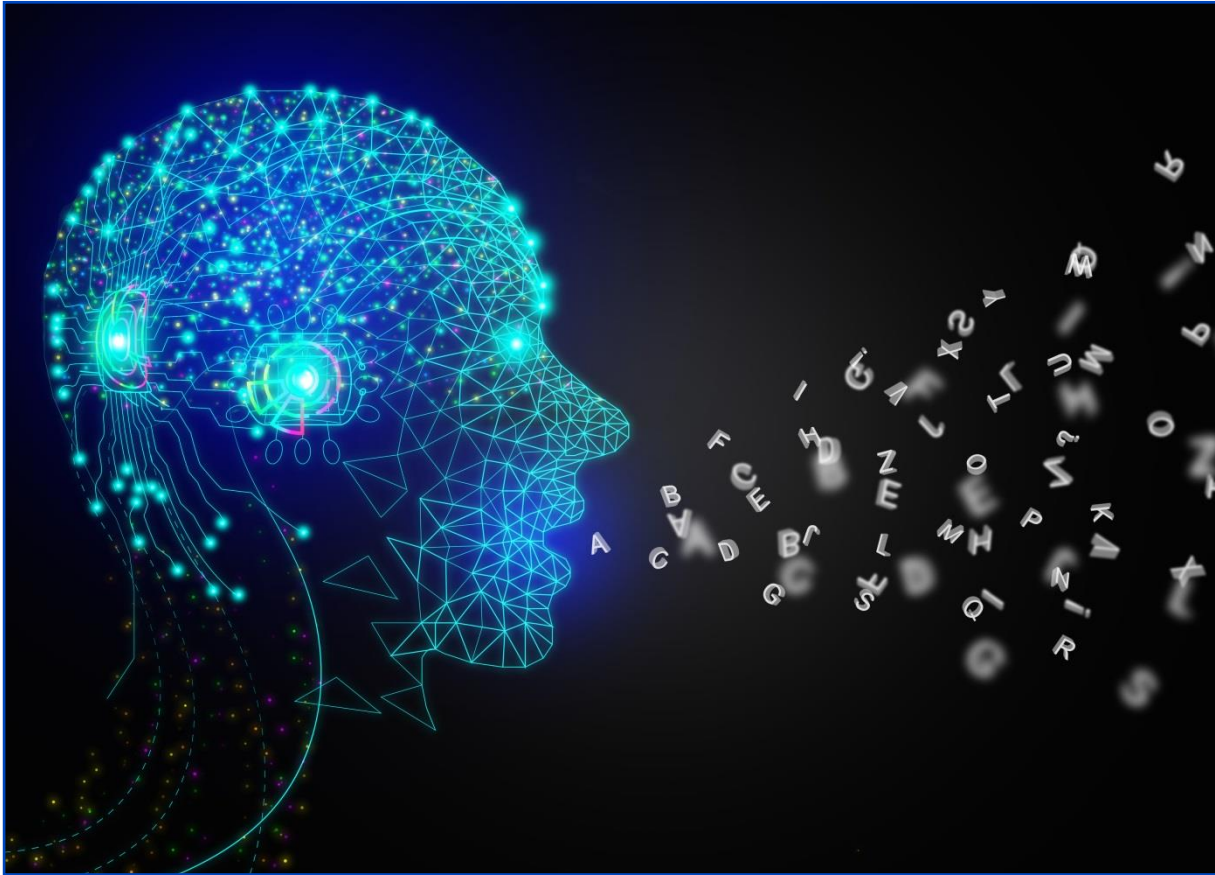


# GENERATIVE KÜNSTLICHE INTELLIGENZ



© iStock/Ole\_CNX

## ZUSAMMENFASSUNG

Seit ChatGPT im November 2022 veröffentlicht wurde, sorgen generative KI-Systeme für Aufsehen und alte Träume, Hoffnungen und Ängste in Bezug auf die Leistungsfähigkeit Künstlicher Intelligenz (KI) werden wieder angefacht. Generative KI verspricht enorme Arbeitserleichterungen und uneingeschränkten Zugang zu Informationen. Bei genauerer Betrachtung offenbaren sich jedoch viele noch ungelöste Schwierigkeiten, die einen Einsatz in den meisten professionellen Umgebungen nicht ratsam erscheinen lassen. Zusätzlich sehen Expert:innen ein großes Sicherheitsproblem auf die Gesellschaft zukommen, da generative KI-Systeme bspw. auch genutzt werden könnten, um tausendfach leicht variierenden Schadcode zu schreiben oder auf Millionen Bürger:innen individuell zugeschnittene Betrugsversuche zu starten.

