

Die Heuristik der Furcht gegenüber der KI

Handout zum Vortrag zu diesem Thema
am 26.01.2023 in Frankfurt am Main
im Seminar Philosophie der KI im WS 2022/23
im Fachgebiet Philosophie
Professur für Theoretische Philosophie

Hans Jonas (1903–1976)

trat dem blinden Fortschrittsglauben und dem unkritischen Einsatz von Technik zur Gestaltung einer vermeintlich besseren, utopischen, Gesellschaft warnend entgegen, wie sie u. a. Bloch¹ marxistisch vorschwebte², was bereits die Wahl des Titels seines Werkes *Prinzip Verantwortung* signalisierte, indem er dessen unsicheren utopischen *Prinzip Hoffnung* weit vor allen anderen die Forderung nach einer nachhaltigen Ethik gegenüberstellte.

Sorge um die Permanenz der menschlichen Existenz

Jonas fürchtete, dass die Menschheit ihre weitere nicht nur individuelle Existenz mit der von ihr geschaffenen globalen und bis in die Zukunft wirkenden Technik zerstören könnte.

„Wir wissen erst, was auf dem Spiele steht, wenn wir wissen, daß es auf dem Spiele steht. Da es dabei nicht nur um das Menschenlos, sondern auch um das Menschenbild geht, nicht nur um physisches Überleben, sondern auch um Unversehrtheit des Wesens,“³

Seine Sorge bezog sich explizit auf die Atomtechnik⁴ und die Gentechnik⁵, besonders hinsichtlich der Veränderung der menschlichen Keimbahn.⁶ Er richtete seine Furcht zwar nicht explizit auf die Computer. Aber er schloss auch nicht aus, dass sie sich auch auf diese ausweiten könnte:

„Wenn der Einsatz von entsprechenden Computersystemen allerdings dazu führen sollte, daß die große Bedeutung des einzelnen Individuums zugunsten einer möglichst reibungslos arbeitenden gesellschaftlichen Maschinerie untergraben würde, wäre das schlimm. Der Verlust des Respekts vor der Subjektivität wäre tatsächlich eine große Gefahr für die Menschheit.“⁷

Notwendigkeit einer neuen Zukunftsethik

Jonas entwickelte eine Ethik, die sich nicht mehr, wie bisher, allein an einem selbst oder den aktuell Lebenden im Umfeld orientiert, sondern die sicherstellt, dass auch „ferne“ und zukünftige Menschen gelingend leben können:

„Im Zeichen der Technologie aber hat es die Ethik mit Handlungen zu tun [...] die eine beispiellose kausale Reichweite in die Zukunft haben, begleitet von einem Vorwissen, das ebenfalls, wie immer unvollständig, über alles ehemalige weit hinausgeht. Dazu die schiere Größenordnung der Fernwirkungen und oft auch ihre Unumkehrbarkeit.“⁸

1 Vgl. Bloch, 1985

2 Jonas, 1979, 276 288

3 Jonas, 1979, 8

4 Vgl. Fränken

5 Vgl. Jonas, 1979, 12

6 Deutscher Ethikrat, 65f.

7 Jonas, 1979, 52

8 Jonas, 1979, 8

Fernstenliebe

Jonas hatte nicht nur, die Eigenliebe oder wie Teile des Christentums, die Nächstenliebe („Du sollst deinen Nächsten lieben wie dich selbst.“⁹) im Blick, sondern darüber hinaus die global oder zeitlich Fernen, als Fernstenliebe, die schon Nietzsche betrachtet hatte.¹⁰

Ökologischer Imperativ

Er erweiterte daher den *kategorischen Imperativ* von Kant ¹¹ zu einem, allerdings noch nicht von ihm so benannten, „ökologischen Imperativ“, um die „*Permanenz echten Lebens auf Erden*“¹² sicherzustellen. Während Kant alle menschlichen Handlungen, ausgehend von der Vernunft der Einzelnen, die dieser besitzt und einsetzen sollte¹³, die Neigungen dominierend zur Grundlage seiner Ethik nahm, hatte Jonas die Menschheit und die Erde als Ganzes im Blick und das über die räumliche und zeitliche Reichweite der aktuell Lebenden hinaus.

