KI – Zwischen Chancen und Gefahren

Bericht zur BCE-Kuratoriumstagung am 24.10.2024 zum Thema Künstliche Intelligenz

Im Zuge der letzten Kuratoriumssitzung der Association of Bioethicists in Central Europe (BCE) am 24.10.2024 in Wien fanden die Planungen für die nächste BCE-Tagung 2025 statt. Inhaltlich wird es um Künstliche Intelligenz (KI) und die damit verbundenen ethischen Fragen gehen. Als ersten Vorgeschmack und auch, um alle Mitglieder von BCE auf einen gemeinsamen Wissensstand zu bringen, wurde ein Reigen interessanter Vorträge zu dem Thema vorbereitet. Nach einem ersten sozialethischen Impuls zur Frage von KI und Bildung durch Alexander Filipovic wurde aus der Perspektive der Informatik über den technischen Stand der Dinge rund um KI und mögliche Weiterentwicklung der näheren Zukunft von Octavian Machidon referiert. Zuletzt beleuchtete Jonas Miklavčič die ethischen Herausforderungen und Prinzipien im Umgang mit KI und da vor allem über die Anwendungsmöglichkeiten in der Medizin.

Alexander Filipovic: Impuls: KI und Bildung – Perspektiven für die Zukunft

Filipovic stellt zunächst fest, dass KI in immer mehr gesellschaftlichen Bereichen von immer größer werdender Bedeutung ist. Damit verbunden sind Fragen nach dem Selbstverständnis des Menschen und seinem Weltverhältnis. KI-Systeme fordern das Verständnis vom Menschen dort heraus, wo es in Abgrenzung zur Leistung von Maschinen definiert wurde. Das Verhältnis zur Welt wird durch KI-gesteuerte Assistenzsystemen herausgefordert, die – menschliche Interaktion imitierend – mit Namen angesprochen werden oder Antworten in einer Qualität wiedergeben, dass sie auf den ersten Blick menschlich erscheinen.

Diese grundsätzlichen anthropologischen Themen führen auch zur Frage, was Bildung bedeuten soll, und was unter einem gebildeten Menschen zu verstehen ist. Filipovic gibt zu bedenken, dass der Begriff der Bildung viele zum Teil gegensätzliche Bedeutungen umfasst. Mit Bildung sind Fragen menschlicher Entwicklungs- aber auch Wissensfähigkeiten verbunden, die sich auch auf das Thema der Verantwortungsfähigkeit beziehen. Hinzu kommen Fragen des guten und gerechten (Zusammen-)Lebens. Hier klingt also die Trias an von Anthropologie, Bildung und Ethik und auf diese sollten Bildungsinstitutionen (natürlich neben anderen Interessen wie der Finanzierung usw.) ausgerichtet sein.

Haben schon Computer und das Internet Bildungsinstitutionen vor neue Herausforderungen gestellt, so ist die Frage nach ihrem Sinn und Zweck in Folge von KI-Anwendungen wie ChatGPT in neuer, umfassenderer Weise gestellt worden. Letztlich, und das ist auch als Chance zu sehen, ist es an der Zeit, über Bildungsziele zu diskutieren und als Bildungsinstitution sich zu hinterfragen, ob man diesen Zielen wirklich gerecht wird. Geht es darum, bloß Inhalte, Fakten und Zahlen wiederzugeben? Ist eine persönliche Reflexion über ein komplexeres Thema wirklich über eine wissenschaftliche schriftliche Arbeit, wie Seminar- oder Bachelorarbeit, abgebildet, wenn ChatGPT diese auf Knopfdruck erstellen kann? Man steht also vor der

Herausforderung, wie man angesichts von KI-Anwendungen kritische, autonome sowie an Beteiligung und Mitgestaltung interessierte Menschen hervorbringt. Hier schwingen auch schon die Bildungsziele mit, wie sie Filipovic vorschlägt, nämlich Autonomie, Beteiligung und Mitgestaltung. Inhalte und Methoden von Bildungseinrichtungen müssen sich folglich selbst hinterfragen, inwiefern sie diesen drei großen Zielen in und angesichts der heutigen Welt gerecht werden.

