PAUL DRAGANOFF

Katholisch Theologische Fakultät der Universität Wien

<https://orcid.org/0009-0003-9147-5780>

paul.draganoff@univie.ac.at

KI – Zwischen Chancen und Gefahren

Bericht der BCE-Kuratoriumssitzung 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. Der thematische Fokus wird auf Künstlicher Intelligenz (KI) und den damit verbundenen ethischen Fragestellungen liegen. Als ersten Vorgeschmack und zur Angleichung des Wissensstands aller Mitglieder wurde eine Reihe spannender Vorträge zu diesem Thema organisiert. Nach einem sozialetischen Impuls zur Frage von KI und Bildung durch Alexander Filipovic, referierte Octavian Machidon über den aktuellen technischen Stand und mögliche zeitnahe Weiterentwicklungen von KI aus der Perspektive der Informatik. Zuletzt beleuchtete Jonas Miklavčič die ethischen Herausforderungen und Prinzipien im Umgang mit KI mit Fokus auf Anwendungsmöglichkeiten in der Medizin.

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

Filipovic stellt zunächst die zunehmende Bedeutung von KI in immer mehr gesellschaftlichen Bereichen fest und die damit verbundenen Fragen nach dem Selbst- und Weltverständnis des Menschen. 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 Assistenzsysteme heraus-

gefordert, mit denen in einer Qualität interagiert werden kann, sodass sie auf den ersten Blick menschlich erscheinen.

Diese grundsätzlichen anthropologischen Themen führen auch zur Frage der Bedeutung von Bildung, 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 von Anthropologie, Bildung und Ethik an, auf welche Bildungsinstitutionen (natürlich neben anderen Interessen wie der Finanzierung usw.) ausgerichtet sein sollten.

Haben schon Computer und das Internet Bildungsinstitutionen vor neue Herausforderungen gestellt, so stellt sich in diesem Zusammenhang die Frage nach deren Sinn und Zweck im Bezug auf KI-Anwendungen wie ChatGPT. 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, lediglich Inhalte, Fakten und Zahlen wiederzugeben? Ist die persönliche Reflexion über ein komplexeres Thema wirklich über eine wissenschaftliche schriftliche Arbeit, wie Seminar- oder Bachelorarbeit, abgebildet, wenn ChatGPT diese auf Knopfdruck erstellen kann? Es besteht also die Herausforderung, wie angesichts von KI-Anwendungen kritische, autonome sowie an Beteiligung und Mitgestaltung interessierte Menschen hervorgebracht werden können. Hier werden Autonomie, Beteiligung und Mitgestaltung als Bildungsziele von Filipovic definiert. Bildungseinrichtungen müssen sich folglich selbst hinterfragen, inwiefern deren Inhalte und Methoden diesen drei großen Zielen in 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 Lernschritt Einfluss nehmen zu können. Ein denkbarer Ersatz für die so agierende analoge Lehrkraft könnte folglich ein KI-Lehrsystem sein, welches 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, da sie Empfänger*innen von Lerninhalten

bleiben, da 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 kann verhindert werden, dass Ungerechtigkeiten/Ungleichheiten 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, sie weiter reproduzieren, sie verstärken oder gar neue hervorrufen.

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

Machidon erläuterte aus der Perspektive der Informatik, welche Konzepte sich hinter dem Begriff der Künstlichen Intelligenz verbergen. Ziel war es, ein vertieftes Verständnis darüber zu fördern, worum es bei Künstlicher Intelligenz tatsächlich geht und welche Missverständnisse in der aktuellen Diskussion über KI verbreitet sind. Künstliche Intelligenz meint allgemein einen Teilbereich der Informatik. Dabei wird zwischen *weak AI* und *general AI* unterschieden. *Weak AI* bezeichnet KI-Systeme, die für eine spezifische Aufgabe entwickelt wurden. Darunter fallen alle KI-Anwendungen, die uns im Alltag begegnen, wie etwa Sprachassistenten auf Smartphones oder selbstfahrende Autos. *General AI* hingegen beschreibt die Vision von KI-Systemen, die menschliche Fähigkeiten nicht nur imitieren, sondern sogar übertreffen und eigenständig agieren können – wie z.B. die fiktive Figur HAL aus 2001: *A Space Odyssey*. Um den aktuellen Stand der Technik präziser abzubilden, wird in der Informatik zunehmend eine alternative Nomenklatur diskutiert. Diese unterscheidet die übermenschlichen Fähigkeiten von *General AI* unter dem Begriff *Artificial Super Intelligence (ASI)*. Anwendungen, die menschliche Intelligenz nachahmen, wie etwa ChatGPT, werden als *Artificial General Intelligence (AGI)* klassifiziert, während *Artificial Narrow Intelligence (ANI)* für KI-Systeme verwendet wird, die nur spezifische Aufgaben bewältigen können. Ein entscheidender Schritt in der rasanten Entwicklung von KI-Systemen ist das sogenannte *Deep Learning*, das durch eine geringere Notwendigkeit menschlicher Eingaben charakterisiert ist. Im Kern dieses Prozesses stehen die sogenannten *Neural Networks* (neuronale Netzwerke), die aus einem vorgegebenen Datensatz durch eine Reihe von Algorithmen neue Verbindungen herstellen und so neue Informationen generieren. Dies zeigt sich in Anwendungen wie KI-gesteuerter Objekterkennung

