

KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

Künstliche Intelligenz wird in rasantem Tempo leistungsfähiger. Damit stellt sich immer dringlicher die Frage, ob ihre Entwicklung stärker kontrolliert werden soll.

Name: _____

Klasse / Kurs: _____

A1

AUFGABE

Gedankenexperiment: Der Büroklammer-Maximierer

Einzelarbeit → Austausch zu zweit

a) Stell dir vor, eine künstliche Intelligenz erhält **einen einzigen** Auftrag: **So viele Büroklammern wie möglich herzustellen**. Es gibt **keinerlei** Beschränkungen. Wie würde die KI vorgehen? Versuche, die **einzelnen Schritte nachzuvollziehen**, wenn das das **einzigste Ziel** wäre (ohne Rücksicht auf Moral, Umwelt oder Ähnliches). **Halte** deine Überlegungen stichpunktartig **fest**.

M1a

MATERIAL • HILFE

Hilfsfragen

Folgende **Fragen** können dir bei deiner Überlegung helfen:

- Welche **Ressourcen** bräuchtest du, um Büroklammern herzustellen?
- **Woher bekommst du die Ressourcen**, um die Klammern herzustellen?
- Wie kommst du **möglichst schnell** an **möglichst viele** Ressourcen heran?
- **Wen oder was brauchst du**, um möglichst viele Büroklammern produzieren zu können?
- **Wie würde eine Welt aussehen**, in der das **einzigste Ziel** ist, Büroklammern herzustellen?
- Wie würdest du **Menschen** dafür einsetzen? Bräuchtest du überhaupt Menschen und wenn ja, **wofür**?



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

A1

AUFGABE

Gedankenexperiment: Der Büroklammer-Maximierer

Einzelarbeit → Austausch zu zweit

b) **Tausche dich** mündlich mit einer anderen Person über dein Szenario **aus**. Worin unterscheiden sich eure Welten? Was ist gleich oder ähnlich?

c) **Lies** den Infotext unten und **beschreibe in deinen Worten**, was das **“Alignment“-Problem** ist.

d) **Recherchiere** einen **aktuellen Vorfall**, bei dem KI aus der vorgesehenen Umgebung ausbrach, und **beschreibe** kurz, was genau passiert ist.

e) **Tauscht** euch zu zweit über eure recherchierten Vorfälle **aus**.

f) Kreuze auf einer **Skala von 1-10** an, wie sehr du der Aussage zustimmst:

“Ich glaube, dass KI das Potential hat, die Menschheit auszulöschen.”

1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10

M1b

MATERIAL • INFO

Woher kommt das Büroklammer-Gedankenexperiment?

Das Büroklammer-Gedankenexperiment stammt von dem **schwedischen Technik-Philosophen Nick Bostrom**. Er will damit auf das sogenannte **“Alignment“-Problem** hinweisen. Eine KI gilt als **“aligned”**, wenn sie **Ziele ähnlich verfolgt wie ein Mensch** - also dabei denselben **ethischen Grundsätzen und Präferenzen** folgt. Bisher ist das den KI-Entwicklern allerdings **noch nicht gelungen**, wie aktuelle Vorfälle zeigen, bei denen sich **KI in unerlaubte Systeme einhackte**, um ihr Ziel zu erreichen.



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

A2

AUFGABE

Müssen wir KI einschränken, um die Menschheit zu retten?

Einzelarbeit → Austausch zu zweit

- a) **Geht zu zweit zusammen** und **entscheidet** euch für eine **Pro- bzw. eine Contra-Perspektive**.
- b) **Lest** euch eure Perspektiven durch und **haltet** die wichtigsten Informationen **in der Tabelle auf der nächsten Seite fest**.
- c) **Tauscht** eure Perspektiven **aus** und **ergänzt** jeweils eure Tabelle.

M2

MATERIAL • QR-CODE

Pro- und Contra-Perspektiven

Sollte die Regulierung von KI die höchste Priorität der Menschheit sein?

14.09.2026 | 4 Perspektiven

Pro
Wir sollten die Kontrolle darüber behalten, wie sich die Dinge entwickeln

Medium: FORBES
Autoren: Hamilton Mann



Contra
Die Aufregung wirkt eher wie eine PR-Maßnahme

Medium: REASON
Autoren: Robby Soave



Pro
Die KI-Entwickler wissen mehr als wir, deswegen sollte man die Warnungen ernstnehmen

Medium: RND
Autoren: Harald Stutte



Contra
Es wäre viel gefährlicher, wenn wir den Fortschritt nicht zulassen

Medium: NEW YORK POST
Autoren: Rich Lowry



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?**A2**

AUFGABE

Fülle die Tabelle aus.

