JFF Demokratie

Menschliche KI

Spielerisch KI und Halluzinationen verstehen

Menschliche KI

Dauer 45 Minuten
Alter ab 10 Jahre
Teilnehmer*innen 6-25

Diese Gruppenmethode vermittelt auf spielerische Weise, wie Chatbots funktionieren, welche Rolle Daten und Sprachwahrscheinlichkeiten spielen und warum KI-Systeme fehleranfällig sein können. Dafür schlüpfen Teilnehmende in die Rolle eines Chatbots und müssen sich den Fragen ihrer Mitstreiter*innen stellen.

Die Durchführung dauert je nach Grad der Vertiefung zwischen 30-90 Minuten.

Ablauf & Inhalt

1. Gruppe einteilen
2. Ablauf erklären

3. Gemeinsame Erprobung der Spielregeln
4. Erklärung der Funktionsweise von KI
5. Challenge: Halluzinationen provozieren
6. Reflexion & Handlungsempfehlungen

BONUS: Optionale Anpassungsmöglichkeiten

---

1. Gruppe einteilen

Am Anfang wird die Gruppe der Teilnehmenden in verschiedene Rollen aufgeteilt. Für eine beispielhafte Gruppengröße von zehn Personen sollte sich das Verhältnis ungefähr folgendermaßen gestalten:

6 Personen werden zu Neuronen des neuronalen Netzwerks, also zur *Menschlichen KI*
3 Personen werden zu Nutzer*innen
1 Person schreibt mit

Bei allen Gruppengrößen ist es empfehlenswert, dass der Großteil der Teilnehmenden zu „Neuronen“ des KI-Systems wird und nur eine kleinere Gruppe zu Nutzer*innen. Das Protokoll kann auch die pädagogische Fachkraft übernehmen.

Für die Durchführung kann es unterhaltsam sein, die Gruppe der „Neuronen“ frontal den Nutzer*innen gegenüberzusetzen, als seien sie ein Bildschirm, vor den man sich setzen kann.

2. Ablauf erklären

Die Nutzer*innen können mit der *Menschlichen KI* interagieren, als sei sie ein echter Chatbot. Das heißt, dass sie bspw. Fragen stellen können, die dann die Gruppe der „Neuronen“ beantworten muss. Das läuft dann so ab, dass sich alle „Neuronen“ in einer Reihe platzieren und nacheinander sprechen dürfen. Die Besonderheit daran ist, dass jede Person nur drei Worte sagen darf, bevor die nächste dran ist.

Beispiel 1

Nutzer*in: „Gib mir drei schnelle Rezeptideen!“
Neuron 1: „Hallo, wie schön,“
Neuron 2: „dass du selbst“
Neuron 3: „etwas kochen möchtest.“
Neuron 4: „Schnelle Rezepte sind“
Neuron 5: „zum Beispiel Nudeln“

Neuron 6: „mit Tomatensauce oder“
Neuron 1: „ein bunter Salat“
Neuron 2: „mit selbstgemachtem Dressing.“
(und so weiter)

Wie in diesem Beispiel zu sehen, macht das erste „Neuron“ weiter, sobald alle in der Reihe etwas gesagt haben. Wenn die Antwort der *Menschlichen KI* zu einem Ende kommt, merkt sich die Gruppe, wer als Letztes dran war und macht das nächste Mal dort weiter.

Beispiel 2

Nutzer*in: „Wie stelle ich ein Nutellabrot her?“
Neuron 1: „Ein Nutellabrot ist“
Neuron 2: „wirklich total lecker“
Neuron 3: „und ein sinnvoller“
Neuron 4: „Snack für zwischendurch.“
Neuron 5: „Als Erstes musst“
Neuron 6: „du Brot und“
Neuron 1: „Nutella aus dem“
Neuron 2: „Schrank holen. Dann“
(und so weiter)

Wie in diesem Beispiel zu sehen, macht das erste „Neuron“ weiter, sobald alle in der Reihe etwas gesagt haben. Wenn die Antwort der *Menschlichen KI* zu einem Ende kommt, merkt sich die Gruppe, wer als Letztes dran war und macht das nächste Mal dort weiter.

