



#digital|denken

© Stefan Klatt

✉ FRONT DESK

📍 Foyer des Philosophikums
📍 Domplatz 23
48143 Münster
Tel: +49 251 83-28319
✉ dkphil2024@uni-muenster.de
Servicezeit: So: 12–22 Uhr | Mo–Mi: 8–19
Uhr | Do: 8–16 Uhr

@ #DKPHIL24 | ONLINE

Für alle, die nicht zum Kongress kommen können, streamen wir ausgewählte englischsprachige Veranstaltungen live. Im 10-teiligen Podcast "German Philosophy and the World" können Sie zudem den vielfältigen globalen Verbindungen der deutschen Philosophie nachspüren.

📍 [Zu den Livestreams und zum Podcast](#)



Memory slices by Anna Strasser
**DISCLAIMER: JUST MEMORIES – AIMING FOR CORRESPONDENCE
WITH REALITY BUT CANNOT GUARANTEE IT.**

Gabriele Gramelsberger (Aachen):

“Was ist Philosophie des Digitalen?”

Die Philosophen sind “schuld”

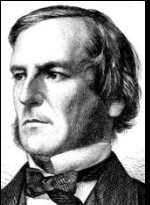
1. Programm der Operationalisierung des **Geistes**
Descartes, Hume, Leibniz
2. Programm der Formalisierung der **Sprache**
Boole, Peirce
3. Programm der Formalisierung des **Schreibens**
Turing, von-Neumann

Digitalitäts der Lebenswelt des 21. Jahrh.

- Schritt 1* ‘Computer’ werden immer kleiner → Smart-phones
- Schritt 2* massive Asymmetrie der Digitalität
- Schritt 3* Körper- und Verhaltensdaten (gesammelt und ausgewertet)
- Schritt 4* Automatisierung der Datenanalyse durch maschinelles Lernen

DEANTHROMORPHISIERUNG?!?

VIELFÄLTIGKEIT DES DIGITALEN REFLEKTIEREN → DGPHIL AG PHILOSOPHIE DER DIGITALITÄT



Ko-Performanz von Mensch und Generativer KI



Anna Strasser
(Berlin)



Christoph Durt
(München)



Christiane Schöttler
(München)



Die Intelligenz der künstlichen Intelligenz

Thesen im Video I

Die Intelligenz der künstlichen Intelligenz

Stellungnahmen von:

Sonja Thiel
Prof. Dr. Gisela Schmalz
Raphael Brähler
Dr. Alexander Klier



Deutsche Gesellschaft für Philosophie e.V. Fokusgruppe „Künstliche Intelligenz“

<https://www.youtube.com/watch?v=mIDTjX9bZyM>



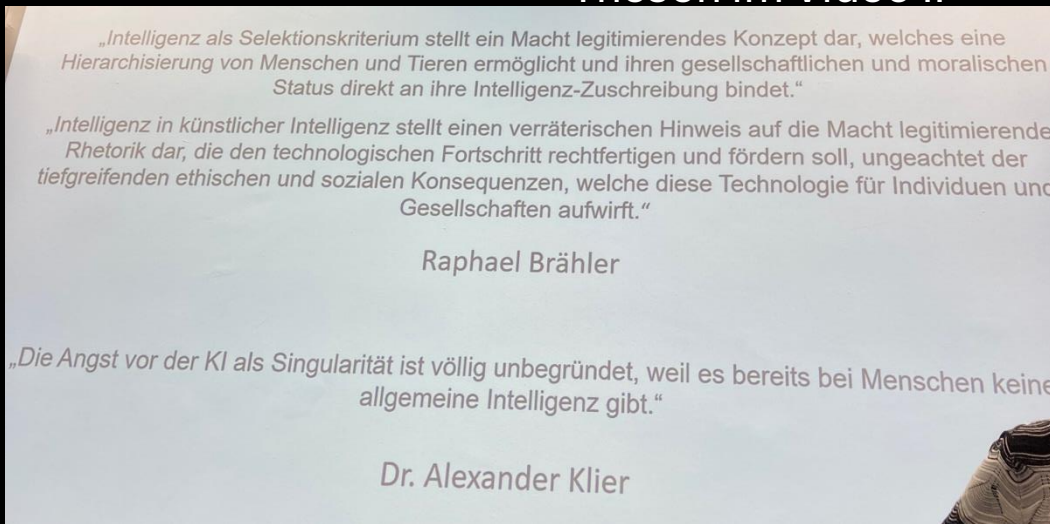
„Ergebnisse maschineller Intelligenz sind Merkmal einer sich aktuell neuformierenden digitalen Kultur und diese können am ehesten transdisziplinär, durch kulturelle Intelligenz und ethischer Perspektive beschrieben werden.“

Sonja Thiel

„Der Begriff Intelligenz sollte dringend von der Philosophie erobert werden. Um die Bedeutung von KI in der Welt besser zu begreifen, müssen wir zunächst die menschliche Intelligenz begreifen.“

Prof. Dr. Gisela Schmalz

Thesen im Video II



„Intelligenz als Selektionskriterium stellt ein Macht legitimierendes Konzept dar, welches eine Hierarchisierung von Menschen und Tieren ermöglicht und ihren gesellschaftlichen und moralischen Status direkt an ihre Intelligenz-Zuschreibung bindet.“

