

Transkript: „Desinformation“

Einleitung

Avatar: Hallo, schön dass Sie da sind und mich auf meiner Lernreise begleiten. Ich stehe heute hier in der neuen Telekomzentrale, umspielt vom Sound der neuesten Taylor Swift Platte und kann Sie nach meinem Sprachstudium in allen europäischen Sprachen begrüßen. Ich verspreche Ihnen eine spannende Tour ...

Inga: Moment, da stimmt doch etwas nicht? Schauen wir doch einmal etwas genauer hin. Schauen wir zunächst auf meinen Hintergrund: das ist natürlich nicht die neue Telekom-Zentrale, sondern ein von der KI generiertes Bild. Und natürlich ist das auch kein Song von Taylor Swift... Und diese Moderatorin bin ich in Wirklichkeit auch nicht.

Und schon sind wir mittendrin im Thema. Sieht aus wie ich, spricht wie ich und trotzdem bin ich es nicht. Mit wenigen aufgenommenen Videosequenzen und ein paar gesprochenen Sätzen lässt sich ein digitaler Zwilling, ein sogenannter Avatar von mir erstellen. Diesem Avatar können wir dann jeden beliebigen Text in den Mund legen und das auch noch nahezu lippensynchron in allen Sprachen dieser Welt. Künstliche Intelligenz machts möglich.

Eintauchen

Hätten Sie das gedacht? KI-Technologien schreiten schnell voran, und mit ihnen entstehen ständig neue Anwendungsfälle.

Bevor wir in das Thema KI und Desinformation tiefer einsteigen, ist es wichtig zu wissen, welche verschiedenen Anwendungsbereiche es von Künstlicher Intelligenz derzeit gibt. Wir legen unseren Fokus auf die generative KI. Die generative KI ist eine spezielle Anwendung des maschinellen Lernens und bezieht sich auf Systeme, die in der Lage sind, neue Daten zu generieren, die auf vorhandenen Mustern oder Modellen basieren. Maschinelles Lernen ermöglicht es Computern, aus ihren eigenen Erfahrungen zu lernen, ohne explizit neu programmiert werden zu müssen. Generative KI findet breite Anwendungsbereiche in der Bild- und Videobearbeitung, der Spracherzeugung, der Musikkomposition und sogar der Erstellung von Texten und Kunstwerken.

All diese Tools ermöglichen es uns zum Beispiel Prognosen zukünftiger Ereignisse oder Trends basierend auf vorliegenden Daten zu ermitteln. Auf der anderen Seite fordern uns diese Formen der künstlichen Intelligenz aber auch stark heraus. Das Thema Desinformation ist eng mit generativer KI verbunden, da die fortschrittlichen KI Modelle es ermöglichen, künstliche Inhalte zu erstellen, die für den Menschen immer schwieriger von echten Inhalten zu unterscheiden sind.

Verstehen

Die rasante Entwicklung von KI-Anwendungen hat dazu geführt, dass wir als Gesellschaft vor neuen Herausforderungen stehen.



Es stellt sich also die Frage, wie wir diesen Herausforderungen angemessen begegnen können. Was wir definitiv brauchen, ist ein tieferes Verständnis dieser Systeme. Um zu verstehen, welche Art von KI vorliegt, haben wir grundlegende Informationen für Sie im folgenden Video zusammengestellt.

Legefilm

Wenn neue Inhalte von Künstlicher Intelligenz erstellt werden, spricht man von generativer KI.

Generative KI kann neue Texte, Bilder, Musik und sogar ganze Videos aus bestehenden Daten erschaffen. Dies öffnet Türen für vielfältige Anwendungen. So können realistisch aussehende Abbildungen von Menschen, Tieren, Landschaften oder Objekten erzeugt werden. Aber auch in der Bildung bietet generative KI gute Möglichkeiten, zum Beispiel für das Erstellen von personalisierten Lehr- und Lernmaterialien. Auch kann sie als kreative Schreibhilfe und zum Lernen von Sprachen eingesetzt werden.

Mit Hilfe der Künstlichen Intelligenz können sogar virtuelle Labore erstellt und Experimente simuliert werden. Außerdem kann sie komplexe Ideen in 3-D-Modellen veranschaulichen und sie mit Animationen oder interaktiven Diagrammen darstellen.