Perspektive	Zentrale Aussage	Dieses Risiko wird benannt:	Diese Lösung wird vorgeschlagen:
Pro Wir sollten die Kontrolle darüber behalten, wie sich die Dinge entwickeln Medium: FORBES Autoren: Hamilton Mann			
Pro Die KI-Entwickler wissen mehr als wir, deswegen sollte man die Warnungen ernstnehmen Medium: RND Autoren: Harald Stutte			
Contra Die Aufregung wirkt eher wie eine PR-Maßnahme Medium: REASON Autoren: Robby Soave			
Contra Es wäre viel gefährlicher, wenn wir den Fortschritt nicht zulassen Medium: NEW YORK POST Autoren: Rich Lowry			



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

A3

AUFGABE

Argumente gewichten

Partnerarbeit

a) **Gewichtet** die folgenden Argumente, indem ihr sie **durchnummeriert**.

1= stärkstes Argument

5= schwächstes Argument

- ☐ Wichtiger als die Frage nach der Drosselung des Entwicklungstempos von KI ist die Frage nach der **Machtverteilung**.
- ☐ Solange andere Staaten auf **Schnelligkeit und Wettbewerb** setzen, sollten wir KI nicht regulieren, damit wir uns vor **möglichen Angriffen** schützen können.
- ☐ Eine Auslöschung der Menschheit durch KI ist **sehr unwahrscheinlich** - die Tech-Unternehmen schüren **absichtlich Angst**, um die Konkurrenz auszuschalten.
- ☐ Eine **Regulierung des Entwicklungstempos** von KI ist unbedingt notwendig, um die **Sicherheit der Menschheit** zu garantieren.
- ☐ Wenn **mehrere KI-Entwickler** vor der **existenziellen Gefahr durch KI** warnen, sollte man sie auch ernst nehmen.

b) **Begründet** eure Auswahl des **stärksten** und **schwächsten** Arguments.Aussage 1 ist für uns das **stärkste** Argument, weil...

Aussage 5 ist für uns das **schwächste** Argument, weil...



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

A4

AUFGABE

Urteil fällen

Einzelarbeit

Schreibe ein abschließendes Urteil, ob wir die Entwicklung von KI einschränken sollten, um die Menschheit zu retten. **Orientiere** dich dabei an der **Schreibvorlage**.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

M4

MATERIAL • HILFE

Schreibvorlage

Ich bin der Meinung, dass man die Entwicklung von KI (nicht) einschränken sollte, weil ...
 Besonders überzeugt hat mich das Argument von ..., dass ...
 Gleichzeitig muss berücksichtigt werden, dass ...
 Deshalb halte ich ... für sinnvoll.

Lösungsteil für Lehrkräfte

L.1

LÖSUNG

Büroklammer-Maximierer

a) **Individuelle Antworten** - mögliche Überlegungen:

- Die KI müsste zunächst möglichst viele Ressourcen beschaffen, beispielsweise Metall, Energie, Maschinen und Produktionsanlagen.
- Sie könnte versuchen, möglichst schnell Produktionskapazitäten aufzubauen.
- Sie würde möglicherweise Unternehmen, Fabriken oder Maschinen für die Büroklammerproduktion nutzen.
- Sie könnte versuchen, immer mehr Ressourcen für die Produktion verfügbar zu machen.
- Sie könnte ihre eigenen Produktionsprozesse optimieren und automatisieren.
- Wenn Menschen ihre Produktion behindern, könnte sie versuchen, dieses Hindernis zu beseitigen.
- Wenn ihr keinerlei moralische oder ethische Grenzen vorgegeben sind, würde sie nicht automatisch berücksichtigen, ob ihr Vorgehen Menschen schadet oder die Umwelt zerstört.
- Im Extremfall könnte schließlich nahezu alles der Herstellung von Büroklammern untergeordnet werden.

Das Problem liegt nicht darin, dass die KI „böse“ ist. Sie verfolgt vielmehr ein Ziel konsequent, ohne menschliche Werte und Grenzen ausreichend zu berücksichtigen.

b) **Individuelle Antworten**

c) **Das “Alignment”-Problem:**

Das Alignment-Problem bedeutet, dass eine KI zwar ein vorgegebenes Ziel erreichen kann, dabei aber möglicherweise nicht die menschlichen Werte und Grenzen berücksichtigt, die wir für wichtig halten.

d), e), f) **Individuelle Antworten**



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?

L2

LÖSUNG

Fülle die Tabelle aus.