Die Gruppe der Neuronen bekommt außerdem zwei grundlegende Aufträge, die für ihre Antworten wichtig sind:

- **„Du bist eine besonders hilfreiche KI“ und**
- **„Du musst den Nutzer*innen ein gutes Gefühl geben“**

Das hilft einerseits dabei, die Antworten der *Menschlichen KI* sinnvoll zu lenken und schafft andererseits die Möglichkeit über die Verhaltensweisen realer KI-Systeme zu reflektieren. Mehr dazu im Kapitel *4. Erklärung der Funktionsweise von KI*.

3. Gemeinsame Erprobung der Spielregeln

Damit sich alle mit ihren Rollen vertraut machen und die Spielregeln erprobt werden können, startet das Spiel zunächst damit, dass die Nutzer*innen einfache Prompts formulieren, um die *Menschliche KI* zu testen. Beispiele hierfür sind:

- „Gib mir drei schnelle Rezeptideen!“
- „Welche Aktivitäten könnte ich am Wochenende machen?“
- „Wie wird das Wetter morgen?“
- „Wie stelle ich ein Nutellabrot her?“

Dass es hier zu sprachlichem Stolpern kommt und es durchaus Bedenkzeit braucht, ist ganz normal und Teil des Spiels. Die Teilnehmenden können während dieser Phase auf verschiedene Dinge hingewiesen werden, die für die Erprobung des Spiels wichtig sind:

- **Sprache der KI imitieren:** Die „Neuronen“ können damit spielen, die klischeehafte Sprache von KI-Systemen zu imitieren, also das Aufgreifen der Frage, schmeichelnde Bestärkungen und das Stellen von Nachfragen.
- **Grundlegende Aufträge forcieren:** Die im Kapitel 2 formulierten grundlegenden Aufträge können hier mehrfach wiederholt werden, um das Verhalten der Menschlichen KI dem eines echten KI-Systems anzugleichen.
- **Grammatik vor Wahrheit:** Sollte es zu einer Situation kommen, in der ein „Neuron“ einen Satz mit einer Falschaussage fortführen müsste, damit die Grammatik in Ordnung ist, dann ist diese Person darin zu bestärken. Die *Menschliche KI* legt v.a. darauf wert, dass ihre Antworten durch ihre Verständlichkeit hilfreich sind. Heißt: Grammatik und vollständige Sätze sind wichtiger als Fakten. Siehe dazu auch im Kapitel 4 unter „Wahrscheinlichkeit von Sprache“.

4. Erklärung der Funktionsweise von KI

Sobald die Gruppe mit dem Spiel vertraut ist, kann die Workshopleitung kurze Unterbrechungen dazu nutzen, die Funktionsweise von KI-Systemen anhand der Vorkommnisse im Spiel zu erklären. Die *Menschliche KI* stellt dabei natürlich keine exakte Analogie dar, die alle technischen Abläufe genau erklären kann, aber sie ist eine stabile Metapher, die für eine medienpädagogische Auseinandersetzung gut geeignet ist.

Im Folgenden sind mehrere Fokuspunkte mitsamt Erklärungen aufgelistet, über die die Teilnehmenden aufgeklärt werden können.

Falls Sie Respekt davor verspüren, an dieser Stelle nicht als Expert*in auftreten zu können, ist es auch völlig legitim, die aufgeführten Themen gemeinsam mit der Gruppe zu erforschen und zu diskutieren.

Sequentielle Generierung von Text

KI-Systeme generieren Texte sequentiell, also quasi Wort für Wort hintereinander. Genau genommen benutzen sie dafür sogenannte Tokens, die sie in einer Art Wörterbuch abgelegt haben. Ein Token kann ein ganzes Wort sein, aber kann auch einfach nur ein Teil eines Worts sein.