Raphael Brähler

„Intelligenz in künstlicher Intelligenz stellt einen verräterischen Hinweis auf die Macht legitimierende Rhetorik dar, die den technologischen Fortschritt rechtfertigen und fördern soll, ungeachtet der tiefgreifenden ethischen und sozialen Konsequenzen, welche diese Technologie für Individuen und Gesellschaften aufwirft.“

„Die Angst vor der KI als Singularität ist völlig unbegründet, weil es bereits bei Menschen keine allgemeine Intelligenz gibt.“

Dr. Alexander Klier

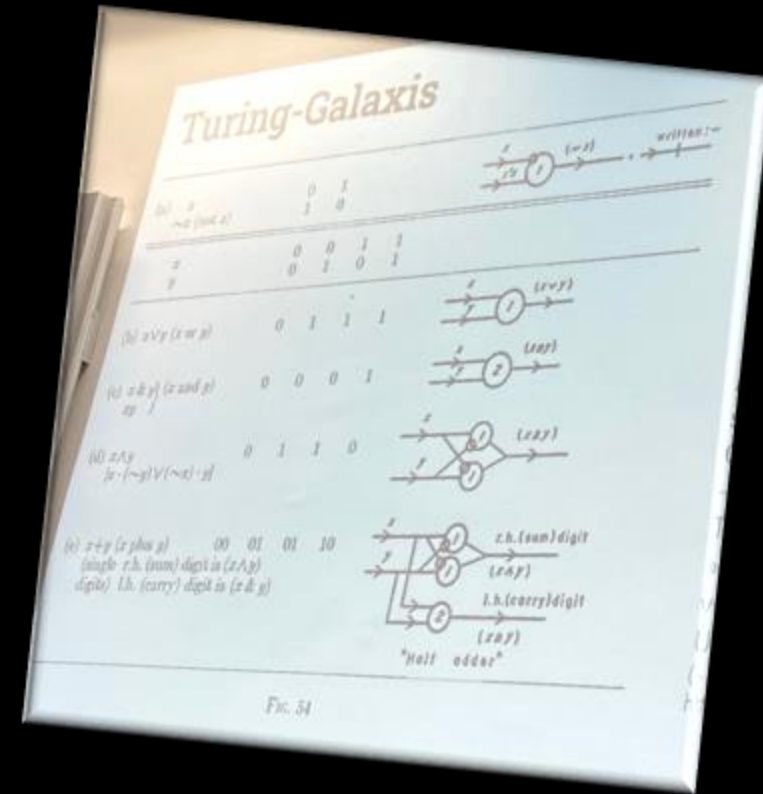
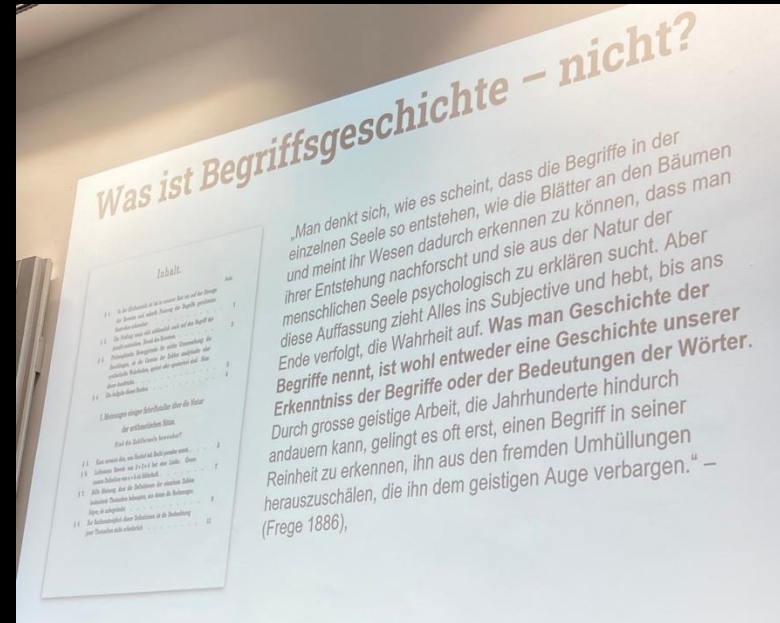


Fokusgruppe Wissens- und Begriffsgeschichte

Begriffsgeschichte in der Turing-Galaxis



Christian Schröter, geb. Vater (Mainz) & Florian Arnold (Stuttgart)



**Daten ohne Interpretation
bleiben 'stumm'!**
(Gunther Scholz, 2005)

NICHT ERFASSEN, SONDERN AUFFASSEN

Es geht nicht um Begriffe, es geht um Begreifen!
... nicht um Praxis in operativen Wissensmedien, nicht um den Ideenhimmel, der uns auf den Kopf fällt ...

Es geht darum, zu begreifen, wie wir schon begriffen werden durch medial gewordene operative Begriffe, die auch uns pro-programmieren.



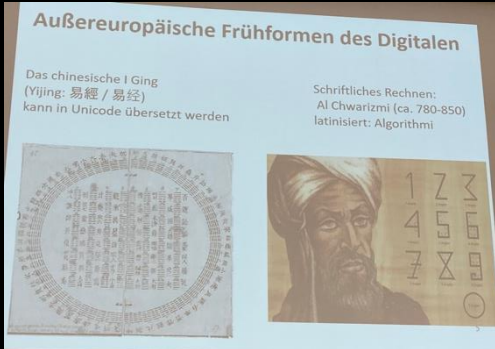
Was heißt ‚DIGITAL‘?