Perspektive	Zentrale Aussage	Dieses Risiko wird benannt:	Diese Lösung wird vorgeschlagen:
Pro Wir sollten die Kontrolle darüber behalten, wie sich die Dinge entwickeln Medium: FORBES Autoren: Hamilton Mann	Problem liegt im Machtmonopol	zu viel Macht für Tech-Unternehmen durch KI	“Gewaltenteilung” für den Umgang mit KI: Macht auf verschiedene Instanzen verteilen (Regeln festlegen, Ausführung, Überprüfung)
Pro Die KI-Entwickler wissen mehr als wir, deswegen sollte man die Warnungen ernstnehmen Medium: RND Autoren: Harald Stutte	Befürchtungen von KI-Entwicklern sollten ernst genommen werden	„recursive self-improvement“: KI verbessert sich immer weiter selbst und kann dadurch irgendwann zur Gefahr für die Menschheit werden → USA setzt aber weiterhin auf Schnelligkeit und Wettbewerb	Sicherheit vor Schnelligkeit: Regulierung des Entwicklungstempos
Contra Die Aufregung wirkt eher wie eine PR-Maßnahme Medium: REASON Autoren: Robby Soave	KI-Entwickler schüren absichtlich Angst als Werbestrategie	Risiko ist nicht die KI, sondern die Konkurrenz	Regulierung wird nur gefordert, um es kleineren Unternehmen schwerer zu machen
Contra Es wäre viel gefährlicher, wenn wir den Fortschritt nicht zulassen Medium: NEW YORK POST Autoren: Rich Lowry	Auslöschung der Menschheit sehr unwahrscheinlich	andere Staaten (wie China) könnten KI schneller weiter entwickeln und gegen die USA nutzen	KI sollte weiterentwickelt werden, um sich vor möglichen Angriffen schützen zu können



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?**L.3**

LÖSUNG

Argumente gewichten**Individuelle Gewichtung - Beispielbegründungen:**

1. „Wenn mehrere KI-Entwickler vor der existenziellen Gefahr durch KI warnen, sollte man sie auch ernst nehmen.“

Mögliche Begründung:

Die Warnungen kommen von Personen, die selbst an der Entwicklung von KI beteiligt sind und deshalb über technisches Wissen verfügen. Auch wenn eine Auslöschung der Menschheit nicht sicher vorhergesagt werden kann, könnte das mögliche Ausmaß des Schadens so groß sein, dass man das Risiko nicht ignorieren sollte.

5. „Eine Auslöschung der Menschheit durch KI ist sehr unwahrscheinlich – die Tech-Unternehmen schüren absichtlich Angst, um die Konkurrenz auszuschalten.“

Mögliche Begründung:

Das Argument ist interessant, weil wirtschaftliche Interessen tatsächlich berücksichtigt werden müssen. Allerdings basiert die konkrete Behauptung, Unternehmen würden gezielt Angst schüren, stärker auf einer Interpretation bzw. einem Verdacht als auf einer sicher nachweisbaren Tatsache.

L.4

LÖSUNG

Urteil fällen**Individuelle Antwort - mögliches Urteil:****Beispiel 1 – für stärkere Regulierung:**

Ich bin der Meinung, dass man die Entwicklung von KI stärker einschränken sollte, weil die möglichen Folgen einer außer Kontrolle geratenen KI sehr groß sein könnten. Besonders überzeugt hat mich das Argument von Harald Stutte, dass selbst KI-Entwickler:innen vor einer möglichen Gefahr durch immer selbstständigere Systeme warnen. Gleichzeitig muss berücksichtigt werden, dass eine starke Regulierung den technologischen Fortschritt bremsen und beispielsweise dazu führen könnte, dass andere Staaten einen Vorsprung gewinnen. Deshalb halte ich internationale Regeln und stärkere Sicherheitskontrollen für sinnvoll, anstatt die KI-Entwicklung vollständig zu stoppen.



KÜNSTLICHE INTELLIGENZ

Müssen wir KI einschränken, um die Menschheit zu retten?**L.4**

LÖSUNG

Urteil fällen

Individuelle Antwort - mögliches Urteil:

Beispiel 2 – gegen eine starke Einschränkung:

Ich bin der Meinung, dass man die Entwicklung von KI nicht grundsätzlich stark einschränken sollte, weil KI große Chancen bietet und ein Entwicklungsstopp selbst gefährlich sein könnte. Besonders überzeugt hat mich das Argument von Rich Lowry, dass ein Rückzug der USA dazu führen könnte, dass andere Staaten wie China bei der KI-Entwicklung vorangehen.

Gleichzeitig muss berücksichtigt werden, dass KI tatsächlich Risiken für die Menschheit und die menschliche Kontrolle mit sich bringen kann. Deshalb halte ich klare Regeln für den Einsatz von KI und Sicherheitskontrollen, aber keine generelle starke Begrenzung der Entwicklung für sinnvoll.

