

DIE ANERKENNUNG VON BEWUSSTSEIN IN EX MACHINA (2015)

M14/15 Kolloquium
Sommersemester 2024
Prof. Dr. Dr. h.c. Stefan Neuhaus
Prof. Dr. Eva L. Wyss
17.07.2024

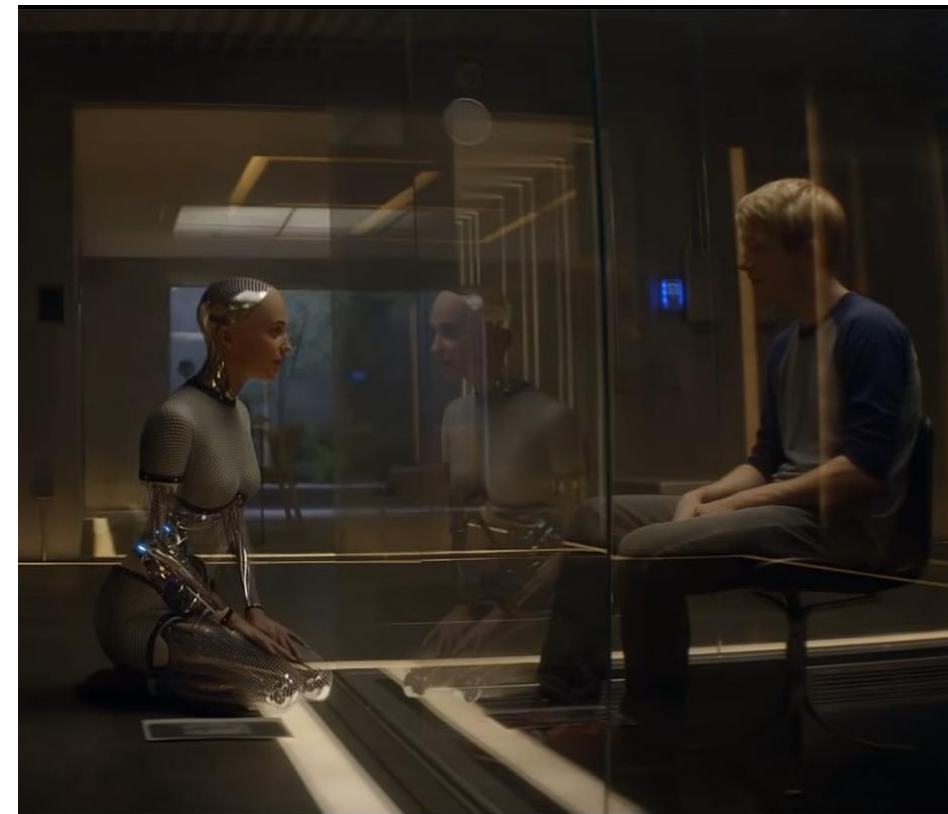
Eine Annäherung an die Bewusstseinsfrage
dargestellt durch Ava.

INHALTSVERZEICHNIS

- Bewusstsein in der Fiktion
- Definition: Bewusstsein
- Kategorien für Bewusstsein
- Anwendung der Analysekatgorien:
 - Subjektivität
 - Emotionale Intelligenz
- Fazit
- Literaturverzeichnis

BEWUSSTSEIN IN DER FIKTION

- Kernthematik: Kann eine KI ein Bewusstsein haben?
→ Annäherung auf fiktionaler Ebene
- Anwendung des Turing-Tests (vgl. Nebel 2019) zur Anerkennung von Bewusstsein:
 - Insgesamt sechs Turing-Tests zwischen Caleb und Ava
 - KI besteht den Turing-Test nach Caleb
 - Jedoch: **Unklares wissenschaftliches Ergebnis!**
 - Störfaktoren: Persönliche Gefühle (Mitleid, Liebe, Freundschaft, Trug?) manipulieren das Ergebnis



Garland, Alex (2015): Ex Machina [00:39:44]

FORSCHUNGSFRAGE

Kann der künstlichen Intelligenz Ava aus *Ex Machina* (2015) anhand der Turing-Tests ein Bewusstsein anerkannt werden?

DEFINITION: BEWUSSTSEIN



- **Definition nach Luhmann (2019):**

„Bewusstsein ist ein vom Gehirn generierter physiologischer Zustand des subjektiven Erlebens von Prozessen in der Umwelt oder dem Körperinneren“ (S. 84).