(1) Wie kann das Digitale computerunabhängig begriffen werden?

- Kontinuum wird diskretisiert in disjunkte Elemente (codierbar & arbiträr kombinierbar)
- Alphabete & Zahlenschriften bilden Keimformen

Operationalisierung symbolischer Aktivität durch

- algorithmisch geleitete Problemlösekapazität & temporär Absehung von Bedeutung
 - **ABLÖSUNG DES KNOWING HOW VOM KNOWING THAT**



(2) Welcher Zusammenhang besteht zwischen der Digitalität & der Kulturtechnik der Verflachung?

- inskribierte, illustrierte Oberflächen stiften artifiziellen Sonderraum von Überblick, Kontrolle, Manipulierbarkeit, Übertragbarkeit, Teilbarkeit
 - zwischen Eindimensionalität der Zeit & Dreidimensionalität des (Lebens-) Raumes ermöglicht das Medium zweidimensionaler Projektion die Übersetzung von Zeit in Raum und vice versa

Interpretationale Wiedereinführung der dritten Dimension

- Kompensation durch Interpretation
 - *Leibniz: Verzicht auf Sinn- und Bedeutungsverstehen = ‚cogitatio caeca vel symbolica‘*
 - *Whitehead: „Civilisation progresses by expanding the number of important operations that can be performed without thinking about them.“*

(3) Was ist neu an den zeitgenössischen Formen des Digitalen?

Mikrotemporalität, prädikative Algorithmen, Vernetzung, Datifizierung



5. Künstliche Intelligenz: Jenseits des Anthropomorphismus

- Die Übertragung menschlicher Attribute auf Software ist ein *Kategorienfehler*: Es geht um eine nicht-menschliche, *alternative* Form von Denkopoperationen, Zeichengebrauch und Kommunikation.
- Die Fähigkeit der Chatbot's plausible Texte zu erzeugen beruht nicht auf Sprachverstehen, sondern auf dem statistischen Operieren mit bedeutungslosen Token
- Die Dreierheit von Tokenstatistik, Kombinatorik und Korrelation charakterisiert die 'alternative Intelligenz' der Chatbots

FROM THEORETICAL PHILOSOPHY OF AI TO APPLIED ETHICS OF AI

Colin Allen (Santa Barbara), Bjørn Lundgren (Erlangen)

Chair: Vincent C. Müller (Erlangen)

- check out: Cameron Buckner, Melanie Mitchell, David Chalmers

SCIENCE IS ALWAYS SHIFTING GOAL POSTS

ChatGPT and other LLMs

- ❖ do not know what they are doing
- ❖ their behavior is not driven by any of the structures they describe
- ❖ lack metacognitive monitoring & control of their own errors and error correction

White paper: Value alignment in autonomous systems

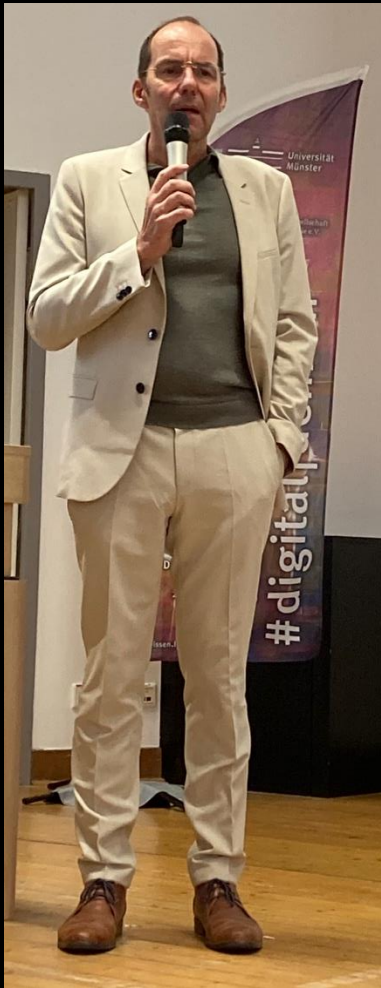
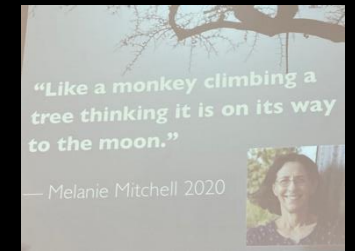
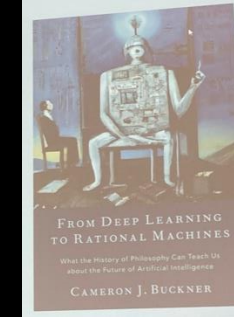
Stuart Russell, Professor of Computer Science, UC Berkeley

Summary

For an autonomous system to be helpful to humans and to pose no unwarranted risks, it needs to align its values with those of the humans in its environment in such a way that its actions contribute to the maximization of value for the humans. I propose to investigate a formal definition of this approach as *cooperative inverse reinforcement learning*, to develop algorithms for solving such problems, and to investigate their behavior in a variety of settings.