Vor diesem Hintergrund referierte Filipovic zuletzt über die Zukunft von Bildung und die zukünftige Gestaltung von Bildungsprozessen. Die Frage, ob KI menschliche Lehrkräfte ersetzen könnte, bejahte er vorsichtig. KI-Anwendungen werden schon heute von Lehrkräften eingesetzt, um anhand von (Leistungs-)Daten der Studierenden gezielt und individueller auf den jeweiligen Bildungsprozess Einfluss nehmen zu können. Ein denkbarer Ersatz für die so agierende analoge Lehrkraft könnte dann ein KI-Lehrsystem sein, das mit einem individuellen KI-Begleiter einen noch individuelleren Bildungsprozess für den aber weiterhin menschlichen Studierenden gestalten soll.

Hier ist aber auch schon eine wichtige Einschränkung zu sehen, nämlich dass Bildungsprozesse auch trotz KI immer eine menschliche Angelegenheit bleiben. Die Menschen bleiben also relevant, insofern sie ja Empfänger*innen von Lerninhalten bleiben, ihre Ergebnisse darüber Auskunft geben, ob die Bildungsziele erreicht wurden und es letztlich auch immer um den Umgang der Menschen untereinander gehen muss. Denn genau das wurde in der darauffolgenden Diskussion, etwa bei den Daten, die KI-Systeme verarbeiten, thematisiert. Wie können Ungerechtigkeiten der Datenerfassung in der menschlichen Welt nicht wiederum durch KI reproduziert werden? Technologien sind eben nicht nur bloße Mittel. Sie bieten die Chance, Ungerechtigkeiten zu beseitigen, können sie aber ebenso verschleiern und weiter reproduzieren oder gar neue hervorrufen.

Octavian Machidon: Insights into how AI works – the current state of play and the prospects for further development

Machidon stellte aus der Perspektive der Informatik vor, was sich alles hinter dem Begriff der Künstlichen Intelligenz verbirgt. Hierdurch soll klar werden, worüber wir eigentlich reden und welche Missverständnisse in der Diskussion über KI vorhanden sind.

Künstliche Intelligenz meint allgemein einen Teilbereich der Informatik. Weiters wird zwischen „weak AI“ und „general AI“ unterschieden. Ersteres meint KI-Systeme, die für eine spezifische Aufgabe konstruiert wurden. Darunter fallen alle KI-Anwendungen, die uns im Alltag begegnen, von Assistenten am Smartphone bis hin zu selbstfahrenden Autos. General AI meint das, was wir vor allem als Antagonisten aus Science-Fiction-Filmen kennen, autonome Computer-Systeme wie HAL in „2001: A Space Odyssey“, die menschliche Fähigkeiten nicht nur imitieren, sondern diese übertreffen und die selbstständig handeln. Um den gegenwärtigen Stand der Technik besser abbilden zu können, wird in der Informatik auch eine alternative Nomenklatur diskutiert, die die übermenschlichen Fähigkeiten der General-AI unter der neuen Bezeichnung einer „Artificial Super Intelligence“ (ASI) fassen. So werden dann Anwendungen, die menschlichen Fähigkeiten nahekommen wie ChatGPT als „Artificial General Intelligence“ (AGI) bezeichnet, im Gegensatz zu „Artifical Narrow Intelligence“ (ANI), die nur spezifische Aufgaben bewältigen kann.

Ein wichtiger Schritt in der rasanten Entwicklung von KI-Systemen ist das sogenannte „Deep Learning“, das sich durch weniger menschliche Inputs auszeichnet. Entscheidend hierfür sind die sogenannten „Neural Networks“ (neurale Netzwerke), die aus einem von Menschen vorgegebenen Datensatz durch eine Serie von Algorithmen neue Verbindungen herstellen und somit neue Informationen kreieren. Dies kommt etwa in Anwendungen wie KI-gesteuerte Objekt-Erkennungen (Face-ID) oder im medizinischen Bereich in Hautkrebs-Screening vor, aber auch in virtuellen Assistenten oder ChatGPT. Der Vorteil von KI-Systemen, die über Deep Learning Daten bearbeiten, ist es, dass sich der menschliche Input vor allem auf die Bereitstellung eines möglichst großen Trainingsdatensatz beschränkt. Danach sind nach einer erfolgreichen Testphase nur mehr fallweise Anpassungen an den Parametern zur Datenauswertung nötig; also eine äußert minimale Moderation durch Menschen.

