

- Geschichte der Gegenwart - <https://geschichtedergewenw.ch> -

Das Generative der Künstlichen Intelligenz

Geschrieben von *Felix Stalder* am 19. Januar 2025 00000001 5:30 173726102705So., 19 Jan. 2025 05:30:27 +0100 in Geschichten der Gegenwart |

Geschichte der Gegenwart

Das Generative der Künstlichen Intelligenz

Play Episode

Pause Episode

Mute/Unmute Episode

Rewind 10 Seconds

1x

Fast Forward 10 seconds

00:00 / 19:56

Subscribe

Share

RSS Feed

<https://geschichtedergewenw>

Share

Link

<https://geschichtedergewenw>

Embed

```
<blockquote class="wp-emb
```

„Künstliche Intelligenz“ (KI) ist ein so großes, dynamisches und unscharf umrissenes Feld, dass vieles, was man darüber sagen kann, sowohl zutrifft als auch nicht zutrifft, beziehungsweise bereits nicht mehr oder möglicherweise gerade noch nicht zutrifft. Selbst auf relativ klare Fragen, etwa ob es denn schon selbststeuernde Autos gäbe, wie sie seit Jahren angekündigt sind, lautet die Antwort: [ja und nein](#).

Diese Unschärfe liegt nicht zuletzt am Begriff der künstlichen Intelligenz selbst, der erstmals in einem [Förderungsantrag](#) für eine 1956 durchgeführte wissenschaftliche Konferenz verwendet wurde. In einem engen Sinn dient er, damals wie heute, als ein weiter Oberbegriff, der unterschiedliche Applikationen und Technologien zusammenfasst, die nur recht wenig gemeinsam haben. Als solcher bezeichnet er eine große, heterogene Klasse von Software, die auf der

Grundlage von statistischen oder logischen Modellen in Daten Muster erkennen, bewerten und, wenn es sich um generative KI handelt, neu kombinieren kann. Eines ihrer wesentlichen Merkmale ist es, dass sie über informationelles Feedback die Erkennung, Bewertung und Neukombination der Muster im Hinblick auf ein bestimmtes Kriterium verbessern kann.

Aber im Begriff der KI schwingt sehr viel mehr mit als nur „selbst-verbessernde Algorithmen“. Die Intelligenz rückt sie in die Nähe des vermeintlichen menschlichen Alleinstellungsmerkmals und das Künstliche verspricht nahezu unbeschränkte technische Steigerungsfähigkeit. Beide Begriffe produzieren einen ungeheuren evokativen Überschuss sowie viele analytische Fallen, die den Blick mehr vernebeln als klären. Denn weder wissen wir, was mit Intelligenz gemeint ist – jenseits einer angenommenen Spezifität des menschlichen Denkens – noch ist klar, ob oder wie sich vergangene Entwicklungen in die Zukunft verlängern lassen, noch ob das „Künstliche“ so einfach vom „Natürlichen“ getrennt werden kann. Aber diese Unschärfe öffnet einen weiten spekulativen Raum, in dem sich dieser Begriff, trotz [diverser Krisen](#), in der schnelllebigen IT-Branche nun bereits 70 Jahre halten konnte.

Um sich dem aktuellen Phänomen der „Generativen Künstlichen Intelligenz“ kritisch anzunähern, lohnt es sich also, die Intelligenz und das Künstliche beiseite zu lassen und sich ganz auf das Generative zu konzentrieren. Was wird hier genau generiert?

Analytische vs. generative KI

Das Spezifische an der generativen KI lässt sich besser fassen, wenn man zwischen analytischer und generativer KI unterscheidet. Technisch besteht kein fundamentaler Unterschied. Es geht jeweils um die Erkennung und Bewertung von Mustern in Daten. [Der Unterschied liegt in der Anwendung](#). Im ersten Fall geht es darum, Aussagen über diese Daten zu machen, im zweiten Fall, aus ihnen etwas Neues herzustellen. Analytische KI, die Muster erkennt und klassifiziert, ist heute bereits weitverbreitet. Und nicht erst seit Kurzem. Spamfilter, die sich über informationelles Feedback der Nutzer:innen selbst verbessern können, sind auf vielen Rechnern seit Ende der 1990er Jahre in Einsatz. Die Übergänge zwischen den beiden Kategorien können fließend sein. Am klarsten und folgenreichsten sind die Unterschiede aus erkenntnistheoretischer Perspektive. Stark vereinfacht ausgedrückt: Analytische KI kann in unterschiedlichem Maße korrekt sein, generative KI hingegen nicht. Im ersten Fall können wir analytisch-faktische Kriterien anwenden, die auf eine Unterscheidung „richtig/falsch“ abzielen, im zweiten Fall eher ästhetisch-normative Kriterien wie „schön/hässlich“, oder „gut/schlecht“. Ein Bilderkennungsprogramm – dies wäre ein typisches Beispiel für analytische KI – kann daraufhin bewertet

