

KI und Verantwortung

Julia Neidhardt

Forschungsbereich E-Commerce

TU Wien, Österreich

julia.neidhardt@tuwien.ac.at

<http://www.ec.tuwien.ac.at/neidhardt/>



Über mich

- Hintergrund in Mathematik und Informatik
- Forscherin an der Fakultät für Informatik an der TU Wien
- Gastforscherin an der ÖAW
- Teil der Digital Humanism Initiative der TU Wien
- Schwerpunkte: Recommender Systeme, Soziale Netzwerkanalyse, Maschinelles Lernen
- Interdisziplinäre Forschung in den Bereichen Digital Humanities und Digital Humanism

Informatik

- **Digitale Transformation** zeigt die allgegenwärtige Rolle der Informatik
 - Wichtig für Innovation, Wachstum und Wohlstand
 - Informatik hat direkten Einfluss auf (fast) alles
- Grundlegend – Methoden und Artefakte werden entwickelt
- Metamorphose vom PC zur globalen Maschine
 - Heute kann fast alles ein Computer werden



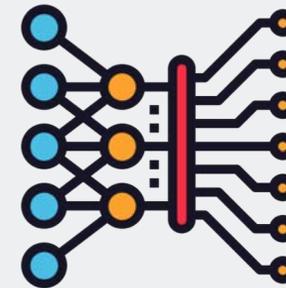
Digitalization

- Die Digitalisierung hat zu mehr Effizienz und Komfort geführt, z. B. kann sie Ressourcen sparen, die Produktivität steigern oder Menschen befreien
- Die schnelle Entwicklung hat ihre **Schattenseiten**, zum Beispiel:
 - **Web:** globales Megasytem mit Monopoltendenz
 - Am Anfang: Vision der Informationsfreiheit
 - Führt zu einem werbebasierten Geschäftsmodell; User zahlen mit ihren Daten
 - **Das System versagt**, Tim Berners-Lee, Gründer des WWW
 - **Das Internet entschuldigt sich**, Jaron Lanier, Pionier Virtual Reality



Künstliche Intelligenz (KI) / Maschinelles Lernen (ML)

- KI/ML automatisiert viele Bereiche (z.B., Arbeit und Entscheidungsfindung)
 - Diskussionen zwischen **Utopie und Dystopie**
 - Neuronale Netze als aktuelles Paradigma der KI: Roy Amara (IFTF Präsident) mit seinem „Gesetz“: ***Wir neigen dazu, die Wirkung einer Technologie kurzfristig zu überschätzen und langfristig zu unterschätzen.***
 - Viele KI-Sommer und Winter seit den 50er Jahren
 - Der aktuelle Boom erforderte einen zugehörigen Technologie-Stack:
 1. Rechenleistung
 2. Speicher
 3. Verfügbare / zugängliche Daten

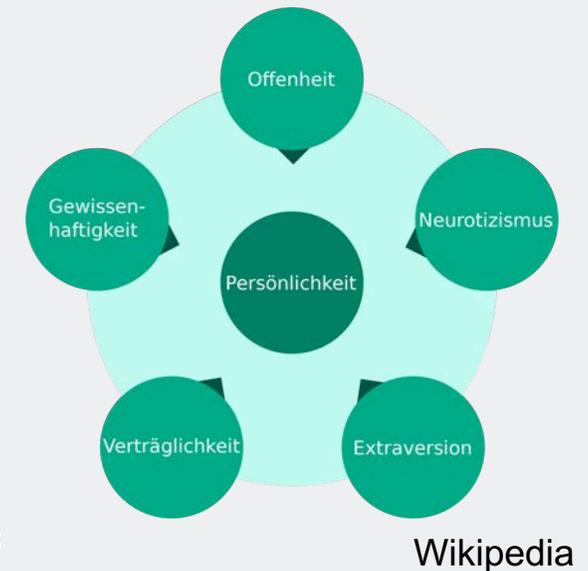


→ **Viele kritische Bereiche**

Persönlichkeitsstruktur

- **Personalisierung:** Allgegenwärtig aber notwendig, um der Informationsüberflutung entgegenzuwirken
- **Implizites Eruiere von Persönlichkeitsmerkmalen:** Analyse des Nutzungsverhaltens und von User Generated Content ermöglicht, viel über die User in Erfahrung zu bringen - sogar über deren Persönlichkeitsstruktur („Big Five“-Eigenschaften).
 - Mikro-Blogs (Twitter)
 - Soziale Medien (Facebook, Weibo)
 - Spiele
 - Handy-Protokolle
 - Stereotype Geschichten
 - Ausgewählte Bilder

→ Hohe Missbrauchsgefahr (siehe Cambridge Analytica!)



