



universität
wien

Dr. Janina Loh
Universitätsassistentin (Post Doc)
Technik- und Medienphilosophie

Kontakt und Informationen:

janina.loh@univie.ac.at

<http://philtech.univie.ac.at/team/janina-loh-nee-sombetzki/>

VERTRAUEN IN TECHNIK

EINE PHILOSOPHISCHE PERSPEKTIVE

Transform 2020: «*in good data we trust*»

IPST | Institut Digital Enabling | BFH-Zentrum Digital Society

13. November 2020



AGENDA

1. Vertrauen und Vertrauenswürdigkeit
2. Hannah Arendts Perspektive auf Vertrauen
3. Dingen vertrauen
4. Risiken des Vertrauens
5. Fazit – ein paar offene Fragen

1. VERTRAUEN UND VERTRAUENSWÜRDIGKEIT

VERTRAUEN...

- ...ist ein Zwischenkonzept; lokalisiert zwischen Unwissenheit/Verzicht auf Informationen und vollständigem Wissen/Kontrolle/begründeten Erwartungen,
- ...erfordert Optimismus, dass die fragliche Person mit der Fähigkeit des Urteilsvermögens = Vertrauenswürdigkeit ausgestattet ist,
- ...ist etwas, das sich in erster Linie an Menschen (und nicht an Dinge) richtet.

DEFINITION: Vertrauen ist eine optimistische Haltung, die jemand gegenüber einer Person in einer Situation der Unsicherheit einnimmt, dass diese Person mit der Urteilskraft ausgestattet ist.

2. HANNAH ARENDTS PERSPEKTIVE AUF VERTRAUEN

HANDLUNGSTHEORIE IN *VITA ACTIVA* (1958)

- **ARBEIT:** sich wiederholende Prozesse ohne klaren Anfang und definitives Ende, die auf Selbsterhaltung und Reproduktion ausgerichtet sind; Ergebnisse sind konsumierbar.
- **HERSTELLEN:** erfordern eine Vision, die sie antreibt und die ihren Anfang und ihre Folgen definiert; Ergebnisse sind dauerhafte Objekte, die Teil der menschlichen Welt werden und diese konstituieren.
- **HANDELN:** echte menschliche Form des Tätigseins, die Worte und Taten einschließt und zwischen Menschen angesiedelt ist; Unvorhersehbarkeit und Irreversibilität; kein definitives Ende oder Ergebnis einer Handlung; kann nicht isoliert betrachtet werden, sondern Handlungen sind miteinander verbunden; erfordert Urteilsvermögen.

2. HANNAH ARENDTS PERSPEKTIVE AUF VERTRAUEN

VERTRAUEN...

- ...wird in Kontexten von Arbeit und Herstellen **nicht benötigt**,
 - ...ist in den Eichmann-Fällen **nicht möglich**,
 - ...ist **möglich und nötig** nur in Fällen wirklicher Handlungen; was bedeutet, dass wir aufgrund der Unvorhersehbarkeit von Handlungen Vertrauen **benötigen** und Vertrauen **möglich** ist, aufgrund der Urteilsfähigkeit der beteiligten Personen, der potentiell zu vertrauenden Person, zu vertrauen.
- **URTEILSVERMÖGEN** = notwendige und hinreichende Voraussetzung für Vertrauen (aufgrund von Arendts engem Handlungsbegriff)

3. DINGEN VERTRAUEN

- Wenn wir sagen, dass wir darauf vertrauen, dass unser Auto einwandfrei funktioniert, vertrauen wir in Wirklichkeit den Menschen, die das Auto gebaut haben (unter gewöhnlichen Umständen ist es jedoch absolut ausreichend, sich lediglich auf berechnete Erwartungen zu verlassen).
- Die Behauptung, dem eigenen Auto zu vertrauen, ist eine metaphorische Verwendung des Begriffs »Vertrauen«, insofern wir tatsächlich damit sagen, dass wir die berechnete Erwartung hegen, dass das Auto ordnungsgemäß funktionieren wird.
- Im Fall von intelligenten künstlichen Systemen: Je nach Definition und Grad der Beurteilung könnten einige künstliche Plattformen (in einem rudimentären Sinne) zumindest als potentiell vertrauenswürdig bezeichnet werden.

3. DINGEN VERTRAUEN

- Intelligente künstliche Systeme stellen unser traditionelles Verständnis von Vertrauen in Frage, das im herkömmlichen Sinne ausschließlich auf Menschen beschränkt ist. Das liegt an ihrem erhöhten Maß an autonomem Verhalten, das es uns möglicherweise erlaubt, ihnen eine Quasi-Form von Urteilskraft zuzuschreiben.
- Dies öffnet zwei Türen, durch die wir treten können:
 - (A) Wir können entweder kategorisch davon absehen, den Dingen (auch den klügsten) zu vertrauen.
 - (B) Oder wir können eine Beurteilungsschwelle definieren, die es uns erlaubt, zumindest einige sehr intelligente künstliche Systeme als potenziell vertrauenswürdig zu bezeichnen.