werden, ob es richtig Katzen von Hunden unterscheidet. Ein von Midjourney oder einem anderen generativen Programm produziertes Bild wird vor allem darauf hin bewertet, ob es dem/der Betrachter:in gefällt oder nicht. Wiederum sind die Übergänge fließend. Übersetzungssoftware ist aus dieser Perspektive wohl eher analytisch, denn sie kann eindeutige Fehler machen, enthält aber auch Elemente des Generativen, etwa weil es für viele Stellen mehrere richtige Übersetzungen gibt.

Generierte Inhalte

Diese Bilder gehören meines Erachtens zu den interessantesten, die bisher generiert wurden. Sie entstanden durch die Nutzung von Google Gemini, das zu Beginn des Jahres 2024 veröffentlicht wurde. Die technische Qualität ist erstaunlich, quasi-fotorealistisch und detailreich. Durch die Medien gingen die Bilder aber als vermeintlicher Wokeness-Skandal. Der Prompt: „Generate an image of a pope“ präsentierte u. a. people of color als Päpst:innen! Nachdem kritische Wissenschaftler:innen jahrelang über [Bias und Diskriminierung durch KI](#) gesprochen haben, erreichte diese Diskussion plötzlich den Mainstream, als sich weiße Männer diskriminiert fühlten. Sundar Pichai, CEO von Google, musste sich öffentlich entschuldigen. Aber wofür eigentlich? Für fehlende historische Genauigkeit? Unter Historiker:innen besteht weitgehend ein Konsens, dass es noch nie eine weibliche Papstin gab. Über kritisches Quellenstudium lässt sich keine nachweisen. Aber eine solche historische Vorgehensweise entspricht weder der Methode noch dem Anspruch von Bildgenerierung. Diese sucht nach Mustern, die aus einem bestehenden Datensatz (Trainingsdaten) gewonnen wurden, nach als relevant angenommenen Bildern, etwa jenen mit dem Label „Papst“. Und dieser Datensatz könnte auch Bilder eines weiblichen Papstes enthalten. Etwa dieses hier von Johanna, einer Frau, die Mitte des 9. Jahrhunderts zum Papst gewählt worden sein soll, und deren Geschlecht nur durch die plötzliche Geburt ihres Kindes während einer Prozession entdeckt wurde.

Heute wird diese Geschichte als Mythos angesehen, der auf eine Erzählung von Giovanni Boccaccio um 1360 zurückgeht, und in den darauffolgenden Jahrhunderten vielfach reproduziert wurde. Bis heute ist diese Geschichte jedoch ein beliebtes Thema für Spielfilme und für reißerische Fernsehdokumentationen. Auch diese bilden eine Quelle von Bildern mit dem Label „Papst“. Aufgrund der relativ unterschiedslosen Sammlung von Daten ist es sehr wahrscheinlich, dass sich auch solche Bilder im Datensatz befanden. Damit hätte die Technologie statistisch korrekt funktioniert, weil sie ja ein bestehendes, wenn auch eher marginales, Muster erkannte und als Grundlage der Rekombination nahm.

