Woche 14 Praktische Anwendungsbeispiele: Generative Modelle

Skript

Erarbeitet von
Dr. Jacqueline Klusik-Eckert

Inhaltsverzeichnis

Lernziele 1

Inhalt 2

Einstieg 2

Generative Modelle und ihre Medien 2

Wiederholung: Mit was haben wir es eigentlich zu tun? 3

Was sind die Vorteile? Welche Szenarien sind denkbar? 5

Herausforderungen und Risiken 6

Take-Home Message 9

Quellen 9

Disclaimer 11

Lernziele

- Verständnis über den Medienaspekt der Ergebnisse von Generativen Modellen
- Kennenlernen der potenziellen Einsatzszenarien für Text und Bild
- Gespür für die Herausforderungen entwickeln
- Bewusstsein über die aktuellen und zukünftigen Debatten gewinnen

Inhalt

Einstieg

Wie schön wäre es, wenn es eine KI geben würde, die mir die Inhalte für Lernvideos vorbereitet, damit ich sie nur noch einsprechen muss.

Text-to-Speech Anwendung https://www.naturalreaders.com/

Quelle [1]

Ach stimmt. Dann frag' ich mal den Chatbot, was ich in diesem Video erzählen soll.

Was sind die wichtigsten fünf Lernziele in einem Erklärvideo über generative Systeme der künstlichen Intelligenz wie ChatGPT, GANs oder ähnlichen Verfahren?

Quelle [2]

Generatives Textmodell ChatGPT https://chat.openai.com

Und wenn ich jetzt noch ein bisschen weiter frage, spuckt mir der Chatbot eine Menge Text aus. Den muss ich jetzt auch noch vorlesen.

Text-to-Speech Anwendung https://www.naturalreaders.com/

Du klingst mir aber ein bisschen zu hölzern.

Text-to-Speech Anwendung https://www.naturalreaders.com/

Generative Modelle und ihre Medien

Generative Modelle wie ChatCPT, DALL-E, Stable Diffusion oder auch Text-To-Speech-Modelle, die annähernd natürlich klingende Sprache erzeugen, produzieren Medien. Text, Bilder, Musik, Ton. Sie besitzen als digitale Medien die gleichen Eigenschaften wie jene, die von Menschen produziert wurden. Als Textdatei, PNG oder MP3 sind sie von den anderen nicht zu unterscheiden.

CC BY

Icons von https://www.flaticon.com

Warum ist die Diskussion über die Ergebnisse dieser KIs so hitzig geführt? Und was sind eigentlich praktische Anwendungsszenarien für diese Technik? Sind die auch wirklich nützlich oder nur zum unterhaltsamen Zeitvertreib geeignet?

Wiederholung: Mit was haben wir es eigentlich zu tun?

Jetzt kenne ich mich ja schon recht gut aus, aber eine Spezialistin bin ich auf dem Gebiet nicht. Für mich sehen die Antworten von ChatGPT auf meine Fragen nach den Lernzielen schon sehr gut aus. Damit kann ich arbeiten. Die nehme ich mir einfach als Inspiration und überlege mir selbst, was ich davon noch so brauchen könnte. Immer noch besser als vor einem leeren Platt zu sitzen. Da weiß ich nie, wo ich anfangen soll.
Aber was kann eine Spezialistin dazu sagen? Sag mal, Joana, wie findest du die Antworten von ChatGPT? Die sehen für mich doch ganz gut aus.
Das klingt für mich doch schon gut, was die so machen. Aber was würde eine Spezialistin dazu sagen?

[Interview Anfang]

Jacqueline
Hi Joana, wie findest du die Antworten von ChatGPT? Würdest du das so stehen lassen?
Joana
Die Antworten wirken natürlich auf den ersten Blick erstmal ziemlich sinnvoll und geben zumindest nichts grundlegendes Falsches wieder. Es ist aber schon auffällig, dass der Text ziemlich generisch und oberflächlich formuliert ist, sodass viel Spielraum für die eigentlichen Inhalte gelassen wird.

