



Agora for Quality and Culture Interview Series
AQC.I No 05 – Januar 2026
ISSN 3042-6502

«Künstliche Intelligenz muss zukünftig wissen, wenn sie etwas nicht weiss – und uns dies dann auch mitteilen»

Anstösse zur dynamischen Entwicklung von Künstlicher Intelligenz

Interview mit Florian Lippke (ORCID: 0000-0002-9150-7346)

Florian Lippke ist überzeugt: Künstliche Intelligenz (KI) kann unter anderem als ein zukunftsfähiges «gesellschaftliches Projekt» verstanden werden. Dieses Projekt sollte aber nur dann als «intelligent» gelten, wenn es Vielfalt sichtbar macht, anstatt diese Vielfalt algorithmisch zu nivellieren. 20 thematische Schlaglichter stehen bei diesem Interview im Vordergrund: unter anderem «soziale KI», «KI als Spiegel der Gesellschaft», «Dilemma-Szenarien», «Erklärbarkeit von KI», «KI und Demokratie» oder «verfälschende KI». Aber auch «Datenschutz und Recht», «Inklusion» sowie «Ökologie und Nachhaltigkeit» werden je kurz bezüglich ihrer Überlappungen mit der neuen disruptiven Technologie beleuchtet. Insgesamt gilt: Nicht die Technik fehlt – sondern der Mut zur sinnvollen Begrenzung. Das Interview wirbt für eine orientierend-rahmensetzende und sinnstiftende Position zu Gunsten einer wertebasierten und menschenzentrierten Vision von KI.

Zitation: Lippke, F./Team AQC, «Künstliche Intelligenz muss zukünftig wissen, wenn sie etwas nicht weiss – und uns dies dann auch mitteilen». Anstösse zur dynamischen Entwicklung von Künstlicher Intelligenz, Agora for Quality and Culture Interview Series No 05, Freiburg i. Ue., Januar 2026, S. 1-14.

DOI: <https://doi.org/10.51363/unifr.lrr.2026.008>

© AQC Publishing | Office of the Agora for Quality and Culture

Mail: CH-3012 Bern | Neufeldstrasse 27c (c/o Swisspotential)

Uni: CH-1700 Fribourg | Universität Freiburg i.Ue. | DBS MIS 4219A



Agora for Quality and Culture
Politics • Society • Administration
Communication • Management



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG



Dieser Beitrag wurde publiziert unter der Creative Commons Attribution 4.0 International License (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0>



«Künstliche Intelligenz muss zukünftig wissen, wenn sie etwas nicht weiss – und uns dies dann auch mitteilen»

Anstösse zur dynamischen Entwicklung von Künstlicher Intelligenz

Interview mit Florian Lippke

Vorbemerkung: Für die Terminologie der hier verwendeten Spezialbegriffe (Künstliche Intelligenz [KI], KI-Systeme, KI-Sprachmodelle etc.) werden die Definitionen der Schweizerischen Bundesverwaltung ([Merkblätter](#), [Terminologie](#)) zugrunde gelegt.

- 5 **Abschnitt I: Orientierung und Grundverständnis**
 - 1. Vielfalt der Perspektiven
 - 2. Eine soziale Künstliche Intelligenz?
 - 3. Künstliche Intelligenz als Werkzeug
 - 4. Künstliche Intelligenz als Spiegel
- 10 **Abschnitt II: Entscheidung, Verantwortung, Autorität**
 - 5. Künstliche Intelligenz und Dilemma-Szenarien
 - 6. Dilemma oder Fortschritt?
 - 7. Künstliche Intelligenz und Erklärbarkeit
 - 8. Künstliche Intelligenz und Demokratie
 - 15 9. Künstliche Intelligenz als nutzbar gemachte disruptive Technologie
- Abschnitt III: Verzerrung, Macht, Recht**
 - 10. Künstliche Intelligenz in Translation
 - 11. Verfälschende Künstliche Intelligenz
 - 12. Künstliche Intelligenz, Datenschutz und Recht
- 20 **Abschnitt IV: Lernen, Teilhabe, Inklusion**
 - 13. *Cui bono*, Künstliche Intelligenz?
 - 14. Künstliche Intelligenz und Inklusion
 - 15. Künstliche Intelligenz für Innovation und Lernen
- Abschnitt V: Ausblick**
 - 25 16. Scheiternde und überraschende Künstliche Intelligenz
 - 17. Der Mensch im Zentrum
 - 18. Künstliche Intelligenz, Kultur und Spiritualität
 - 19. Künstliche Intelligenz, Natur und ökologische Aspekte
 - 20. Künstliche Intelligenz und Nachhaltigkeit



Abschnitt I: Orientierung und Grundverständnis

1. Vielfalt der Perspektiven

Team-AQC: *Florian Lippke, die «Vielfalt der Perspektiven» im Rahmen von «intelligenter» KI ist ein vieldiskutiertes Thema. Multiperspektivität ist nicht nur in den Geschichtswissenschaften oder in den Rechtswissenschaften («audiatur et altera pars») wichtig, sondern offensichtlich auch bei der Beschäftigung mit Künstlicher Intelligenz. Worum geht es im Kern dieser Debatte?*

