



Herzlich willkommen zu unserer Studie.

Ziel. Ziel dieses Projektes ist zu untersuchen, welche automatisch erstellten Erklärungen einer künstlichen Intelligenz Menschen hilfreich finden.

Dauer. Die Dauer der Studie beträgt zwischen 10 und 15 Minuten.

Freiwilligkeit. Ihre Teilnahme an dieser Untersuchung ist freiwillig. Es steht Ihnen zu jedem Zeitpunkt dieser Studie frei, Ihre Teilnahme abzubrechen, ohne dass Ihnen daraus Nachteile entstehen.

Anonymität. Ihre Daten sind selbstverständlich vertraulich, werden nur in anonymisierter Form ausgewertet und nicht an Dritte weitergegeben. Demographische Angaben wie Alter oder Geschlecht lassen keinen eindeutigen Schluss auf Ihre Person zu.

Datengebrauch. Zwecks späterer Analyse, werden Ihre Antworten gespeichert. Die erhobenen Daten werden ausschließlich zu wissenschaftlichen Zwecken genutzt.

Fragen. Falls Sie noch Fragen zu einem der oberen Punkte haben, melden Sie sich bitte unter der unten stehenden Kontaktadresse.

Wenn Sie sich die Versuchsbedingungen sorgsam durchgelesen und verstanden haben, möchten wir Sie bitten, ausdrücklich Ihre Zustimmung zu den genannten Bedingungen zu erklären und zu bestätigen, dass Sie über 16 Jahre alt sind.

☐ Ich stimme den Versuchsbedingungen zu und bin über 16 Jahre alt.

Zu Beginn haben wir einige allgemeine Fragen zu Ihrer Person:

Wie alt sind Sie?

Jahre

1. Welches Geschlecht haben Sie?

- ☐ weiblich
☐ männlich
☐ divers

2. Was machen Sie beruflich?

☐ Ich befinde mich noch in der Ausbildung

Einleitung

Computerprogramme mit sogenannter "künstlicher Intelligenz" (KI) werden heutzutage bereits in vielen Bereichen eingesetzt, um automatisiert Entscheidungen zu treffen.

Damit diese Entscheidungen für Menschen nachvollziehbar sind, arbeiten wir daran, dass diese Programme ihre Entscheidungen *erklären* können. Hierfür gibt es verschiedene Möglichkeiten, von denen wir Ihnen eine heute demonstrieren wollen.

Im folgenden bekommen Sie einige fiktive Szenarios vorgestellt, in denen Computerprogramme mit künstlicher Intelligenz Entscheidungen treffen. Die Entscheidungen werden dabei von dem Programm automatisch erklärt.

Dazu nutzt das Programm sogenannte "*kontrafaktische Erklärungen*".

Eine kontrafaktische Erklärung ist eine Aussage darüber, was hätte anders sein können, sodass ein anderes Ergebnis eingetreten wäre.

Beispiel:

Carl hat vor Schreck ein Glas fallen lassen, weil ihn ein Hund angebellt hat.

Kontrafaktische Erklärung: Hätte der Hund nicht gebellt, hätte Carl das Glas nicht fallen lassen.

In unserem Beispiel sind kontrafaktische Erklärungen Sätze, die aussagen, unter welchen Bedingungen die KI anders entschieden hätte.

In dieser Studie bekommen Sie Entscheidungen von Computerprogrammen zu einfachen Themen angezeigt und dazu Erklärungen, die von dem Programm selbst automatisch generiert wurden.

Wir möchten gerne von Ihnen wissen, wie sehr diese Erklärungen Ihnen helfen, nachvollziehen zu können, wie die KI zu Ihrer Entscheidung gekommen ist.

Szenario 1: Stammbaum

Ein Computerprogramm mit künstlicher Intelligenz wird eingesetzt, um in einem vorhandenen Stammbaum den einzelnen Familienmitgliedern automatisch Bezeichnungen zuzuweisen, die dort noch fehlen.