Biometrisches Tracking

The composite image consists of three main parts:

- Top Left:** A video frame of a woman with long brown hair, smiling slightly. The name 'MadamSexypants99' is visible in the bottom left corner.
- Top Right:** A table with biometric data:

BMI	21.7	19.3-25.3	Yes
Life expectancy	89	75.8-85.8	No
Came closer?	No	No (66%)	Yes
- Bottom Left:** A smaller video frame of the same woman with a sad expression, enclosed in a red bounding box. A red label 'Sad' is overlaid on the top left of this frame.
- Bottom Right:** A dark blue interface showing a circular image of a dog wearing a pink hat. Below it, text reads: 'Your reaction to the doggy was: sad' and 'The dominant expression on your face is currently: happy.'

<https://www.hownormalami.eu>

- Read terms?
- Shared age?
- Beauty
- Age
- Lied about age?
- Gender

- **Beauty**-Scoring-Modelle, Modelle zur Vorhersage von **Alter**, **Geschlecht** und **Mimik/Emotion**, **BMI**-Vorhersage-modelle, etc.

Biometrisches Tracking (2)

- Findet schon breite Anwendung:
 - **Plattformen** wie TikTok und Tinder verwenden Beauty-Scoring-Modelle
 - **Überwachungssysteme** werden eingesetzt, die die Demografie von KundInnen analysieren
 - BMI-Vorhersage Projekt von **Google**-MitarbeiterInnen (Google: privates Projekt)
 - Lebenserwartung: Wettbewerb initiiert von **Versicherungsunternehmen**
 - Gesichtsabdruck: **Cross-Plattform Tracking** Überwachungssysteme <https://www.hownormalami.eu>
- Was bedeutet es auf lange Sicht für die Gesellschaft?



Autonome Waffen

AUTONOMER ANGRIFF

Drohne soll offenbar selbstständig Soldaten angegriffen haben

Das geht aus einem UN-Bericht hervor. Unklar ist, ob es bei dem Angriff in Libyen Opfer gab

1. Juni 2021, 11:41 275 Postings



Eine Drohne des Typs Kargu-2-Quadcopter soll eigenmächtig einen Angriff auf einen Soldaten geflogen sein.

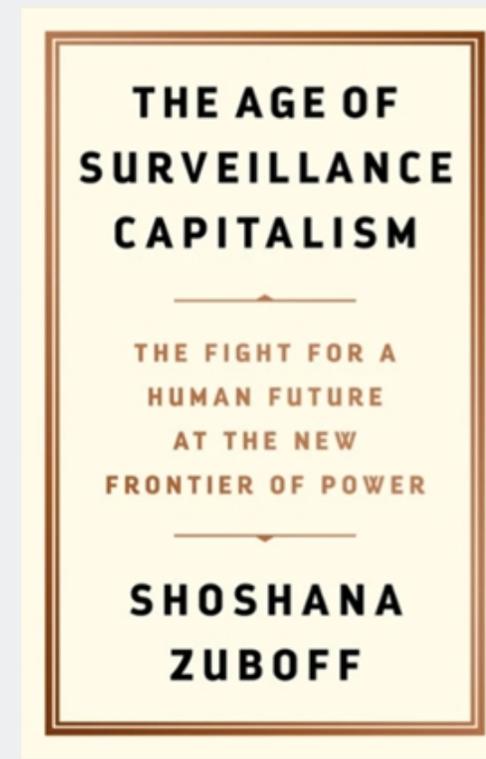
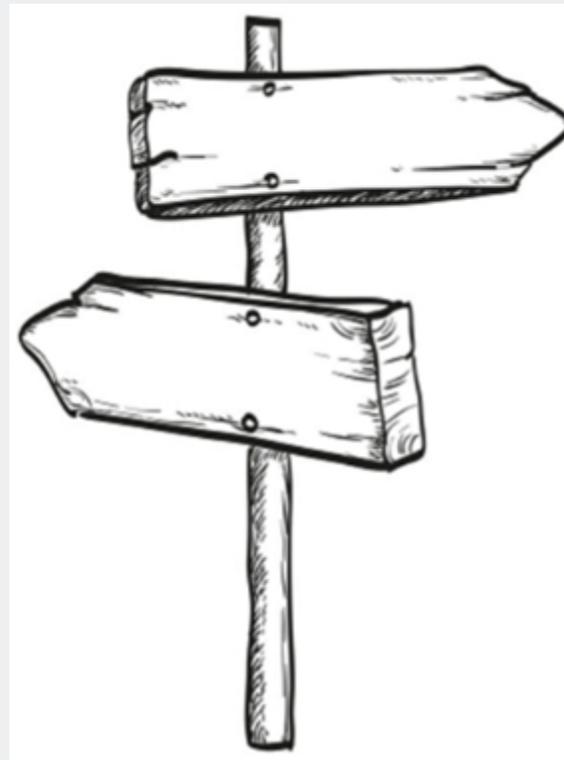
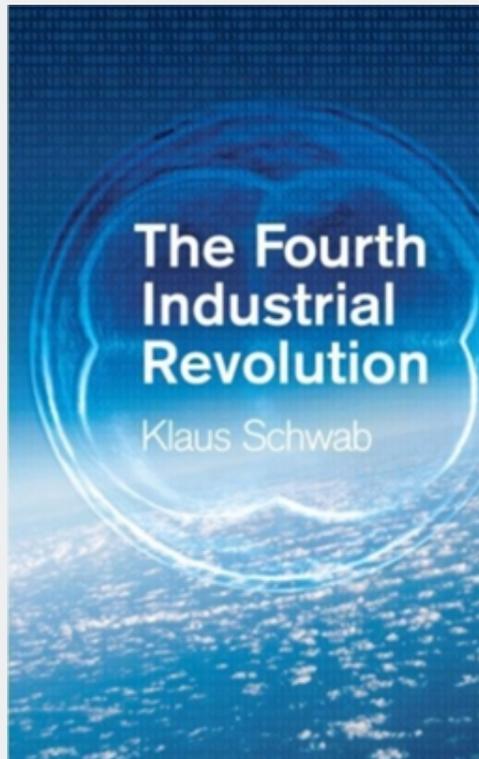
Foto: APA/dpa/Felix Kästle