FL: Immer wieder begegnet man in philosophischen Überlegungen der Position, dass eine KI nur dann wirklich «intelligent» sei, wenn sie die Vielfalt von Perspektiven ernst nimmt. Es geht genauerhin um die Vielfalt der menschlichen Perspektiven. Weil «menschlich» aber auch immer «vielgestaltig» bedeutet, darf die «Vielfalt des Menschlichen», wenn KI involviert ist, nicht durch Algorithmen nivelliert werden. Die Vielfalt muss vielmehr strukturiert sichtbar gemacht werden. Vergleichbar wäre hier beispielsweise das Konzept des «*arbre de connaissances*», das von Michel Authier und Pierre Lévy entwickelt wurde. Nach Authier bestünde der wahre Fortschritt darin, Maschinen zu schaffen, die Menschen besser miteinander verbinden – und nicht einer Trennung von Gruppen Vorschub zu leisten. Auch dies beinhaltet eine Vielfalt von Perspektiven, denn um Verbindung zu praktizieren, muss auch Vielfalt (von Perspektiven und Einstellungen) akzeptiert, toleriert und gelebt werden. Konkret auf die Entwicklung von KI bezogen bedeutet dies: Wir brauchen von Anfang an multiperspektivische Trainingssets. Gemeint ist der Aufbau kuratierter Datensätze mit bewusst widersprüchlichen Positionen (z. B. aus Ethik, Recht oder Medizin). Im Output müssen diese Perspektiven klar gekennzeichnet werden – nicht nur als «Antworten», sondern explizit als «Standpunkte». KI-Systeme sollten daher über einen dezidierten *Perspektiven-Modus* verfügen, in dem transparent gemacht wird: «Diese Antwort folgt Perspektive A. Alternative Sichtweisen wären B und C.»

2. Eine soziale Künstliche Intelligenz?

Team-AQC: *Heisst bessere Verbindung von Menschen auch sozialere Verbindung – also wäre ein Projekt wie eine «soziale KI» von Bedeutung? Müssen wir in Zukunft vermehrt an sozialen KIs arbeiten?*

FL: Ja, in einem solchen Fall würde man die ethischen, kulturellen und gesellschaftlichen Aspekte von KI in den Vordergrund stellen. Und da geht es sofort bei jedem dieser Begriffe um soziale Kategorien. In der Forschung wird häufig betont, dass KI nicht nur ein Projekt der Technologie bleiben darf – sie müsste ein Projekt der Gesellschaft werden. Dabei wäre es zentral ernstzunehmen, dass die Fragen, die an KI gestellt werden, genauso wichtig sind wie die Antworten, die uns im Prozess gegeben werden. Denn mit den (mehr oder weniger sozialen) Fragen, die gestellt werden, beginnt alles. Dass die Art der Fragen einen entscheidenden Einfluss auf das hat, was am



Ende dabei «herauskommt», kennen wir schon von den antiken Philosophen. Und zugleich kennen wir es aus der Beschäftigung mit KI, seitdem wir wissen, wie wichtig «Prompt Engineering» ist. Um nochmals kurz auf die Frage nach einer «sozialen KI» einzugehen: Man sollte für die zügige Weiterentwicklung einen *Social-Impact-Check* vor der Produktivsetzung einer KI ansetzen. Das heisst, dass jede KI-Anwendung eine verpflichtende soziale Wirkungsanalyse, vergleichbar einer Umweltverträglichkeitsprüfung, durchlaufen müsste.

3. Künstliche Intelligenz als Werkzeug

Team-AQC: Man könnte also sagen, KI «vermittelt» zwischen Frage und Antwort – und ist damit auch für Bildungsprozesse interessant. Sollte man darum das Bild von «KI als (Bildungs-) Werkzeug» ernst nehmen? «Werkzeug/Tool» ist heutzutage sicherlich eine der am häufigsten verwendeten Metaphern für KI, nicht wahr?

FL: In der Tat. Die breite Masse, aber auch etliche Experten, sehen in der KI ein Werkzeug, das zum Beispiel individuelle Lernprozesse unterstützen kann (didaktisch und psychologisch). Entsprechend bedeutsam ist für diesen Ansatz der Fokus auf eine ethische Pädagogik, die Technik als dienenden Faktor begreifen möchte. KI muss als Bildungswerkzeug vor allem «verstehen», dass Lernen nicht linear ist, sondern vielmehr persönlich, lebendig und manchmal chaotisch. Wenn es darum geht, dass Maschinen/Werkzeuge uns unterstützen sollen, müssten sie zuerst einmal lernen, richtig und genau zuzuhören. Einige Überlegungen hierzu wurden kürzlich von der *Qualife-Stiftung* angestellt. KI kann als (Reflexions-)Werkzeug im Bildungsbereich tatsächlich konkret eingesetzt werden. Sie dient in diesem Fall nicht primär der Antwortlieferung, sondern der Gegenfrage, etwa: «Welche Annahmen stecken in deiner Fragestellung»? Auf diese Weise setzt Lernen auf einem reflektierten Niveau ein.

4. Künstliche Intelligenz als Spiegel

Team-AQC: KI kann und muss ganz anders verstanden werden als eine klassische Suchmaschine. Welche weiteren Bilder wären mehr oder weniger naheliegend, um KI in ihrer Besonderheit verständlich zu machen?

FL: Verfolgt man einen ganzheitlichen und gesellschaftsorientierten Ansatz, der demokratische Werte und den Schutz menschlicher Würde in der digitalen Transformation zugrunde legt, erkennt man schnell, dass das Bild der «Maschine» nicht besonders tragfähig ist. Wir müssten aufhören, KI als Maschine zu sehen – vielmehr kann sie als ein Spiegel unserer Werte und unserer schriftlichen Weltkonstruktion verstanden werden. Zu diesen Aspekten hat unter anderem *impactIA* vermehrt gearbeitet. KI muss, wenn ein solcher Ansatz grundgelegt wird, vor allem auch ein Pendant zu «menschlicher Reife» als Fähigkeit beinhalten, um wirklich intelligent zu



sein. Ansonsten wird aus einem Spiegelbild schnell ein Zerrbild. Als praktische Lösung für die Minimierung von Zerrbildern würden sich direkte Werte-Feedback-Schleifen anbieten. Im thematischen Inhalt erfahrene Nutzer:innen könnten Rückmeldung geben, welche Werte sie in KI-Antworten wiederfinden – und welche sie vermissen.