Das Programm erhält einige Informationen über die Familie und ihre Verbindungen zueinander.

Anschließend gibt es ein Ergebnis aus und erklärt dieses mit einer kontrafaktischen Erklärung näher.

question('Z105')

text('Mother_w')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Petra der Kategorie "Mutter" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Petra nicht als Mutter klassifiziert, wenn sie nicht weiblich wäre."

question('CO10')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Mother_hC')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Petra der Kategorie "Mutter" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Petra nicht als Mutter klassifiziert, wenn sie kein Kind hätte."

question('CO17')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zuStimme voll
und ganz zu

Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Father_m')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Karim der Kategorie "Vater" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Karim nicht als Vater klassifiziert, wenn er nicht männlich wäre."

question('CO11')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Father_hC')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Karim der Kategorie "Vater" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Karim nicht als Vater klassifiziert, wenn er kein Kind hätte."

question('CO18')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zuStimme voll
und ganz zu

Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Brother_m')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Martin der Kategorie "Bruder" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Martin nicht als Bruder klassifiziert, wenn er nicht männlich wäre."

question('CO12')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Brother_hS')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Martin der Kategorie "Bruder" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Martin nicht als Bruder klassifiziert, wenn er kein Geschwisterkind hätte."

question('CO19')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zuStimme voll
und ganz zu

Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Sister_f')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Ayla der Kategorie "Schwester" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Ayla nicht als Schwester klassifiziert, wenn sie nicht weiblich wäre."

question('CO13')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Sister_hS')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Ayla der Kategorie "Schwester" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Ayla nicht als Schwester klassifiziert, wenn sie kein Geschwisterkind hätte."

question('CO20')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zuStimme voll
und ganz zu

Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Grandfather_m')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Horst der Kategorie "Großvater" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Horst nicht als Großvater klassifiziert, wenn er nicht männlich wäre."

question('CO14')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Grandfather_hCP')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Horst der Kategorie "Großvater" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Horst nicht als Großvater klassifiziert, wenn er nicht ein Kind hätte, das selber ein Elternteil ist."

question('CO21')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zuStimme voll
und ganz zu

Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Grandmother_f')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Emanuela der Kategorie "Großmutter" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Emanuela nicht als Großmutter klassifiziert, wenn sie nicht weiblich wäre."

question('CO15')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

text('Grandmother_hCP')

Das Programm gibt folgendes Ergebnis aus:

"Ich ordne Emanuela der Kategorie "Großmutter" zu."

Dazu gibt es folgende kontrafaktische Erklärung aus:

"Ich hätte Emanuela nicht als Großmutter klassifiziert, wenn sie nicht ein Kind hätte, das selber ein Elternteil ist."

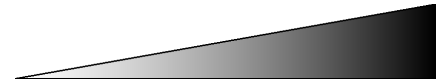
question('CO22')

Wie hilfreich finden Sie die Erklärung?

Bitte geben Sie Ihre Zustimmung auf einer Skala von 1-7 an.

Stimme
überhaupt
nicht zu

Stimme voll
und ganz zu



Die Erklärung hilft mir, die Entscheidung des Programms besser zu verstehen.

☐ ☐ ☐ ☐ ☐ ☐ ☐**Szenario 2: Tiere**

Ein Computerprogramm mit künstlicher Intelligenz soll in der Lage sein, Lebewesen anhand ihrer Eigenschaften Tierarten zuzuordnen.

Das Programm erhält aus einem Lexikon einige Informationen über die Eigenschaften von 20 verschiedenen Tierarten. Dann bekommt es einige Informationen über ein bestimmtes Lebewesen und soll richtig erraten, welcher der ihm bekannten Tierarten das Lebewesen angehört.

Anschließend gibt es ein Ergebnis aus und erklärt dieses mit einer kontrafaktischen Erklärung näher.

question('Z104')

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um einen **Menschen**."