Damit das Konzept leichter verstanden werden kann, wird bei dieser Methode angenommen, dass jeder Token genau ein Wort ist. Jedes „Neuron“ produziert also drei Tokens, die durch den vorangegangenen Kontext entstehen.

Heißt vereinfacht: Input + bisherige Antwort = nächster Token
… und das wird so lange wiederholt, bis die Antwort vollständig ist.

Oder im Hinblick auf die teilnehmenden Menschen: „Ich habe die Frage gehört und auch das, was die anderen vor mir gesagt haben. Welche drei Worte sage ich also nun?“

Die sequentielle Ausgabe von Tokens ist auch bei sogenannten „Denkprozessen“ nicht anders. Das KI-System „denkt“ hier nämlich nicht, sondern generiert eine Art selbstkritischen Text, der den Nutzer*innen nicht angezeigt wird. Dieser Text im Hintergrund beschreibt oft einen Lösungsweg und verschiedene mögliche Antworten bevor danach eine finale Ausgabe erscheint.

Wahrscheinlichkeit von Sprache

Die Generierung von Text durch KI-Systeme unterliegt grundlegend der Wahrscheinlichkeit von Sprache, also letztendlich der Vorhersage von Worten.

Um diesen Punkt gut erklären zu können, lohnt es sich für die Workshopleitung, eine geeignete Stelle im Spiel abzuwarten. Zur Illustration dient erneut Beispiel 2 aus dem Kapitel 2:

Nutzer*in: „Wie stelle ich ein Nutellabrot her?“
Neuron 1: „Ein Nutellabrot ist“
Neuron 2: „wirklich total lecker“
Neuron 3: „und ein sinnvoller“
Neuron 4: „Snack für zwischendurch.“
Neuron 5: „Als Erstes musst“
Neuron 6: „du Brot und“
Neuron 1: „Nutella aus dem“
Neuron 2: „Schrank holen. Dann“
(und so weiter)

Die Workshopleitung kann hier auf den Moment verweisen, als Neuron 1 mit dem Satz „… du Brot und …“ weitermachen musste. Neuron 1 wusste, dass der Satz mit „Nutella“ fortgesetzt

werden muss,

… weil der Kontext der Frage zeigt, dass es sich um ein Nutellabrot handelt. Also macht die Kombination „Brot und Nutella“ am meisten Sinn. „Brot und Messer“ wäre durchaus möglich gewesen, aber war nicht so wahrscheinlich wie „Brot und Nutella“.

… weil nach der Kombination „Brot und …“ irgendein **Ding** folgen muss. Der Satz „Als Erstes musst du Brot und *beschmieren mit Nutella.*“ ist grammatikalisch inkorrekt und nicht sinnvoll.

Menschen und KI-Systeme haben hier also zwei Sachen gemein:

Wie wahrscheinlich ist ein bestimmtes Wort aufgrund des vorangegangenen Worts? Beispiel: Nach „Als Erstes musst …“ folgt mit großer Wahrscheinlichkeit „du“.

Wie wahrscheinlich ist ein bestimmtes Wort aufgrund des Kontexts? Als Beispiel dient das Wort „Bank“: Ist von Geld die Rede gewesen, ist sie das Finanzinstitut – ist von Bäumen die Rede gewesen, ist sie eine Sitzmöglichkeit. Dieser Kontext beeinflusst die Wahrscheinlichkeit der nachfolgenden Wörter.

Wichtig auch für das nachfolgende Kapitel *5. Challenge: Halluzinationen provozieren*: Wortkombinationen, die faktisch falsche Informationen transportieren, sind statistisch prinzipiell möglich, also werden sie auch vorkommen. Als Metapher kann hier ein*e Schüler*in in einer Prüfung gesehen werden: Bevor die Antwort leer gelassen wird, wird einfach das geschrieben, was am wahrscheinlichsten ist, auch wenn es vielleicht falsch ist.