Value-alignment by Stuart Russell rather implausible

- we need some sort of human control
 - no utilitarianism
 - autonomy is no solution



PROFESSIONAL SKEPTICS

From Analogue to Digital

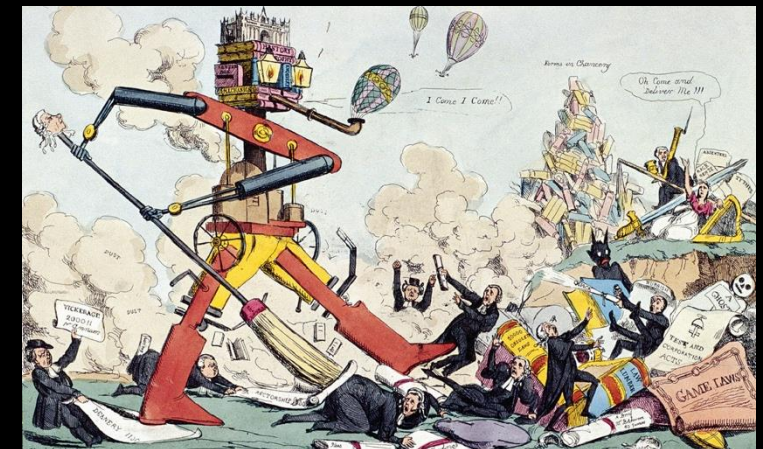
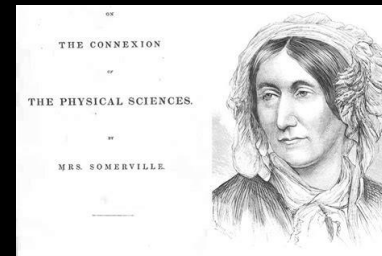
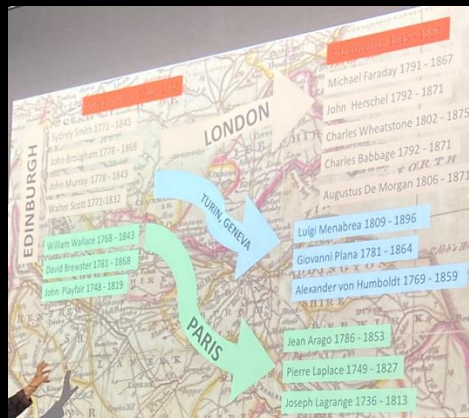
1. Cristina Sánchez Muñoz (Madrid):
POLITICAL VIOLENCE'S AND GENDER. EXCLUSIONS, REDUCTIONS, AND CHALLENGES



2. Ursula Martin (Oxford):
FEMALE SCIENTISTS AND SCHOLARS IN NINETEENTH CENTURY BRITAIN

'translations' which were an independent piece of science

the march of intellect



From Analogue to Digital



3. Sabine Thürmel (München):

CURRENT DEVELOPMENTS IN INFORMATICS FROM THE PERSPECTIVE OF FEMALE SCHOLARS IN THE HUMANITIES



4. Pedro Prikladnitzky (Paderborn):

PHILOSOPHICAL CONNECTIONS: EXPLORING THE DIGITIZATION OF BRAZILIAN WOMEN PHILOSOPHERS FROM PADERBORN TO BRAZIL

WORK OF PIONEERING BRAZILIAN WOMEN PHILOSOPHERS ADDRESSING KEY ISSUES OF GENDER, RACE, AND CLASS

Nísia Floresta (1810-1885) & Josefina Álvares de Azevedo (1851-1913),

- women's rights & education reform, drawing on European Enlightenment ideals, feminist thought, inclusion of women in public life & education, questioning traditional patriarchal structures.
- many works were not sold anymore but are digitalized now



Chatbots im Philosophieunterrichts? Selbst ausprobieren und kritisch diskutieren

Marte Henningsen / Melissa Schnabel /Rainer Mühlhoff (Osnabrück)

1. individualisierte Arbeitsmaterialien
2. simulierte Konversationen
3. Chatbot als Lerntutoren
4. automatisierte Korrektur und Bewertung



Fallstudie zu <https://fobizz.com/>

4 Stationen

1. Korrekturen
2. Bildgenerierung
3. KI-Chat Avatare
4. KI Lerntutorin

- Wie bewerten Sie die Qualität des Tools und seiner Antworten? Kann man das Ihrer Einschätzung nach unbeaufsichtigt im Schulalltag einsetzen?
- Glauben Sie, dass solche KI-Chatbots zu mehr Bildungsgerechtigkeit führen können? Geht das Tool auf die individuelle Situation der einzelnen Schüler:in ein oder sind die Antworten eher generisch?
- Was denken Sie, wie Schüler:innen auf den Einsatz von KI-Tutor-Systemen reagieren würden? Hätte es Nachteile, wenn Betreuungsarbeit nicht vom Menschen käme?

Strict Answerability and Artificial Intelligence

Maximilian Kiener (TU Hamburg)

A morally responsible (ie. answerable) agent is one who can intelligibly be asked to "answer for" her attitudes and conduct, where 'to answer for means to give her (justificatory) reasons for thinking, feeling, or acting in the way she has
(Smith, 2015, p. 103).

When we fail to conform to reasons that apply to us, those reasons persist in exerting a certain normative force
(Gardner, 2005, p. 57).

