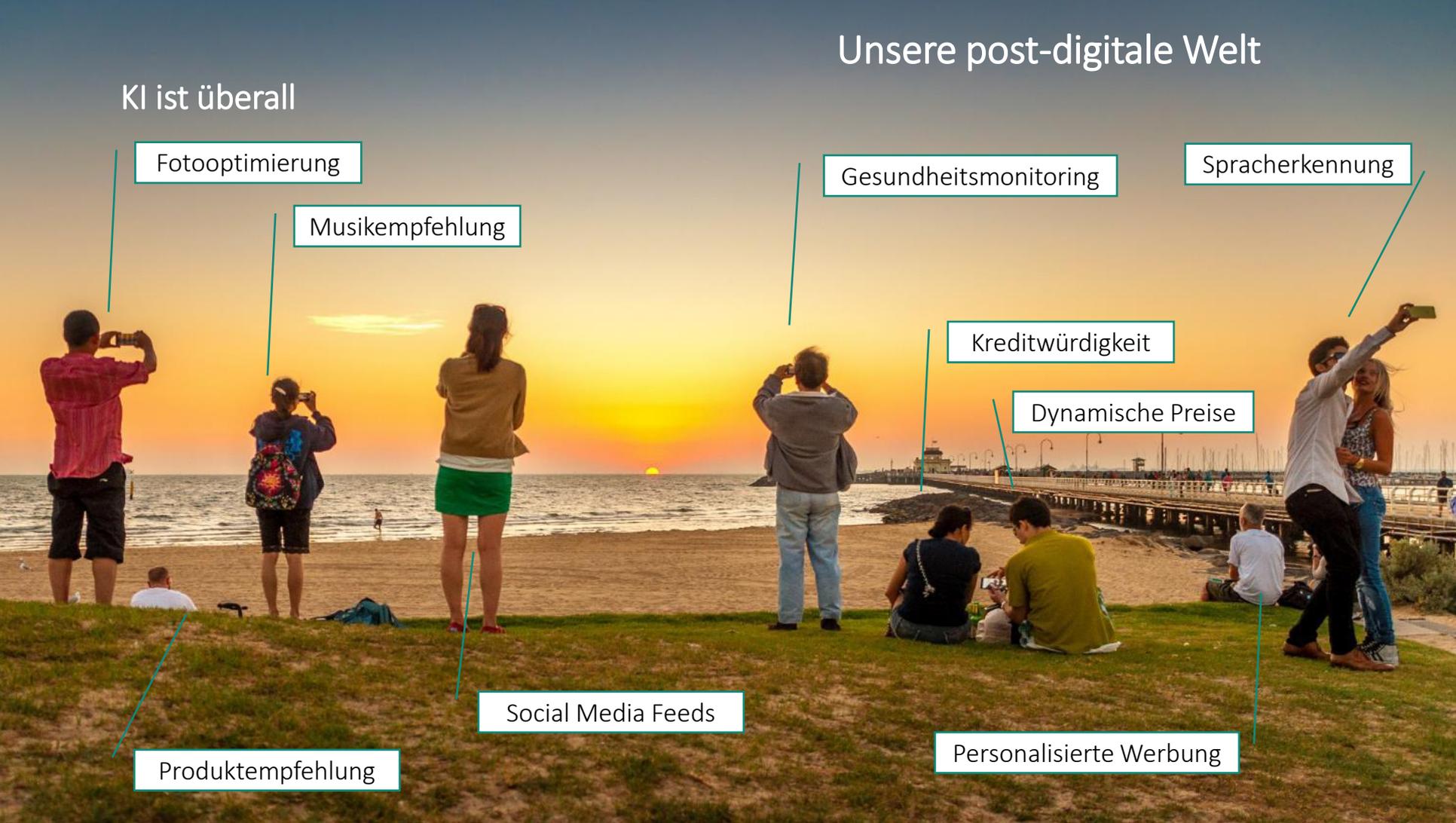
Kreditwürdigkeit

Dynamische Preise

Social Media Feeds

Produktempfehlung

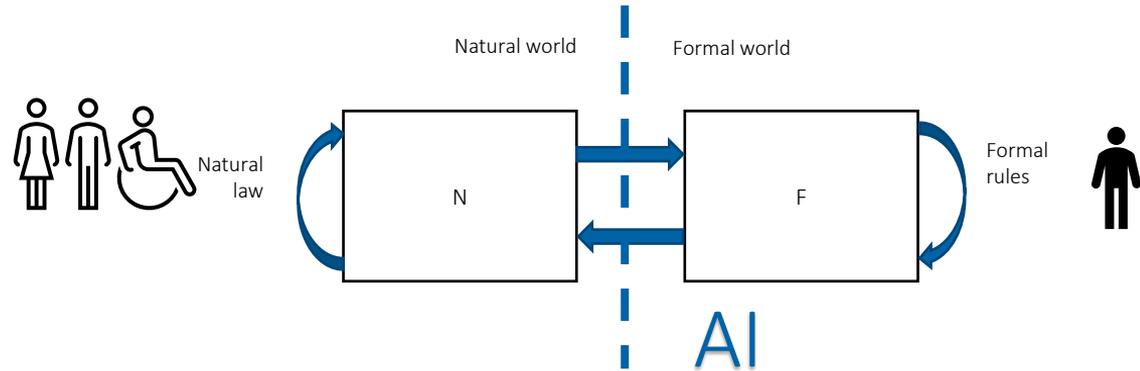
Personalisierte Werbung



# Menschen mit KI modellieren

Unsere KI-Modelle haben Grenzen, was sie wissen, vorhersagen oder kontrollieren können.

In Bezug auf Menschen sind wir auch begrenzt darin, was wir tun sollten.



Sollen wir von Personen wissen

- Geschlecht, Religion, sexuelle Orientierung
- Suche im Internet
- Einkäufe in der Apotheke

