



Nullen und Einsen – Ethik und Jugendschutz

Peter Lähn im Gespräch mit Michael Mayer

Welches Weltbild steckt in künstlicher Intelligenz (KI)? Auch dem Jugendschutz stellt sich die Frage, in welchem Verhältnis menschliche Urteilskraft und algorithmische Beurteilung zueinander stehen. Der Berliner Medienphilosoph Prof. Dr. Michael Mayer forscht derzeit am Institut für Theorie (ith) der Zürcher Hochschule der Künste (ZHdK) zu Fragen des Verhältnisses künstlicher und künstlerischer Kreativität und gibt vorläufige Antworten und Einschätzungen.

In der Fantasieproduktion ist künstliche Intelligenz seit jeher präsent: Über mittelalterliche Mythen und zu industrieller Gegenwartskultur bis in die nahe Zukunft bereichern Vorstellungen von KI die Kulturgeschichte: Nun verwirklichen Nullen und Einsen nicht nur die Fantasieproduktion. Wie ist das Verhältnis analog/digital medienhistorisch einzuordnen?

Theoretisch wurde Digitalität im mathematischen Verfahren einer binären Algebra im 18. und vor allem im 19. Jahrhundert vorgedacht und kann bis zu dem deutschen Philosophen Gottfried Wilhelm Leibniz zurückverfolgt werden. Es wurde eine Mathematik entworfen, die auf ihren bislang natürlichen Referenten verzichtet, nämlich auf ein ausgewiesenes Zahlensystem, wie wir es

beispielsweise von Null bis Zehn kennen. Digitalität führt im Grunde genommen ihre gesamten mathematischen Operationen in einer Opposition von Eins oder Null, An oder Aus, Energie oder Nichtenergie durch. Aber erst mit dem Stromkreislauf der Elektrizität und schließlich mit der Erfindung des Computers wurde möglich, etwas auf den Weg zu bringen, was man mit dem Begriff der Digitalität heute, einer digitalen Kultur verbinden kann.

Das Moment, das alles zum Laufen bringt, ist doch der Algorithmus, also gewissermaßen der Grundbaustein digitalen Geschehens?

Mit einem gewissen Selbstverständnis gehen wir davon aus, dass Algorithmen mit technischer Kultur notwendig etwas zu tun haben. Algorithmen aber sind im Grunde genommen mindestens so alt wie die Mathematik, wie formale Logik überhaupt. Unter Algorithmus kann man eine Entscheidungs- oder Handlungsroutine verstehen. Sie ermöglicht mir, unabhängig vom jeweiligen materiellen Input ganz bestimmte Entscheidungsprozesse in Einzelschritte zu zerlegen, zu standardisieren und zu automatisieren. Automatisierung ist das, was wir heute auch bei der sogenannten künstlichen Intelligenz und selbstlernenden Maschinen diskutieren. Aber Algorithmen als standardisierte Problemlösungsstrategien sind als Verfahren innerhalb der Mathematik tatsächlich schon sehr lange bekannt. Der Ausdruck selbst soll übrigens auf einen persischen Gelehrten und Mathematiker des 9. Jahrhunderts n. Chr. verweisen, dessen für westliche Zungen schwer aussprechbarer Name latinisiert Algorismi lautete.

Der Vater der KI, Norbert Wiener, verhielt sich ihrer Entwicklung gegenüber schon nach kurzer Zeit sehr kritisch. Als die Forschung zu KI immer greifbarer wurde, sah er „einen bedrohlichen neuen Faschismus“ aufkommen, „der von der Maschine als Regent bedingt ist“. Ist er vielleicht ein früher Prophet dystopischer Fantasieproduktion?