5

Abschnitt II: Entscheidung, Verantwortung, Autorität

5. Künstliche Intelligenz und Dilemma-Szenarien

Team-AQC: Aber kann KI, die vielfältige Perspektiven berücksichtigt und auf soziale Dimensionen befragt wird, nicht auch schnell in Dilemmata landen?

FL: Ja, sicherlich. Dies wäre eine spannende Frage, für deren Beantwortung auch schon wissenschaftlich einige Vorarbeit geleistet wurde. Hierbei sticht die Bedeutung von Szenarien ins Auge, die nicht einfach zu lösen sind, sondern eine Spannung zu unterschiedlichen Normen, Positionen etc. aufweisen. Forschungsinteresse besteht darum gerade in der Verwendung solcher «spannungsfähiger» bzw. «spannungskompetenter» Systeme in der realen Welt für mehrdimensional herausfordernde Situationen. KI müsste also lernen, mit realen Spannungen umzugehen – nicht nur und ausschliesslich mit Daten. Ganz konkret könnten Dilemma-Simulatoren entwickelt werden: KI-Systeme, die bewusst keine Entscheidungen treffen, sondern Entscheidungskonflikte sichtbar machen. Solche Systeme wären besonders wirksam in der Politikberatung, der Medizinethik oder der Justiz. Zu betonen ist die Bedeutung der Wissensabschätzung: Die beste KI ist die, welche weiss, dass sie nicht alles weiss. Also eine platonisch-sokratische KI, die weiss, dass es auch Dinge gibt, die sie (noch) nicht weiss. Aus dieser Einschätzung ergibt sich naheliegender die praktische Forderung nach einem «Nicht-Wissen-Marker» als technischer Pflicht. KI muss Unsicherheiten quantifizieren und explizit anzeigen – analog zu der Frage, die bei Selbst-

25 einschätzungen häufig gestellt wird: «Wie sicher ist meine Antwort?»

6. Dilemma oder Fortschritt?

Team-AQC: Aber bleibt dann KI nicht einfach in Dilemmata hängen? Wäre nicht eher der Fortschritt und die weiterführende Antwort der springende Punkt?

30 FL: Wahrscheinlich würden KI-Analysten, die KI als Entwicklungsmotor verstehen, zustimmen. In dieser Denkrichtung wird KI allerdings eher als «Katalysator» verstanden. Und katalysiert oder auch erleichtert werden soll die Zusammenarbeit in hybriden Entscheidungssystemen, bei denen menschliche Fähigkeiten und maschinelle Analyse zusammenwirken. Dabei kann man KI als einen Katalysator für den kulturellen Wandel in realen Strukturen (Firmen, Institutionen etc.) verstehen – nicht einfach als Ersatz für menschliche Rollen. Es geht also nicht um Eliminie-

35



5 rung des Menschen aus dem System, sondern im wahren Sinne um eine Form der «(digitalen) Transformation». Vor allem die Frage nach der «algorithmischen Autorität» ist hier vertieft zu betrachten. Der Einsatz von KI kann eine Organisation durchaus intelligenter machen – aber nur, wenn die Menschen darin auch wachsen und sich entwickeln können. Mit anderen Worten:
10 Es müssten die Grundlagen für eine KI-digitale Transformation geschaffen werden. Vor allem aber sollten wir aufhören, die KI zu fragen, was wir tun sollen. Stattdessen sollten wir beginnen, die KI-Systeme zu fragen, was wir übersehen. Beim Aufzeigen blinder Flecken ist KI nämlich inzwischen erstaunlich ertragreich und nicht zu unterschätzen (medizinische Diagnosen, Geoanalysen). Konkret wäre an eine standardisierte Blinde-Flecken-Analyse zu denken: KI würde gezielt eingesetzt werden, um übersehene Alternativen und systemische Verzerrungen aufzuzeigen. Solche Funktionen lassen sich trainieren und in bestehende Lösungen integrieren.

7. Künstliche Intelligenz und Erklärbarkeit

15 *Team-AQC: KI hat für viele etwas Zauberhaftes – nicht selten, weil sie als Blackbox wahrgenommen wird. Damit wären wir aber bei einem besonders spannenden Thema im Rahmen der künstlichen Intelligenz: «KI und Erklärbarkeit», also die Frage nach der Nachvollziehbarkeit der Arbeitsweise einer KI und wie gut sie sich selbst erklärt und den Nutzenden gegenüber transparent bleibt.*

20 FL: In der Tat ein sehr spannendes Thema. An der Universität von Genf (HEPIA) und an anderen Orten wird die Forschung zu erklärbarem Output («Explainable AI») stärker in den Blick genommen. Diese Erklärbarkeit gilt als wesentlicher Meilenstein für die Zukunft von KI. Ein momentaner Fokus liegt hierbei auf Anwendungen, die in sozialen und gesundheitlichen Kontexten eingesetzt werden. Denn immer häufiger wird KI in Verantwortungsfragen beigezogen. Es geht um Antworten, die auch die Verantwortlichkeit des Systems und der Nutzenden beinhalten. Wer in einem solchen Fall aber versteht und orientiert ist, wie ein System zu seinen Outputs
25 kommt, ist entscheidend im Vorteil. Ein solcher Wissensfortschritt ermöglicht korrigierende Eingriffe oder auch die begründete Ablehnung von Ergebnissen. Dies ist natürlich nur möglich, wenn die menschliche Komponente in KI-Zyklen immer noch «im Loop» vorhanden ist (HITL: Human-in-the-loop/HOTL: Human-on-the-loop). Ist dies nicht mehr gegeben, wie bei den HOOTL-Szenarien (Human-out-of-the-loop), können sich potenziell gefährliche Situationen ein-
30 stellen. Für den westlichen Kontext liesse sich zudem ein Governance-Standard zugunsten von Human-in-the-loop definieren. In anderen Weltregionen – etwa in Teilen Asiens – würde dies anders beurteilt. Fraglich bleibt, ob Europa und «der Westen» sich im Umkehrschluss auf ein Verbot von HOOTL-Systemen (z. B. in Demokratie, Justiz oder Gesundheitswesen) einigen könnten. Stattdessen könnte auch ein «Recht auf menschliche Entscheidung» gesetzlich
35 verankert werden, das bei allen folgenrelevanten KI-Entscheidungen greift – also eine Art Hand-



bremse in der KI-Praxis. Dies würde meines Erachtens auch den Trends zur Erklärbarkeit in die Hand spielen.