Nach der technischen Einführung zeigte Machidon auf, welche Interessen hinter der Entwicklung von KI stehen. Zunächst ist gemäß Moore’s Law alle zwei Jahre mit einer Verdoppelung der Anzahl von Transistoren auf Computer-Chips zu rechnen, was die Rechenleistung entsprechend erhöht. Diese Produktivitätssteigerung der Computer-Chips spiegelt auch die wirtschaftliche Forderung nach allgemeiner Produktivitätssteigerung wider. Dementsprechend bestehen massive finanzielle Anreize, die Entwicklung von KI zu fördern, auch und gerade weil in praktisch allen Lebensbereichen Anwendungsmöglichkeiten gegeben sind. Enorme Potentiale werden beispielsweise in verbesserten Klimamodellen, in der Entdeckung neuer medizinischer Behandlungsmöglichkeiten sowie im individuellen Bereich von Smart-Homes gesehen. Jedoch sind auch Ängste verbreitet, die zum einen wohl popkulturellen Darstellungen einer Übernahme der Menschheit durch KI geschuldet sind. Zum anderen aber auch weil die ethischen Debatten bisher den Entwicklungen hinterhergehinkt sind. Insofern ist die geplante Konferenz 2025 ein wichtiger Beitrag, interdisziplinär und vor allem proaktiv und nicht mehr bloß reaktiv über die ethischen Implikationen dieser sich immer rasanter entwickelnden Technologie zu reflektieren.

Den Abschluss bildete bei Machidon eine kurze Einschätzung, wie nahe wir an der sogenannten Singularität sind, also der Entwicklung einer selbstbewussten KI. Machidon meint zwar, dass wir schon sehr nahe herangekommen sind mit jenen technischen Anwendungen, die sich auf generative KI beziehen, wie etwa Chat-GPT. Viele Aufgaben können in einer Qualität bewältigt werden, die mit menschlichen Fähigkeiten vergleichbar sind oder sie sogar übertreffen. Jedoch bestehen noch einige Einschränkungen: Zum einen sind diese KI-Systeme von der Qualität des ursprünglichen Trainingsdatensatzes abhängig. Zum anderen bestehen Schwierigkeiten, sich an geänderte Bedingungen anzupassen. Daneben ist fraglich, ob etwa ChatGPT wirklich kreativ sein kann, und die Fragen des Datenschutzes sind ebenso noch offen.

Innerhalb einer fünf-stufigen Skala der KI-Entwicklung, so Machidon, befinden wir uns schon auf der dritten Stufe. Nachdem es schon seit ein paar Jahren Chat-Bots gibt und manche KI-Anwendungen mit menschlichen Fähigkeiten mithalten können, sind gegenwärtige KI-Systeme zu eigenständigen Tätigkeiten fähig, das heißt ohne zusätzlichen Input. In den zwei noch folgenden Stufen müsste KI eigene Erfindungen erstellen können (echte Kreativität) und zuletzt auch ganze Organisationen ersetzen können, die die Kooperation verschiedenster Menschen mit unterschiedlichen Hintergründen koordinieren.

Dies macht es folglich umso wichtiger, einerseits Forschung an KI voranzutreiben, andererseits diese zu regulieren und schließlich auch ein Bewusstsein über KI zu schaffen, was wiederum an Bildung und an ethischen Reflexionen hängt, die proaktiv erfolgen sollten.

Jonas Miklavčič: Ethics of Artificial Intelligence: Challenges and Principles

Jonas Miklavčič setzte sich mit den Fragen auseinander, welche ethischen Problemstellungen sich im Zusammenhang mit KI ergeben und welche Prinzipien angewendet werden können. Das Themengebiet der KI-Ethik ist in der Technikethik angesiedelt. Durch die Entwicklung des Maschinenlernens und noch mehr durch die Anwendungen von Deep Learning haben sich viele Fragestellungen der Maschinenethik, Roboterethik oder Algorithmenethik massiv erweitert. Fragen der Vorhersagbarkeit von Ergebnissen der KI, aber auch die Herausforderung immer größerer Datenmengen sowie die Genauigkeit und Personalisierungsmöglichkeiten, die in der Arbeit mit KI nun möglich sind, stellen die Ethik vor neue Herausforderungen.