Norbert Wiener hat sich ja in verschiedenen Zusammenhängen kritisch zu jener Wissenskultur, die er mit hervorgebracht und geprägt hat, gestellt. Tatsächlich ist dieses Bonmot von Wiener bereits 1948 gefallen, in seinem Buch *Cybernetics*. Die Warnung, die er hier aus-

spricht und die nur kurz nach dem Ende des Zweiten Weltkrieges auch nachvollziehbar ist, zeigt einerseits ein gutes Gespür für ein Problem, das sich hier auftut. Andererseits aber zeigt sie auf eine fast verstörende Weise eine Unfähigkeit, Begriffe und Ausdrücke zu entwickeln, die dieses Problem tatsächlich beschreiben. Vielleicht sollte man statt von „Faschismus“ von einer Art „digitalem Totalitarismus“ sprechen, verstanden als Implementierung einer techno-ökonomischen Kultur der totalen Transparenz, Kontrolle und Rechenbarkeit sozialer Interaktionen. Unsere Gedanken, Wünsche und intimsten Regungen werden permanent durchleuchtet und bewertet.

Hat die künstliche Intelligenz ein eigenständiges Weltbild oder ist dieses letztendlich dadurch geformt, was wir dieser Maschine in Auftrag geben?

Man kann tatsächlich von einer Art Weltbild des Digitalen sprechen. Sie geht einher mit einer Wertigkeit und Wertsetzung. Das Digitale gilt per se als das Neue, das Fortschrittliche, das Hippe. Das ganze Projekt der Digitalisierung wird mit der Aura eines Aufbruchs ins Neue aufgeladen, in eine neue Zeit atemberaubender Möglichkeiten und Gelegenheiten. Demgegenüber erscheint das sogenannte Analoge als zurückgeblieben, rückschrittlich und minderwertig – als Rest, den es nach dem Vorbild des Digitalen zu optimieren gilt. Etwas Ähnliches zeigt sich bei der künstlichen Intelligenz. Wenn man von einer künstlichen Intelligenz spricht, stellt man sie implizit bereits in Opposition und schließlich in Konkurrenz zur natürlichen Intelligenz, der Intelligenz von uns Menschen. Ist künstliche Intelligenz aber tatsächlich *Intelligenz* im Sinne des Lesens und Verstehens von Zusammenhängen, der Einsicht in sie? Oder ist die künstliche Intelligenz doch etwas ganz anderes, dessen radikale Fremdheit wir mit Begriffen verdecken, die uns vertraut scheinen? Vielleicht bekommen wir diese Fremdheit erst in den Blick, wenn wir auf solche binären Operatoren wie digital – analog oder künstlich – natürlich verzichten und damit auf einen Vergleich von per se Unvergleichlichem.

Wir kennen ja Beschränkungen des Handelns durch das, was wir Ethik nennen. Kann man einer Maschine unterstellen, dass sie ethisch funktioniert oder ethisch handeln kann?

Diese Frage stellt sich u. a. bei durch KI selbstfahrenden Autos und berührt ethische Probleme. Die KI bräuchte ethische Entscheidungsrouninen, die dann in Alltagssituationen, womöglich in Konfliktsituationen, in dramatischen Situationen tief greifende Entscheidungen treffen müssen. Die Frage ist natürlich keine Science-Fiction mehr. Was geschieht, wenn ein selbstfahrendes autonomes Transportsystem einen Unfall produziert? Wer ist dann dafür verantwortlich, der Programmierer, der Hersteller oder die Insassen des selbstfahrenden Autos?

Könnte das dann nicht wiederum ein Algorithmus bestimmen?