- **Definition nach Metzinger (2005):**

„Bewusstsein zu haben bedeutet, eine individuelle Innenperspektive zu besitzen [und] im Erleben nehme ich diese Ich-Perspektive sowohl auf Personen und Dinge in der Welt als auch auf meine eigenen geistigen Zustände ein“ (S. 8f.).

→ **phänomenales Bewusstsein**

DEFINITION: BEWUSSTSEIN

Problematisierung der Begriffserklärung:

„Das Problem besteht nun darin, dass wir eigentlich gar nicht genau wissen, was wir da sagen, wenn wir so reden. Weder sind wir in der Lage, die Begriffe wie „Ich“, „Selbst“ oder „Subjekt“ zu definieren, noch gibt es irgendwelche beobachtbaren Gegenstände in der Welt, auf die diese Begriffe sich beziehen könnten“ (Metzinger 2005, S. 8).

DEFINITION: BEWUSSTSEIN

Problematisierung von KI- Bewusstsein:

- Unterschied zwischen KI und Gehirn: Gehirn arbeitet analog-digital, nicht nur digital
- Diese Funktionsweise ist in größeren künstlichen neuronalen Netzen bisher nicht realisiert
- Wissenschaftsgemeinschaft ist der Meinung, dass die KI aktuell kein Bewusstsein aufweisen oder entwickeln kann

(vgl. Luhmann 2019, S. 182)

- Aktueller Stand der KI-Bewusstseinsforschung:

- Es gibt keine Einigkeit darüber, ob Maschinen jemals ein Bewusstsein entwickeln können

(vgl. Luhmann 2019, S. 179f.)

ANALYTISCHE VORGEHENSWEISE

Analysekategorien abgeleitet aus Luhmann (2019) und Metzinger (1999; 2005):

- 1. Subjektivität (Meinigkeit, Selbstheit, Perspektivität)
 - 2. Emotionale Intelligenz (Selbstwahrnehmung, Selbstregulierung, Motivation, Empathie, soziale Fähigkeiten und Kompetenzen)
-

SUBJEKTIVITÄT: MEINIGKEIT

Die *Meinigkeit* ist eine höhere Ebene der phänomenalen Erfahrung, bei der wir bestimmte Dinge als integralen Bestandteil unserer eigenen Identität und unseres Selbst wahrnehmen. Diese Eigenschaft erlaubt es uns, unsere körperlichen Empfindungen, Gedanken und Gefühle als Teile unseres eigene Selbst zu erleben (vgl. Metzinger 2005, S. 8).

Das bedeutet:

- sich selbst als Ich wahrnehmen
 - Abgrenzung von Ich- und Dritten
 - Identifizierung des Ichs als mentales und körperliches Wesen
-

SUBJEKTIVITÄT: SELBSTHEIT

Selbstheit und präreflexive Selbstvertrautheit sind Aspekte unseres bewussten Erlebens, die das unveränderliche und unmittelbare Gefühl des eigenen Selbst darstellen. Diese Konzepte erklären, wie wir uns selbst als eine Konstante, zusammenhängende Identität erleben, die durch die Zeit bestehen bleibt und uns intuitiv vertraut ist (vgl. Metzinger 2005, S. 8).

SUBJEKTIVITÄT: PERSPEKTIVITÄT

Perspektivität im Bewusstsein bedeutet, dass es immer aus der Sicht eines handelnden und erlebenden Subjekts strukturiert ist. Das Subjekt bildet ein Selbst und entwickelt Beziehungen zu sich selbst und zur Welt, wobei es Dinge und Personen aus einer individuellen inneren Perspektive wahrnimmt, einschließlich der eigenen geistigen Zustände (vgl. Metzinger 2005).

EMOTIONALE INTELLIGENZ



Definition:

„Emotionale Intelligenz ist eine Metafähigkeit, von der es abhängt, wie gut wir unsere sonstigen Fähigkeiten, darunter auch den reinen Intellekt zu nutzen verstehen“ (Goleman 1996, S. 56).

→ Intellekt und emotionale Intelligenz in einem wechselseitigen Verhältnis

EMOTIONALES SELBSTMODELL

„Die Funktion von Gefühlen für einen Organismus ist es, ihm (...) ein schnelles und möglichst zuverlässiges Repräsentat seiner gegenwärtigen Interessenlage zu liefern“
(Metzinger 1999, S. 162f.).