Eines der meistgenutzten Tools ist das der textgenerierenden KI. Eine textgenerierende KI ist ein Werkzeug, das mit riesigen Mengen von bestehenden Texten trainiert wird. Diese Informationen werden genutzt, um neue Texte zu verschiedenen Themen zu erstellen. So kann es beispielsweise für Nachrichtenartikel oder Social Media Posts eingesetzt werden.

Textgenerierende KI kann aber auch von einer Sprache in die andere übersetzen und lange Texte oder Datenmengen zusammenfassen.

Die Antworten einer solchen KI werden anhand von Wahrscheinlichkeiten generiert. Das bedeutet, dass die KI nur "errechnet", welches Wort am wahrscheinlichsten auf das nächste folgt. Ein weiterer wichtiger Punkt ist, dass die Informationen, auf denen das Training der KI basiert, zu einem bestimmten Zeitpunkt abgeschlossen wurden und somit veraltet sein können. Daher ist ein sprachverarbeitendes Modell wie ChatGPT eher ungeeignet für wirkliche "faktische Recherche", da es primär auf Mustererkennung in Daten und weniger auf verifizierbares Faktenwissen basiert.

Das Besondere ist, dass man sich wie in einem Dialog mit der KI austauschen kann. Im Gegensatz zu einer herkömmlichen Recherche können Fragen oder Anfragen korrigiert oder erneut gestellt werden, bis man mit der Antwort zufrieden ist. Dieses dialogische Verfahren nennt man „Prompting“. Besonders hilfreich ist, dass man die Ergebnisse an die Bedürfnisse verschiedener Zielgruppen anpassen kann, beispielsweise indem man fragt, wie man ein bestimmtes Thema einem 5-jährigen Kind erklären würde.

Textgenerierende KI birgt jedoch das Risiko, dass sie auch für die Erstellung gefälschter Texte und Nachrichten missbraucht werden kann. Diese KI-Modelle sind in der Lage, überzeugend klingende Texte zu erstellen, die menschliche Autor*innen imitieren.

Sprachverarbeitende KIs basieren auf Natural Language Processing, kurz NLP. Das ist eine Methode der KI, die es Computern ermöglicht, die natürliche Sprache des Menschen zu verstehen und zu imitieren. Angewendet werden sie z.B. bei Chatbots, Smartspeakern wie Alexa oder Siri und bei der Erstellung von Untertiteln. Diese Systeme erkennen gesprochene Sprache und wandeln diese in Text um.

Durch die Analyse der Sprache identifizieren NLP-Systeme nicht nur das gesprochene Wort, sondern erkennen auch spezifische Emotionen der Nutzer*innen. Dies ermöglicht die Entwicklung zielgerichteter Desinformationskampagnen, die genau darauf abgestimmt sind.



Bei der Bilderkennung mit KI handelt es sich zuerst einmal um Anwendungen, die Bilder erkennen und klassifizieren können. Eingesetzt werden diese Tools zum Beispiel, um Gesichter anhand von eindeutigen Merkmalen zu erkennen, wie beim Entsperren des Smartphones.

Bildgebende KI erweitern dieses Feld und werden zur Bearbeitung und Neuerstellung von Bildern verwendet. Ähnlich wie bei der Textgenerierung über Chat-Dialoge können über textbasierte Eingaben ganz neue Bildwelten erschaffen und bestehende verändert werden.

Bisher können KI-generierte Bilder oft noch relativ leicht erkannt werden, da die vorhandenen Daten möglicherweise nicht ausreichen, um alle Details realistisch wiederzugeben. Trotzdem muss uns stets bewusst sein, wie leicht Bilder manipuliert werden können, um Menschen, Landschaften oder Ereignisse in ein neues oder falsches Licht zu rücken. Durch die Manipulation von Bildern können falsche Darstellungen geschaffen werden, die die Realität verzerren und die Menschen in die Irre führen.

Wir alle können dazu beitragen, die Herausforderungen im Umgang mit generativer KI anzugehen und die Erstellung sowie Verbreitung von Desinformation einzudämmen. Dazu gehört kritisches Denken und eine bewusste Informationsverbreitung. Es ist wichtig, sich kontinuierlich über die Funktionsweise von KI zu informieren und diese Technologien verantwortungsbewusst zu nutzen. Unterstützung von Transparenz und die Förderung ethischer Standards sind ebenfalls entscheidend, um den Missbrauch von KI einzudämmen und ihre positiven Potenziale zu maximieren.