(z.B. Face-ID), in der Medizin (z.B. Hautkrebs-Screening), aber auch in virtuellen Assistenten oder ChatGPT. Der Vorteil von KI-Systemen, die *Deep Learning* verwenden, liegt darin, dass der menschliche Input weitgehend auf die Bereitstellung eines umfangreichen Trainingsdatensatzes beschränkt bleibt. Nach einer erfolgreichen Testphase sind lediglich punktuelle Anpassungen der Parameter zur Datenauswertung notwendig, was bedeutet, dass der menschliche Einfluss auf den Prozess minimal bleibt.

Nach dieser technischen Einführung beleuchtete Machidon die verschiedenen Interessen, die hinter der Entwicklung von KI stehen. Zunächst ist es gemäß Moore's Law zu erwarten, dass sich die Anzahl der Transistoren auf Computer-Chips alle zwei Jahre verdoppelt, was zu einer entsprechenden Erhöhung der Rechenleistung führt. Diese Produktivitätssteigerung der Chips spiegelt auch die wirtschaftliche Forderung nach einer allgemeinen Produktivitätssteigerung wider. Dementsprechend gibt es massive finanzielle Anreize, die KI-Entwicklung zu fördern, nicht zuletzt aufgrund der Vielzahl an Anwendungsmöglichkeiten in nahezu allen Lebensbereichen. Besonders großes Potenzial wird dabei in Bereichen wie verbesserten Klimamodellen, der Entdeckung neuer medizinischer Behandlungsmethoden und im privaten Sektor, etwa bei *Smart Homes*, gesehen.

Dennoch gibt es auch weit verbreitete Ängste, die einerseits durch popkulturelle Darstellungen von KI und deren potenzieller Übernahme der Menschheit genährt werden, andererseits jedoch auch durch die Tatsache, dass ethische Debatten bislang oft hinter den technologischen Entwicklungen zurückgeblieben sind. In diesem Kontext stellt die für 2025 geplante Konferenz einen wichtigen Beitrag dar, um interdisziplinär und vor allem proaktiv – und nicht nur reaktiv – über die ethischen Implikationen dieser sich schnell entwickelnden Technologie nachzudenken. Den Abschluss bildete eine Einschätzung darüber, wie nahe wir der sogenannten Singularität sind – der Entwicklung einer selbstbewussten KI. Machidon ist der Ansicht, dass wir mit technischen Anwendungen wie generativer KI, beispielsweise ChatGPT, bereits sehr nahe an diesem Punkt sind. Viele Aufgaben können heute in einer Qualität erledigt werden, die mit menschlichen Fähigkeiten vergleichbar ist oder diese sogar übertrifft. Dennoch gibt es noch einige wesentliche Einschränkungen: Erstens hängen diese KI-Systeme stark von der Qualität des ursprünglichen Trainingsdatensatzes ab. Zweitens haben sie Schwierigkeiten, sich an geänderte Bedingungen oder unerwartete Szenarien anzupassen. Zudem bleibt fraglich, ob Systeme wie ChatGPT wirklich kreativ sein können, und auch die Frage des Datenschutzes ist noch nicht abschließend geklärt.