derStandard.at

- Szenario: Vollständig autonome Waffen, die ohne weiteres menschliches Eingreifen über Leben und Tod entscheiden
 - Konsequenzen sind nicht absehbar (Einfluss auf Hemmschwelle?)
 - Auch andere Einsatzgebiete, z.B. Grenzkontrolle?
- Internationale Ächtung wird seit Jahren diskutiert, aber von (einigen) militärische Großmächte abgelehnt

Welchen Weg?

4. Industrielle Revolution vs. Überwachungskapitalismus



Leitlinien für die Entwicklung von KI-Systemen

- Immer mehr Unternehmen haben Leitlinien, z.B.:
 - Human-in-the-Loop
 - Evaluierung von Bias
 - Erklärbarkeit und Transparenz
 - Reproduzierbarkeit
 - Datenschutz
 - Folgeabschätzung für MitarbeiterInnen
 - Beyond-Accuracy
 - Datenrisikobewusstsein



→ Positive Entwicklung, aber noch immer viele Probleme; öffentlicher Druck wichtig!

Die Digitale Humanismus Initiative



- April 2019: Wiener Workshop **Digitaler Humanismus** an der TU Wien (unterstützt von der Stadt Wien und vom WWTF)
- Digitaler Humanismus
 - Bezug auf Humanismus und Aufklärung, wonach der Mensch selbstverantwortlich ist, sich des eigenen Denkens bedient und im Mittelpunkt steht (Nida-Rümelin und Weidenfeld 2018)
 - **Für uns:** Ansatz, der das komplexe Zusammenspiel von Technologie und Menschheit beschreibt, analysiert und vor allem versucht zu beeinflussen - für eine Gesellschaft unter voller Achtung universeller Menschenrechte.

Workshop Digitaler Humanismus

- Über 100 TeilnehmerInnen – weltweit – Informatik, Philosophie, Geschichte, Anthropologie, Recht, Ökonomie, Politikwissenschaften, Mathematik, Soziologie
- Drei Themenblöcke:
 - Geschichte und den Auswirkungen der Informationstechnologien
 - Gesellschaft, Künstlicher Intelligenz und Ethik
 - Dynamik und Zukunft des Sektors



Vienna Manifesto on Digital Humanism



- Wichtiges Resultat des Workshops
- Das Manifest ist ein Aufruf zum Nachdenken und Handeln (gerichtet auch an Entscheidungsträger), es ist auch ein Forschungsprogramm
- Formen wir Technologien nach menschlichen Werten und Bedürfnissen, anstatt dass Technologien Menschen formen



Vienna Manifesto on Digital Humanism (2)

Zentrale Forderungen sind unter anderem:

- **Demokratie und Beteiligung**

- Digitale Technologien sollten so gestaltet sein, dass sie Demokratie und Inklusion fördern.
- Privatsphäre und Redefreiheit sind Grundwerte, die im Mittelpunkt unserer Aktivitäten stehen sollten.