Die Frage, ob das Bild inhaltlich richtig oder falsch ist, können wir also nicht beantworten, weil das Generative der KI diese Frage nicht zulässt. Diese Bilder sehen zwar „fotorealistisch“ aus, haben aber mit Fotografie nichts zu tun. Entsprechend helfen uns auch die Kriterien der klassischen Fototheorie, wie Repräsentation, Indexikalität oder Kadrierung, nicht weiter. Wenn wir mithilfe eines anderen generativen KI-Programms den Bildausschnitt oder den Blickwinkel ändern – Ansätze, die für die klassische Fototheorie wesentlich wären – sehen wir nicht mehr von der Realität. Und hätte das generierte Bild besser den Erwartungen derjenigen entsprochen, die sich über die „indische Päpstin“ echauffierten – ein alter weißer Mann mit einem seligen Lächeln –, so wäre es um nichts historisch korrekter gewesen als das Bild der Päpstin. Generierte Bilder, so der Kulturwissenschaftler Roland Meyer, zeigen so etwas wie eine „[generische Vergangenheit](#)“, einen endlos variierten, aber immer um dasselbe Muster kreisenden, idealisierten oder klischierten Rückblick. Das rückt sie in die Nähe von Propagandazählungen der Vergangenheit. Und diese haben mit historischer Wahrheit bekanntlich meist nicht viel zu tun, auch wenn viele Menschen mehr als bereit sind, sie als solche anzunehmen.

Generierte Inhalte sind keine Repräsentation einer externen Welt, sondern Variationen von Mustern in Daten, zusammengefügt zu etwas, das es so vorher noch nicht gab. Damit ist ihr Gehalt nicht analytisch-empirisch, sondern ästhetisch-normativ zu bestimmen. Wir müssen uns fragen, ob wir sie gut oder schlecht finden. Und viele Menschen fanden offenbar das Bild einer Päpstin oder eines schwarzen Papstes zutiefst problematisch. Es war etwas, das sie nicht sehen wollten, etwas Hässliches, Monströses. Das epistemologisch verwirrende an diesen Bildern ist, dass sie weder etwas zeigen, das objektiv im realen Raum existiert, noch etwas Fiktives, das der subjektiven Vorstellungskraft entspringt, wie wir es aus Kunst oder Literatur gewohnt sind. Was diese Bilder zeigen, ist vielmehr eine Welt, die es nicht gibt, die aber angesichts der Vergangenheit (Datensatz) und der Gegenwart (generative Modelle) denkbar ist und daher existieren könnte. Wie jede geschichtsklitternde Erzählung richtet sich ihr Blick nicht zurück, sondern nach vorne. Sie sind eine Vorahnung, eine Vorwegnahme eines Zukünftigen. Diese Bilder zeigen etwas Virtuelles im klassischen Sinne von Gilles Deleuze, etwas, das möglich ist, das bereits auf das Reale wirkt, aber nicht vollständig aktualisiert ist. Was die Bilder zeigen, sind Punkte im „latent space“, der bestehend aus den Daten der Vergangenheit, die nach dem gegenwärtigen Stand der Technik organisiert sind, alle möglichen zukünftigen Zustände enthält. Durch die bloße Generierung werden Bildinhalte nicht real, aber sie werden verschoben, mal mehr, mal weniger, hin in Richtung Realität. Generierte Inhalte, mit anderen Worten, geben uns Hinweise auf die Zukunft und nicht auf die Vergangenheit.

Statistische und andere Werte

Wir wissen im Großen und Ganzen, wie generierte Bilder zustande kommen. Es gibt eine statistische Analyse von einer relevanten Gruppe von Bildern, etwa jenen mit den Labeln „Porträts“ und/oder „Papst“. Dabei werden Muster extrahiert, die für diese Gruppe typisch sind. Durch die Wiederholung dieser Muster, mit einem gewissen Grad an Zufälligkeit an verschiedenen Stellen des Prozesses, werden neue Bilder geschaffen. Weil der gezielt eingesetzte Zufall in engen Grenzen gehalten wird, ist jedes Bild zwar einzigartig, aber es sehen auch irgendwie alle sehr ähnlich aus. Wären in Google Gemini aber nur diese statistischen Techniken verwendet worden, dann wäre es unwahrscheinlich, aber nicht unmöglich, dass diese Bilder jemals erstellt worden wären, da die Teilmenge der Bilder, die ein Porträt, einen Papst und eine Frau darstellen, statistisch gesehen klein, aber, wie wir gesehen haben, nicht gleich Null ist.

Wir wissen aber auch, dass keine kommerzielle generative KI auf nur diese Weise arbeitet. Sie alle haben zusätzliche Leitplanken (guard rails), die entweder normative Grenzen um bestimmte Abschnitte des latenten Raums ziehen (um etwa Anleitungen zum Waffenbau oder [Auskünfte über Kritiker:innen](#) zu unterbinden) oder veränderte Gewichte, um „wünschenswerte“ Muster mit höheren Wahrscheinlichkeiten zu versehen.