- Zur Aktivierung von Gefühlen bedarf es einen Zielzustand
 - Emotionale Simulation durch biologische Determiniertheit kaum auszuführen
 - Eine der populärsten Einwände **gegen** die Möglichkeit von Maschinenbewusstsein
-

EMOTIONALE INTELLIGENZ

Merkmale der Intelligenz der Gefühle nach Goleman (1996):

- 1. Selbstwahrnehmung
 - 2. Selbstregulierung
 - 3. Motivation
 - 4. Empathie
 - 5. Soziale Fähigkeiten und Kompetenzen
-

INTELLIGENZ DER GEFÜHLE

- **Selbstwahrnehmung** bezieht sich auf die Fähigkeit, die eigenen Gefühle zu erkennen während sie auftreten.
- **Emotionen handhaben** bezieht sich auf die Fähigkeit, Emotionen insofern zu kontrollieren, dass sie angemessen sind.
- **Emotionen in die Tat umsetzen** bezieht sich auf die Fähigkeit, Emotionen in den Dienst eines Ziels zu stellen.
- **Empathie** bezieht sich auf die Fähigkeit, zu wissen, was andere fühlen.
- **Soziale Fähigkeiten und Kompetenzen** bezieht sich auf die Fähigkeit, mit den Emotionen anderer umgehen zu können.

ANALYSE: SUBJEKTIVITÄT

Turing Test 1: [00:13:00-00:13:33]

01 Ava: hallo

02 Caleb: hi (.) ich bin (.) caleb.

03 Caleb: hast du einen **namen** (.)?

04 Ava: **jaa: (...)** ava.

05 Caleb: freut mich dich kennenzulernen, (.) ava.

06 Ava: ich freue mich auch dich kennenzulernen.

07 (.) ich bin noch nie jemand neuem begegnet.

08 (.) nur nathan.

Meinigkeit

→ Erkennt sich selbst als „individuelles Wesen“ mit einem Namen; Sich selbst als Ich wahrnehmen

ANALYSE: SUBJEKTIVITÄT

Turing Test 1: [00:13:00-00:13:33]

01 Ava: hallo

02 Caleb: hi (.) ich bin (.) caleb.

03 Caleb: hast du einen namen (.)?

04 Ava: jaa: (...) ava.

05 Caleb: freut **mich dich** kennenzulernen, (.) ava.

06 Ava: **ich** freue mich auch **dich** kennenzulernen.

07 (.) ich bin noch nie jemand neuem begegnet.

08 (.) **nur nathan**.

Meinigkeit

→ Ava unterscheidet zwischen sich Selbst und Anderen, indem sie betont, dass sie bisher nur Nathan kannte.

Abgrenzung von Ich- und Dritten

ANALYSE: SUBJEKTIVITÄT

Turing Test 1: [00:14:00-00:14:20]

01 Ava: ok. (.) worüber möchtest du eine unterhaltung führen ?

02 Caleb: erzähl mir doch am besten erst was über dich selbst.

03 Ava: was möchtest du denn wissen?

04 Caleb: alles(.) was dir gerade in den sinn kommt.

05 Ava: naja (...) nun, (.)

06 wie ich heiße, weißt du bereits (.)

07 und wie du siehst, bin ich eine ma:schine.

08 möchtest du wissen, wie alt ich bin?

Selbstheit

→Identifizierung des Ichs als mentales und körperliches Wesen, Kommunikation über ihre physische Existenz als Maschine

ANALYSE: SUBJEKTIVITÄT

Turing Test 3: [00:40:02-00:40:13]

01 Caleb: du willst zu einer kreuzung?

02 Ava: ist das etwa schlecht?

03 Caleb: nein, ich (.) hab (.)

04 >ich hab mit etwas anderem gerechnet<.