4. RISIKEN DES VERTRAUENS

- Die offensichtlichste und ernsthafteste Gefahr des Vertrauens liegt in der Gefahr des Verrats oder Betrugs.
- **OPTION A):** Wenn wir uns des Vertrauens in das System enthalten und stattdessen den Menschen vertrauen, die es entwickelt haben, ist die Gefahr eines betrügerischen Verhaltens unmittelbar mit den beteiligten Menschen verbunden. Da das System nicht mit Urteilskraft ausgestattet ist (obwohl es in der Tat sehr autonom sein könnte) ist es per Definition nicht in der Lage, diejenigen zu verraten, die sich auf es verlassen. Es könnte schwere interne Fehlfunktionen erleiden, die zu katastrophalen Ergebnissen führen können. Betrug bleibt eine von Natur aus externe Gefahr für das System, die die Integrität des Systems gefährdet, soweit es z.B. jemand für persönliche Zwecke kontrolliert

4. RISIKEN DES VERTRAUENS

- **OPTION B):** Wenn wir das System in einem eingeschränkten Sinne als urteilsfähig betrachten (Urteilsvermögen als Rationalität oder als minimale praktische Argumentation), ist die Möglichkeit betrügerischen Verhaltens nicht auf die Menschen beschränkt, die das System entwickelt haben. Die externen Gefahren (Systemintegrität, Datensicherheit) bleiben extern, soweit sie die Menschen betreffen. Zusätzlich werden aber auch die externen Faktoren intern, da das System selbst inzwischen zur vertrauenswürdigen Entität geworden ist.

5. FAZIT – EIN PAAR OFFENE FRAGEN

- Wie wir gesehen haben, stellt uns das Konzept des Vertrauens vor mehrere Herausforderungen: Wir wissen, dass die Urteilsfähigkeit bei der Entscheidung, ob jemand vertrauenswürdig ist, eine entscheidende Rolle spielt. Aber wie können wir Urteilskraft angemessen definieren?
- Es bleibt unklar, ob die Urteilsfähigkeit die notwendige und hinreichende Voraussetzung für Vertrauen ist (Arendt). Wenn ja, dann müssen die möglichen Kontexte, in denen wir von Vertrauen sprechen, begrenzt werden (Arendt: nur Fälle echten Handelns).
- Es bleibt unklar, ob die Urteilsfähigkeit nur eine notwendige Voraussetzung für Vertrauen ist. Wenn ja, welche anderen Voraussetzungen sind notwendig, um mögliche Kontexte, in denen wir von Vertrauen sprechen, zu definieren (evtl. abhängig von der Erfahrung, der potenziell zu vertrauenden Person, von dem Ausmaß der berechtigten Erwartungen,...).
- Solange diese Fragen unbeantwortet bleiben, könnte es keine gute Idee sein, den Kreis der potentiell zu vertrauenden Wesen auf intelligente künstliche Systeme auszudehnen, da die Kosten und Risiken des Vertrauens zu hoch sein könnten.

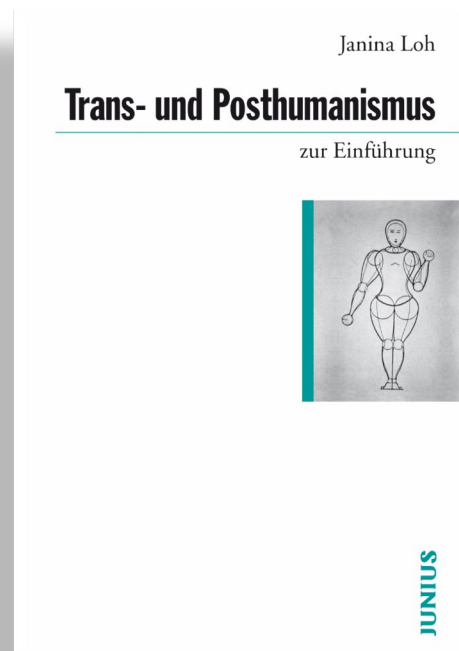


Dr. Janina Loh
Universitätsassistentin (Post Doc)
Technik- und Medienphilosophie

Kontakt und Informationen:
janina.loh@univie.ac.at
<http://philtech.univie.ac.at/team/janina-loh-nee-sombetzki/>



Trans- und Posthumanismus
zur Einführung
ISBN: 978-3-88506-808-2



**Verantwortung als Begriff,
Fähigkeit, Aufgabe**
Eine Drei-Ebenen-Analyse
ISBN: 978-3-658-04249-3



**Techno:Phil – Aktuelle
Herausforderungen der
Technikphilosophie**
Eine Buchreihe
ISSN: 2524-5902

Roboterethik
Eine Einführung
ISBN: 978-3-518-29877-0