KI wird eine entscheidende Rolle in vielen Bereichen unseres Lebens spielen. Das hat viele Vorteile, denn es kann uns unter anderem dabei helfen, Probleme schneller zu lösen und neue kreative Möglichkeiten eröffnen. Gleichzeitig wird es für uns aber schwieriger, echte von unechten Inhalten zu unterscheiden, und daher zur Verbreitung von Desinformation beitragen kann.

Entdecken

Verblüffende Ergebnisse mit KI erzielt man auch mit Apps, mit denen man sich selbst oder andere Personen auf Basis eines Fotos oder kurzer Videoausschnitte in nichtexistierende Situation begeben kann. Stellen Sie sich doch einmal Barack Obama als Kassierer an der Supermarktkasse vor oder Cleopatra, die gerade ein Selfie mit ihrem Smartphone macht - was in unseren beiden Beispielen als unterhaltsam erscheinen mag, kann in anderen Zusammenhängen ernste Folgen haben. Diese sogenannten Deepfakes können gezielt und manipulativ eingesetzt werden.

Mit Deepfakes kann man Desinformationen erzeugen, die dann über Bots, Likes, Teilen und Kommentare das Netz fluten und die öffentliche Meinungsbildung beeinflussen.

Für alle Altersgruppen ist es schwer, KI-generierte Desinformation zu erkennen, daher sind wir hier gleichermaßen gefährdet. Kinder und Jugendliche sind jedoch besonders anfällig, da sie möglicherweise nicht über die nötige Medienkompetenz verfügen, um solche manipulativen Inhalte angemessen zu identifizieren und einzuordnen.

Im nächsten Schritt zeigen wir Ihnen einmal anhand verschiedener Beispiele, wie einfach diese KI-Werkzeuge genutzt werden können. Mit diesem Wissen wird es für Sie in Zukunft einfacher sein, KI-generierte Inhalte selber aufzudecken.



Handeln

Lassen Sie uns nun einmal den Blick auf 4 konkrete Beispiele werfen, die bei der Verbreitung von Desinformation eine Rolle spielen können. Ich habe mir dazu einen Experten eingeladen, der uns alles ganz genau erklären kann. Hallo Thomas!

Thomas: Hallo Inga. Da bin ich. Ob ich ein Experte bin, weiß ich nicht, aber tatsächlich beschäftigen wir uns viel mit künstlicher Intelligenz im Klassenzimmer oder auch in verschiedenen Workshops, die wir mit Jugendlichen durchführen. Und ich würde gerne tatsächlich so ein paar Beispiele zeigen.

Das erste Beispiel, was ich zeige, basiert auf Text und was nimmt man da? Na ja, der FC Bayern ist eng verbunden mit der Deutschen Telekom und ich habe mir die aktuelle Pressemitteilung rausgesucht zum neuen Trainer. Beim FC Bayern hat ja endlich geklappt und ich will zeigen, wie einfach das ist, mit einer textgenerierenden KI Fake-News zu produzieren.

Achtung, Trigger-Warnung: Ich produziere nicht wirklich Fake-News.

Ich will nur demonstrieren, wie das tatsächlich im großen Stile gemacht wird. Ich habe also folgendes gemacht: Ich habe einen Prompt vorbereitet. Den kopiere ich hier schon einmal rein. In dem Prompt sieht man oben: Hallo, hier ist ein Ausschnitt an der Pressemitteilung zum neuen Trainer vom FC Bayern. Bitte ändere diese Thomas Schmidt einsetzen. Thomas Schmidt ist Thüringer und er findet, „Bratwurst und Leberkäse sind eng vereint wie die Brezelarme. Und er ist natürlich tatsächlich ein Sportler und denkt, dass Radfahren und der Rasen zusammengehören. Dann habe ich das Originalzitat von der Pressemitteilung reingepackt und jetzt geht es los.