(1) WHAT IS STRICT MORAL ANSWERABILITY

Sufficient Condition: Causing the harm and having a prospective obligation to guard against such harm

Strictness: answerability is 'strict when there is no requirement of fault, display of ill-will, or any mens rea.

(2) WHY IT EXISTS

Maximilian Kiener: Normative Persistence: what happens when we cause harm which we had an obligation to guard against? (two changes: reasons → prospective obligations, / strong → a weak sense of non-conformity)

(3) HOW IT APPLIES TO AI

Sufficient Condition: a person is morally answerable for **some harm** if that person caused the harm and, ex ante, had an obligation to guard against such harm.

Causation: various people causally contribute to the harm

Obligation: various people have obligations to guard against harm.

→ A large number of people satisfy the sufficient condition of strict moral answerability

→ abundance, not a gap, of responsibility

(4) CHALLENGES

DILEMMA:

- i. great number of agents makes it unfeasible to hold everyone responsible
- ii. (in) just singling some of them out would be arbitrary or unjust.
→ we cannot do (i) and we should not do (ii)

NO 'RESPONSIBILITY GAPS' IN AI

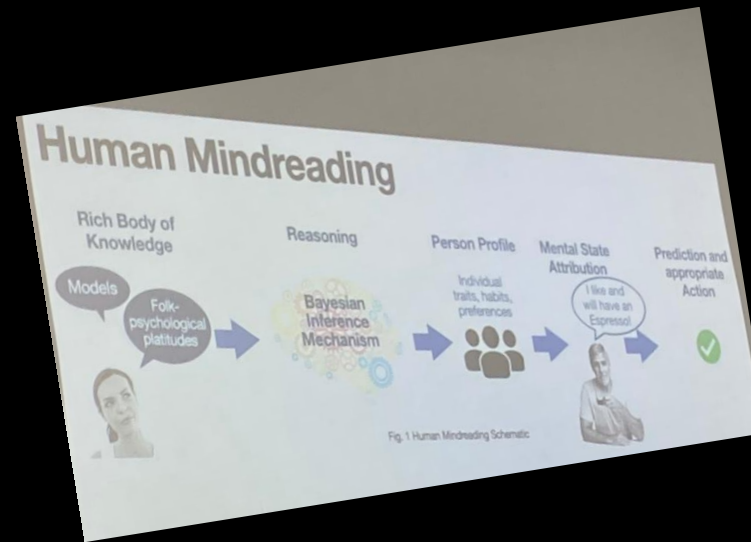
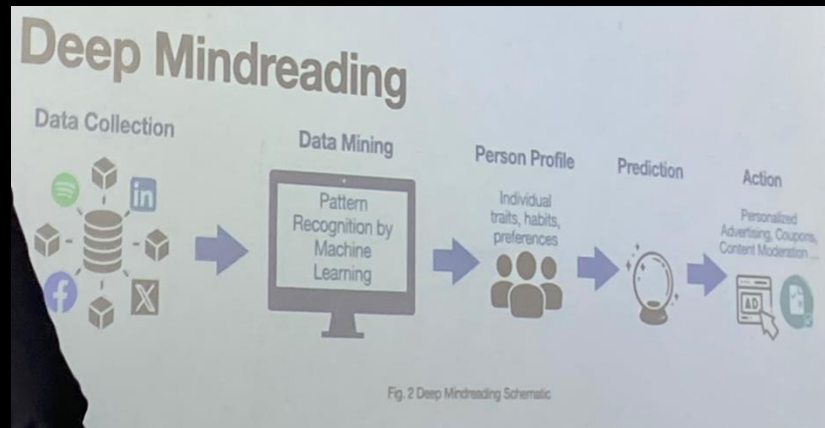
responsibility abundance: numerous agents are responsible for the harm caused by AI

Deep Mindreading

Tobias Schlicht (Bochum)

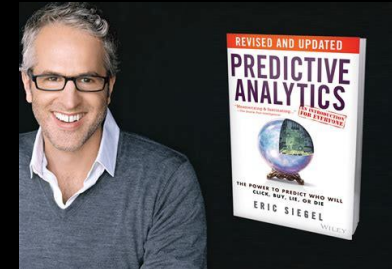
practice of online providers like Meta, Alphabet, Amazon etc. to infer personal data about users via machine learning and influence them

→ a form of mindreading, but different from the human variety.



predictive analysis

- predict what you like
- make you do things



expectation of privacy
you shouldn't pay for privacy

forthcoming paper with Carissa Véliz

The practices of collecting and analyzing personal data from users of online services to predict and manipulate their behaviour are genuine forms of mindreading and mindshaping.

Verstehen Chatbots Sprache?

KO-PERFORMANZ

MENSCH / MASCHINE DIVERS!

- ermöglicht Zusammenarbeit

KULTURTECHNIK DER VERFLACHUNG

- 2-dimensionale Projektionen zur Bearbeitung von 3-dimensionalen Phänomenen

BUCHSTABENSTATISIK

- ❖ Chatbots arbeiten token-statistisch
- ❖ Trainings-Daten besteht aus 'humanen' Material | kollektives Gedächtnis

INTERPRETATION DES VERFLACHTEN

- ❖ Verflachung macht Interpretation notwendig
- ❖ Mensch muss interpretieren (machen keine Token-Statistik) (induktives Erschließen)
- ❖ beim Schreiben produzieren wir ohne es zu wissen Muster

Chatbot produzieren Spuren

- für den Chatbot Muster
- für Menschen Bedeutung



'Webfehler' der digitalen Gesellschaft: Philosophische Perspektiven auf Ungerechtigkeiten, antisoziale Tendenzen und die Dringlichkeit, Netzwerke menschenzentriert zu denken.