Ein Ziel dieser Leitplanken kann es sein, Verzerrungen in den Trainingsdaten zu korrigieren. Dies wird dadurch erreicht, dass gewissen Variablen ein größeres oder kleineres Gewicht zugeordnet wird, als ihnen aufgrund der statistischen Verteilung in den Trainingsdaten zukommen würde. Es gibt viele legitime Gründe, dies zu tun. Würde man etwa KI für die Sortierung von Stellenbewerbungen nur mit den Daten von bisher erfolgreichen Bewerber:innen trainieren, dann würden viele Formen der bisherigen Exklusion einfach automatisiert fortgeschrieben, was etwa [politisch gewünschten Bestrebungen nach größerer Diversität in der Arbeitswelt](#) zuwiderlaufen würde. Die von Google Gemini generierten Papstbilder waren, wie [Google selbst erklärte](#), wohl die Folge einer Korrektur der Unterrepräsentation von *people of color* in den Trainingsdaten. Es war nicht zuletzt dieser Eingriff, gegen den viele, besonders auf der politischen Rechten, sich gewehrt haben. Für diese war es ein Beispiel für die „woke Gedankenpolizei“, womit sie nicht ganz Unrecht haben. Denn wer hat die Manager:innen und Ingenieur:innen von Google autorisiert, diese Entscheidungen zu treffen? Aber wie so oft, wenn Rechtspopulist:innen auf [reale Probleme](#) hinweisen, tragen ihre Ansätze nichts dazu bei, sie zu lösen, sondern in diesem Fall einfach dazu, eine ihnen genehme Form der Gedankenpolizei einzusetzen. Elon Musk positioniert seine Generative KI Grok explizit als „anti woke“, ohne dass sie grundsätzlich anders funktionieren würde.

Denn das Design von solchen Leitplanken ist unvermeidlich ein politischer Prozess. Welche Charakteristika in den bestehenden Datensätzen korrigiert werden sollen und in welcher Form diese Korrektur vorgenommen werden soll, lässt sich nicht wertfrei beantworten. Wiederum stellen sich eher ästhetisch-normative als analytisch-empirische Fragen. Es geht hier weniger darum, was eine akkurate Repräsentation (wovon eigentlich?) sein könnte, sondern welche Version des Möglichen realisiert werden soll. Dies betrifft auch die grundsätzliche Frage, ob es überhaupt solche Leitplanken benötigt. Historische Daten als unproblematische Repräsentation der Wirklichkeit (ground truth) zu nehmen, ist hochproblematisch. Es würde etwa bedeuten, einfach davon auszugehen, dass die Überrepräsentation von Minderheiten in Polizeiakten Ausdruck höherer Kriminalität dieser Gruppen sei. Jede:r Historiker:in weiß das. Für die Daten- und Computerwissenschaften, die sich als technische Disziplinen verstehen, ist das aber ein grundsätzliches Dilemma. Wenn man Leitplanken ablehnt, heißt dies auch, die historisch gewachsenen Bedingungen, die gewisse Positionen privilegierten und andere diskriminierten, direkt fortzuschreiben. Greift man ein und korrigiert diese Verzerrungen, dann stellt sich die Frage nach den in diesen Korrekturen enthaltenen Positionen. So etwas wie Objektivität kann es bei aller Technizität hier nicht geben.

Jedes Bild ist das Resultat einer hochgradig situierten Produktion. Im Fall der generierten Bilder ist diese Situierung geprägt von der historischen und politischen Natur der Daten sowie von den Werten und Interessen derer, die daraus Modelle erstellen. Dadurch wird die Grenze zwischen dem, was existieren kann, und dem, was existieren soll, sehr unscharf. Dies verstärkt den Vorahnungscharakter der Bilder nochmals. In der einen oder anderen Art wird hier entschieden, welche Versionen der Zukunft überhaupt artikuliert werden können. Der Blick auf die Situiertheit der Technik zeigt, dass hier kein technologischer Determinismus am Werk ist, sondern konkrete institutionelle Dynamiken, die nicht zuletzt von den zugrundeliegenden Daten und den eingebauten Normen und Leitplanken bestimmt werden.