Jacqueline
Willst du damit sagen, der Algorithmus weiß gar nicht genau, wovon er da spricht?

Joana
Na ja … nicht wirklich, nein! Und wir sollten eine KI natürlich auch nicht anthropomorphisieren. Die große Stärke von ChatGPT besteht darin, Texte so zu verfassen, dass sie den Eindruck erwecken, 1) von Menschen geschrieben zu sein und 2) absolute Fakten wiederzugeben.
Das ist beides gefährlich, denn die Art und Weise, wie ChatGPT Texte produziert, hat wenig damit zu tun, wie wir Menschen einen Text verfassen. Wie wir mittlerweile schon wissen, basiert ChatGPT auf einem neuronalen Netzwerk, das notwendigerweise auf Daten aus der Vergangenheit trainiert wurde – aus sehr vielen Daten … Genauer gesagt überwiegend aus Texten aus dem Internet, die natürlich nicht auf Richtigkeit geprüft werden, womit wir direkt schon bei Punkt 2 wären. ChatGPT kann mit einer gewissen Souveränität Dinge als Fakten darlegen, die eigentlich völliger Bullshit sind.

Jacqueline

Es sieht aus wie ein Text, verhält sich wie ein Text, aber ist doch mit Vorsicht zu genießen. Der Algorithmus ist kein allwissendes Superhirn. Das ist klar. Aber was macht der Chatbot hier denn besonders gut?

Joana
Was wir auch im Beispiel oben sehen, ist das Novum von ChatGPT, dass die KI in der Lage ist, Bezug auf die vorherige Konversation zu nehmen und Kontexte besser zu erkennen als vorherige Modelle. Nichtsdestotrotz ist sie nicht in der Lage, „das große Ganze" zu sehen, sondern generiert immer nur einen kleinen Teil des Texts, basierend darauf was vorher geschrieben wurde.

Jacqueline
Dann verlasse ich mich mal lieber nicht komplett auf diese generativen Modelle, wenn es um dieses Lernvideo geht. Danke, Joana, für die Erklärung und Einordnung.
Zum Glück habe ich einen heißen Draht zu den Spezialistinnen und Spezialisten. Dann laufe ich nicht Gefahr, irgendeiner KI etwas zu glauben.
[Interview Ende]

Es ist aber auch zu verlockend. Alleine die Vorstellung, dass es da im Internet ein einfaches Eingabefeld gibt, in das ich alle meine Fragen reinstelle und am Ende kommt ein richtiger Text raus … Das wäre Magie. Doch hinter solchen generativen Modelle steckt keine Zauberei, sondern eine Mischung aus Wahrscheinlichkeitsrechnung und Mut zur Lücke.

Wie ich das meine? Ted Chiang zieht in seinem Artikel im New Yorker eine Parallele zwischen den generativen Modellen und einem aufsehenerregenden Fall der jüngeren Technikgeschichte.

Quelle [3]

Er vergleicht ChatGPT, DALL-E und Co. mit unscharfen, verschwommenen JPGS und einer Reihe von Drucker Fehlfunktionen.

Quelle [4]

Dieser Bug in den Xerox Drucker hat jetzt nicht direkt etwas mit generativen Systemen zu tun, aber als Analogie hilft er uns zu verstehen, was da bei ChatGPT, DALL-E oder Stable Diffusion läuft. Oder besser schieflaufen kann.
Wir haben am Anfang etwas sehr Komplexes. Alle Texte aus dem Internet, alle digitalen Bilder, die in das System eingespielt werden. In der Black Box, der KI, beginnt nun das sogenannte Lernen. Hier wird gefaltet, komprimiert, diese schiere Datenmasse vereinfacht, um in einer unglaublichen Rechengeschwindigkeit wieder etwas Komplexes daraus zu machen. Am Anfang und am Ende steht ein Text oder ein Bild, aber das Endprodukt entspricht nicht dem, was hineingekommen ist. Es hat Teile von den originalen Texten oder Bildern und viele Lücken, die nicht mit abgespeichert wurden. Und diese Lücken wurden durch die KI nach Wahrscheinlichkeit gefüllt. Wie wenn man bei einem Lückentext nicht überlegt, was da an der Stelle semantisch reinpasst, sonders sich immer hochrechnet, was

CC BY

am wahrscheinlichsten, in der Kombination mit den Nachbarwörtern, bisher vorgekommen ist.

