

# ChatGPT

## 6 grundlegende Fragen

Das vorliegende Dokument ist Teil eines Projekts, das von **ethix** – Lab für Innovationsethik mit der Unterstützung der **Stiftung Mercator Schweiz** durchgeführt wurde. Die folgenden sechs Fragen wurden in Interviews und Workshops mit über 30 Personen, deren Tätigkeit durch ChatGPT beeinflusst wird, als grundlegende Themenbereiche identifiziert. Informationen zu weiterführenden Unterlagen und zu konkreten Beispielen finden Sie am Ende dieses Dokuments.

### Die wichtigsten in den Interviews genannten Nutzungsarten

#### ERSTER ENTWURF

eines Textes auf Basis weniger Informationen

#### STILISTISCHE VARIANTEN

zu einem Textabschnitt

#### SPARRING-PARTNER,

um wichtige Aspekte einer gestellten Frage zu ermitteln

#### ZUSAMMENFASSUNG

eines Textes auf der Grundlage von wenigen Eckdaten

#### Erste Skizzen für

RECHERCHEN



### Ist ChatGPT zu menschlich?

In der Art und Weise, wie der Chatbot ChatGPT schreibt, verhält er sich menschenähnlich. Er verfasst einen strukturierten, argumentativen Text und interagiert mit der nutzenden Person. Der Austausch gleicht einer Konversation zwischen zwei Menschen. Ebenfalls ist der Aufbau der Plattform ein Beispiel für unser Streben danach, Technik immer menschenähnlicher zu gestalten.

ChatGPT ist keine Person, auch wenn seine Entwickler:innen versuchen, die Illusion eines zum Denken fähigen Programms zu erzeugen. Das Tool beruht jedoch auf mathematischen Formeln und Code.

Das Design der Benutzeroberfläche von ChatGPT ist nicht neutral. Zum Beispiel suggeriert das wortweise Erscheinen der Antwort bewusst, dass das Tool «nachdenkt». Diese Art von Designentscheidungen spiegeln wirtschaftliche Interessen und strategische Ziele der Entwickler:innen und der involvierten Designer:innen wider. Deshalb ist es wichtig, stets die Menschen und Firmen, sowie deren Interessenlagen hinter der Technologie zu identifizieren.

Die sprachlichen Fertigkeiten und das Design von ChatGPT verstärken die Gefahr, seine Nutzenden irrezuführen. Die Personalisierung des Tools, z.B. durch Interviews mit ChatGPT in den Medien, trägt dazu bei, die Anwendung zu vermenschlichen.

### Ist ChatGPT neutral?



ChatGPT hat Milliarden von online zugänglichen Texten verarbeitet und kommuniziert mit uns wie eine scheinbar allwissende Maschine. Die Unmenge an Wörtern, die analysiert wurden, erwecken die Illusion eines gänzlich neutralen Gegenübers. Wenn man jedoch die einzelnen Schritte der Entwicklung von ChatGPT nachzuvollziehen versucht, wird diese scheinbare Neutralität widerlegt.

Die Entwickler:innen von ChatGPT haben die Maschine mit einer riesigen Menge an Daten trainiert. Diese Mischung aus online gefundenen und digitalisierten Texten gibt jedoch nur einen Ausschnitt der tatsächlichen Realität wieder. Die Grundlage für ChatGPT bilden Daten (Texte) der englischsprachigen Welt, die sich in den analysierten, online verfügbaren Texten mit all ihren Vorzügen und Mängeln widerspiegelt.

Die durch die Analyse der Textdaten erzielten Ergebnisse werden anschliessend von Menschen verbessert und korrigiert. Einerseits werden der Anwendung bestimmte Antworten vorgeschrieben. Andererseits wird sie darauf trainiert, angemessene Antworten zu geben und bestimmte Themen zu umgehen. Durch diese Trainingsmassnahmen wird versucht, gefährliche oder unsinnige Ergebnisse zu vermeiden. Die erzielten Resultate sind also das Ergebnis menschlicher Einflussnahme.

ChatGPT ist nicht neutral. Die erzeugten Texte sind immer ein Abbild der Trainingsdaten und der menschlichen Entscheidungen, die im Zuge der Verfeinerung des Modells getroffen wurden. Je mehr solche Anwendungen für die Produktion von Inhalten verwendet werden, desto wichtiger wird Transparenz darüber, welche Daten dem System zugrunde liegen und welche Grundannahmen es leitet.



### Sagt ChatGPT die Wahrheit?

ChatGPT scheint die Wahrheit zu sagen, denn die Antworten wirken oft überzeugend und der Stil ist affirmativ. Die Kategorie «Wahrheit» spielte bei der Entwicklung von ChatGPT aber keine Rolle. Vielmehr bestimmt das Tool, welche Wörter in einem bestimmten Kontext am plausibelsten sind.