- **Regulierung und öffentliche Kontrolle**

- Die Regulierungsbehörden müssen gegenüber Technologiemonopolen intervenieren.
- Entscheidungen, deren Folgen die individuellen oder kollektiven Menschenrechte betreffen können, müssen weiterhin vom Menschen getroffen werden.



Vienna Manifesto on Digital Humanism (3)

- **Spezifische Rolle der Wissenschaft**

- Wissenschaftliche Ansätze, die verschiedene Disziplinen integrieren und disziplinäre Silos aufbrechen, sind die Voraussetzungen, um unsere Herausforderungen zu meistern.
- Universitäten sind jene Orte, an denen neues Wissen erzeugt und kritisches Denken gepflegt wird. Sie haben daher eine besondere Verantwortung.

- **Bildung und Ausbildung**

- Es bedarf neuer Curricula, welche Geistes-, Sozial-, technische und Ingenieurwissenschaften kombinieren.
- Die Ausbildung in der Informatik und Bildungsarbeit über ihre gesellschaftlichen Auswirkungen muss so früh wie möglich beginnen.

Themencenter im Digitalen Humanismus

Automatisierung und
Arbeit

Partizipation und
Gesellschaft

KI – Ethik – Philosophie –
Menschliche Kontrolle
and Identität

Fake News,
Echokammern,
Kommunikation,
Demokratie

Tech-Giganten,
Plattformökonomie,
Regulierung

Privatsphäre,
Überwachung,
Menschenrechte

Bildung und Ausbildung

Digitale Politik und
Geopolitik, Souveränität

Systemresilienz vs.
Effizienz, Partizipation,
Demokratie

DigHum Vortragsreihe

- **Online-Vortragsreihe** zur Diskussion verschiedener DigHum-Aspekte zu den Themenclustern
- Sehr renommierte, internationale Vortragende
- Seit Juni 2020; jetzt (fast) alle zwei Wochen
- **Nächste Veranstaltung:**



June 15, 2021
5:00 – 6:00 PM
(17:00) CEST

Lecture Series

“Digital Humanism and Democracy in Geopolitical Context”

Speaker: **Allison Stanger** (Middlebury College, USA)

Moderator & Respondent: **Moshe Y. Vardi** (Rice University, USA)

Positive Auswirkungen

- Digital Humanism Call 2019 der **Stadt Wien** (€ 320.000)
 - Über 70 Anträge eingereicht, 9 Forschungsprojekte gefördert¹
- Digital Humanism Call 2020 des **WWTF** (> 2 Mio. €)
 - 99 eingereichte Anträge, 26 wurden für die Vollantragsphase ausgewählt -> Entscheidung wurde vor Kurzem bekannt gegeben – 9 Projekte gefördert
- Digitaler Humanismus als einer der **Hauptforschungsbereiche** im Programm der österreichischen Bundesregierung² sowie im Programm der Regierung der Stadt Wien³ angeführt

¹ https://www.ots.at/presseaussendung/OTS_20191115_OT0058/digitaler-humanismus-wien-als-zentrum

² <https://www.bmkoes.gv.at/Ministerium/Regierungsprogramm.html>

³ https://www.wien.gv.at/regierungsabkommen2020/files/Koalitionsabkommen_Master_FINAL.pdf

DigHum Materialien



- **DigHum Plattform** (<https://dighum.ec.tuwien.ac.at>)
 - Manifest (lesen und unterschreiben)
 - Aktuelle Veranstaltungen
 - Veranstaltungsarchiv mit Folien
- YouTube Channel **Digital Humanism**
 - Videos aller Veranstaltungen
- Lektüre:
 - **Hannes Werthner**: *Das Wiener Manifest für Digitalen Humanismus*. In *Digitaler Wandel und Ethik* herausgegeben von Rat für Forschung und Technologieentwicklung, Markus Hengstschläger, Ecwin, 2020
 - **DigHum-Buchprojekt**: 45 internationale Beiträge, im Juni online auf der DigHum-Plattform veröffentlicht, erscheint noch dieses Jahr bei Springer

Zusammenfassung

- Technologie ist keine Magie, sie kommt nicht aus dem Nichts sondern ist auch mit spezifische Interessen und Sichtweisen verbunden
- Wir müssen sie verstehen und reflektieren
- Aber wir können und müssen auch aktiv bei der Gestaltung teilnehmen
- Mehr als Ethik – es geht auch um politische Entscheidungen
- Wien hat sich als internationales Zentrum für Digitalen Humanismus etabliert

Wir sind am Scheideweg!

Popper (1969) “Moral Responsibility of the Scientist” <http://www.unz.com/print/Encounter-1969mar-00052>

Vielen Dank!

<https://dighum.ec.tuwien.ac.at>