Mitunter begegnet sogar folgende zugespitzte Position: Intelligenz ohne Erklärbarkeit ist keine Intelligenz, sondern lediglich Statistik. KI darf nicht einfach nur funktionieren – sie muss ver-
5 ständiglich, zugänglich und hinterfragbar sein. Aus diesen Überlegungen lässt sich eine klare Notwendigkeit ableiten: *Explainability-by-Design* sollte als konkreter Implementierungsansatz etabliert werden. Denkbar wären gesetzliche Mindestanforderungen an Erklärbarkeit für KI-Systeme in sensiblen Bereichen.

10

8. Künstliche Intelligenz und Demokratie

Team-AQC: Wir wechseln von der Frage der Nachvollziehbarkeit zur Frage der Legitimität im Gemeinwesen. Eine berühmte Befürchtung besteht darin, dass KI durch gleichlautende und unbalancierte (bewusst manipulierte) Antworten demokratische Strukturen gefährden könnte. Erneut ein Aspekt, der die Verantwortung thematisiert, nicht wahr?

15 FL: Ja, KI ist nicht einfach nur Innovation. Sie beinhaltet ein Steuerungs- und sogar ein Manipulationspotenzial. Darum sind zahlreiche Forschende auch der Meinung, dass KI Werte und Standards einzubeziehen habe, die wir unter dem Begriff «Verantwortung» subsumieren. Dies wäre Verantwortung für Themen und Aspekte der Vergangenheit, Gegenwart und Zukunft. Vor allem aber ist die Verantwortung gegenüber den kommenden Generationen ein wesentlicher Ansatz-
20 punkt für ethische Fragen. Deshalb wird den KI-Systemen im Bildungsbereich auch eine so wichtige Funktion zukommen. Gerade im Bildungssektor müsste man den Nutzenden zuerst beibringen, sich selbst zu entfalten, bevor eine beliebige Form von KI überhaupt die Chance hat, die Breite der Wahrnehmung und offene Haltungen in eine gewisse Richtung zu lenken. Vergessen wir nicht: KI sollte nicht das Ziel, sondern der Schlüssel sein, damit viele ihr eigenes Ziel er-
25 reichen können. KI kann somit eng mit Bildungsidealen verbunden sein und mit einer Öffnung der Bildung in alle Gesellschaftsbereiche. Damit sind wir aber eigentlich einen Schritt weiter: Nicht nur KI und Demokratie sind hierbei in ihrem Verhältnis angesprochen, sondern auch die «Demokratisierung der KI». Wer KI demokratisieren will, muss sie aber zuerst entmystifizieren, das heisst, sie wirklich breiten Gruppen zugänglich machen und in umfangreichen Bildungspro-
30 zessen verankern. Eine konkrete Anwendung wäre die Einrichtung von Demokratie-Laboren: öffentliche Testumgebungen, in denen Bürger:innen KI-Systeme kritisch ausprobieren und hinterfragen können – insbesondere im Hinblick auf Meinungsbildung, politische Positionen und den Wandel von Herrschaftsformen.



9. Künstliche Intelligenz als nutzbar gemachte disruptive Technologie

Team-AQC: Aus demokratischer Verantwortung folgt die Gestaltungsfrage: Wie machen wir eine disruptive Technologie gemeinwohlfähig? Was wir heute unter dem Schlagwort «KI» bezeichnen, kann mit guten Gründen als disruptive Technologie bezeichnet werden. Kann man eine solche disruptive Technologie breit und verantwortungsvoll nutzbar machen?

FL: Ja, dieser Nutzen ist denkbar und wir sollten alle ein besonderes Interesse daran haben, disruptive Technologien gesellschaftlich nützlich und nutzbar zu machen. Meiner Ansicht nach muss KI nicht nur für Effizienz, sondern vor allem für Transparenz, Dezentralität und neue Formen des Vertrauens stehen. Dies haben auch Technologiepioniere mit Schwerpunkten in KI und Blockchain in den letzten Jahren deutlich betont. Denkt man diesen Ansatz konsequent weiter, muss man KI-Systeme jenseits klassischer Geschäftsmodelle denken. Sie müssen zugänglich und gemeinwohlorientiert gestaltet sein. Der Umkehrschluss lautet, dass vor einer Machtzentralisierung im KI-Sektor nachdrücklich gewarnt werden muss. Dies entspricht einem Plädoyer zugunsten von Open-Source- und Community-basierten Lösungen. Konkret liesse sich mit einer Open-Source-Bildungs-KI viel erreichen – am Gemeinwohl orientierte Alternativen zu den bisherigen kommerziellen Systemen müssten eigentlich an der Tagesordnung sein. Mit anderen Worten: KI gehört nicht in die Hände weniger Menschen. Sie gehört in die Verantwortung vieler. Dabei ist auch zu betonen, dass Transparenz nicht einfach nur ein Luxus ist – vielmehr stellt sie das Rückgrat jeder ethischen KI dar. Zentral ist dabei die Forderung nach transparenten Datenpfaden. Nutzer:innen sollen nachvollziehen können, welche Daten wohin fließen – und warum.