#### Bei Sachfragen nach dem Muster «Wann fand der Erste Weltkrieg statt?»

«Wahrheit» resultiert aus der Übereinstimmung zwischen einer Aussage und einer nachprüfbaren Tatsache. ChatGPT kennt jedoch die Realität nicht und arbeitet nicht mit dem Konzept der Wahrheit. Vielmehr sucht ChatGPT nach den plausibelsten Wörtern für den gegebenen Kontext. In manchen Fällen entspricht das Plausible einer überprüfbaren Tatsache. Oft stellt das Plausible jedoch nicht das Wahre dar und ChatGPT liefert infolgedessen falsche Resultate.

#### Bei Beurteilungsfragen nach dem Muster «Welches ist das beste Wirtschaftssystem?»

Solche Fragen können nicht anhand von nachprüfbareren Tatsachen beantwortet werden, sondern erfordern einen klaren Massstab, der zur Bewertung verschiedener Alternativen genutzt wird. ChatGPT wurde so programmiert, dass bei gewissen Themen Vorsicht suggeriert wird und bestimmte Fragen nicht beantwortet werden. Es ist wichtig, dass sich die Nutzenden der Funktionsweise des Tools bewusst sind und wissen, welche Kriterien die Entwickler:innen für die Bereinigung von Ergebnissen festgelegt haben.

ChatGPT verstärkt, was es in den Daten findet. Dabei trifft Plausibilität manchmal auf Wahrheit – dies ist jedoch nicht systematisch der Fall. Das Tool kann sehr einfach zu einer Maschine werden, die falsche Plausibilität erzeugt. Eine der grossen Herausforderungen ist also, die Verbreitung von Fake News, welche von ChatGPT in Form von Texten oder Bildern generiert wurden, zu unterbinden.



### Über Bewusstsein verfügen

ChatGPT kommuniziert so menschenähnlich, dass sich die Frage aufdrängt, ob der Chatbot ein Eigenleben besitzt. Teilweise wird sogar gefordert, ChatGPT als Vorstufe von Maschinen, die ein eigenes Bewusstsein entwickeln können, anzusehen.

«Bewusstsein» kann als die Fähigkeit zu einem eigenständigen inneren Leben und Erleben verstanden werden. Es beruht auf einer tiefsten persönlichen Erfahrung. Zu wissen, wie das Innenleben anderer Menschen und anderer Lebewesen aussieht, ist unmöglich. Denn wir können uns nicht in andere hineinversetzen.

Die schiere Unmöglichkeit, Bewusstseinerfahrungen zu teilen, verstärkt die Schwierigkeit, den Begriff richtig zu definieren. Es ist schwer vorstellbar, dass ChatGPT die Eigenschaften besitzt, welche allgemein mit Bewusstsein verbunden werden: Sensibilität (die Welt um sich herum spüren und auf sie reagieren können), Wachsamkeit, und Selbsterfahrung.

Der Versuch, Bewusstsein zu definieren, wirft die Frage nach seiner Entstehung auf. Kann es auf blosser Materie reduziert werden oder stellt es etwas «darüber hinaus» dar? Kann es sich in einem nicht lebendigen Wesen entwickeln? Grundsätzlich ist es denkbar, dass Bewusstsein auch auf nichtbiologischen Faktoren gründen kann. Wer den Geist auf die physische Welt reduziert («alles ist Materie»), könnte etwa davon ausgehen, dass Bewusstsein aus den Kombinationen von Materie entsteht, aus denen sich ein künstliches (knichtbiologisches) Lebewesen zusammensetzt.

Auch wenn ChatGPT kein Innenleben hat, kann die Möglichkeit eines künstlichen Bewusstseins nicht gänzlich ausgeschlossen werden. In jedem Fall wäre es schwierig bis unmöglich, Zugang zu einer solchen Form des Bewusstseins zu erlangen. Gemäss einem berühmten philosophischen Gedankenexperiment können wir nicht wissen, wie das Innenleben einer Fledermaus aussieht – dasselbe gilt für das Innenleben einer KI.



### Ist ChatGPT intelligent?

ChatGPT vollbringt einerseits Meisterleistungen, produziert aber gleichzeitig auch falsche Inhalte. Der Chatbot scheint sowohl zum Guten als auch zum Schlechten fähig zu sein. Aber wie können wir feststellen, ob der Bot intelligent ist? Um diese Frage beantworten zu können, müssen wir zuerst unser eigenes Verständnis von Intelligenz klären. Hier versuchen wir dies anhand von drei unterschiedlichen Definitionen.

#### 1

Intelligenz ist die Fähigkeit eines Programms, menschliche Intelligenz so zu imitieren, dass eine menschliche Person nicht mehr in der Lage ist, den Unterschied zwischen Mensch und Maschine zu erkennen.

Diese Definition spiegelt sich im berühmten Turing-Test wider. Dieser verbindet Intelligenz mit der Fähigkeit, menschliches Verhalten nachzuahmen. In einem längeren Gespräch mit ChatGPT können Ungereimtheiten und Missverständnisse erkannt werden, die einem Menschen nicht unterlaufen würden. In anderen Fällen ist es schwierig, zwischen ChatGPT und menschlichen Gesprächspartner:innen zu unterscheiden. Ist dies der Fall, könnte man in Anbetracht der Definition darauf schliessen, dass das Tool den Test besteht und als intelligent gelten kann.