Machidon ordnete den aktuellen Stand der KI-Entwicklung in einer fünfstufigen Skala ein, wobei wir uns seiner Einschätzung nach bereits auf der dritten Stufe befinden. Seit einigen Jahren gibt es Chatbots, und viele KI-Anwendungen können mittlerweile Aufgaben ausführen, die menschliche Fähigkeiten nachahmen. Gegenwärtige KI-Systeme sind in der Lage, eigenständige Tätigkeiten zu übernehmen, das heißt, ohne ständigen menschlichen Input. In den beiden verbleibenden Stufen müsste KI jedoch in der Lage sein, echte kreative Erfindungen zu machen und schließlich ganze Organisationen zu ersetzen, die die Zusammenarbeit von Menschen mit unterschiedlichen Hintergründen koordinieren.

Angesichts dieser Entwicklungen wird es immer wichtiger, die Forschung im Bereich der KI weiter voranzutreiben, gleichzeitig aber auch ihre Regulierung zu gewährleisten. Ebenso muss ein Bewusstsein für die Auswirkungen von KI geschaffen werden, was sowohl eine stärkere Bildung als auch eine proaktive ethische Reflexion erfordert.

Jonas Miklavčič: *Ethics of Artificial Intelligence: Challenges and Principles*

Jonas Miklavčič setzte sich intensiv mit den ethischen Fragestellungen auseinander, die im Zusammenhang mit Künstlicher Intelligenz (KI) entstehen, und den Prinzipien, die zur Lösung dieser Herausforderungen herangezogen werden können. Das Themenfeld der KI-Ethik ist in der Technikethik angesiedelt und hat sich insbesondere durch die Entwicklung von Maschinenlernen und *Deep Learning* erweitert. Diese Technologien werfen neue ethische Fragen auf, etwa bezüglich der Vorhersagbarkeit von Ergebnissen, der Handhabung immer größerer Datenmengen sowie der Genauigkeit und Personalisierungsmöglichkeiten, die durch KI möglich werden.

Trotz vieler positiver Berichterstattung über KI-Anwendungen im medizinischen Bereich, etwa bei Augenuntersuchungen oder Hautkrebs-Screenings, werden misslungene Beispiele selten thematisiert. In den letzten Jahren sind Fehler und Unfälle aufgrund fehlerhafter Ergebnisse von KI-unterstützten Untersuchungen gestiegen. Dies lässt sich jedoch relativieren, da KI mittlerweile häufiger eingesetzt wird und Fehler daher auch verstärkt gemeldet werden.

Miklavčič identifizierte vier zentrale ethische Herausforderungen im Bereich der medizinischen KI, die auch auf andere Anwendungsgebiete übertragbar sind. Erstens geht es um die Frage der Privatsphäre und den Umgang mit personenbezogenen Daten. KI kann auch ohne direkte Zustimmung der Patient*innen Daten generieren, die ursprünglich freiwillig zur Verfügung gestellt wurden. Zweitens

muss der Bias, der durch die Trainingsdatensätze entsteht, berücksichtigt werden. Ein Beispiel ist das Hautkrebs-Screening, das bei Menschen mit heller Hautfarbe sehr gut funktioniert, bei Menschen mit dunklerer Haut jedoch eine höhere Fehlerquote aufweist. Der Grund dafür liegt in den Trainingsdaten, die überwiegend aus Datensätzen von Menschen mit weißer Hautfarbe bestanden. Auch geschlechtliche Vorurteile in der medizinischen Forschung, die immer noch den Mann mittleren Alters als Modell-Patienten annehmen, könnten durch KI verstärkt werden; diese könnten allerdings auch überwunden werden, wenn diese Verzerrung durch Programmierung korrigiert wird.

Ein weiteres Problem stellt die Transparenz und Erklärbarkeit der KI-unterstützten Diagnosen dar. Wenn KI-Systeme in der Lage sind, Hautkrebs mit einer Genauigkeit zu erkennen, die die menschlichen Fähigkeiten übersteigt, können selbst Dermatolog*innen die Ergebnisse nicht nachvollziehen. Dies könnte zu einem Vertrauensverlust bei den Patient*innen führen, da Therapieentscheidungen schwerer zu erklären wären. Schließlich stellt sich die Frage der Verantwortlichkeit bei Fehlern der KI, insbesondere wenn diese die Fähigkeiten des medizinischen Personals übertrifft und unklar bleibt, wer letztlich für die getroffenen Entscheidungen verantwortlich ist.