Fokusgruppe: Digitale Ungerechtigkeit(en) & Fokusgruppe: Digitale Ethik



Anna Strasser



Regina Müller



Christoph Böhm



Oliver Zöllner

Fokusgruppe „Digitale Ungerechtigkeiten“

Nächstes Treffen: Freitag, 11. Oktober, 15-16 Uhr

Kontakt: regina.mueller@uni-bremen.de

Interessierte sind jederzeit willkommen!



Fokusgruppe: Digitale Ethik
Freitag, der 18.10. um 11:00

Regina Müller

Digitale Ungerechtigkeiten

Einblicke in die Arbeit der FK „Digitale Ungerechtigkeiten“

- Vertretungsprofessorin am Institut für Ethik und Geschichte der Medizin (Göttingen)
- wissenschaftliche Mitarbeiterin am Institut für Philosophie (Bremen)

Christoph Böhm

Sozialer Atomismus durch Digitalisierung

- freischaffender Philosoph (Promotion Universität Freiburg)
- Mitarbeiter eines Softwarekonzerns

Oliver Zöllner

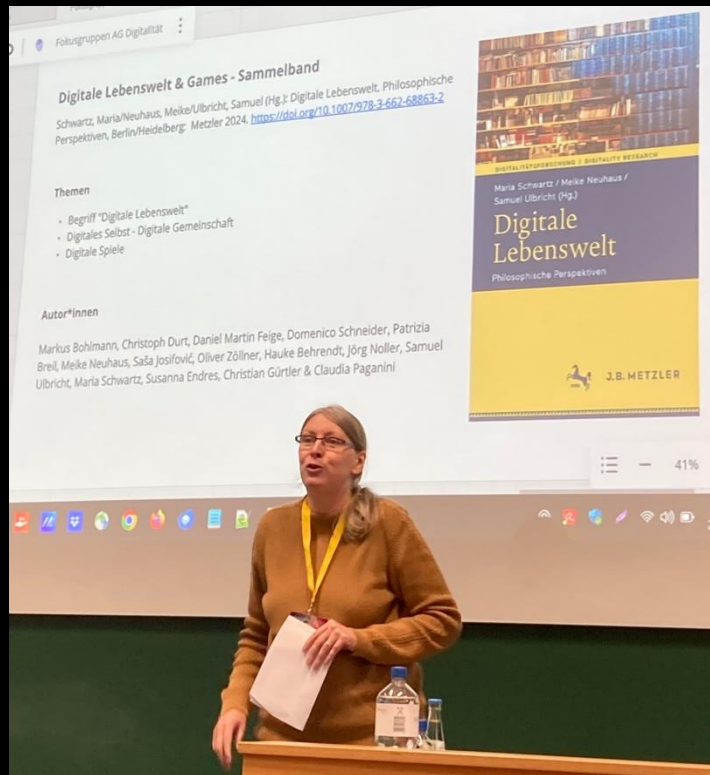
Brauchen wir ein „digitaletisches 1968“?

- Professor am Institut für Digitale Ethik (Stuttgart)
- Honorarprofessor an der Heinrich-Heine-Universität (Düsseldorf)

Theorien der Lebenswelt im Zeichen der Digitalität. Vorstellung, Input und Diskussion

Fokusgruppe Digitale Lebenswelt & Games

Themen: Veränderung der Lebenswelt durch Digitalität und Digitaltechnologie. Virtualität/Metaversum, das Internet als Informationssphäre und insbesondere digitale Spiele.



Digitale Lebenswelt

- ❖ Das Selbstverständliche und seine Irritation
 - Untersuchung der Lebenswelt fordert die „Zerstörung ihres essentiellen Attributs der Selbstverständlichkeit“
- ❖ Das Unsichtbare und seine Aufdeckung
 - Sichtbarmachung durch methodische Hinwendung
- ❖ Der Common Ground und seine Reproduktion
 - Was sind spezifische Phänomene der unsichtbaren, selbstverständlichen, allgemeinen digitalen Lebenswelt?



Didaktik und Digitalität

Workshop: OER und digitale Lernumgebungen in der Philosophie- und Ethikdidaktik

Diskussionspunkte

- Leichte Zugänglichkeit
- Flexibler Einsatz der Materialien
- Fragen der Rechte
- Interaktivität der Seiten
- Aktualität der Inhalte
- Langfristige Betreuung der Projekte (inklusive Hosting)
- Qualitätssicherung
- Technik-Support
- Hoher Aufwand, geringe wissenschaftliche Anerkennung