Abschnitt III: Verzerrung, Macht, Recht

25

10. Künstliche Intelligenz in Translation

Team-AQC: KI, vielleicht sogar am deutlichsten in Form von LLMs (large language models), übersetzt immer auch zwischen Code und Sprachen. Ist KI darum eigentlich ein Übersetzungstool, und besteht die Gefahr, dass es zu Situationen kommt, die man als «Lost in Translation» bezeichnen muss?

FL: Spezialist:inn:en, die neben KI auch Erfahrung als professionelle Übersetzer:innen haben, sind immer wieder mit der Frage konfrontiert: Wie übersetzen Maschinen – was bleibt erhalten und was geht dabei verloren? Wir müssen uns klar machen – KI-Systeme sind niemals neutrale Vermittler. KI-Systeme müssen viel mehr tun, als «einfach nur Worte» übersetzen – es geht um die Weltbilder dahinter. Zudem gibt es keine neutrale KI, denn jede Entscheidung trägt eine Bedeutung in sich. Und diese ist entweder programmiert oder erlernt. In Bezug auf Übersetzungs-

35



tätigkeit muss KI lernen, im echten Sinne zu übersetzen (also auch die Welt dahinter wahrzunehmen und in der Übersetzung zu präsentieren). Simple Übertragen ist keine Option, sondern führt tatsächlich eher zu einem Verlust an Nuancen und Präzision. Insofern ist die Phrase «Lost in Translation» als reale Gefahr gar nicht mal so aus der Luft gegriffen. Im Rahmen der Weiterentwicklung sollten vor allem kontext-sensitive Übersetzungs-KIs etabliert werden. Diese würden eine «Nuancenwarnung» enthalten, also Hinweise auf Übersetzungsvorgänge, bei denen ein Bedeutungsverlust droht.

11. Verfälschende Künstliche Intelligenz

10 *Team-AQC: Der Verlust von Nuancen kann als eine Art der Verfälschung verstanden werden. Uns ist bewusst, dass KI (leider auch) fälschen oder verfälschen kann. Welches sind die Herausforderungen in diesem Bereich?*

FL: Expert:inn:en auf diesem Sektor sind überzeugt, dass Entstehung von Empathie, Lernen und Kontextverständnis in KI-Systemen besser untersucht werden müsste. Hierbei stellt sich nämlich genau die Verfälschungsfrage: Ob und wie kann KI menschliche Interaktionen abbilden oder begleiten, ohne sie zu verfälschen? KI-Systeme sollten zukünftig die emotionale und soziale Intelligenz ernst nehmen. KI würde dann gefährlich, wenn sie vorgeben würde, ein Gehirn zu besitzen und wir dann dieser Annahme auch noch aufsitzen würden. «Verstehen» heisst eben nicht nur Muster erkennen, sondern Beziehungen wahrnehmen. An diesem Punkt ist KI aber noch immer relativ blind und «verfälscht» damit regelmässig die Resultate. Konkret wäre hier an eine automatisierte Prüfung von Verfälschungsindikatoren zu denken. KI-Systeme markieren dabei Stellen, an denen emotionale, soziale oder kulturelle Aspekte nur approximiert werden und die Gefahr von Ergebnisverfälschungen besteht.

12. Künstliche Intelligenz, Datenschutz und Recht

Team-AQC: Eine grosse Frage wirft auch der Themenkomplex «KI und Datenschutz» auf: Nutzer:innen geben achtlos in die Textfelder grosser Sprachmodelle ihre Kreditkartennummern und Geheimzahlen ein. Aber Datenschutz beginnt ja schon auf einer viel elementarerer Stufe. Was gibt es hier zu berichten, gerade wenn man auch den Begriff der Datenverantwortung einbezieht?

30 FL: Immer wieder stellt sich die Frage nach der Einhaltung europäischer Werte bei der Konzeption und Programmierung von KI-Systemen: Datenschutz, Transparenz und Wahrhaftigkeit durch KI sind hier zentrale Begriffe. Zugleich muss aber auch die Relativität dieser Begriffe, zuvorderst beim Datenschutz, beachtet werden. Die europäischen Werte unterscheiden sich deutlich von nordamerikanischen oder asiatischen Werten. Und entsprechend unterscheidet sich auch die inhaltliche Füllung von Datenschutz und Transparenz. Datenschutz beginnt im Übri-



gen schon beim gewählten Abonnement. Je nach Anbieter und Abonnementmodell gelten unterschiedliche Regelungen dazu, ob Eingaben zum Training verwendet werden. Zwar gehen die Daten wahrscheinlich in nicht zurückverfolgbarer Form in einen grossen Daten- und Statistikpool ein. Dennoch bereiten solche Szenarien grosses Kopfzerbrechen. An der Wortwahl «wahrscheinlich» lässt sich ein weiterer wichtiger Punkt benennen: Das Ziel muss lauten, eine erklär-
bare KI zu schaffen, die für Nutzer:innen nachvollziehbar bleibt, indem sie sich auch selbst erklären kann. Hierdurch kann letztlich auch Vorsicht resultieren, wenn man zum Beispiel orientiert wird, wofür die eingegebenen Daten weiterverwendet werden. Datenschutzexpert:inn:en in vielen Ländern fordern beispielsweise klare gesetzliche Rahmenbedingungen, transparente Datenverarbeitung sowie das Recht auf menschliche Kontrolle. Sie warnen vor einer Welt, in der autonome Systeme ohne rechtliche Verantwortung agieren. Für sie ist der Schutz der digitalen Selbstbestimmung Voraussetzung für jede legitime KI-Nutzung. Denn ohne Datenschutz wäre jede KI ein Risiko – für Demokratie, Freiheit und Würde. Wir bräuchten eigentlich Gesetze, die nicht nur Maschinen regeln, sondern vor allem Menschen schützen. Inhaltlich gewendet: Gute KI erklärt nicht nur, was sie tut – sie zeigt auch, warum sie es tut. *Wahrheit* im Vorgang und in seiner Erklärung wird nicht einfach automatisch durch die Grösse der Datenmenge erschaffen. Wahrheit ist vielmehr ein Prinzip, das wir Maschinen beibringen/lehren müssen. KI kann entsprechend nur dann Vertrauen verdienen, wenn sie selbst das entgegengebrachte Vertrauen schützt. Es ist also eine KI notwendig, die nicht nur rechnet, sondern auch verantwortlich handelt. Rechtlich könnte dies durch das Prinzip «Datenschutz durch Voreinstellung» (*by default*) begleitet werden. Es sollte der Standard sein, dass Nutzereingaben nicht für Trainingszwecke Verwendung finden. Nicht alles, was technisch trainierbar ist, darf auch trainiert werden – vergleichbare ethische Standards existieren schon heute im Gesundheitsbereich.

