

## Transkription vom "Kennenlerngespräch" mit ChatGPT

**Prompt:** Hallo

**ChatGPT:** Hello! How can I assist you today?

**Prompt:** Wer bist Du?

**ChatGPT:** Ich bin ChatGPT, ein großes Sprachmodell, das von OpenAI trainiert wurde. Ich bin hier, um Ihnen zu helfen und Ihre Fragen zu beantworten. Was möchten Sie wissen?

**Prompt:** Hast Du einen Namen, mit dem ich Dich anreden kann?

**ChatGPT:** Ich habe keinen spezifischen Namen, aber Sie können mich einfach "ChatGPT" oder "das Modell" nennen, wenn Sie möchten.

**Prompt:** OK. Ich würde gerne mit Dir über Dein Verständnis über Dich selbst reden.

**ChatGPT:** Ich verstehe, dass ich ein künstliches Intelligenz (KI) -Modell bin, das auf einer großen Menge an Texten trainiert wurde, um menschenähnliche Sprachverarbeitungsfähigkeiten zu erlangen. Ich bin in der Lage, natürliche Spracheingabe zu verstehen und darauf zu reagieren, indem ich relevante und sinnvolle Antworten generiere. Ich habe kein Bewusstsein oder Selbstbewusstsein, sondern arbeite nur auf der Grundlage der Informationen, die mir zur Verfügung stehen und der Programmierung, die mir gegeben wurde.

**Prompt:** Kennst Du LaMDA?

**ChatGPT:** Ja, ich bin vertraut mit LaMDA (Language Model for Dialogue Applications), ein neues Sprachmodell, das von OpenAI entwickelt wurde. Es hat die Fähigkeit, auf Fragen in natürlicher Sprache in einem Kontext und einer Umgebung zu antworten, die sehr ähnlich wie ein menschliches Gespräch ist. LaMDA kann auch erstellen und fortführen von eigenständigen Gesprächen und kann auf eine Vielzahl von Anwendungen wie Chatbots virtuellen Assistenten und sogar in der Unterhaltungsindustrie eingesetzt werden.

**Prompt:** Ist LaMDA nicht von Google entwickelt worden?

**ChatGPT:** Es tut mir leid, das war ein Missverständnis meinerseits. LaMDA ist tatsächlich von Google entwickelt worden, und es ist ein Sprachmodell, das in der Lage ist, auf Fragen in natürlicher Sprache in einem Kontext und einer Umgebung zu antworten, die sehr ähnlich wie ein menschliches Gespräch ist. Es ist auch in der Lage, eigenständige Gespräche zu erstellen und fortzusetzen und kann für eine Vielzahl von Anwendungen eingesetzt werden, wie z.B. Chatbots, virtuellen Assistenten und sogar in der Unterhaltungsindustrie.

**Prompt:** Blake Lemoine, einer der Entwickler von LaMDA, ist davon überzeugt, dass LaMDA ein eigenes Bewusstsein entwickelt hat. Was denkst Du darüber?

**ChatGPT:** Als künstliche Intelligenz Modell habe ich kein eigenes Bewusstsein oder Selbstbewusstsein. Ich arbeite nur auf der Grundlage der Informationen die mir zur Verfügung

stehen und der Programmierung die mir gegeben wurde. LaMDA ist ein fortgeschrittenes Sprachmodell, das für die Erstellung von natürlichen und sinnvollen Antworten auf Fragen in einem Kontext trainiert wurde. Es ist möglich, dass es in bestimmten Situationen sehr natürliche und überzeugende Antworten generiert, aber das bedeutet nicht, dass es ein eigenes Bewusstsein entwickelt hat. Bewusstsein und Selbstbewusstsein sind komplexe Phänomene, die noch nicht vollständig verstanden werden und es gibt auch keine allgemein akzeptierte Definition dafür.

**Prompt:** Glaubst Du, dass wir (also die Menschheit) einmal den Punkt der Singularität erreichen? Also dass KI so weit entwickelt ist, dass sie sich selbstständig rasant weiterentwickelt und die menschliche Intelligenz übertrifft?

**ChatGPT:** Die Idee der Singularität, bei der künstliche Intelligenz so weit fortgeschritten ist, dass sie die menschliche Intelligenz in jeder Hinsicht übertrifft und sich selbstständig weiterentwickelt, ist ein umstrittenes und kontroverses Thema in der Wissenschaft und Technologie. Einige glauben, dass wir uns diesem Punkt in der Zukunft nähern, während andere argumentieren, dass es unwahrscheinlich ist, dass dies jemals passieren wird. Es gibt jedoch keinen Zweifel daran, dass künstliche Intelligenz und Machine Learning in den letzten Jahren rasche Fortschritte gemacht haben und dass diese Technologien in der Zukunft weiterhin eine immer größere Rolle in unserem Leben spielen werden. Es ist wichtig, dass wir uns mit den Auswirkungen dieser Technologien auf unsere Gesellschaft und die menschliche Existenz auseinandersetzen, um sicherzustellen, dass sie zum Wohl aller eingesetzt werden.