Bei all den Jubelmeldungen über den übermenschlichen Erfolg von KI-Anwendungen im medizinischen Bereich, etwa bei Augenuntersuchungen oder Hautkrebs-Screening, werden die missglückten Beispiele oft nicht erwähnt und vor allem sind in den letzten Jahren die Fehler und Unfälle aufgrund von fehlerhaften Ergebnissen solcher KI-unterstützen Untersuchungen gestiegen. Dies kann jedoch relativiert werden, insofern einerseits KI deutlich häufiger zum Einsatz kommt und andererseits nun auch vermehrt Fehler gemeldet werden.

Konkret macht Miklavčič für den Anwendungsbereich der KI in der Medizin vier große ethischen Herausforderungen aus, die aber auch in anderen Gebieten nützlich sein dürften:

Zum einen die Frage nach der Privatsphäre, also dem Umgang mit personenbezogenen Daten, die von einer KI auch ohne konkrete Zustimmung der Patient*innen aus den wenigen Daten, die freiwillig zur Verfügung gestellt wurden, erstellt werden können.

Zum anderen ist aufgrund des Trainingsdatensatzes, den Menschen der KI zu Beginn geben müssen, immer auch die Frage des Bias zu beachten. Welche Menschen zur Erstellung der Trainingsdaten herangezogen werden, kann die daraus entstandenen Diskriminierungen durch KI verstärken. So funktionierte ein KI-unterstütztes Hautkrebsscreening bei Menschen mit weißer Hautfarbe sehr gut, während sie bei Menschen mit dunkleren Hautfarben eine deutlich höhere Fehlerquote aufwiesen. Grund hierfür waren die Trainingsdaten der KI, die vor allem mit Datensätzen von Menschen mit weißer Hautfarbe befüllt waren. Auch kann man die geschlechtliche Dimension miteinbeziehen, dass immer noch die medizinische Forschung als Modell-Patient von einem Mann Mitte 30 ausgeht. Solche Vorurteile könnten durch KI reproduziert werden, oder sie könnte, wenn man dies bedenkt und bewusst dagegen programmiert, diese Bias überwinden.

Des Weiteren stellen die Transparenz und Erklärbarkeit KI-unterstützter Diagnosen eine Herausforderung dar. Wenn beim Hautkrebs-Screening die Erkennungserfolge der KI schon jetzt deutlich über den menschlichen Fähigkeiten liegen, sodass Dermatolog*innen die Ergebnisse nicht mehr nachvollziehen können, müssen sie der KI, auch bestärkt durch die hohe Erfolgsrate, im wahrsten Sinne des Wortes blind folgen. Etwaige Therapieentscheidungen könnten dann schwerer zu erklären sein, was wiederum das Vertrauensverhältnis auf Seiten der Patient*innen strapazieren könnte.

Zuletzt ist noch die Frage der Verantwortlichkeit bei Fehlern der KI zu klären. Dies verschärft sich umso mehr, wenn, wie oben beschrieben, KI die Fähigkeiten des medizinischen Personals dermaßen übertrifft, dass man sich fragen muss, wer letztlich wirklich die Therapieentscheidungen getroffen hat.

Daraus ergeben sich für Miklavčič drei große Probleme: Einerseits ist das mangelnde Vertrauen in KI insofern ein ethisches Problem, als dass die Nutzung von KI im medizinischen Bereich infrage gestellt wird, obwohl sie deutlich bessere Ergebnisse bei bestimmten Untersuchungen ermöglicht. Andererseits stehen die neuesten KI-Anwendungen nicht allen Menschen zur Verfügung, was generell aufgrund der Gewinnorientierung von KI-Unternehmen und medizinischen Unternehmen bei der Anwendung neuer medizinischer Technologien eine Herausforderung darstellt. Zuletzt ist durch den erhöhten Einsatz von Technologien und insbesondere von KI eine zunehmende Dehumanisierung von Care-Tätigkeiten zu beobachten und kritisch zu hinterfragen.