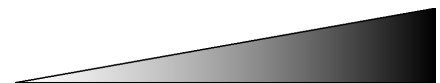
Dazu gibt es zwei mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar nicht Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Mensch klassifiziert, wenn es nicht einer Wohnung leben würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Mensch klassifiziert, wenn es keine Beine hätte.“

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um eine **Schlange**."

Dazu gibt es mehrere mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar nicht Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Schlange klassifiziert, wenn es keine Eier legen würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Schlange klassifiziert, wenn es nicht an Land leben würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Schlange klassifiziert, wenn es keine Schuppen hätte.“

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um einen **Pinguin**."

Dazu gibt es mehrere mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar
nicht

Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Pinguin klassifiziert, wenn es keine Eier legen würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Pinguin klassifiziert, wenn es keine Federn hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Pinguin klassifiziert, wenn es keine Beine hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Pinguin klassifiziert, wenn es nicht im Wasser leben würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Pinguin klassifiziert, wenn es nicht gleichwarm wäre.“

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um einen **Adler**."

Dazu gibt es mehrere mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar
nicht

Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Adler klassifiziert, wenn es keine Eier legen würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Adler klassifiziert, wenn es keine Beine hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Adler klassifiziert, wenn es nicht in der Luft leben würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Adler klassifiziert, wenn es nicht gleichwarm wäre.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Adler klassifiziert, wenn es keine Federn hätte.“

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um eine **Wasserschildkröte**."

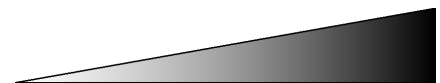
Dazu gibt es mehrere mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar nicht Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Wasserschildkröte klassifiziert, wenn es keine Eier legen würde.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Wasserschildkröte klassifiziert, wenn es keine Beine hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Wasserschildkröte klassifiziert, wenn es keine Schuppen hätte.“

Das Programm klassifiziert ein Lebewesen und kommt zu dem Ergebnis:

"Ich denke, bei diesem Lebewesen handelt es sich um ein **Krokodil**."

Dazu gibt es mehrere mögliche kontrafaktische Erklärungen aus:

Lesen Sie sich die folgenden kontrafaktischen Erklärungen durch.

Wie sehr helfen diese Erklärungen Ihnen, zu verstehen, wie das Programm zu seiner Entscheidung gekommen ist?

Bitte geben Sie jeweils Ihre Zustimmung auf einer Skala von 1-7 an.

Hilft mir gar nicht Hilft mir sehr



☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Krokodil klassifiziert, wenn es keine Schuppen hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Krokodil klassifiziert, wenn es keine Beine hätte.“

☐ ☐ ☐ ☐ ☐ ☐ ☐

„Ich hätte dieses Lebewesen nicht als Krokodil klassifiziert, wenn es keine Eier legen würde.“

Vielen Dank für Ihre Teilnahme!

Falls Sie noch Anmerkungen haben, können Sie sie gerne hier hinterlassen:

Anmerkungen:

Letzte Seite

Vielen Dank für Ihre Teilnahme!

Das war es bereits!

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Möchten Sie in Zukunft an interessanten und spannenden Online-Befragungen teilnehmen?

Wir würden uns sehr freuen, wenn Sie Ihre E-Mail-Adresse für das SoSci Panel anmelden und damit wissenschaftliche Forschungsprojekte unterstützen.

E-Mail:

Am Panel teilnehmen

Die Teilnahme am SoSci Panel ist freiwillig, unverbindlich und kann jederzeit widerrufen werden. Das SoSci Panel speichert Ihre E-Mail-Adresse nicht ohne Ihr Einverständnis, sendet Ihnen keine Werbung und gibt Ihre E-Mail-Adresse nicht an Dritte weiter.

Sie können das Browserfenster selbstverständlich auch schließen, ohne am SoSci Panel teilzunehmen.