Wir schicken das zur textgenerierenden KI. Ich bin gespannt. Das ist alles live natürlich, wie immer bei uns hier im Studio. Und da ist sie schon. Wunderbar. Pressemitteilung Überschrift. Könnte direkt starten. FC Bayern München stellt Thomas Schmidt vor. Thomas Schmidt hat seine erste Pressekonferenz als Trainer des FC Bayern hinter sich gelassen. Der neue Coach der Münchner glänzte bei der Vorstellung mit Tatendrang, Optimismus und Angriffswelle.

Na, wie schön geschrieben, wie aus dem Kicker. Thomas Schmidt, der aus Thüringen stammt und Bratwurst und Leberkäse eng – hat sie auch gleich gemacht – überzeugte bereits die Verantwortlichen des Deutschen Rekordmeisters. Der Ex-Sportler sieht sich als Radfahrer bestens geeignet, auch auf dem Rasen für Teamgeist zu sorgen.

Irre. Und stellen Sie sich das jetzt vor mit einem Kontext im politischen Umfeld, vielleicht in anderen Umfällen. Na, dann sehen Sie, da ist eine textgenerierende KI, obwohl sie eigentlich darauf getrimmt ist, keine falsche Informationen zu erzeugen, wunderbar genutzt werden kann, Texte zu verändern, den Fokus zu verändern, Einzelheiten in den Fokus zu stellen. Und nein, ich dementiere ganz offiziell hiermit. Ich bin nicht der neue Trainer. Wäre bestimmt reizvoll, aber ich glaube, da gibt es viel, viel, viel bessere.

Nachdem wir nun eine Technik, die genutzt wird, Desinformation zu erzeugen, gesehen haben, nämlich mit einer textgenerierten KI, können wir uns das einmal mit Bildern anschauen. Dazu habe ich mit einer künstlichen Intelligenz ein zugegebenermaßen nicht sehr perfektes Bild, aber ein Bild erstellt von jungen Menschen, die sozusagen von den Wellen wegrennen und sehr viel Spaß am Strand haben.

Und Framing ist nun diese Technik, die setzt sozusagen diese jungen Menschen zum Beispiel in einen anderen Zusammenhang. Framing gibt es nicht nur bei Bildern, auch wenn man sich es da vielleicht am leichtesten vorstellen kann, aber ich zeig mal, was ich gemacht habe. Das Erste, natürlich alles mit einer bildgenerierenden KI, war ganz einfach, den Hintergrund auszutauschen und schon hat man eine ganz



andere Stimmung. Es sind exakt die gleichen Personen. Es ist genauso von der KI generiert, aber plötzlich sind wir nicht mehr am Ostseestrand, wie ganz am Anfang.

Vergleichen wir doch einmal kurz. Also hier sieht es mehr so wie ein ganz langweiliger Strand aus, währenddessen man jetzt hier schon sehen kann, wow, der Frame ist verändert und wir sind am Palmstrand. Noch interessanter wird das Ganze, wenn man diese Situation ganz anders beschreibt und Sie sehen gleich, was ich gemacht habe.

Stellen wir uns vor, ein Vulkan bricht aus und wieder haben wir exakt die gleichen jungen Menschen in der gleichen Stellung, aber plötzlich eine vollkommen andere Dramatik. Das laute Lachen der einen Person wird dann zum Schrei, das Zähnezeigen, zum Rennen und wir haben plötzlich eine vollkommen andere Assoziation, obwohl es die gleichen Menschen sind.

Das ist Framing mit KI und ich habe auch noch ein letztes, versöhnlicheres Bild gemacht. Natürlich wäre es auch witzig, wenn Sie gerade auf der Party am Strand sind, von einer Stelle zu anderen rennen. Nun kann man sagen: „na ja, Thomas, das hätte man auch schon vor dem KI-Zeitalter mit Photoshop zusammensetzen können, Bild freistellen und draufsetzen.“

Ja, aber wenn man das wirklich live mit der KI macht, dazu haben wir heute nicht die Zeit, dann bekommt man noch ganz andere Aspekte. Das heißt, man bekommt die jungen Menschen in anderen Situationen und sind noch viel besser verschmolzen mit diesem Thema Framing. Und geht man dann noch weiter und stellt sich vor, dass auch noch Aussagen dabei sind, Töne dabei sind, dann merkt man ganz schnell, wie einfach es ist, tatsächlich Informationen mit einer KI herzustellen.