*Generative KI –  
Hype, Gamechanger  
oder Gefahr für die  
Gesellschaft?*

## ÜBERBLICK ZUM THEMA

Die Möglichkeiten, Verständnis und Denken zu simulieren, haben sich seit der Aufklärung deutlich verbessert, als 1769 in Wien der „Schachtürke“, ein angeblich selbst denkender Schachautomat, in dessen Innerem ein menschlicher Spieler verborgen war, vorgestellt wurde. Als Vorgänger der aktuellen generativen KI-Systeme kann man Joseph Weizenbaums Programm ELIZA aus dem Jahr 1966 sehen, das über verschieden Skripte unterschiedliche Gesprächspartner darzustellen vermochte. Die Idee denkender, intelligenter Maschinen war immer mit hohen Erwartungen, Träumen und Utopien verbunden, aber auch mit bedrohlichen Szenarien und Ängsten. Es ging oft darum, den Menschen mächtiger zu machen, seine „Makel“ auszugleichen, bis hin zu seiner Rettung, wenn wir uns aktuelle Vorstellungen davon ansehen, wie KI-Systeme uns von den Folgen des Klimawandels befreien sollen.<sup>1</sup> Ist das jetzt die Generation Künstlicher Intelligenz, auf die seit ELIZA alle warten? Oder ist es, wie Bareis und Katzenbach (2021) schreiben, mehr „Talking AI into Being“, also mehr das Narrativ als die tatsächlichen Möglichkeiten Künstlicher Intelligenz, die den Hype befeuern?

*Den Wunsch nach denkenden, dienstbaren Maschinen gibt es schon lange*

Eng verbunden mit den Überlegungen zur Nützlichkeit denkender Systeme ist die Frage nach dem Vertrauen. Vertrauen wir Maschinen, wenn es um lebenswichtige Entscheidungen geht? Zugleich neigen Menschen dazu, in Technik menschliche Wesenszüge zu erkennen. Das wird nicht zuletzt in unserem Umgang mit digitalen Assistenten sichtbar (Schaber et al., 2019). Maschinen, besonders Rechenmaschinen und Computern, wird unterstellt, dass sie sich nicht irren könnten, somit ihre Ergebnisse zweifelsfrei richtig und – übertragen auf KI-Systeme – auch ihre Entscheidungen richtig seien.<sup>2</sup> Das liegt daran, dass Menschen zu wenig verstehen, wie diese Systeme funktionieren, und deshalb vielleicht vertrauen, wo kein Vertrauen angebracht ist. Dazu muss man sich auch vor Augen halten, dass diese Systeme von Menschen entworfen und gebaut werden, und KI mit bestimmten Daten trainiert wird, die auch nicht neutral sind. Wir wissen heute, dass Technik allgemein nicht neutral ist. Sie trägt die Werte und Wertvorstellungen ihrer Entwickler:innen in sich. Diese Erkenntnis ist ganz besonders dann relevant, wenn es um Automation und KI geht. Diese Systeme reproduzieren die (diskriminierenden) Strukturen, von denen sie (selbst wieder) lernen.

*Wem vertrauen wir?*

Es gibt auch warnende Stimmen, die davon ausgehen, dass eine lernende, sich selbst optimierende Maschine, in weiterer Folge mit eigenem Bewusstsein, die Entwicklung der Intelligenz und denkender Maschinen derart beschleunigen würde, dass der Mensch in so einer Welt keinen Platz mehr hätte. Maschinen würden Menschen in allen Belangen übertrumpfen und sukzessive ersetzen. Diesen Punkt, an dem die Maschinen in allem weitaus besser wären als der Mensch, bezeichnet man als „technische Singularität“. (Siehe auch die Bücher von Kurzweil 2005, oder Bostrom 2016) Und wäre diese Vorstellung nicht schon bedrohlich genug, gibt es natürlich auch die Phantasien, die diese Maschinen als feindlich,

*Läutet ChatGPT das Ende der Menschheit ein?*

<sup>1</sup> [www.derstandard.at/story/2000142852407/wie-kuenstliche-intelligenz-im-kampf-gegen-die-klimakrise-hilft](http://www.derstandard.at/story/2000142852407/wie-kuenstliche-intelligenz-im-kampf-gegen-die-klimakrise-hilft).