Seine neue, die Tradition und Gegenwart transzendierende Zukunftsethik, entstand weit vor dem heutigen Bemühen um Nachhaltigkeit¹⁴ und ist im Kern, wie die Ethik Kants, deontologisch, jedoch nicht allein an der dem Menschen vermeintlich gegebenen Vernunft ausgerichtet, sondern auch an dem weiteren sich dadurch ergebenden utilitaristischen Nutzen für die Menschheit¹⁵ die das „*größtmögliche Glück der größtmöglichen Zahl*“¹⁶ von Menschen als Ziel der Ethik vorgab:

„Kants kategorischer Imperativ sagte: »Handle so, daß du auch wollen kannst, daß deine Maxime allgemeines Gesetz werde.« Das hier angerufene »kann« ist das der Vernunft und ihrer Einstimmung mit sich selbst: [...] Es liegt aber kein Selbstwiderspruch in der Vorstellung, daß die Menschheit einmal aufhöre zu existieren, und somit auch kein Selbstwiderspruch in der Vorstellung, daß das Glück gegenwärtiger und nächstfolgender Generationen mit dem Unglück oder gar der Nichtexistenz späterer Generationen erkaufte wird. [...] »Handle so, daß die Wirkungen deiner Handlung verträglich sind mit der Permanenz echten menschlichen Lebens auf Erden«; oder negativ ausgedrückt: »Handle so, daß die Wirkungen deiner Handlung nicht zerstörerisch sind für die künftige Möglichkeit solchen Lebens«; oder einfach: »Gefährde nicht die Bedingungen für den indefiniten Fortbestand der Menschheit auf Erden«; oder, [...]daß wir aber nicht das Recht haben, das Nichtsein künftiger Generationen wegen des Seins der jetzigen zu wählen oder auch nur zu wagen.“

Heuristik der Furcht

Diese diente Hans Jonas als analytisches Werkzeug dazu, bei begrenztem Wissen und großem Zeitdruck zu einer möglichst „guten“ Lösung, Entscheidung und letztlich Handlung zu gelangen. Dabei spielen Erfahrungen, deren Bewertungen und auch erworbene irrationale Intuition eine große Rolle. Sie basiert also nicht primär auf Rationalität, Logik, Wahrscheinlichkeit und empirischen Wahrnehmungen. Als Werkzeug schlug er statt dessen vor:

„Heuristik der Furcht«: Erst die vorausgesehene Verzerrung des Menschen verhilft uns zu dem davor zu bewahrenden Begriff des Menschen. [...]Da es dabei nicht nur um das Menschenlos, sondern auch um das Menschenbild geht, nicht nur um physisches Überleben, sondern auch um Unversehrtheit des Wesens, so muß die Ethik, die beides zu hüten hat, über die der Klugheit hinaus eine solche der Ehrfurcht sein“¹⁷

9 Bibel, 22:37-39

10 Vgl. Nietzsche

11 Kant, 1781

12 Jonas, 1979. 36

13 Kant, Immanuel, Was ist Aufklärung?, 1784 in: Berlinischen Monatsschrift, 12: „Aufklärung ist der Ausgang des Menschen aus seiner selbstverschuldeten Unmündigkeit. Unmündigkeit ist das Unvermögen, sich seines Verstandes ohne Leitung eines anderen zu bedienen. Selbstverschuldet ist diese Unmündigkeit, wenn die Ursache derselben nicht am Mangel des Verstandes, sondern der Entschliefung und des Muthes liegt, sich seiner ohne Leitung eines anderen zu bedienen. Sapere aude! Habe Muth, dich deines eigenen Verstandes zu bedienen! ist also der Wahlspruch der Aufklärung.“

14 Vgl. von Carlowitz

15 Vgl. Bentham

16 ebd.