Das wäre die Frage nach einer algorithmisch automatisierten moralischen Urteils kraft: Was aber ist Urteils kraft? Was bedeutet es, ein moralisches Urteil zu fällen? Was bedeutet es, wenn wir die Idee des moralischen Urteils tatsächlich als etwas beschreiben, das durchgängig formalisiert zu werden vermag? Was haben wir da bereits stillschweigend vorausgesetzt, was haben wir unterschwellig bereits entschieden? Und was muss eigentlich, wenn man diesen Zusammenhang zu thematisieren versucht, dringend reflektiert werden? Ein moralisches Urteil ist niemals unabhängig der jeweiligen Kontexte und ist niemals sinnvoll ohne die genaue Betrachtung des Einzelfalles. Wir haben bei den moralischen Urteilen, gerade in Konfliktsituationen, manchmal Situationen, in denen wir als Menschen Entscheidungen treffen müssen, ohne sie noch ausreichend begründen zu können. Damit müssen wir leben. Das ist die *Conditio humana*. Das Projekt einer künstlichen Intelligenz scheint uns hier die Vision einer künstlich generierbaren ethischen Urteils kraft auf der Basis allgemein verbindlicher, fixer und gültiger Urteilsregeln zu offerieren. Ich halte diese Vision für fragwürdig. Und ich bin mir sicher, dass bereits der Versuch ihrer Realisation für tief greifende Verwerfungen sorgen wird.

Norbert Wiener unterschied auch zwischen guter und böser KI. Die Automatisierung hilft beispielsweise, riesige Datenmengen in den Griff zu bekommen. So wird die Fahndung nach Kinderpornografie im Netz durch Algorithmen erst erfolgreich intensiviert. Algorithmen erleichtern hier menschliche Arbeit, indem sie vorsortieren oder aussortieren, sodass sich die Strafverfolgung auf die eigentlichen Taten konzentrieren kann.

Selbstverständlich sind die positiven Auswirkungen enorm – und faszinierend. Es geht nicht darum, kritische Einwände dergestalt zu pflegen, um kulturkonservativ ein Hohelied auf das Analoge zu formulieren. Es geht auch nicht darum, die Idee der algorithmischen Mathematisierbarkeit in der wissenschaftlich-technischen Kultur in Bausch und Bogen zu verdammen. Eher ist zu fragen, wo die Grenzen einer algorithmischen Rationalität sind. Was wir bräuchten, wäre eine Kritik der algorithmischen Rationalität, eine Kritik im kantischen Sinne als Frage nach der Reichweite der algorithmischen Rationalität. Am Beispiel der Kinderpornografie helfen algorithmische Routinen enorm bei der Sichtung des Materials, aber sie entscheiden nicht, ob letztlich ein Straftatbestand vorliegt. Die menschliche Urteils kraft ist hier entscheidend, sie auszuschließen, wäre nicht nur im juristischen Sinne fragwürdig.

Sehen Sie ähnliche Beschränkungen, wenn Jugendschutzkriterien automatisiert werden?

Um die Aufgabe des Jugendschutzes, Entscheidungen zu fällen, welche Filme oder Medien wie einzuordnen sind, beneide ich niemanden. Sie ist und bleibt schwierig. Wenn wir aber glauben, die entsprechenden Entscheidungskriterien algorithmisch formalisieren zu können, um damit maschinell letztinstanzlich entscheiden zu lassen, was wie jugendschutzwürdig ist und was nicht, sehe ich mehrere Probleme auf uns zukommen. Zum einen haben wir es mit der Frage der moralischen Urteils kraft zu tun, die die Kontextualisierung und Reflexion des Einzelfalles erfordert, der nicht immer bruchlos unter eine allgemeine Regel subsumiert werden kann. Zum anderen sehe ich hier so etwas wie eine Performativität hinter unseren Rücken am Werke. Das heißt, algorithmische Entscheidungsrouninen schlagen auf unsere eigene Urteilsfähigkeit und Rezeption wie auch auf unsere eigene Produktivität zurück. Wenn eine Filmregisseurin oder ein Filmregisseur vermeiden will, unter algorithmische Verbotsschranken zu fallen, die das eigene Produkt negativ bewerten könnten, könnte das eine Art von vorseilender Anpassung an algorithmisch vorgegebene Standards vorbereiten, die peu à peu die künstlerische Freiheit beschneidet. So könnte etwa entschieden werden, eine offensichtlich oder vermeintlich als Gewaltdarstellung eindeutig identifizierbare und durch automatisierte Suchsysteme entsprechend eindeutig