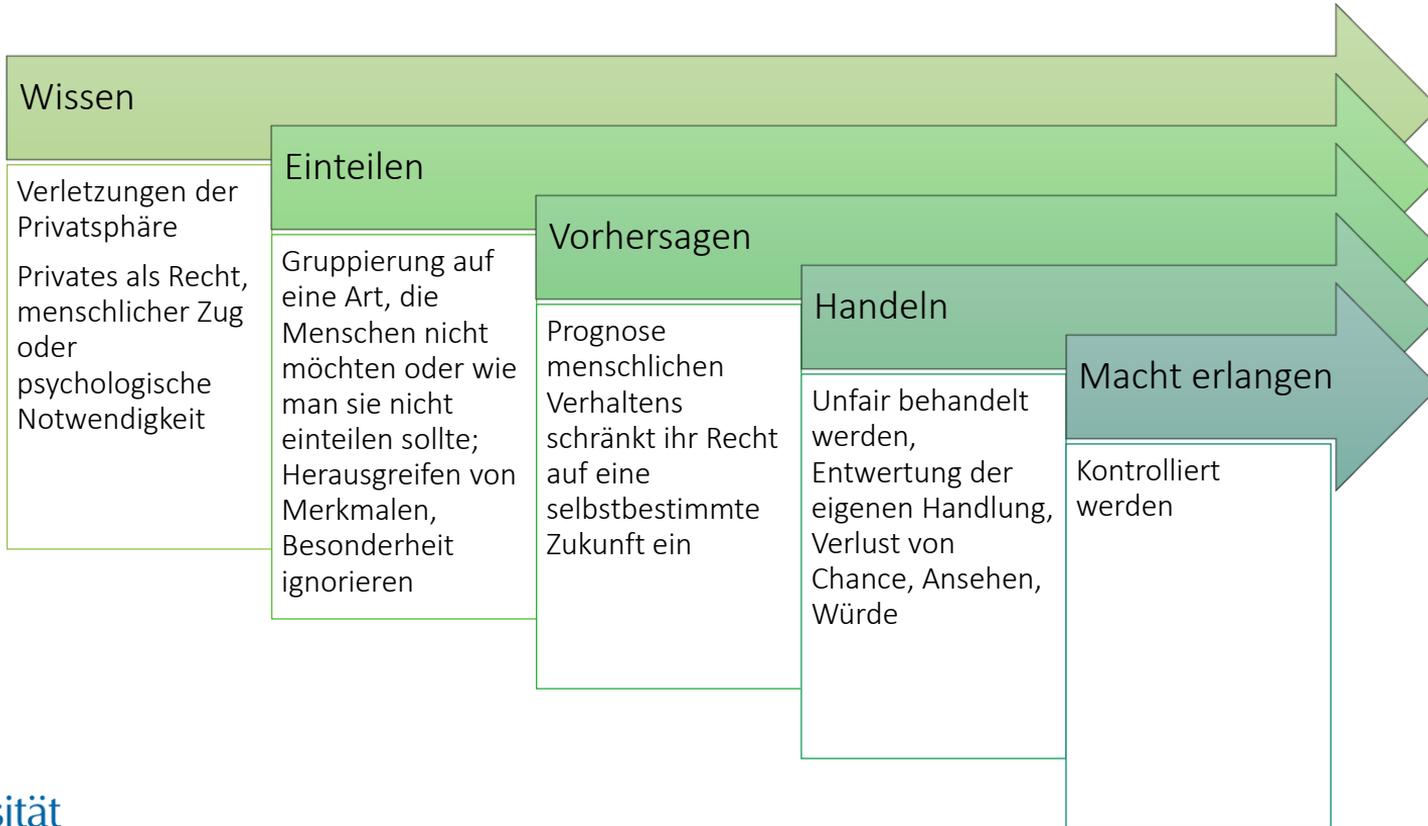
Sollen wir über Personen vorhersagen

- Talent
- Todeszeitpunkt
- Krankheit

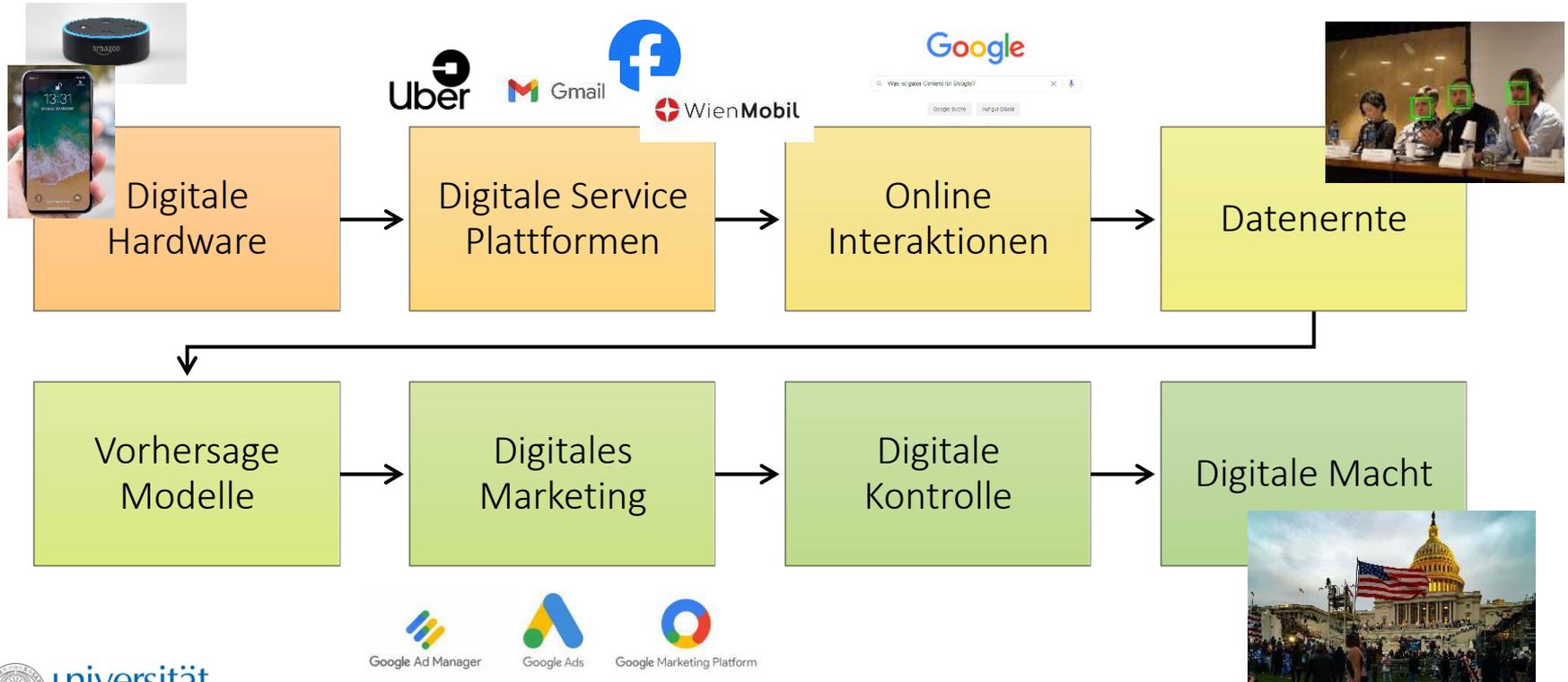
Sollen wir Personen kontrollieren bei

- Autofahrten
- Diät
- Bewegung
- Kommunikation

# Antihumanistische Tendenzen im Digitalen



# Eine dynamische digitale Prozesskette erzeugt neue Phänomene von Überwachung, Prognose und Kontrolle



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

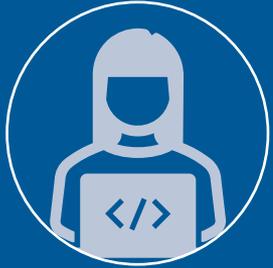
Digitaler Humanismus wurzelt in den europäischen Werten der Aufklärung(\*): Freiheit, Demokratie und Menschenrechte. Digitale Technologie soll Menschen in ihrer Autonomie stärken, ihre Unterschiede anerkennen und ihre Entscheidungsfähigkeit unterstützen.

Sie soll Solidarität, soziale und kulturelle Errungenschaften und sinnbehaftete Arbeit fördern.



*La Liberté guidant le peuple (Eugène Delacroix)*

# Fünf Botschaften des digitalen Humanismus



Digitale Technologie verändert uns und unser Leben

- Folgen
- Fortschritt
- Macht
- Wirtschaft



Digitale Technologie muss den Menschen und seine Umwelt schützen

- Menschenwürde
- Autonomie
- Fairness
- Umwelt
- Diversität



Digitale Technologie muss Demokratie und Gesellschaft stärken

- Demokratie und Menschenrechte
- Freiheit
- Solidarität
- Kultur
- Ausbildung



Technologie ist gestaltbar

- Design
- Kreativität
- Zusammenarbeit Industrie und Wissenschaft
- Internationale Zusammenarbeit
- Demokratische Regulierung



Maschinen sind keine Menschen

- und umgekehrt
- Dienst für Menschen
- Menschliche Eigenheiten

# Klassifikation menschlichen Verhaltens

Sind persönliche Daten andere? Warum?  
Wie viel *sollen* wir über andere wissen?