17 Jonas, 1979, 8

Diese rettete die Menschheit bereits

Am 26.09.1983 in einer gefährlich eskalierten Phase des Kalten Krieges (Kubakrise ab 14.10.1962, NATO-Doppelbeschluss am 12.12.1979¹⁸, Abschuss einer südkoreanische Boeing 747 mit 269 Insassen durch die UdSSR¹⁹ am 1.9.1983, Einmarsch der USA am 25. 10 1983 in Grenada und am 9. November 1983²⁰ das NATO Manöver "Able Archer "(fähiger Bogenschütze)²¹, bei dem so realistisch wie noch nie zuvor ein Atomkrieg simuliert wurde, meldete der zur Überwachung der USA eingesetzte sowjetische Riesencomputer der Sowjets dem leitenden Offizier in der Kommandozentrale der sowjetischen Satellitenüberwachung Stanislaw Jewgrafowitsch Petrow²², einen Angriff der USA mit nuklearen Interkontinentalraketen auf die UdSSR. Er hatte nur eine Viertelstunde Zeit, um zu entscheiden, ob er, wie vorgesehen, seinen Befehlen folgend, den sowjetischen Gegenangriff und damit den globalen Atomkrieg startete.²³ Er stufte diese Meldung zurecht und glücklicherweise als einen Fehlalarm ein und löste den globalen Atomkrieg nicht aus. Er folgte hier nicht der Logik und der damals noch „schwachen“ „Künstlichen Intelligenz“ seines Computers, sondern der eigenen analogen heuristischen Furcht.

Gefährlichkeit der Künstlichen Intelligenz

Jonas befasste sich auch mit der Computertechnik, die zu seiner Zeit aber noch nicht so weit entwickelt wie heute. Er kam zu der Einschätzung, dass Maschinen niemals ein Bewusstsein und damit auch keine dem Menschen gleiche künstliche Intelligenz haben können.²⁴ Dass sich deren Wirkung durch enorme hochfrequente, amivalente, globale Vernetzungen, wechselwirkende Verknüpfungen mit weiteren omnipotenten Techniken und Maschinen noch immens steigern würden, auch zu Lasten der Ökologie, konnte er damals noch nicht erkennen.

Ihr digitaler Geist wirkt bereits jetzt in alle Ritzen der Erde und des menschlichen Lebens hinein, schädigt die Natur und gefährdet die Permanenz der menschlichen Existenz, trägt dazu bei, dass die Menschheit mit rasender Geschwindigkeit den Nutzen dieser Technik transzendierend dem kollektiven Suizid entgegen eilt.²⁵ Diese Datenkrake wird bereits heute immer weniger, nur och teilweise vom Menschen kontrolliert und ist auf dem Weg, sich selbst von der Menschheit entkoppelt, autark auszubreiten.

Es gibt Denkmodelle, nach denen die KI den Menschen ablösen wird, es eine Welt mit Robotern und ohne Menschen geben könnte²⁶ und der Mensch dies mit der KI selbst herbeiführen wird, ggf. als Transhumanismus, und es die vom Menschen geschaffenen Maschinen und Roboter sind, welche die Menschheit existenziell ablösen.²⁷ Auf den subjektiven letztlich biologischen bedingten menschlichen Geist, der sich als dialektische Synthese aus einem totalen Geist und einem absoluten Geist heraus gebildet haben könnte, würde dann ein elektronischer Geist folgen, der nicht mehr an ein biologisches menschlich begrenztes Leben gebunden, sondern von diesem befreit autark und souverän bis zum Ende des Universums existieren würde.

Doch noch könnte die Menschheit mit der Heuristik der Furcht die KI streng beäugen und ihr notfalls rechtzeitig den Stecker ziehen. Die Zukunftsethik und die Heuristik der Furcht von Hans Jonas sind also aktueller denn je.