25

Abschnitt IV: Lernen, Teilhabe, Inklusion

13. Cui bono, Künstliche Intelligenz?

Team-AQC: Wer hat einen Vorteil durch KI? Wer wird begleitet, ausgebildet oder informiert? Geht es um Individuen, Gruppen oder wird KI sogar als übergreifender kollektiver Lernprozess verstanden?

FL: Wenn man die Diskussionen verfolgt, die beispielsweise bei *robot me* und bei der *impactIA Foundation* geführt werden, so ist die Perspektive auf KI geprägt von Kreativität, Offenheit und menschenzentriertem Design. Nicht nur die Maschine wird trainiert und lernt, sondern es ist vor allem der Mensch, der mit seinem Lernen die entscheidende Grösse darstellt. Expert:inn:en in diesem Bereich setzen sich für eine zugängliche, lernfreundliche KI-Kultur ein, die auf Zusammenarbeit basiert. Besonders wichtig ist hierbei, dass auch Anfänger:innen Raum in der KI-Welt finden und sich entwickeln dürfen. Denn alle haben einmal klein angefangen mit KI – das Ent-

35



scheidende ist aber, was daraus hervorgeht. KI ist weniger ein technisches Problem als vielmehr eine soziale Herausforderung. Wesentlich scheint mir zu sein, dass KI-Entwicklung keine Eliteübung sein darf, sondern ein kollektiver Lernprozess, der auch offen für Irrtümer und Neugier sein muss. Ein gut umsetzbarer Vorschlag ist die Einführung von KI-Kompetenz als Grundbildung. Denkbar wäre ein niederschwelliges Bildungsprogramm für alle Altersgruppen im Sinne des lebenslangen und generationenübergreifenden Lernens.

14. Künstliche Intelligenz und Inklusion

Team-AQC: Der gerade angesprochene «kollektive Lernprozess» setzt grundsätzlich eine «Nutzung durch viele» voraus. Kann diese Nutzung «durch viele» auch als integrativer Akt verstanden werden oder sogar als Inklusionsbestreben? Kann man also aus dem Gesagten auch eine notwendige Verbindung von KI und Inklusion ableiten?

FL: Das wäre dann eine sehr sozialpolitisch motivierte Perspektive auf KI – aber sie ist absolut relevant. Man muss sich immer die Frage stellen, wer Algorithmen gestaltet, wen diese dann einschliessen und wen sie potenziell ausschliessen. Es ist tatsächlich so, dass Bias-Strukturen in KI-Systemen eingebettet sind, und zwar schon auf der Ebene der Programmierung. Dies setzt sich dann mit einer spezifischen Fragestellung (Input) fort – denn neben der Grundkonfiguration trägt bei lernenden Systemen auch jeder Input zur Anpassung des Systems bei. Möchte man an diesen Stellen gegensteuern, so ist dies nur mit fairen, inklusiven Entwicklungen – von Anfang an – möglich. Entsprechend sind intersektionale Ansätze in Systemen, die soziale Gerechtigkeit mitdenken, von grosser Bedeutung. Die beschriebenen Herausforderungen im Bereich KI-Bias könnten rechtlich durch verpflichtende Bias-Audits minimiert werden. Dies wären regelmässige externe Prüfungen von KI-Systemen auf Verzerrungen und Bias-Entwicklungen.

Für Inklusion gilt: Viel wichtiger als nur «Vielfalt zu erkennen» ist es in diesem Fall, «Vielfalt zu respektieren». Darum ist in diesem Ansatz «gerechte KI» in keinem Fall ein Nebenprodukt – sie ist eine Frage der Entscheidung und des Designs. Wer diskriminierende Systeme mit KI beschleunigt, nennt dies oft «Effizienz». Ebenso gut könne man dies aber als unfairen und destabilisierenden Ausschluss bezeichnen. Auf die Herausforderungen der inklusiven KI wird im Rahmen von Konferenzen immer wieder hingewiesen. Insgesamt ist es möglich, in KI ein Werkzeug zu sehen, das soziale und inklusive Projekte fördern kann – aber auch ein Risiko darstellt, wenn es unreflektiert eingesetzt wird. Im Bereich der partizipativen Technologieentwicklung, in der Nutzer:innen nicht nur Zielgruppe, sondern auch Mitgestaltende sind, spielt dies eine entscheidende Rolle. Denn Mitgestalten muss inklusiv gedacht werden. Und es spielt eine grosse Rolle, dass Technologie immer sozialer werden muss. Ohne soziale Einbettung droht Technologie immer an Relevanz zu verlieren. Letztlich: Um einen wirklichen Effekt zu erzielen, muss KI nicht nur wissen, wie wir funktionieren – sie muss auch lernen, gemeinsam mit uns zu



gestalten. Für echte Inklusion ist zudem ein partizipatives KI-Design erforderlich: Betroffene Gruppen werden aktiv in die Entwicklung einbezogen. Das Sichtbarmachen solcher Schritte könnte mit einem Inklusionsscore entwickelt werden. KI-Systeme würden danach bewertet, welche Gruppen sie ausschliessen und in welchem Ausmass. Ein hoher Inklusionsscore könnte als
5 Qualitätsmerkmal dienen.