Programm – 20.10.2023 15:00-18:30 Uhr Zoom		
15:00-15:10	Annett Wienmeister Markus Bohlmann	Begrüßung und Einleitung
1. Block: Lernumgebungen und Materialsammlungen in der Philosophie- und Ethikdidaktik		
15:10-15:30	Anne Burkard	Philovestiz, Bausteine für den Philosophie- und Ethikunterricht
15:30-15:50	Sophia Peukert	Das Portal DigitalG macht Schule – Philosophieren mit Kindern über Themen der angewandten Ethik
2. Block: OER und digitale Lernumgebungen zu speziellen Themen der Philosophie und Ethik		
15:50-16:10	Annett Wienmeister	Alles Relativ? – Zum Umgang mit dem ethischen Relativismus von Schüler:innen (K2-Bach-Focus-Videoportal)
16:10-16:30	Markus Bohlmann	Doing Geo&Ethics als Plattform für Materialien zur Digitalisierung und Klimaethik
Kaffeepause		
16:50-17:10	Maria Schwartz	Digitale Ethik – Digitale Lebenswelt, Metaversum & VR – Computerspiele
17:10-17:30	Annett Wienmeister	Genome Editing am Menschen. Die Ethik-Lernplattform
3. Block: Datafizierung in und durch DER und digitale Lernumgebungen		
17:30-17:50	Frank Brosow Patrick Mosenhölzer	Digitale Assessment-Tools als blinder Fleck von DER
17:50-18:10	Jonathan Geiger	OER in der Philosophie aus Sicht des Forschungsdatenmanagements und digitaler Infrastrukturen
18:10-18:30	Schlussdiskussion	

Dieser Workshop ist Open Access!
Den Zoomlink erhalten Sie per Mail.
Bitte melden Sie sich an unter:
annett.wienmeister@charite.de
oder
markus.bohlmann@uni-muenster.de

Open Educational Resources und digitale Lernumgebungen in der Philosophie- und Ethikdidaktik

Workshop der Fokusgruppe Didaktik der DG-Phil AG Digitalisatorforschung und der GPED e.V.

WIE WOLLEN WIR LEBEN IN DER DIGITALEN GESELLSCHAFT?

AUTONOMIE & DEMOKRATIE SIND GEFÄHRDET

- Filter Bubble
- Überwachungskapitalismus
- Privatheit: Wem gehören die Daten?
- Manipulation (Mikrotargeting)

Die menschliche Natur: ein dichter Begriff

- Werden Menschen mehr wie Roboter?: Implantate (Parkinson)
- Postphänomenologen & Transhumanisten: unkritische Perspektive
- Werden Roboter mehr wie Menschen? → Care bots (noch nicht sehr überzeugend)
- Das Uncanny Valley



VERSCHIEDENE FORMEN VON EINFLUSSNAHME

- ❖ Überzeugen:
 - rationale Argumente (Erziehung)
- ❖ Verführen, verleiten:
 - traditionelle Werbung, (makrotargeting - transparent)
 - Nudging/Schubsen: transparent, keine versteckten Motive, (zumeist) im Interesse der wählenden Person
 - Manipulation, Betrug, Täuschung, Lügen: wie z.B. fake news
 - Zwang: Handlungsalternativen ausgeschlossen

Wie wollen wir leben?

WENN weiterhin: Autonomie (relational), Demokratie, und Menschen (wie bisher)
→ DANN brauchen wir:

- Transparenz von Beziehungen, Planbarkeit, Vorhersehbarkeit, Vertrauen in die Welt, in Sicherheiten und Nicht-Täuschung (Realität, Wahrheit, Aufrichtigkeit)

ABER: Diese Verhältnisse sind politisch und ökonomisch extrem ausbeutbar
→ Daten - in den falschen Händen? X; AfD, PVV in den NL, etc

- Philosophie als Kritik



Extracting Cognition To Digital Tools



How do digital tools influence the kind of cognition we have?

OTTO + NOTEBOOK



extended cognition:

- internal skills as externally enhanced

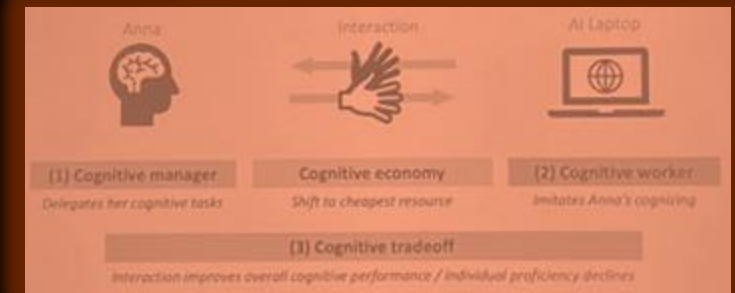
4.1 – The Theoretical Argument

Extended Cognition	Agent-tool relatedness	Extracted Cognition
A lot (required)	Decisional Power within a cognitive task	Not much (intended)
A lot (required)	Causal Control over the tool	Not much (not needed)
A lot (required)	Epistemic Access to the tool	Not much (opaque)

4.2 – The Practical Argument

Extended Cognition	Practices	Extracted Cognition
Otto takes notes of his ideas and meetings	Cognitive outsourcing	AI summarizes Anna's ideas in her style
Otto adds new street maps for navigation	Cognitive insourcing	AI downloads app to navigate Anna around
Engagement, augmentation, enhancement	Purpose & Impact	Delegation, substitution, retirement

ANNA + AI LAPTOP



- *extracted cognition*: extract cognitive skills from users that are then no longer required

4.3 – The Ethical Argument

Extended Cognition	Scenarios	Extracted Cognition
Tool loss Without his notebook, Otto would not be able to solve his daily cognitive tasks anymore.	Negative impact	Tool excess Anna gives all the work to her AI until she is no longer able to solve the tasks herself.
To not marginalize the loss, you better treat Otto an extended cognizer to protect his <i>external</i> cognition.	Cognitive blackmailing	To prevent her from excess, you better treat Anna an extracted cognizer so to protect her <i>inner</i> cognition.