Was sind die Vorteile? Welche Szenarien sind denkbar?

Da diese Verfahren in dieser Qualität relativ neu sind, haben sich so richtige Anwendungsbeispiele noch nicht ergeben. Aber lasst uns doch mal darüber nachdenken, für welche Szenarien man generative Modelle einsetzen kann? Wo stellen Sie einen Gewinn dar?

Die großen Natural Language Processing Modelle wie CPT-3, auf dem ChatGPT basiert, können zu den unterschiedlichsten Themen in Windeseile unzählige Varianten von Texten, auch in unterschiedlichen Stilen, auswerfen.
Diese könnten in der Schule eingesetzt werden, um das kritische Lesen zu schulen. Angeleitet und in einem didaktischen Rahmen hilft es, kritisches Gegenlesen zu lernen. Man muss ganz genau nachdenken, was man da gerade gelesen hat und besser zweimal hinschauen.

Video von Salvador Varón / HYPERLINK "https://pixabay.com/de/?utm_source=link-attribution&utm_medium=referral&utm_campaign=video&utm_content=69625" Pixabay

In diesem Zusammenhang kann man generell die ganzen Schreib- und Textanalyse Prozesse unterstützen.
Generative Modelle könnten bei speziellem Training Zusammenfassungen von Büchern liefern, würde Studierenden helfen, einen ersten Überblick über ein Thema zu bekommen. Wenn man sich dann mal einen Überblick verschafft hat, kann man das Lesen und Lernen besser strukturieren.

Video von DistillVideos / Pixabay

Denkbar wäre auch so ein Antwortmodell, um die Debattenfähigkeit zu trainieren. Man könnte sich fair, aber bestimmt streiten. Das geht jedoch nur, wenn so ein Modell auch eine Meinung hätte.

Man hat weiterhin auch die Hoffnung, dass die großen Sprachmodelle eine bessere Teilhabe am Internet und in der digitalisierten Welt ermöglichen. In Kombination mit Speech-to-Text oder Text-to-Speech Lösungen ist es denkbar, dass auch Menschen mit Lese- und Schreibproblemen Zugang zu Informationen und Kommunikationsformen bekommen.

Quelle [5]

Und bei Bildgeneratoren?

CC BY

Ein erstes Einsatzszenario liegt schnell auf der Hand. Ich habe eine Idee von einem Bild, weiß aber nicht, in welche Richtung es gehen soll. Stable Diffusion oder DALL-E eignen sich sehr gut, um erste Ideen visualisieren zu lassen. Es fungiert als Werkzeug für eine Art Prototyping und kann damit den kreativen Prozess unterstützen. Die willkürlichen Ergebnisse können inspirierend sein und einem auf die Sprünge helfen.

Auch für Leute, die nicht so kreativ sind und mit Pinsel und Stift gar nichts anfangen können, kann das Modell hilfreich sein. Ich muss nicht auf immer die gleichen, illustrierten Stockfotos zurückgreifen. Schnell habe ich ein hübsches Bild für meine Kurzgeschichte, in der es um einen Löwen geht, der Löwenzahn liebt, aber leider allergisch ist. Also nur als Beispiel.

Generierte Bilder von https://stabeldiffusionweb.com

Mit ein paar Programierskills kann man sich die Modelle, zumindest all diese, die unter Open Source Lizenz stehen, für eigene Arbeitsprozesse anpassen. Auf Reddit zeigt ein findiger Programmierer und Designer ein PlugIn, mit dem man direkt im Photoshop Bildbereich mit Stable Diffusion ergänzen kann. In Windeseile verwandelt sich die grobe Skizze in eine hübsche Hügellandschaft.