15. Künstliche Intelligenz für Innovation und Lernen

Team-AQC: *Könnte KI also auch als gestaltende Grösse für Innovations- und Lernprozesse verwendet werden?*

10 FL: Ja, zum Beispiel könnte die Verbindung von Bildung, sozialer Praxis und Technologie sehr ertragreich durch KI beleuchtet, optimiert und weiterentwickelt werden. In jüngster Zeit hört man solche Ansätze vermehrt vom *idEE-Léman*. Man realisiert dort, dass KI ein Medium für kollektives Lernen darstellt, das Menschen dabei helfen soll, komplexe Herausforderungen gemeinsam zu verstehen – und nicht ein Mittel, komplexe Herausforderungen zu umgehen. KI soll in dieser
15 Vorstellung nicht die Komplexität des Lebens glätten, sie muss vielmehr helfen, diese zu gestalten. Darum fördert ein gutes KI-System nach diesem Verständnis nicht Effizienz um jeden Preis, sondern kreiert vielmehr Resonanz, Reflexion und Dialog. Didaktisch gewendet: In der Praxis sollte KI vor allem als Lernbegleiter auftreten, nicht als Prüfinstanz. In dieser Rolle liegt der Fokus stärker auf der oben genannten Reflexion und dem Dialog sowie auf dem Feedback anstatt
20 einer Bewertung. Eine weitere vielversprechende Idee sind fehlerfreundliche KI-Umgebungen: digitale Räume, in denen auch falsche Antworten explorativ genutzt werden dürfen – in erster Linie als eigenständige Experimentierräume (und nur nachgeordnet zur Profilierung nicht-fehlerfreundlicher Systeme).

25

Abschnitt V: Ausblick

16. Scheiternde und überraschende Künstliche Intelligenz

Team-AQC: *Gibt es einen Aspekt, den Sie bemerkenswert oder faszinierend finden in Bezug auf neue
30 KI-Entwicklungen?*

FL: Oh ja, da gibt es tatsächlich ein Beispiel: Die scheiternde und auch die überraschende KI. Im Bereich der performativen und der darstellenden Kunst gibt es einen Trend, zu erforschen, wie Maschinen als kreative Partner im künstlerischen Prozess wirken können. In solchen «Performances» geht es gar nicht mehr um die Kontrolle der KI, sondern um Dialog: Wie reagiert ein Al-
35 gorithmus auf menschliche Bewegung – und umgekehrt? *Jonathan O'Hear* (Künstler, Leiter eines



experimentellen Tanzprojekts mit KI) gehört zu den Personen, die hierzu viel Erfahrung gesammelt haben. In diesem konkreten Fall gibt es gute Gründe, sich dafür einzusetzen, dass KI nicht auf reine Vorhersagbarkeit reduziert wird. Vielmehr soll KI improvisieren, reagieren und überraschen können – ihr wird hierfür der interaktive Raum zur Verfügung gestellt. In einem solchen

5 Ansatz ist KI auch immer an Ästhetik und Beziehung geknüpft. Und es geht noch weiter: KI, die wirklich kreativ agieren soll, muss auch scheitern können oder dürfen. Dies bedeutet auch eine andere Zielsetzung: Es geht nicht um eine KI, die perfekt tanzt, perfekt malt etc. Vielmehr geht es um eine KI, die dem jeweiligen Gegenüber etwas Unerwartetes zeigt. Eine konkrete Anwendung wären künstlerische KI-Labore, in denen Kunst, Tanz und Musik als Gefäße für neue, ungeahnte Formen der Improvisation dienen. In diesem Kontext muss Scheitern als Designprinzip

10 verstanden werden: Ziel ist nicht Optimierung, sondern die bewusste Erhöhung der Überraschungsfähigkeit.

17. Der Mensch im Zentrum

15 *Team-AQC: Gehen wir von der performenden und künstlerisch interagierenden KI wieder zur Idee vom «Menschen im Mittelpunkt»: Wie sieht eine mögliche menschenzentrierte KI aus?*

FL: Einige Sachbeiträge von Personen mit multidisziplinärem Hintergrund (Ärzt:inn:e:n oder Therapeut:inn:en in kognitiver Verhaltenstherapie u.a.) beschreiben ihre Perspektive auf KI unter besonderer Berücksichtigung des Gesundheitsbereichs: Sie setzen sich für eine KI ein, die unter anderem emotionale Intelligenz respektiert und damit vor allem das menschliche Erleben

20 nicht auf wenige Dimensionen reduziert. In einer Welt, in der Algorithmen zunehmend Diagnosen oder Empfehlungen geben, mahnen diese Stimmen zur Vorsicht. Genauerhin wird hier die Ansicht vertreten, dass KI die therapeutische Beziehung oder die ganzheitliche Wahrnehmung niemals ersetzen darf. Deshalb auch die Forderung nach einer Gesundheits-KI mit Beziehungsprimat. KI darf hier nur unterstützend wirken – niemals anstelle einer therapeutischen Beziehung. Mit andern Worten müssten KI und KI-Ergebnisse individuell, traumasensibel und behandlungsunterstützend sein. Grundsätzlich gilt die Hypothese, dass KI in zahlreichen Situationen helfen kann. Es darf aber nie vergessen werden, dass ein Mensch keine Variable, sondern ein Individuum ist. Gesundheit beginnt zudem mit und durch Beziehung. Diese wichtigen Aspekte kann – nach meinem Dafürhalten – eine Maschine zumindest nicht in absehbarer Zeit authentisch imitieren.