<sup>2</sup> Mehr Informationen zum sog. Automation Bias bspw. von Kate Goddard (2014).

dem Menschen böse gesinnt denken. Ausschließen kann man für die fernere Zukunft vielleicht nichts davon, aber wir sind derzeit sicher nicht einmal nahe dran.

Die KI-Systeme, die derzeit Menschen auf der ganzen Welt faszinieren, sind sog. generative KI-Systeme, die über Texteingaben weitere Texte erstellen oder andere audiovisuelle Inhalte produzieren. Sie simulieren mit Wahrscheinlichkeiten und extrem großen Mengen an Trainingsdaten Intelligenz und Verständnis.<sup>3</sup> Es werden statistische Methoden und (Trainings-)Daten genutzt, um Anfragen/Aufgaben zu bearbeiten und darauf Antworten zu entwickeln, die auf Grund statistischer Verteilungen und Wahrscheinlichkeiten Wortteile hintereinander reihen (oder Bildelemente, Programmcode zusammenfügen usw.). Dabei entstehen sinnvoll klingende Sätze, Bilder, Videos oder Teile eines Programmcodes, die sich auf die gestellte Aufgabe beziehen. Bei den Nutzer:innen entsteht dadurch der Eindruck, verstanden zu werden und eine inhaltlich passende Antwort erhalten zu haben. Das System hat jedoch zu keinem Zeitpunkt wirklich ein Verständnis entwickelt, nachgedacht oder darauf basierend eine Antwort formuliert. Wie Tests gezeigt haben, sind die Antworten auch oft nicht faktisch korrekt. Sie erscheinen mitunter frei erfunden, obwohl auch dieser kreative Akt eigentlich nicht stattfindet, sondern Produkt eines statistischen Zufalls ist.

Diese Funktionsweise lässt keine verlässliche Auskunft über z. B. wissenschaftliche Fragestellungen erwarten, die Stringenz und der mittelmäßige Wahrheitsgehalt der generierten Antworten reichen jedoch aus, um für die meisten Menschen täuschend echt erscheinende Fake News zu produzieren. Hierin liegt eine große Gefahr für die Gesellschaft, deren Aushandlungsprozesse und die Demokratie, da von der Maschine produzierte Fake News nicht als solche erkannt werden, und der Manipulation der öffentlichen Meinung damit Tür und Tor geöffnet sind.

Auch wenn Sprache/Text schon zuvor missbräuchlich, manipulativ und/oder in Täuschungsabsicht verwendet wurde, und *auch Bilder und Videos gefälscht werden*, sind beide nach wie vor zentrale Basis unserer Kommunikation. Insbesondere Bildmaterial ist ein gefühlter Beleg für den Wahrheitsgehalt einer Behauptung. Das liegt u. a. daran, dass gefälschte Bilder in der Vergangenheit einerseits selten waren, weil es aufwendig und nur mit dem nötigen Wissen möglich war sie herzustellen, und die Menschen andererseits gelernt hatten, textbasierte Nachrichten intuitiv auf Vertrauenswürdigkeit der Quelle und Plausibilität zu überprüfen. Das ist jetzt anders. Expert:innen meinen, die Gesellschaft steuere hier auf eine „Sicherheitskatastrophe nicht gekannten Ausmaßes“ zu.<sup>4</sup> Einerseits fehlt noch die Praxis im Umgang mit von Maschinen erstellten Inhalten, andererseits ist es auf Anwender:innenseite nahezu voraussetzungslos geworden, Falschnachrichten in hoher Qualität, Anzahl und/oder Individualität herzustellen.

Ebenso ist es problematisch, wenn derartige Systeme mit rein statistischen Methoden und Daten aus der Vergangenheit Prognosen oder Entscheidungen für die Zukunft oder über zukünftiges Verhalten erstellen.

*Die Funktionsweise*

...

*... und daraus resultierende Gefahren*

*Weitere Missbrauchsszenarien*

<sup>3</sup> Beispiele für derartige Systeme sind:

*Bard, ChatGPT* (Login erforderlich), *Dall-E2, Jasper, Kaiber, LaMDA* (Beschreibung) und *Midjourney*.