Für die Bearbeitung solcher ethischen Fragen schlägt Miklavčič folgende sieben Prinzipien vor:

1.) Respekt gegenüber der Privatsphäre und Datensicherheit: Patient*innen muss klar gemacht werden, welche Daten wie verarbeitet werden und welche Daten die KI generieren wird. Erst durch eine entsprechende Aufklärung kann hierzu ein informed consent erteilt werden.

2.) Unparteilichkeit und Fairness: Bias in den (Trainings-)Daten sind ernst zu nehmen, um entsprechend gegensteuern zu können.

3.) Transparenz: hängt eng mit der Unparteilichkeit zusammen, insofern offengelegt werden sollte, welche Trainingsdaten verwendet wurden.

4.) Klare Verantwortlichkeiten: wer ist für Fehler der KI bis zu welchem Grad verantwortlich?

5.) Vertrauen in KI-Systeme: ist aufgrund des enormen potenziellen und schon realisierten Nutzens von KI im medizinischen Bereich zu fördern.

6.) Allgemeine Zugänglichkeit: KI-Systeme sollten kein Luxus für Menschen mit besonderen Krankenversicherungen oder des globalen Nordens sein.

7.) Menschenzentriertheit: gegen eine Dehumanisierung von Care gilt es, Aufgaben des Pflegepersonals, die sie von der Arbeit mit Patient*innen abhalten, wie etwa die Dokumentation, von KI übernehmen zu lassen.

Zusammenfassend geht Miklavčič davon aus, dass Ärzt*innen nicht von KI ersetzt werden, aber Ärzt*innen, die KI-Anwendungen nutzen, werden jene ersetzen, die sie nicht nutzen. Auch er plädiert dafür, KI als Werkzeug und nicht als Ersatz zu verstehen. Bestimmte KI-Anwendungen sollten für bestimmte Aufgaben und zu bestimmten Zwecken genutzt werden. Bei all den vielen positiven Möglichkeiten, die sich durch KI ergeben, ist keine Angst vor KI, aber sicherlich eine gewisse Vorsicht angesagt.

Abschließende Gedanken zur BCE-Konferenz 2025

Mit diesen drei spannenden Vorträgen sind die Weichen gestellt für eine fruchtbare Konferenz von BCE im nächsten Jahr. Die Stimmung zwischen kritischer Neugier und Vorsicht, die während den Diskussionen bei der Kuratoriumssitzung vorherrschte, kann auch für die Konferenz leitend sein. KI ist schon jetzt eine Realität und die Anwendungsmöglichkeiten sind vielfältig. Sowohl im Bereich der Medizin, aber auch im Bereich der Bildung sollte KI als Werkzeug verstanden werden und nicht bloß als Ersatz.

Positiv kann man KI auch als Chance sehen, Althergebrachtes zu hinterfragen. Auch könnte man infrage stellen, ob der Wert der Effizienz, der als Hauptargument für den Einsatz von KI oft herangezogen wird, wirklich der einzige Leitwert sein sollte. Sollte nicht viel mehr die Menschenzentriertheit, wie sie Miklavčič als ein Prinzip vorschlägt, zumindest mit Effizienz zusammengedacht werden? Insofern könnte dann KI jene Bereiche, die einer Effizienzsteigerung an Menschlichkeit im Wege stehen (etwa Dokumentation im Pflegebereich) übernehmen, um dann mehr Zeit für die persönliche Patient*innenarbeit zu verschaffen. Einer Effizienz der Menschlichkeit würde es folglich widersprechen, wenn die frei gewordenen Arbeitsstunden zu einem Personalabbau oder zu einer entsprechenden Steigerung des Patient*innenaufkommens führen würde.

KI stellt uns also vor große und neue Herausforderungen. Die Konferenz im kommenden Jahr ist eine Chance, proaktiv und nicht mehr bloß reaktiv auf diese Herausforderungen zu reagieren. KI könnte auch eine Chance sein, bisher leitende ethische Prinzipien zu hinterfragen und dort nachzuschärfen, wo ein Mehr an Menschlichkeit nötig ist.