rubrizierbare Szene zu streichen, unabhängig ihres Sinns, ihrer ästhetischen und auch moralischen Funktion, die erst gar nicht mehr diskutiert wird. Im Grunde genommen müssen aber aus jedem Film selbst heraus die Kriterien seiner Beurteilung und Kritik entwickelt werden – was im Falle des Jugendschutzes, wo es um genuin moralische und auch juristische Fragen geht, nicht eine Person allein machen kann. Was den Jugendschutz nicht nur institutionell, sondern wesentlich auszeichnet, ist, dass die Prüferinnen und Prüfer aus unterschiedlichen sozialen, kulturellen und politischen Milieus kommen. Sie setzen sich miteinander über die Sache auseinander und suchen im Streitgespräch einen Konsens. Zu glauben, dass diese schwierigen Fragen auf der Ebene einer algorithmischen Standardisierung entschieden werden können, halte ich, um ein fast naives Wort zu verwenden, für fahrlässig.

Kann Deep Learning eventuell weiterhelfen?

Ich möchte und kann hier nicht vertiefen, was unter Deep Learning eigentlich verstanden werden kann. Nur so viel: Bei Deep Learning habe ich ein ähnliches Problem wie bei der künstlichen Intelligenz auch. Wir setzen bereits einen Begriff des Lernens voraus und glauben, dass maschinelle Intelligenz nicht nur intelligent, sondern auch zum Lernen in der Lage sei. Dass es da mittlerweile ganz erstaunliche Leistungen gibt, ist gar nicht die Frage, vielmehr, ob wir das als Lernen beschreiben können. Das ist etwas, was wir generell erst einmal thematisieren müssten. Was ist überhaupt Lernen und wie verändert sich das Lernen im Kontakt mit digitalen Medien? Ich würde dafür plädieren, für all diese Phänomene wie Deep Learning, künstliche Intelligenz, künstliche Kreativität etc. andere Begriffe zu suchen. Provisorien, um uns klar zu werden, dass wir es mit einer Fremdheit zu tun haben, die fundamental ist. Digitale Medien treten uns als etwas gegenüber, das mit nichts, mit dem wir bislang konfrontiert waren, vergleichbar ist. Und diese Fremdheit sollten wir nicht unter der Hand dadurch nivellieren, dass wir mit Begriffen wie Lernen, wie Intelligenz oder dergleichen operieren. Es sind bereits begriffliche Vorentscheidungen und theoretische Vorannahmen getroffen, die wir stillschweigend akzeptieren. Sie müssten aber erst einmal für sich selbst betrachtet und analysiert werden.

Die Geschichte des Jugendschutzes ist auch eine Geschichte der Fortentwicklung von Beurteilungskriterien, die sich in immer kürzeren Zeitabständen teilweise massiv verändert haben. Wie finden gesellschaftliche Entwicklungen ihren Niederschlag in algorithmischen Beurteilungen?

Der Anspruch der künstlichen Intelligenz und der ganzen digitalen Kultur ist im Grunde genommen in einem Werbeslogan einer bekannten Suchmaschine enthalten. Da soll einmal mit dem Slogan geworben worden sein, dass man die Antworten auf Fragen habe, von denen der User noch nicht einmal wisse, dass er sie haben werde. Eine algorithmische Rationalität scheint mit dem Anspruch verbunden, prognostizieren zu können, was wir in naher Zukunft denken, fühlen und fragen werden. Kann es aber sein, dass diese algorithmische Rationalität solche Entwicklungen deshalb prognostizieren kann, weil sie sie initiiert? Ist diese algorithmische Rationalität tatsächlich ein neutrales Instrumentarium, das in irgendeiner Weise prognostische Aussagen über kommende gesellschaftliche Entwicklungen, soziale Wertpräferenzen und politische Einstellungen erlaubt? Oder ist sie ein Medium, das seine Message ist, sprich: diese Entwicklungen ins Werk setzt? Eine solche Frage sollten wir nicht nur dystopisch gestimmten Sci-Fi-Autorinnen und Autoren überlassen. Wir müssen sie sehr ernst nehmen.