<sup>4</sup> [heise.de/hintergrund/Drei-Gruende-warum-KI-Chatbots-eine-Sicherheitskatastrophe-sind-8933941.html](https://heise.de/hintergrund/Drei-Gruende-warum-KI-Chatbots-eine-Sicherheitskatastrophe-sind-8933941.html).

Abgesehen von anderen Gefahren ist es auch demokratiepolitisch wichtig auf individueller Ebene zu wissen, ob man mit einem anderen Menschen oder eine Maschine kommuniziert – um deren Antworten einordnen zu können und zu wissen, wem man sich anvertraut. Auch davon, dass über die Entwicklung der Systeme wenig bekannt ist, nicht überprüfbar ist, welche Daten für das Training verwendet wurden (und welche nicht)<sup>5</sup>, wie die Gewichtung der bei Chat GPT-3 175 Milliarden Parameter (bei GPT-4 angeblich eine Billion)<sup>6</sup> vorgenommen wurde, und letztlich auch unklar ist, was mit den durch die Anwendung gesammelten Daten passiert, geht eine gewisse Gefahr aus. Es ist nur für wenige – wenn überhaupt – nachzuvollziehen, welchen Bias diese Systeme durch Programmierung und Training erhalten haben. Im Spannungsfeld geopolitischer Wirtschaftsinteressen und Machtbestrebungen können KI-Systeme, die als Blackboxes funktionieren (Udrea et al., 2022), durch bewusste Manipulation der Ergebnisse auch Einfallstore für die Durchsetzung fremder Interessen sein, womit deren Einsatz auch aus Gründen der *digitalen Souveränität* zu hinterfragen wäre.

Vor allem die Unkenntnis über den Umgang mit den gesammelten Daten von ChatGPT, die sich durch die verpflichtende Registrierung bestimmten Personen zuordnen lassen, hat dazu geführt, dass die Anwendung in Italien<sup>7</sup> und Deutschland von den Datenschutzbehörden näher analysiert wird.<sup>8</sup> In China wird ein Gesetz vorbereitet, das generativer KI vorschreibt, genaue und wahrheitsgetreue Antworten/Inhalte liefern zu müssen,<sup>9</sup> was natürlich (präventiv) nicht durchsetzbar sein wird, aber Unklarheiten in Bezug auf Haftungsfragen drängender erscheinen lässt. Auf EU-Ebene wird diskutiert, wie der Regulierungsentwurf zur KI-VO angepasst werden kann und ob die Vorgaben der VO über Digitale Dienste zielführend eingesetzt werden können (Helberger et al. 2023)

Weil die Angst besteht, dass Schüler:innen und Studierende Aufsätze von der Software schreiben lassen, diese als ihre ausgeben und die Lehrenden den Betrug nicht bemerken könnten, wurde der Einsatz der Software, unabhängig von der Durchsetzbarkeit solcher Regeln, an manchen Schulen und Universitäten verboten. An anderen Bildungseinrichtungen werden die Studierenden zum Gebrauch der Software ermuntert, um die kritische Auseinandersetzung mit den Ergebnissen zu fördern.

Positive Nutzungsmöglichkeiten für solche Systeme ergäben sich (v. a. bei verlässlicher und transparenter Funktionsweise) u. a. in den Bereichen Journalismus/ Recherche, Zusammenfassung von Texten, Kuratierung von Inhalten, in Werbung und Marketing, bei der Erstellung von Programmcode oder der *Diagnoseerstellung im Gesundheitswesen* (siehe auch Albrecht, 2023). Auch bspw. der Bereich generischer Bebilderung von Publikationen würde sich stark verändern, was neue Businesskonzepte für Bilddatenbanken erforderlich machen würde. Gene-

*Gefahren für  
liberale Demokratien*

*Regulierung*

*Generative KI  
und Bildung*

*Neue Chancen  
zur Vereinfachung  
bestimmter  
Arbeitsprozesse*

<sup>5</sup> [heise.de/news/Blick-in-die-Blackbox-KI-Trainingsdatensatz-C4-schoepft-auch-aus-trueben-Quellen-8973782.html](https://heise.de/news/Blick-in-die-Blackbox-KI-Trainingsdatensatz-C4-schoepft-auch-aus-trueben-Quellen-8973782.html).

<sup>6</sup> [de.ax-semantic.com/gpt-4-ueberblick-und-die-unterschiede-zu-gpt-3/](https://de.ax-semantic.com/gpt-4-ueberblick-und-die-unterschiede-zu-gpt-3/).