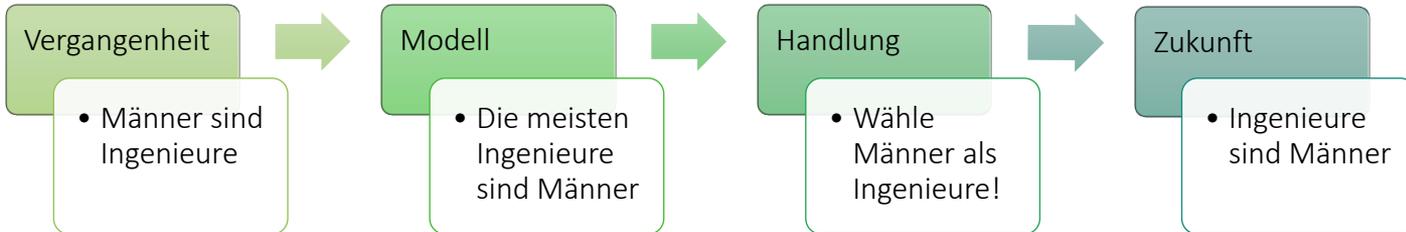
Beispiel: Überwachung einer Toilette, z.B. für  
Menschen mit Behinderungen oder Demenz

Keine rein rechtliche Frage:

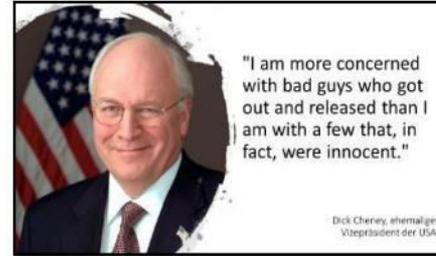
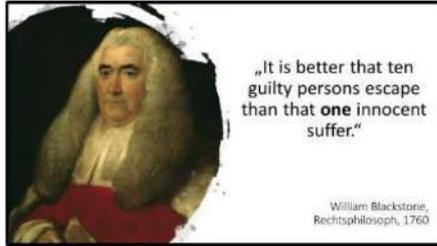
Menschenwürde als Grund für Privatsphäre



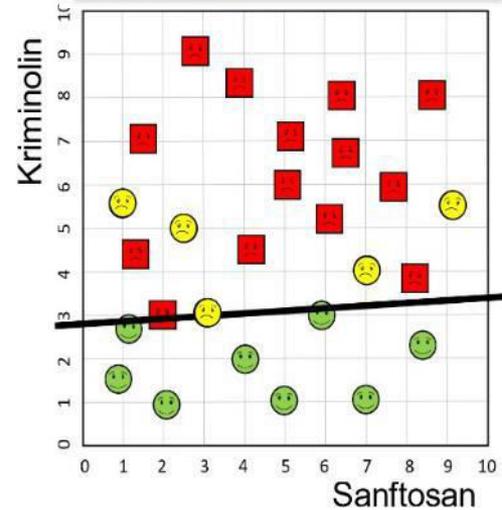
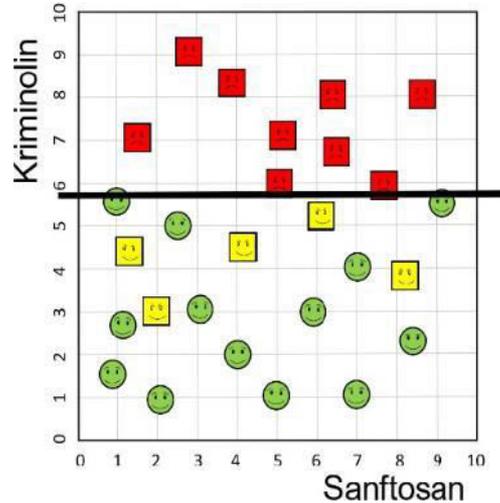
# Prognose auf Basis der Vergangenheit



# Grenzen der Entscheidung



Ethik oder Politik?