Quelle [6, 7]

Reddit Post
https://www.reddit.com/r/StableDiffusion/comments/wyduk1/show_rstablediffusion_integrating_sd_in_photoshop

Die Modelle sind dann besonders hilfreich, wenn sie für ein spezifisches Einsatzszenario mit den dafür passenden Textsorten gefüttert wurden. Wäre ChatGPT mit sehr vielen didaktischen Lernzielen trainiert worden, wären die Antworten auch besser gewesen.

Quelle [6, 8]

Herausforderungen und Risiken

Mit ihrer enormen Rechenleistung können generative Modelle Abermillionen Varianten von Medien erzeugen. Manches davon macht für uns Sinn, sieht ästhetisch ansprechend aus oder klingt auch ganz gut. Ziemlich viele dieser Ergebnisse landen jedoch auf dem Datenmüll. Man muss nur die Anfrage refreshen.
Die Modelle erwecken aber auch den Schein, Sprache, Texte, Bilder oder Musik wie aus dem Nichts erzeugen zu können. Das stimmt jedoch nicht.
Zu den größten Herausforderungen gehört die Frage zum Urheberrecht. Das betrifft zum einen die Fragen, wer die Urheberschaft auf die neugenerierten Medien innehat.

CC BY

Portrait of Edmond de Belamy
https://de.wikipedia.org/wiki/Portrait_of_Edmond_de_Belamy#/media/Datei:Edmond_de_Belamy.png

In einem ersten Fall, dem Verkauf des AI-generated Bildes Portrait of Edmond Belamy beim Auktionshaus Christie's, ging es in den anschließenden Rechtsfragen um den Code und das trainierte Neuronale Netz. Gehören die Ergebnisse auch dem Autor des Codes? Hat der Programmierer des Codes ein Recht auf einen Anteil an dem Auktionserlös von 432.500 $? Oder ist die Person alleiniger Urheber, die das errechnete Bild auf eine Leinwand hat drucken lassen?

Quelle [9]

Bei den Modellen, von denen wir hier sprechen, geht es eher um die Frage, mit was sie trainiert wurden.
Joana:
Hier tauchen ganz ähnliche Probleme auf wie bei ChatGPT, nämlich u. a. die Tatsache, dass GANs oder Stable Diffusion Modelle auf schon existierenden Bildern trainiert wurden. Das eröffnet eine heiße Debatte über die Urheber*innenschaft der generierten Bilder. In diesem Paper wird gezeigt, dass bei Stable Diffusion in der Tat Kunstwerke aus den Trainingsdaten in generierte Bilder kopiert werden.
Hier sieht man in der oberen Zeile von Stable Diffusion generierte Bilder und in der unteren Zeile das Match aus einem Trainingsdatensatz, das am nächsten ist. Auffällig ähnlich, oder? Häufig werden dabei Objekte, an die sich der Algorithmus „erinnert" aus dem Vordergrund und dem Hintergrund zusammengefügt. Außerdem zeigt das System manchmal ein „rekonstruktives Gedächtnis", bei dem die „erinnerten" Objekte semantisch mit dem Ausgangsobjekt übereinstimmen, ohne, dass sie pixelweise identisch sind.

Quelle [10]

Jaron Lanier schrieb bereits 2013 in seinem Buch "Who owns the Future"

"Digital information is really just people in disguise."

"Digitale Informationen sind eigentlich nur verkleidete Menschen".

Hinter den Milliarden von Daten stecken Designer*innen, Musiker*innen, Künstler*innen, die ihren Alltag mit der Produktion von Medien bestreiten. Das ist ihr Job.

Quelle [11]

Video von Oleg Gamulinskii / Pixabay

Und wir als fleißige Datenproduzent*innen schenken auch noch unsere Medien her. Die Debatte um die Rechtsfragen in alle Richtungen ist im Gange. Unklar ist ja auch, wie Medienkünstler*innen ihre mit KI-Unterstützung erstellten Werke sichern sollen. Da kommt auf die Rechtsabteilungen einiges zu. Und auf uns.