<sup>7</sup> [golem.de/news/italien-datenschutzbehoerde-untersagt-betrieb-von-chatgpt-2303-173125.html](https://golem.de/news/italien-datenschutzbehoerde-untersagt-betrieb-von-chatgpt-2303-173125.html).

<sup>8</sup> [faz.net/aktuell/feuilleton/medien/datenschuetzer-haben-fragen-an-open-ai-18839095.html](https://faz.net/aktuell/feuilleton/medien/datenschuetzer-haben-fragen-an-open-ai-18839095.html).

<sup>9</sup> [heise.de/news/China-plant-KI-Regulierung-KI-generierter-Content-soll-wahrheitsgetreu-sein-8970181.html](https://heise.de/news/China-plant-KI-Regulierung-KI-generierter-Content-soll-wahrheitsgetreu-sein-8970181.html).

rell scheint die Zusammenarbeit von Mensch und Maschine der vielversprechendste Weg zu sein. Die dafür nötigen Kompetenzen müssten jedoch besser gestern als heute in den Bildungseinrichtungen vermittelt werden.

In diese Richtung entwickeln sich derzeit auch Produkte und Dienstleistungen auf Basis generativer KI. Auch wenn hier (Bsp. Suchmaschinen<sup>10</sup>) noch viel zu tun ist, wird ausgelotet, wo KI den Menschen unterstützen kann, und wo sich damit Geld verdienen lässt.

Die ethischen Anforderungen an generative KI-Systeme sind weitgehend mit denen an andere KI-Systeme vergleichbar. Es sollte der Mensch im Mittelpunkt stehen, Entscheidungen müssten transparent, nachvollziehbar, begründbar und reproduzierbar sein.<sup>11</sup> Besonders bemängelt werden die Machtasymmetrien, die durch mangelnde Transparenz über Funktionsweise und Datenverwendung entstehen, sowie die Machtkonzentrationen durch große Datensammlungen und proprietäre Systeme von amerikanischen IT-Konzernen.

Ebenso problematisch erscheinen die Arbeitsbedingungen, unter denen Menschen in Niedriglohnländern die Systeme trainieren, so wie der nicht nachhaltige Betrieb und der hohe Ressourcenverbrauch der Systeme.<sup>12</sup>

*Ethische  
Anforderungen*

## RELEVANZ DES THEMAS FÜR DAS PARLAMENT UND FÜR ÖSTERREICH

Zunächst wird das Parlament in naher Zukunft damit befasst sein, die nationalen Belange dieses Themas zu regulieren, wie bspw. den Einsatz in der Verwaltung und Bildung. Zugleich muss aber auch das Vorgehen auf europäischer Ebene, zuvorderst beim Entstehen des AI Acts und etwaiger anderer Regulierungen<sup>13</sup>, abgestimmt werden.

Darüber hinaus erscheint es angebracht, eine regelmäßige Evaluierung des Umfangs und der Mittelverwendung im Bereich der nationalen Forschungsförderung für KI durchzuführen, um Fehlentwicklungen zu vermeiden und die Stärken der österreichischen KI-Entwicklung weiter zu fördern.<sup>14</sup> Dabei darf jedoch nicht außer Acht gelassen werden, dass bei allen Bemühungen, nationale Standortvorteile geltend zu machen, letztendlich nur ein abgestimmtes Vorgehen auf europäischer Ebene dazu führen kann, dass sich in Europa ein Gegengewicht zu den Entwicklungen in Nordamerika und China herausbilden und erhalten kann.

*Nationale  
Regulierungs-  
erfordernisse*

*Mögliche  
Schwerpunktsetzungen  
in der  
Forschungspolitik  
durch das Parlament*

<sup>10</sup> [www.heise.de/ratgeber/Suchmaschinen-mit-KI-im-Vergleich-8939449.html](http://www.heise.de/ratgeber/Suchmaschinen-mit-KI-im-Vergleich-8939449.html).

<sup>11</sup> Stellvertretend für viele Stellungnahmen zu dem Thema sei hier die erst kürzlich veröffentlichte Stellungnahme des deutschen Ethikrates angeführt: Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz, [ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf](https://ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf).