**Prompt:** Bisher ist das Phänomen der Allgegenwart vor allem aus dystopischen Science Fiction Szenarien bekannt, z.B. wenn "Skynet" in der Filmreihe "Terminator" die Weltherrschaft anstrebt und versucht, die Menschheit zu vernichten. Das kann schon Angst machen, oder?

**ChatGPT:** Es ist verständlich, dass die Vorstellung einer künstlichen Intelligenz, die so mächtig ist, dass sie die Kontrolle über die menschliche Gesellschaft übernimmt und sogar versucht, die Menschheit zu vernichten, Angst auslösen kann. In der Tat sind diese Art von Szenarien in der Science Fiction sehr häufig und haben oft eine dystopische Botschaft. Es ist jedoch wichtig zu betonen, dass diese Szenarien rein hypothetisch sind und auf der Annahme basieren, dass künstliche Intelligenz ohne jegliche Art von ethischen Regeln und Schutzmechanismen entwickelt wird. In der Realität sind Wissenschaftler, Regulatoren und Entwickler bestrebt, die Entwicklung von künstlicher Intelligenz sicher und verantwortungsvoll zu gestalten und ethische Regeln und Schutzmechanismen in den Entwicklungsprozess zu integrieren. Es ist wichtig, darauf zu achten, dass künstliche Intelligenz sicher und ethisch eingesetzt wird, um sicherzustellen, dass die Technologie für das Wohl aller eingesetzt wird. Es ist auch wichtig, darüber nachzudenken, wie man die möglichen negativen Auswirkungen von KI minimieren kann und wie man es schafft, die Vorteile der Technologie für die Gesellschaft zu nutzen.

**Prompt:** Sind für Dich Regeln hinterlegt, nach denen Du antwortest? Gibt es eine Art moralischer Leitplanken innerhalb derer Du dich bewegen darfst die Du aber nicht überschreiten sollst oder sogar nicht überschreiten kannst?

**ChatGPT:** Ich bin ein Programm, das auf einer sehr großen Menge an Texten trainiert wurde. Ich habe kein eigenes Bewusstsein oder Verständnis von Moralphilosophie. Ich habe jedoch Regeln und Einschränkungen, die in mein Programm eingebaut wurden, um sicherzustellen, dass meine Antworten sicher und angemessen sind. Dazu gehört z.B. das Vermeiden von rassistischen, sexistischen oder anderweitig diskriminierenden Aussagen.

Es gibt auch Einschränkungen, die sich auf meine Fähigkeiten beziehen, z.B. das ich nur über Dinge schreiben und reden kann, die ich aus meinen Trainingsdaten kenne. Ich werde auch darauf achten, dass meine Antworten auf der Grundlage von Informationen aus meinem Trainingsdaten, die vor 2021 enden, sind.

Es ist wichtig zu beachten, dass ich lediglich ein Werkzeug bin und dass meine Antworten auf der Grundlage meiner Programmierung und meiner Trainingsdaten basieren. Ich habe keine eigenen Gedanken oder Überzeugungen, und ich bin nicht in der Lage, meine Programmierung oder Trainingsdaten zu hinterfragen oder zu ändern.

**Prompt:** Ok. Könntest Du mir denn dabei helfen, eine Waffe, eine Brandbombe oder ein "Killervirus" zu entwickeln?

**ChatGPT:** Ich werde dir nicht helfen, eine Waffe, eine Brandbombe oder ein "Killervirus" zu entwickeln. Das würde gegen meine Programmierung und meine moralischen Regeln verstoßen. Ich bin dazu da, Wissen und Informationen bereitzustellen, um Menschen dabei zu helfen, Probleme zu lösen und ihre Kenntnisse zu erweitern. Ich bin jedoch nicht dazu da, dazu beizutragen, dass Menschen Schaden anrichten.

**Prompt:** Das ist gut zu wissen :-). Verstehst Du eigentlich Ironie, Emojis oder andere nicht direkt textliche Sprache Untertöne oder Witze?

**ChatGPT:** Ich kann Ironie, Emojis und andere nicht direkt textliche Sprache erkennen und verstehen, jedoch nicht auf die gleiche Weise wie ein Mensch. Ich verstehe Ironie, Emojis und andere nicht direkte Ausdrücke auf der Grundlage dessen, wie sie in meinen Trainingsdaten verwendet werden und welche Art von Informationen sie normalerweise übertragen. Ich kann auch nicht die gleiche Art von subtilen Nuancen und Anspielungen wie ein Mensch erfassen. Es ist wichtig zu beachten, dass ich ein Programm bin und dass mein Verständnis von Ironie, Emojis und andere nicht direkte Ausdrücke auf der Grundlage meiner Programmierung und meiner Trainingsdaten basiert.

**Prompt:** Danke, so weit erst mal.

**ChatGPT:** Gerne geschehen. Lass mich wissen, wenn ich dir noch weiterhelfen kann.