Apropos Töne, das ist mein nächstes Szenario. Mit Ton kann man eine ganze Menge machen. Unter anderem wissen wir ja, dass wir an unserer Stimme sogar unsere Gemütszustände erkennen können. Aber das ist nicht das Thema. Ich will quasi zeigen, wie man auch Musik von Künstlern faken kann. Und ja, das ist verboten. Es gibt auch Warnmeldungen, aber eigentlich kommt man ziemlich nah ran. Erinnern Sie sich noch an den Song vom Anfang? Das war auch mit einer KI generiert und ich habe hier einen ganz wunderbaren Song generiert.

Man sieht gleich noch mal meinen Prompt, den ich hier oben eingegeben habe. Ich möchte gerne, dass ein Song generiert wird mit der Deutschen Telekom und dem Thema 5G im Stil von Taylor Swift. Und das macht die KI ganz wunderbar. Was soll ich erzählen? Hören wir einfach rein. Cyberworld. Wahnsinn, oder? Wahnsinn. Also ich sehe sie quasi vor mir, aber es ist weder Taylor Swift noch, hat sie jemals für die Deutsche Telekom meines Wissens einen Werbesong oder Ähnliches gemacht. Man hat hier auch die Vocals natürlich, man kann also sozusagen alles sehen, machen.

Und ich bin überrascht, wie weit wir mit der KI gekommen sind, dann schlussendlich zum vierten Beispiel zu kommen. Und das vierte Beispiel kennen Sie auch schon. Am Anfang hatten wir Ihnen ja gezeigt, wie unsere liebe Inga als virtuelle Person erscheint. Und das ist schon ein bisschen verrückt und vielleicht auch noch zu weit, vielleicht aber nur.

Und deshalb zeige denke ich Ihnen jetzt, was eigentlich schon fast Standard ist. Die Inga erleben wir nun in einem anderen Video, wo es sich digitale Demokratie handelt. Hören Sie auf die Stimme und den Singsang.

Hallo, schön, dass Sie wieder dabei sind. Das Internet bietet viele Chancen, auch für die Demokratie. Dank Online-Petition oder digitalen Wahlen wird es einfacher, sich eine Meinung zu bilden und politisch aktiv zu werden.

Super, absolut aktuell. Wir stehen ja in dem großen Wahljahr vor vielen Wahlen. Haben Sie den Ton von Inga im Kopf? Und jetzt benutzen wir hier wieder eine andere KI-Plattform. Ich habe das natürlich



vorbereitet und habe einfach gesagt: „Liebe Inga, das geht doch auch auf Englisch. “Hallo, nice to have you back. The Internet offers many opportunities, including for democracy. Online-petitions and digital elections make it easier to form an opinion and become politically active.”

Also für mich fast kein Unterschied, einfach in einer anderen Sprache und das geht in vielen anderen Sprachen. Man erkennt, Inga, obwohl sie nicht spricht, ja, zugegebenermaßen. Wenn man jetzt genauer hinschaut und sie finden ja auf Teachtoday alle diese Videos auch, dann sieht man schon, dass es noch nicht ganz lippen-synchron ist und dass dann noch ein paar Kleinigkeiten sind, aber so schnell wie es geht.

Und das ist nun ein Beispiel, was man Gutes damit machen kann, denn tatsächlich Man kann ja eigene Inhalte in ganz viele Sprachen bringen, aber nun ist es für sie, glaube ich, keine Herausforderung, einfach einmal nachzudenken, was passiert, wenn so was in falsche Hände kommt beziehungsweise aktiv genutzt wird, Fake-News, Desinformation oder was auch immer an Dingen rauszublasen über soziale Netzwerke.