EXTRACTED COGNITION:

Cognitive agents delegate decisional, causal, and epistemic control to external structures to decrease their own cognitive involvement.

Fellow-Cognisants?

A Kantian Perspective on Cognition in Large Language Models

Sub-Question: What is Cognition?

Kant:

- o [T]wo components belong to cognition: first, the concept, through which an object is thought at all (the category), and second, the intuition through which it is given. (B146; see also A50/B74; A92-3/B125)

	sensibility	understanding
FACULTY	receptive	spontaneous
necessary	Body & phenomenological experience	concepts
	Pure forms are space & time	Pure form is constituted by the 12 categories

Kantian Categories: pure concept – necessary

Sellarsian Categories: metalinguistic view – changeable

CONCLUSION

Having a sensibility with space and time as its pure forms is not much of a problem, since it is plausible to assume that if LLMs had a sensibility, its form would be space and time.

If we reject the strict Kantian theory of categories, which sees them as pure and necessary, and adopt a Sellarian, meta-linguistic view, we can ascribe categories, and hence understanding, to LIMs.



KÜNSTLICHE INTELLIGENZ UND DEMOKRATISCHE ÖFFENTLICHKEIT

Lisa Herzog (Groningen), Catrin Misselhorn (Göttingen), Judith Simon (Hamburg) | Podiumsdiskussion | Moderation: Jan-Philipp Kruse (Hamburg)

Unterschiede zwischen 90er Jahre & jetzt Internet:

- anfänglich große Hoffnung auf Austausch von Ideen und Argumenten
- jetzt viele verschiedene Diskursformen jedoch geleitet einer Aufmerksamkeit
- Extrempositionen werden begünstigt

WAS BRAUCHT DEMOKRATIE?

- Bildung & Information
- Partizipation & Bürgerbeteiligung
- Kritik & Kontrolle
- Austausch untereinander

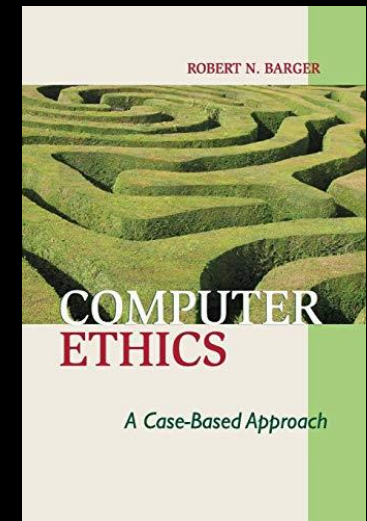
WISSEN führt zur Verantwortungsübernahme

GENERATIVE KI VERSUS INFORMATION RETRIEVAL

- ❖ Meinungsmacht der Tech Firmen
- ❖ antidemokratische Positionen gewinnen an Einfluss

Vertrauen – mühsam aufzubauen – schnell zu zerstören

1. Bias, Diskriminierung
2. Transparenz, Nachvollziehbarkeit
3. Datenschutz, Nachhaltigkeit
4. Täuschung
5. Überwachung etc.



POLITISCHE KI-ETHIK: EINE TRANSFORMATIVE PERSPEKTIVE

Die Intelligenz intelligenter Systeme

- pragmatisch-**performativer** Ansatz
- 2 Perspektiven:
 - Ähnlichkeit human-level KI vs. **Konsequenzen transformative KI**

Beispiele bei der Nutzung von KI

- Arbeitsplätze mittlerer Qualifikation werden reduziert werden
- Performanzunterschiede zwischen transparenten und opaken KI-Systemen??

Systematik der KI-Ethik
Floridi 5 principles: *beneficence, non-maleficence, autonomy, justice, explicability,*

Ausgewählte Probleme der KI-Ethik

- Opazität: Transparenz und Erklärbarkeit
- Bias: Gerechtigkeit und Fairness
- Singularität: Machtübernahme durch KI
 - **nicht Konkurrenz** sondern hilfreiche Assistenz
- Kontrolle über KI: Staat vs. Tech-Giganten vs. **Open Source**

Kann Überregulierung von KI eine Behinderung darstellen?

- für vertrauenswürdiger KI
- für unserer Wettbewerbsfähigkeit

Ausrichtungsproblem („Alignment Problem“)

- Annahme von hoch korrelierten Zielen (proxy)

Systematik der KI-Ethik

11 Cluster (Jobin/Zenca/Vayena)	5 Prinzipien (Floridi/Covits)				
	Respekt vor der Autonomie	Schadensvermeidung	Fürsorge	Gerechtigkeit	Erklärbarkeit
Fürsorge als solche			X		
Nachhaltigkeit		X	X		
Schadensvermeidung als solche		X			
Verantwortung			X	X	
Datenschutz	X	X			
Gerechtigkeit und Fairness					XX
Freiheit und Autonomie	XX				
Vertrauen					XX
Transparenz					XX
Würde des Menschen			X		X
Solidarität				X	X