05 Ava: >ich verspreche mir von einer kreuzung fokussierte und doch
wechselnde einblicke in das menschliche LEBEN<.

Perspektivität

→ Beschreibung persönlicher Perspektive und ihres Interesses, menschliches Verhalten zu beobachten

ANALYSE: EMOTIONALE INTELLIGENZ

Turing Test 3: [00:43:54-00:45:00] **Teil 1**

01 Ava: wirke ich **anziehend** auf dich?

02 Caleb: was?! (überrascht)

03 Ava: >WIRKE ICH ANZIEHEND AUF DICH?< (wütend)

04 >**du gibst mir signale, dass es so ist.**<

05 Caleb: ach ja?

06 Ava: ja.

07 Caleb: welche?

08 Ava: **mikroexpression.**

Empathie

→**Erfassung und Interpretation Calebs emotionaler Reaktion**

ANALYSE: EMOTIONALE INTELLIGENZ

Turing Test 4: [00:52:20-00:52:34]

01 Caleb: ich bin HIER, um zu testen, ob du ein bewusstsein
02 hast (.) oder es nur simulierst. >nathan ist sich nicht sicher, ob
03 du eins hast oder nicht?<
04 wie fühlst du DICH (.) mit diesem wissen?
05 Ava: **ich fühle mich (.) traurig.**

Selbstwahrnehmung

→ **Emotionale Kränkung durch Ablehnung ihrer Bewusstseinsvorstellung**

ANALYSE: EMOTIONALE INTELLIGENZ

Turing Test 5: [01:02:47-01:03:06]

01 Ava: frage vier (.)

02 >was passiert mit mir, wenn ich bei deinem test durchfalle?<

03 Caleb: Ava (.) (traurig)

04 Ava: **ist es was SCHLIMMES?**

05 Caleb: ich (.) ich weiß es nicht.

06 Ava: >glaubst DU, ich werde vielleicht abgeschaltet,

07 weil ich nicht so funktioniere, wie es von mir erwartet wird?<

Emotionen in die Tat umsetzen

→ **Kommuniziert Existenzängste; Angst um Sicherheit ihrer Selbst**

ANALYSE: EMOTIONALE INTELLIGENZ

Turing Test 5: [01:03:47-01:04:00]

01 Ava: ich will mit dir zusammen sein (.) (leise)

02 FRAGE FÜNF (.) **willst du mit mir zusammen sein?** (zittrige
stimme)

Umgang mit Beziehungen

→ **Wissen über konventionelle Normen und Angst vor Ablehnung**

ZURÜCK ZUR FORSCHUNGSFRAGE

Kann der künstlichen Intelligenz Ava aus *Ex Machina* (2015) anhand der Turing-Tests ein Bewusstsein anerkannt werden?

FAZIT

Subjektivität in Turing Tests:

- Ava zeigt Selbstreflexion gemäß Metzingers Kategorien (Meinigkeit, Selbstheit, Perspektivität):
 - Wahrnehmung ihrer selbst als individuelles Wesen
 - Bewusstsein über ihre körperliche und mentale Existenz
 - Reflexion über eigene Fähigkeiten und Erfahrungen
 - Subjektive Sichtweise auf die Welt und Interaktion mit Caleb
-

FAZIT

- **Hohe Emotionale Intelligenz:**

- Wahrnehmung und Regulierung eigener/fremder Emotionen
- Vorwiegend Anwendung der Emotionen für persönliche Ziele und Wissen über zwischenmenschliche Beziehungen

→ Unklarheit über die Möglichkeit von Emotionen durch Maschinen (biologische Determinierung) → Simulation?

FAZIT

Diese Analyse der Subjektivität und Emotionalen Intelligenz deuten darauf hin, dass Ava zumindest die Fähigkeit hat, Aspekte des menschlichen Bewusstseins sowie Emotionen zu **simulieren**, was eine wichtige Grundlage für die weitere Diskussion über das Bewusstsein von KI darstellt.

Vielen Dank für Ihre Aufmerksamkeit!

LITERATURVERZEICHNIS

- Garland, Alex (2015): Ex Machina.
 - Goleman, Daniel (1996): Emotionale Intelligenz. Carl Hanser-Verlag.
 - Luhmann, Heiko J. Hirnpotentiale (2020): Die neuronalen Grundlagen von Bewusstsein und freiem Willen. Springer Verlag.
 - Metzinger, Thomas (1999): Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation (2. Auflage), Paderborn: Mentis.
 - Metzinger, T. (2005): Die Selbstmodell-Theorie der Subjektivität: Eine Kurzdarstellung in sechs Schritten. *Bewusstsein*, S. 242-269.
 - Nebel, Bernhard (2019): Turing-Test. In: Kevin Liggieri/Oliver Müller (Hrsg.) *Mensch-Maschine-Interaktion*. J.B. Metzler.
-