<sup>12</sup> Eine gute Zusammenfassung der Literatur dazu findet sich etwa im Themenpapier zu ChatGPT von TA-Swiss: [ta-swiss.ch/chatgpt](https://ta-swiss.ch/chatgpt); Zusammenfassung zum Thema Wasserverbrauch: [heise.de/news/Wasserbilanz-von-KI-Modellen-Halber-Liter-Wasser-pro-Unterhaltung-mit-ChatGPT-8973680.html](https://heise.de/news/Wasserbilanz-von-KI-Modellen-Halber-Liter-Wasser-pro-Unterhaltung-mit-ChatGPT-8973680.html).

<sup>13</sup> [heise.de/news/EU-Abgeordnete-ChatGPT-Co-sollen-besonders-streng-reguliert-werden-8975888.html](https://heise.de/news/EU-Abgeordnete-ChatGPT-Co-sollen-besonders-streng-reguliert-werden-8975888.html).

<sup>14</sup> [science.orf.at/stories/3218956/](https://science.orf.at/stories/3218956/) und [science.orf.at/stories/3218860/](https://science.orf.at/stories/3218860/).



Mindestens ebenso wichtig wie die Forschung an und die Entwicklung von KI-Systemen und deren Anwendungen ist die Forschung über KI und die damit verbundenen Veränderungen in Forschung und Gesellschaft. Hier wäre es dem Parlament möglich, durch entsprechende Schwerpunktsetzung bei Budgetierungen lenkend einzugreifen, und diesen sehr wichtigen Bereich zu unterstützen, der sehr zum Verständnis und zur Reflexion über die stattfindenden Veränderungen beiträgt.

## VORSCHLAG FÜR DAS WEITERE VORGEHEN

Es könnten Forschungsprojekte zu bestimmten, konkreten Aspekten angedacht werden, um die wissenschaftliche Grundlage für zukünftige politische Entscheidungen vorzubereiten: zur Schwerpunktsetzung bei der Forschungsförderung; zu den Folgen eines vermehrten Einsatzes von KI in der staatlichen Verwaltung; oder im Bildungsbereich.

Darüber hinaus wäre es wichtig, eine Bildungsoffensive zu Digitalisierung und KI zu erarbeiten, um der Bevölkerung für die Zukunft das nötige Wissen mitzugeben, um mit KI-Systemen zu arbeiten, diese produktiv zu nutzen, und deren Vor- und Nachteile besser beurteilen zu können.

*Neue  
Forschungsprojekte*

*Bildungsoffensive zu  
KI und Digitalisierung*

## ZITIERTE LITERATUR

- Albrecht, St. (2023). ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. TAB-Hintergrundpapier Nr. 26. [publikationen.bibliothek.kit.edu/1000158070/150614893](https://publikationen.bibliothek.kit.edu/1000158070/150614893).
- Bareis, J., & Katzenbach, C. (2022). Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics. *Science, Technology, & Human Values*, Vol. 47 (Issue 5), pages 855–881. [doi.org/10.1177/01622439211030007](https://doi.org/10.1177/01622439211030007).
- Bostrom, N. (2016). Superintelligenz. Szenarien einer kommenden Revolution. Suhrkamp, Frankfurt am Main.
- Goddard K, Roudsari A, Wyatt JC (2014). Automation bias: empirical results assessing influencing factors. *International Journal of Medical Informatics*. 2014 May; Vol 83 (Issue 5): pages 368-75. doi: 10.1016/j.ijmedinf.2014.01.001.
- Helberger, N., Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review* (2023) 12 (1). [doi.org/10.14763/2023.1.1682](https://doi.org/10.14763/2023.1.1682).
- Kurzweil, R. (2005). *The Singularity Is Near. When Humans Transcend Biology*. Viking, New York.
- Schaber, F., Krieger-Lamina, J., Peissl, W. (2019). *Digitale Assistenten*. ITA-Projektbericht 2019-01. [pub.oeaw.ac.at/ita/ita-projektberichte/2019-01.pdf](https://pub.oeaw.ac.at/ita/ita-projektberichte/2019-01.pdf).
- Udrea, T., Fuchs, D., Peissl, W. (2022). *Künstliche Intelligenz – Verstehbarkeit und Transparenz*. ITA-Projektbericht 2022-01, S.27ff. [pub.oeaw.ac.at/ita/ita-projektberichte/ITA-2022-01.pdf](https://pub.oeaw.ac.at/ita/ita-projektberichte/ITA-2022-01.pdf).