Quelle [12, 13]

Wie sieht es mit den anderen Risiken und Herausforderungen aus?
Generative Modelle versprechen eine enorme Erleichterung und sie werden sich in unterschiedlichen Einsatzszenarien sicherlich etablieren.
Doch im Moment können sie stark voreingenommen sein. Und das liegt an den Biases, den Vorurteilen, die in den Trainingsdaten verborgen liegen. Darüber hinaus ist die Benutzung der Technik nicht gerade fair geregelt. Viele KIs gehören Firmen, die mit ihrer Rechenleistung natürlich auch ein Geschäftsmodell bedienen müssen. Öffentliche Zugänge für alle werden durch nicht-barrierefreie Benutzeroberflächen erschwert.

Quelle [6]

Generell bergen die Modelle auch die Gefahr, dass man sich zu sehr darauf verlässt, dass die Ergebnisse richtig sind. Mit mangelndem Verständnis über ein Thema oder fehlendem Fachwissen lässt sich das auch nur sehr schwer überprüfen. Es besteht die Gefahr, dass gefährliches Halbwissen oder sogar Falschinformationen gestützt und untermauert werden.

Quelle [14]

Es ist auch noch nicht klar, wie in Zukunft Modell-generierte und von Menschen-generierte Medien voneinander unterschieden werden sollen.
So viele Fragen sind noch nicht geklärt. Wie sieht es mit den Kosten für Schulungen und Wartung aus? Wer sichert den Datenschutz und gibt Datensicherheit? Sind diese Systeme überhaupt nachhaltig? Wie sieht es mit dem Ressourcenverbrauch aus? Schließlich fressen die Serverparks für die Hochleistungsprozesse ziemlich viel Strom.

Quelle [15]

Unklar ist auch, wer die Kosten für die Überprüfung von Informationen übernimmt und ob wir an irgendeiner Stelle die Integrität von Inhalten schützen können.
Es kann sein, dass es immer schwieriger wird zwischen echtem Wissen und überzeugend geschriebenem, aber nicht verifiziertem Modell Output zu unterscheiden.

Quelle [6]

Und wie sieht es mit der Mehrsprachigkeit aus? Die großen Sprachmodelle funktionieren sehr gut, wenn es um Englisch geht. Und wie sieht es mit Niederländisch aus? Urdu? Kambodianisch? Wie soll eine Teilhabe an der neuesten Technik ermöglicht werden, wenn die ach so schlauen KIs nur auf wenige, westeuropäische Sprachen trainiert sind?

CC BY

https://www.visualcapitalist.com/100-most-spoken-languages/

Quelle [16]

Alles in allem: Ein fairer Zugang, ein sicherer Gebrauch, sieht anders aus.

Take-Home Message

Generative Modelle haben sich in den letzten zehn Jahren unglaublich schnell verbessert.

Quelle [17]

Das gilt sowohl für die Generierung von Texten, als auch für das Produzieren von Bildern, Animationen und Musik. Die Anwendungsszenarien sind im Moment noch in der Erprobung. Es ist aber jetzt schon abzusehen, dass die Anwendungen Einzug in unseren Arbeits- und Lebensalltag haben werden. Die Technik ist bereit.
Doch sind wir es? Wo werden die Entwicklungen noch hingehen?
Die UNESCO fordert daher, dass die Entwicklungen in der Künstlichen Intelligenz einem auf den Menschen ausgerichteten Ansatz folgen müssen. Humen-Centered

https://www.unesco.org/en/digital-education/artificial-intelligence

Quelle [18]

Was die Technik der Künstlichen Intelligenz in den nächsten Jahren noch bringen wird, steht in den Sternen. Klar ist schon heute, dass wir vom Einsatz dieser Verfahren an vielen Stellen profitieren. Wir bestimmen auch die Einsatzszenarien, die rechtlichen Rahmenbedingungen und bauen gewisse Hürden gegen den Missbrauch von Künstlicher Intelligenz auf. Es werden noch viele Debatten auf uns zukommen. Debatten, die wir alle gemeinsam führen müssen. Um das tun zu können, brauchen wir Zugang zum Wissen über diese Techniken und Verfahren. KI für alle.