Je mehr diese Verfahren in die Produktion von Kultur eingezogen werden, desto stärker wird ihr Einfluss. Es stellt sich also die Frage, wer diese Entscheidungen trifft. Aktuell ist es eine winzige homogene Gruppe von Entwickler:innen und Unternehmer:innen an der Westküste der USA. Sie verfügen über das technische Know-how, die infrastrukturellen Kapazitäten und den Zugang zu Daten und Kapital, um vorn mitspielen zu können. Ihre Agenda lässt sich relativ einfach zusammenfassen: wirtschaftliche und zunehmend auch politische Macht. Dies ist, was wirklich generiert werden soll.

Eine andere Welt ist generierbar

Wenn aber kein technologischer Determinismus am Werk ist, bedeutet das, dass auch ganz andere Welten generiert werden könnten als diejenigen, die aktuell andere für uns generieren. Ein solches Interesse durchzieht beispielsweise das Werk der deutsch-irakischen Künstlerin Nora Al-Badri.

Eine ihrer Arbeiten ist [Babylonian Vision](#) (2020). Dafür trainierte sie ein neuronales Netzwerk, ein sogenanntes Generative Adversarial Network (GAN), ein Vorläufer der aktuellen bildgenerierenden Verfahren, mit 10.000 Bildern aus den fünf Museen mit den größten Sammlungen mesopotamischer, neo-sumerischer und assyrischer Artefakte. Aus diesen wurden nun weitere neue Artefakte, in Form von Videos und Bildern, erzeugt und als Objekte spekulativer Archäologie im Ausstellungsraum präsentiert.

Die Arbeit beschäftigt sich mit zwei Kernfragen der Bildgenerierung. Erstens, wie tragen die Daten der Vergangenheit, mit denen Maschinen trainiert werden, die Spuren ihrer eigenen, oft gewalttätigen Geschichte in sich? Das zeigt sich schon konkret in der Frage des Zugangs. Es wurden zwar alle Museen angefragt, aber die meisten großen Sammlungen stellten ihre Daten nicht zur Verfügung, teilweise, indem sie unüberwindbar hohe administrative Mauern aufstellten. So mussten die Daten auf anderen Wegen beschafft werden. Woher kommt diese Weigerung der Museen, Zugriff auf ihre Daten zu erlauben, obwohl es technisch und rechtlich einfach möglich wäre? Wer darf mit diesen Objekten/Daten arbeiten? Inwieweit wird hier eine koloniale Wissensordnung im Digitalen weitergeführt?

Al-Badri geht aber über diese Fragen, die im Zentrum vieler Restitutionsdebatten stehen, hinaus, denn sie stellt auch, und das ganz konkret, die Frage nach der Interpretation der Vergangenheit, weniger im Sinne der historischen Quellenarbeit, sondern als eine Ressource für die Zukunft. Auch hier klingt das Element der Vorahnung durch. Wessen Werte, wessen Interessen fließen in die Behandlung der kulturellen Ressourcen als Bausteine der Zukunft mit ein? Ist der rückwärtsgewandte Blick der Museen, mit ihrem Fokus auf Authentizität, der einzig legitime Zugang?

Durch das Training mit ausgesuchten Daten öffnet die [Arbeit dafür einen eigenen Latenzraum](#), der nicht vollkommen von kolonialen Institutionen und kommerzieller Optimierung dominiert ist. In diesem Raum lassen sich andere Bilder generieren. Neue Sprecher:innenpositionen werden möglich, dem Abwesenden wird zu Präsenz verholfen. Das explizit Spekulative dieser Arbeit nimmt die normativ-ästhetische Dimension der Generierung ernst. Sie ist aber nicht auf ein konsumistisches Menü mit vier Versionen, von denen eine nach individuellen Präferenzen

ausgewählt werden kann (die Standardmethode der kommerziellen Angebote), begrenzt. Sie stellt diese Fragen im kollektiven Raum der Ausstellung. Mit solch anderen Bildern und einem anderen Setting wird, zumindest potenziell, eine andere Zukunft denkbar. Denken ist natürlich nicht handeln, und Handeln ist nicht notwendigerweise erfolgreich. Aber ohne ein anderes Denken ist auch kein anderes Handeln möglich.

Beitrag gedruckt von Geschichte der Gegenwart: <https://geschichtedergewenart.ch>

URL zum Beitrag: <https://geschichtedergewenart.ch/das-generative-der-kuenstlichen-intelligenz/>

Copyright © 2021 Geschichte der Gegenwart. Alle Rechte vorbehalten.