18 Vgl. Deutscher Bundestag

19 Vgl. Aviation Safety Network

20 Vgl. Tageschau

21 Vgl. Dba, Der Spiegel

22 Vgl. Bidder

23 Vgl. Leffers

24 Lossau, 49-52

25 Weller, 328

26 Vgl. Heß

27 Vgl. Tuck

1 Literaturverzeichnis

- Bentham, Jeremy. 1789/1789.** *Introduction to the Principles of Morals and Legislation.* 1789/1789.
- Bidder:, Benjamin. 2010.** Der Mann, der den Dritten Weltkrieg verhinderte. *einestages.* [Online] 21. 04 2010. zit. n. https://de.wikipedia.org/wiki/Stanislaw_Jewgrafowitsch_Petrow.
- Bloch, Ernst. 1985.** *Das Prinzip Hoffnung.* Frankfurt am Main, 1985. Bd. 3.
- . 1959. Muße als unerläßliches, erst halb erforschtes Ziel, S. 1039 f. *Das Prinzip Hoffnung.* s.l. : Suhrkamp, 1959, Bd. 3.
- . 1961. *Naturrecht und menschliche Würde,* S. 25-29. s.l. : Suhrkamp Vverlag, 1961.
- Carlowitz, Carl von. 1713.** *Sylvicultura oeconomica.* 1713.
- dba, der Spiegel. 2013.** So nah kam die Welt 1983 einem Atomkrieg. [Online] Der Spiegel, 3. 11 2013. [Zitat vom: 5. 1 2023.] <https://www.spiegel.de/politik/ausland/kalter-krieg-nato-manoever-fuehrte-1983-beinahe-zum-atomkrieg-a-931489.html>.
- Deutscher Bundestag. 1979.** Historische Debatten (9): Nato-Doppelbeschluss. [Online] 1979. [Zitat vom: 05. 01 2023.] <https://www.bundestag.de/dokumente/textarchiv/natodoppelbeschluss-200098>.
- Deutscher Ethikrat. 2019.** Eingriffe in die menschliche Keimbahn. [Online] 05 2019. [Zitat vom: 05. 01 2023.] <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-eingriffe-in-die-menschliche-keimbahn.pdf>.
- Fränken, Annika. 2022.** Hans Jonas und die Atomkraft. [Online] 2022. [Zitat vom: 05. 01 2023.] <https://www.philomag.de/artikel/hans-jonas-und-die-atomkraft,4.1.2023>.
- Heß, Melanie. 2017.** Tagesschau.de. *Künstliche Intelligenz – Löst der Roboter den Menschen ab?* [Online] 22. 11 2017. [Zitat vom: 05. 01 2023.] <https://www.tagesschau.de/jahresueckblick/meldung376038.html>.
- Jay Tuck. 2016.** Künstliche Intelligenz: Wird sie uns eines Tages töten? *Computerwoche.* 08. 09 2016.
- Jochen Leffers. 2017.** Der Mann, der die Welt rettete. <https://www.spiegel.de/geschichte/stanislaw-petrow-der-mann-der-die-welt-rettete-ist-tot-a-1168721.html>. [Online] 19. 09 2017. [Zitat vom: 05. 01 2023.] <https://www.spiegel.de/geschichte/stanislaw-petrow-der-mann-der-die-welt-rettete-ist-tot-a-1168721.html>.
- Jonas, Hans. 1979.** *Das Prinzip Verantwortung.* Frankfurt am Main : Suhrkamp Verlag, 1979.
- . 1993. *Dem Bösen Ende näher.* Frankfurt am Main : s.n., 1993.
- . 1986. *Technik, Medizin und Ethik.* Kassel : s.n., 1986.
- Kant, Immanuel. 1781.** *Critik der reinen Vernunft, AA IV, 430.* Königsberg : s.n., 1781.
- Loewenstein, Bedřich. 2009.** *Der Fortschrittsglaube - Geschichte einer europäischen Idee.* Osnabrück : V&R unipress , 2009.
- Lossau, Norbert. 1993.** Maschinen werden niemals ein Bewusstsein haben können. [Buchverf.] in: Hans Jonas. *Dem bösen Ende näher.* Frankfurt am Main : Suhrkamp, 1993, S. 49f.
- Network, Aviation Safety. 1978.** Thursday 20 April 1978. *Flight 902.* [Online] 20. 04 1978. [Zitat vom: 05. 01 2023.] <http://aviation-safety.net/database/record.php?id=19780420-1>.
- Nietzsche, Friedrich. 1893.** *Also sprach Zarathustra.* [Hrsg.] Privatdruck. Leipzig : s.n., 1893.
- Tagesschau.de. 2010.** <https://www.tagesschau.de/jahresueckblick/meldung376038.html>. *Invasion auf Grenada.* [Online] 15. 12 2010. [Zitat vom: 05. 01 2023.] <https://www.tagesschau.de/jahresueckblick/meldung376038.html>.
- Tuck, Jay. 2016.** *Künstliche Intelligenz: Wird sie uns eines Tages töten?* [Online] 08. 09 2016. [Zitat vom: 05. 01 2023.] <https://www.computerwoche.de/a/kuenstliche-intelligenz-wird-sie-uns-eines-tages-toeten,3260673,4>.
- Weller, Achim. Mutter Natur leidet.** Bad Homburg, 2022.