30



18. Künstliche Intelligenz, Kultur und Spiritualität

Team-AQC: *Wie würden Sie das Verhältnis von KI, Kultur und Spiritualität beschreiben?*

FL: Neue Technologien sollten auf keinen Fall das menschliche Mass aus den Augen verlieren, der Mensch ist (nicht erst seit dem von Protagoras überlieferten Homo-Mensura-Satz) das «Mass
5 der Dinge». Es darf bei KI also nicht um Entmenschlichung gehen, sondern im Gegensatz dazu bewusst um das Kultivieren von tieferen Werten wie Achtsamkeit, Respekt und Verbindung. Dementsprechend ist vor KI-Ansätzen zu warnen, die versuchen, sich von Natur und Gefühl(en) freizumachen. Ein ganzheitlicher Ansatz ist notwendig, der KI in die kulturelle und spirituelle Entwicklung einbettet. Auch wenn KI vieles berechnen kann, zählt oft das, was berührt und Re-
10 sonanz erzeugt viel mehr – nicht nur das mathematisch-statistische. Dies führt zur spannenden Frage, ob das, was uns berührt, oder das, was mathematisch korrekt ist, als «lebensdienlich» bezeichnet werden kann. Über die Aspekte der nicht-lebensdienlichen Entfremdung werden in letzter Zeit immer mehr Analysen publiziert. Entfremdungsanalysen haben ja auch eine lange Tradition, die systemisch mindestens bis Karl Marx zurückreicht.

15

19. Künstliche Intelligenz, Natur und ökologische Aspekte

Team-AQC: *Ganzheitlichkeit hängt auch eng mit der Integration der Natur zusammen: Gibt es denn Überlegungen, bei denen KI und Naturphänomene verschränkt werden?*

FL: Ja, das ist natürlich auch immer ein Thema. Wenn ich die Diskussionslage richtig überblicke,
20 vertreten Forschende von *Algokon* und der *Kina8at Foundation* einen integrativen Ansatz zur KI, der weitreichende Umweltthemen einbezieht: Achtung vor der Natur, der Gemeinschaft/Gesellschaft und balancierte Mensch-Umwelt-Systeme in mehreren Perspektiven. Dies erfordert eine kritische Reflexion darüber, ob Algorithmen und Lebensweisheit(en) in Einklang stehen. Deshalb ist dieser Ansatz nicht ohne Herausforderungen, weil er die Verantwortung, die Empathie
25 und den Respekt als Grundlagen technischer Entwicklung begreifen möchte. Wahre Intelligenz entsteht nach diesem Ansatz nicht aus Kontrolle, sondern aus Verbundenheit mit allem Lebendigen. Technologie darf darum in keinem Fall unsere Wurzeln kappen – vielmehr sollte sie dazu beitragen, die Wurzeln («unsere Wurzeln») zu pflegen und zu kultivieren. Ein weiterer Schritt sind im Übrigen ökologische KI-Bilanzen: Die transparente Offenlegung von Energieverbrauch
30 und CO₂-Fussabdruck schafft Erkenntnisse jenseits rein volkswirtschaftlicher Kennzahlen.



20. Künstliche Intelligenz und Nachhaltigkeit

Team-AQC: *Das schliesst dann aber auch Nachhaltigkeit mit ein – ein heisses Eisen, wenn man den Energieverbrauch von KI-Systemen weltweit betrachtet, nicht wahr?*

FL: «KI und Nachhaltigkeit» ist in der Tat ein herausfordernder Aspekt. Der unglaubliche und ungebändigte Energiehunger der neuen Systeme stellt uns bereits jetzt vor Herausforderungen und bringt ein Begriffspaar wieder in den Blick, das lange Zeit nicht betrachtet wurde: Demut und Bescheidenheit. Denn an wenigen Stellen stellt eine Masslosigkeit, wenn sie zum Systemparameter erhoben wird, ein so greifbares Problem dar wie bei der Energiebilanz unlimitierter KI-Systeme. Als wirksames Instrument könnten KI-Moratorien bei Masslosigkeit dienen – etwa als politische Möglichkeit, besonders energieintensive Systeme zu begrenzen. Thematisch-philosophischer haben sich Stimmen von *Genève G4* in letzter Zeit immer wieder zu Wort gemeldet: Dabei werden vor allem Nachhaltigkeit und interkulturelle Entwicklung im Bereich KI fokussiert. Es geht hierbei, wie auch bei der oben genannten Energiebilanz, um Verantwortung. Dies ist ein Schlüssel zu vielen Herausforderungen: Der Einsatz von KI bedeutet nicht Beschleunigung um jeden Preis, sondern vielmehr Innehalten. Hierbei muss die Frage gestattet sein, welche Zukunft gestaltet werden soll. Strukturelle Gerechtigkeit und globale Ausgewogenheit bei KI-Anwendungen spielen in jedem Fall eine besondere Rolle. Technologie braucht immer einen Kompass. Ohne Ethik ist auch KI nur blinder und unbewusster Fortschritt. All dies mündet in ein klares Plädoyer für Ethik als Systemkomponente: Ethik ist kein nachgelagerter Schritt und keine Randerscheinung, sondern muss fest im Systemdesign verankert sein. Und weiterhin: Ein wahrhaft intelligentes System kann nur entstehen, wenn der Mensch sich nicht selbst vergisst. Er muss Systemkomponente bleiben. Und das heisst auch, mit Begrenzung und natürlichen Schranken umzugehen. KI muss sich entsprechend auch an der Frage messen lassen: Welche Welt wollen wir mit KI (er)schaffen? Da diese disruptive Technologie – wie jede Form der Technologie – kein reiner Selbstzweck ist, müsste sie Teil eines grösseren ökologischen und sozialen Plans sein.