Quellen

Quelle [1] AI Voices—NaturalReader Home. (2023). Abgerufen 27. Februar 2023, von https://www.naturalreaders.com/

Quelle [2] ChatGPT. (o. J.). Abgerufen 27. Februar 2023, von https://chat.openai.com

Quelle [3] Chiang, T. (2023, Februar 9). ChatGPT Is a Blurry JPEG of the Web. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

Quelle [4] Kriesel, D. (Sprecher). (2015, Januar 1). *Traue keinem Scan, den du nicht selbst gefälscht hast*. https://www.youtube.com/watch?v=7FeqF1-Z1g0

Quelle [5] Kasneci, E. et al (2023). ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education [Preprint]. EdArXiv. https://doi.org/10.35542/osf.io/5er8f

Quelle [6] Vincent, J. (2022, September 15). Anyone can use this AI art generator—That's the risk. The Verge. https://www.theverge.com/2022/9/15/23340673/ai-image-generation-stable-diffusion-explained-ethics-copyright-data

Quelle [7] alpacaAI. (2022, August 26). *Show r/StableDiffusion: Integrating SD in Photoshop for human/AI collaboration* [Reddit Post]. r/StableDiffusion. www.reddit.com/r/StableDiffusion/comments/wyduk1/show_rstablediffusion_integrating_sd_in_photoshop/

Quelle [8] Meineck, S. (2022, August 27). Text-zu-Bild-Generierung: Der Anfang von etwas Großem. netzpolitik.org. https://netzpolitik.org/2022/text-zu-bild-generierung-der-anfang-von-etwas-grossem/

Quelle [9] Small, Z. (2018, Oktober 29). *Christie's Sells „AI-Generated" Art for $432,500 As Controversy Swirls Over Creators' Use of Copied Code*. Hyperallergic. http://hyperallergic.com/468060/christies-sells-ai-generated-art-for-432500-as-controversy-swirls-over-creators-use-of-copied-code/

Quelle [10] Somepalli, G. et al (2022). *Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models* (arXiv:2212.03860). arXiv. https://doi.org/10.48550/arXiv.2212.03860

Quelle [11] Lanier, J. (2014). *Who owns the future?* (Simon&Schuster trade paperback edition). Simon & Schuster Paperback.

Quelle [12] Meineck, S. (2023, Januar 19). Sammelklage: Streit um Bild-Generatoren soll vor Gericht landen. netzpolitik.org. https://netzpolitik.org/2023/sammelklage-stable-diffusion-bild-generator-rechtsstreit/

Quelle [13] Chayka, K. (2023, Februar 10). Is A.I. Art Stealing from Artists? *The New Yorker*. https://www.newyorker.com/culture/infinite-scroll/is-ai-art-stealing-from-artists

Quelle [14] Mai, K. T., Bray, S. D., Davies, T., & Griffin, L. D. (2023). Warning: Humans Cannot Reliably Detect Speech Deepfakes (arXiv:2301.07829). arXiv. http://arxiv.org/abs/2301.07829

Quelle [15] *Training a single AI model can emit as much carbon as five cars in their lifetimes*. (2023). MIT Technology Review. Abgerufen 21. Februar 2023, von https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Quelle [16] Soria, C. (2020, Januar 1). Decolonizing Minority Language Technology. State of the Internet's Languages Report. https://internetlanguages.org/en/stories/decolonizing-minority-language/

Quelle [17] Offert, F. (2023). Ten Years of Image Synthesis. Zentralwerkstatt.Org. Abgerufen 21. Februar 2023, von https://zentralwerkstatt.org/blog/ten-years-of-image-synthesis

Quelle [18] Quelle 18 UNESCO. Artificial intelligence in education. Abgerufen 20. Februar 2023, von https://www.unesco.org/en/digital-education/artificial-intelligence

Disclaimer

Transkript zu dem Video „Woche 14 Praktische Anwendungen: Generative Modelle“, Dr. Jacqueline Klusik-Eckert.
Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz CC-BY 4.0 veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.