Für Miklavčič ergeben sich daraus drei wesentliche ethische Probleme: Erstens das mangelnde Vertrauen in KI, trotz ihrer nachweislich besseren Ergebnisse in bestimmten Bereichen, was die Nutzung von KI im medizinischen Sektor infrage stellt. Zweitens ist der Zugang zu den neuesten KI-Technologien ungleich verteilt, was zu einer Herausforderung wird, wenn KI-Systeme nur für bestimmte gesellschaftliche Gruppen zugänglich sind. Drittens wird durch den vermehrten Einsatz von KI eine Dehumanisierung in der Pflege erkennbar, die es zu hinterfragen gilt.

Zur Lösung dieser ethischen Herausforderungen schlägt Miklavčič sieben Prinzipien vor:

1. Respekt gegenüber der Privatsphäre und Datensicherheit: Patient*innen müssen über die Verwendung ihrer Daten und die von der KI generierten Daten informiert werden; ein *informed consent* muss eingeholt werden.
2. Unparteilichkeit und Fairness: Bias in den Trainingsdaten muss erkannt und korrigiert werden, um Diskriminierungen zu vermeiden.
3. Transparenz: Es muss offengelegt werden, welche Trainingsdaten verwendet wurden, um Transparenz und Unparteilichkeit zu gewährleisten.
4. Klare Verantwortlichkeiten: Es muss definiert werden, wer im Falle von Fehlern der KI verantwortlich ist.

5. Vertrauen in KI-Systeme: Aufgrund des großen Potenzials der KI im medizinischen Bereich muss Vertrauen aufgebaut werden.
6. Allgemeine Zugänglichkeit: KI-Systeme sollten nicht nur den Wohlhabenden oder Menschen im globalen Norden zugänglich sein, sondern für alle verfügbar.
7. Menschenzentriertheit: Die Dehumanisierung von Care-Tätigkeiten sollte vermieden werden, indem KI die Aufgaben übernimmt, die Pflegepersonal von der direkten Arbeit mit Patient*innen abhalten, wie etwa die Dokumentation.

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. Er plädiert dafür, KI als Werkzeug und nicht als Ersatz zu verstehen. Bestimmte KI-Anwendungen sollten für spezifische Aufgaben und Zwecke eingesetzt werden. Angesichts der vielen positiven Potenziale von KI hält Miklavčič eine gewisse Vorsicht für angebracht, ohne jedoch eine grundsätzliche Angst vor KI zu schüren.

Abschließende Gedanken zur BCE-Konferenz 2025

Mit diesen drei spannenden Vorträgen sind die Weichen für eine fruchtbare Konferenz von BCE im Herbst 2025 bereits gestellt. Die Atmosphäre, die während der Diskussionen bei der Kuratoriumssitzung zwischen kritischer Neugier und Vorsicht herrschte, könnte auch als richtungsweisend für die Konferenz dienen. Künstliche Intelligenz ist längst eine gelebte Realität, und ihre Anwendungsmöglichkeiten sind vielfältig. Sowohl im medizinischen Bereich als auch in der Bildung sollte KI vor allem als Werkzeug verstanden werden – nicht als bloßer Ersatz.

Es bietet sich an, KI auch als Chance zu begreifen, bestehende Normen und Arbeitsweisen zu hinterfragen. In diesem Kontext könnte man überlegen, ob der Effizienzgedanke, der häufig als Hauptargument für den Einsatz von KI angeführt wird, wirklich der alleinige Maßstab sein sollte. Sollte nicht vielmehr die Menschenzentriertheit, wie sie Miklavčič als grundlegendes Prinzip vorschlägt, als gleichwertiger Wert neben der Effizienz gedacht werden – als Effizienz der Menschlichkeit? KI könnte gerade in Bereichen, in denen Effizienzsteigerung auf Kosten der Menschlichkeit geht – wie etwa der Dokumentation im Pflegebereich – die nötige Entlastung schaffen, um mehr Zeit für die persönliche Arbeit mit den Patient*innen zu gewinnen. Einer so gedachten Effizienz der Menschlichkeit wider-

spräche es, wenn die gewonnene Zeit als Argument für Personalabbau oder einer Erhöhung des Patient*innen-Aufkommens erhalten müsste.

Künstliche Intelligenz fordert uns somit mit großen und neuen Herausforderungen heraus. Die Konferenz im kommenden Jahr stellt eine wertvolle Gelegenheit dar, proaktiv an diese Herausforderungen heranzugehen, anstatt nur reaktiv zu reagieren. Sie könnte zudem ein Moment sein, etablierte ethische Prinzipien zu hinterfragen und dort nachzuschärfen, wo mehr Menschlichkeit gefragt ist.