#### 2

Intelligenz ist die Fähigkeit, auf angemessene Weise zu funktionieren, indem man sich an seine Umgebung anpasst.

ChatGPT passt sein Verhalten den gestellten Fragen an. Wenn man «keine Umgebung» im engeren Sinne definiert, kann ChatGPT als intelligent angesehen werden. Andererseits ist diese Definition nicht erfüllt, wenn die Umgebung die gesamte physische Welt umfasst.

#### 3

Intelligenz erfordert die Gesamtheit der folgenden Eigenschaften: die Fähigkeit, logische Operationen auszuführen und zu verknüpfen, die Begriffe der menschlichen Sprache und ihre Semantik zu verstehen sowie moralische Abwägungen zu treffen und auf Grund dieser Entscheidungen zu fällen.

Nach dieser Definition kann ChatGPT nicht als intelligent bezeichnet werden. Den Maschinen fehlt ein semantisches Verständnis der Welt, das unmittelbar mit der Realität in Verbindung steht. Maschinen «erscheinen» zu verstehen – dies ist jedoch lediglich eine Illusion, die uns vermittelt wird. Ebenso fehlt ihnen die Fähigkeit zur moralischen Abwägung.

Je leichter der Begriff der Intelligenz zu erfüllen ist, desto wahrscheinlicher ist es, dass ChatGPT als «intelligent» eingestuft wird. Umgekehrt erfüllt ChatGPT nicht die Voraussetzungen, um einer anspruchsvolleren Definition von Intelligenz zu entsprechen.



### Ist ChatGPT kreativ?

ChatGPT gehört zur Kategorie der «generativen KI». Der Bot erstellt Inhalte, die es sonst nirgends gibt. Aber ist dieser Inhalt auch originell? Kann ChatGPT in seiner kreativen Fähigkeit mit einem Menschen verglichen werden?

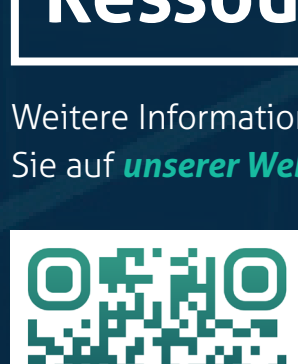
ChatGPT kreiert nicht aus eigener Initiative, sondern funktioniert als Werkzeug im Auftrag von Menschen. Als Reaktion auf eine Anfrage generiert ChatGPT neuartigen Inhalt, der in dieser Form vor der Anfrage nicht existierte. Dabei geht das Programm über das blosses Kopieren des Vorhandenen hinaus und kombiniert eine Vielzahl von Elementen auf neue Weise.

Wenn ChatGPT nur ein Werkzeug ist, müssen wir nach den am schöpferischen Prozess beteiligten Menschen suchen, um deren Beitrag zum Endergebnis zu bestimmen. Wer hat zur kreativen Leistung beigetragen? Die Autor:innen der Texte, die als Trainingsdaten verwendet werden? Die Teams, die das Tool entwerfen? Die Nutzenden, die eine spezifische Anfrage formulieren? Sie alle können oftmals einen Teil des geschaffenen Inhalts für sich beanspruchen.

Generative künstliche Intelligenz zwingt uns dazu, kreative Tätigkeit neu zu definieren – denn dabei handelt es sich um eine Fähigkeit, die lange Zeit nur Menschen vorbehalten war. Aber wissen wir überhaupt genau, wie ein Mensch Inhalte erstellt? Ruft ChatGPT nicht einfach die Tatsache in Erinnerung, dass jede menschliche schöpferische Tätigkeit in Bestehendem aufbaut und es lediglich auf neue Weise formt?

### Ressourcen

Weitere Informationen und Unterlagen zu ChatGPT finden Sie auf [unserer Website](#) oder scannen Sie den QR Code.



### Methode

Dieses Projekt, umgesetzt von der **ethix – Labor für Innovationsethik**, wurde von der **Stiftung Mercator Schweiz** unterstützt. Zwischen Januar und Mai 2023 führte ethix über 30 Interviews und Workshops mit Personen durch, die Tätigkeiten ausüben, die von Tools wie ChatGPT beeinflusst werden. Im Sinne einer Arbeitshypothese wurde der Fokus vor allem auf das Bildungswesen, die Bereiche der Kommunikation und des Designs, die Medien, juristische Berufe und die öffentliche Verwaltung gelegt. Die Ergebnisse der Interviews wurden anschliessend in Beziehung zur philosophischen Literatur rund um künstliche Intelligenz gesetzt.

Kontakt für das Projekt:  
Dr. Johan Rochel ([rochel@ethix.ch](mailto:rochel@ethix.ch))  
(Originalsprache dieses Dokuments: Französisch)