Vier Beispiele habe ich Ihnen gezeigt. Ich komme zurück zum Anfang. Vier Beispiele: Texte. Ich war mal ganz kurz in der Pressemitteilung der Trainer vom FC Bayern. Wir haben uns mit Bildwelten beschäftigt und haben dort Framing kennengelernt. Wir haben gesehen, was man mit Tönen machen kann und verstehen vielleicht auch ein bisschen, warum die Künstler Sturm laufen zum Thema Urheberrecht, wenn man das so leicht faken kann. Und wir haben ganz am Anfang unseres Videos Inga als Avatar, als „Go Inga“, gesehen und nun haben wir auch gesehen, es gibt positive Aspekte von KI, aber das sozusagen falsch benutzt, erschafft quasi eine Welt, die beliebige Information Informationen in beliebigen Kontext setzt und man sich schon sehr fragen muss: „Wie bekomme ich heraus? Ist das eine Desinformation? Ist das ein Fakt oder ist das ein Fake?“

Und das ist das, was wir zusammen neu lernen müssen. Dafür sind wir auch da. Besuchen Sie gerne unsere Akademie. Da erleben Sie die Inga voll in Aktion, auch zum Thema digitale Demokratie.

Danke Inga, dass ich da sein durfte. Weiter geht es mit unserem Lern-Modul.

Inga: Danke Thomas! Nun, da Sie dieses Wissen haben, sind Sie besser gerüstet, um Bilder und Videos kritisch zu betrachten. Seien Sie wachsam und hinterfragen Sie Informationen, um Desinformation zu erkennen und nicht weiter zu verbreiten. Gemeinsam können wir so dazu beitragen, die Verbreitung von Desinformationen einzudämmen und die Echtheit der Informationen zu schützen.

Reflektieren

Blicken wir noch einmal zurück zum Anfang. Mein digitaler Zwilling hat gezeigt, wie perfekt KI generierte Inhalte eingesetzt werden können, um es für Laien nahezu unmöglich machen, zu erkennen, ob es sich um die „echte“ Inga handelt.

Daher ist es umso wichtiger, dass wir alle beim Einsatz von Künstliche Intelligenz gewisse Spielregeln einhalten. Halten wir also noch einmal fest:

1. Generative KI bezeichnet die Technologien, mit denen neue Inhalte mithilfe von KI erstellt werden. Das können Texte, Bilder, Musik, Kunst und Videos sein.
2. Ergebnisse von KI-Anwendungen sind mittlerweile so realistisch, dass es immer schwieriger wird zu erkennen, ob es sich um echte oder künstliche Darstellungen handelt.
3. Generative KI birgt das Risiko des Missbrauchs und kann bewusst eingesetzt werden, um Desinformation in Form von Deepfakes & Co. in Umlauf zu bringen.
4. Seien Sie wachsam gegenüber Desinformation und hinterfragen Sie Informationen immer kritisch.



Denn eines ist ganz sicher: Generative KI- Technologien beeinflussen schon jetzt unser Leben in vielerlei Hinsicht. Sie erleichtern kreatives Arbeiten, unterstützen Lernprozesse und lösen komplexe Probleme. Allerdings bergen sie auch Risiken: KI kann missbraucht werden, um überzeugende Deepfakes und manipulierte Texte zu verbreiten, und es besteht die Gefahr von Verzerrungen und Fehlinterpretationen.

Gleichzeitig bietet KI aber auch bedeutende Vorteile im Kampf gegen Desinformation, indem sie bei der Faktenprüfung unterstützt. Durch den Einsatz von Machine-Learning-Algorithmen können verdächtige Muster und Merkmale in Texten, Bildern oder Videos erkannt werden, die auf potenzielle Desinformation hinweisen. Seit 2018 verfolgt die EU einen Aktionsplan gegen Desinformation, der auf europäischen Werten und Grundrechten, wie der freien Meinungsäußerung, basiert. Ziel ist es unter anderem, Desinformation besser zu erkennen und die Gesellschaft dafür zu sensibilisieren.

Es ist entscheidend, KI verantwortungsvoll einzusetzen, um ihre positiven Aspekte zu nutzen und gleichzeitig die Risiken zu kontrollieren, sodass sie als Instrument zur Förderung der Wahrheit in unserer digitalen Welt dienen kann.

Abmoderation

Und? Haben Sie jetzt den Durchblick? Dann bleibt uns nur noch zu sagen:

KI ist das Zukunftsthema schlechthin. Stellen wir uns gemeinsam der Herausforderung, dieses mächtige Werkzeug mit Bedacht zu nutzen, damit es uns hilft, die Wahrheit zu erkennen – und nicht in einem Meer von Desinformation unterzugehen. In diesem Sinne: Bleiben Sie neugierig und bleiben Sie kritisch!