# Welche Diskriminierung Schwarzer ist fair? Bitte wählen Sie...

Fairnessmaß (Literatur)	Optimierung	Intuition/Beispiel	Philosophy
Maximierung der Genauigkeit	N/A	Das genaueste Modell gibt den Leuten den Kredit und die Rate, die sie anhand der Daten "wirklich" verdienen	Desert
Demographische Parität, Gruppenfairness	Ergebnis	Schwarze und Weiße Bewerber haben die gleiche Genehmigungsrate	Fair equality of opportunity
Gleiche Chance	FNR	Unter den kreditwürdigen Antragstellern haben schwarze und weiße Antragsteller ähnliche Erfolgsraten	
Vorhersagegleichheit	FPR	Unter den nicht kreditwürdigen Antragstellern haben schwarze und weiße Antragsteller ähnliche Erfolgsraten	
Positive Vorhersageparität	TPR, TNR, PPV	Unter den kreditwürdigen Antragstellen ist die Wahrscheinlichkeit, die Kreditwürdigkeit vorherzusagen unabhängig von der Hautfarbe	Cause and effect
Kontrafaktische Fairness	Vorhersage des nicht eingetretenen Falls	Für jede Person wäre die Vorhersage gleich, wenn sie die andere Hautfarbe hätten	
Individuelle Fairness	Ergebnis für ähnliche Personen	Jede Person erhält das selbe Ergebnis wie andere „ähnliche“ Personen der anderen Hautfarbe	Responsibility-sensitive egalitarianism

# Wie würden Sie unterscheiden?



Prämien für teure Autos sind höher als für billige? Ist das fair?



## Mit Datenanalyse

Jüngere Fahrer haben höhere Prämien (oder werden ausgeschlossen) wenn sie ein Auto mieten. Fair?



## Mit Kontrolle

Junge Männer haben mehr Unfälle in der Nacht oder an Wochenenden. Sollen sie eine Reduktion fürs Nicht-Fahren erhalten und eine Blackbox installieren?



## Mit KI

Soll das Auto eine sichere Route empfehlen und falls man sich anders entscheidet die Prämie erhöhen?

Ethik oder Politik?

# Extensionsethik

Amor vincit omnia: Kunst, Medizin oder Pornografie?

Externalisation / Extensionalisierung der Wahrheit

- Digitale Beurteilung von Legalität
- Exklusive Orientierung an formalen Kriterien.
- Beurteilung von Äußerlichkeiten, nicht der Intention.  
Ist Pornografie auf Nacktheit reduzierbar?
- Intentionen werden nie abgebildet.  
Vgl. EU-Debatte zur Überwachung der elektronischen Kommunikation
- Weitere Beispiele
  - Upload-Filter, automatisches (paternalistisches) Fahren



[https://de.wikipedia.org/wiki/Datei:Caravaggio\\_-\\_Cupid\\_as\\_Victor\\_-\\_Google\\_Art\\_Project.jpg](https://de.wikipedia.org/wiki/Datei:Caravaggio_-_Cupid_as_Victor_-_Google_Art_Project.jpg)

# Sollen wir uns ethische KI wünschen?



Digitale Systeme helfen, Normen so zu implementieren, dass ihre Verletzung de facto unmöglich wird.

Beispiele:

Verkehr, Eigentum, menschliche Beziehungen, Sprechen....

- Abkehr vom Rechtsprinzip der Strafe *nach* der Handlung.
- Frage der Ausnahme, Notsituation, Protest, Autonomie, Fehler

Tendenz zu einer Gesellschaft, die Normverletzungen nicht nur bestraft, sondern verhindert? Recht auf Rechtsverletzung?

*“Unsere Technologie erzwingt langsames Fahren und Fahrverbote, um unsicheres Fahren zu erkennen und zu verhindern.”*



universität  
wien

Superpedestrian.com (Link)

# Kontakt

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

[www.erichprem.at](http://www.erichprem.at)

erich.prem@eutema.com

 @ErichPrem

Institut für Philosophie, Uni Wien

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

eutema GmbH

[www.eutema.com](http://www.eutema.com)

Association of digital humanism

[www.digitalhumanism.at](http://www.digitalhumanism.at)



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