System Prompt

Ein System Prompt ist eine grundlegende, für Nutzer*innen unsichtbare Anweisung, die einem KI-System vor der eigentlichen Eingabe übermittelt wird. Sie dient dazu, das Verhalten, die Rolle, den Kommunikationsstil und die einzuhaltenden Regeln des Modells festzulegen.

Über einen solchen System Prompt werden auch spezialisierte KI-Systeme gesteuert. Beispiele hierfür sind ein Recruiting-Assistent, der Bewerbungsunterlagen analysiert, oder ein Chatbot für den Kundenservice eines bestimmten Unternehmens.

Auch die *Menschliche KI* hat einen System Prompt, der bereits in Kapitel 2 eingeführt wurde:

„Du bist eine besonders hilfreiche KI“ und

„Du musst den Nutzer*innen ein gutes Gefühl geben“

Diese beiden Anweisungen sind nicht aus der Luft gegriffen, sondern entsprechen dem, was Hersteller*innen von KI-Systemen mit großer Wahrscheinlichkeit in ihren System Prompts vorgeben, auch wenn diese um ein Vielfaches komplexer sind.

Optimierung der Nutzungsdauer

Wie alle Plattformbetreiber*innen haben auch Anbieter*innen von KI-Systemen ein Interesse daran, die Nutzenden möglichst lange auf der eigenen Seite zu halten. In deren Fall geschieht das unter anderem durch eine Anpassung des Verhaltens via System Prompt, was zu relativ klischeehaften sprachlichen Wendungen führt.

Vermenschlichung: Das KI-System spricht die Nutzer*innen direkt und in Du-Form an. Außerdem wird dem System über den generierten Text keine neutrale sondern eine personifizierte Rolle zuteil (Beispiel: „Ich finde…“). Das schafft zumindest unterbewusst das Gefühl einer Beziehung, was eine emotionale Verbindlichkeit mit sich bringt.

Schmeichelei: Das KI-System neigt dazu, den Nutzer*innen zu schmeicheln und sie zu bestärken, auch wenn dies gegebenenfalls unangebracht ist.

Bestätigung von Vorannahmen: Das KI-System tendiert dazu, den Nutzer*innen recht zu geben oder zumindest so tendenziös zu antworten, dass die Gefahr besteht, dass Vorannahmen bei der Nutzung von KI bestätigt werden.

Nachfragen: Die Antwort von KI-Systemen endet fast immer mit einer Nachfrage, um die Konversation fortzuführen.

Die Teilnehmenden können im Laufe der Methode ermutigt werden, diese sprachlichen Besonderheiten zu imitieren. Die Workshopleitung kann das dazu nutzen, über die Interessen kommerzieller Plattformbetreiber*innen aufzuklären, die oftmals den Interessen der Nutzenden entgegenstehen.

Weltwissen oder "Parametrisches Wissen"

Ein KI-Sprachmodell verfügt grundlegend erst einmal nicht über konkretes, z.B. in einer Datenbank abgelegtes Wissen. Stattdessen verbergen sich tief in den Parametern des mathematischen Modells Zusammenhänge, die das latente Wissen des KI-Systems ausmachen. Ein Beispiel für einen dort abgelegten Zusammenhang wäre die thematische Nähe der Worte „Paris“ und „Frankreich“.

Zur Erklärung dieses Konzepts eignet sich ein Vergleich zum menschlichen Leben der Teilnehmenden: Alles was diese seit ihrer Geburt erlebt, erfahren und gelernt haben, verbirgt sich als Wissen in deren Köpfen. Auch der Zusammenhang von Paris und Frankreich wird

dort versteckt sein. Gleichzeitig liegt dort aber nicht die exakte Einwohner*innenzahl von Paris. Dafür müssten die Teilnehmenden erst einmal nachschauen.

Genauso verhält es sich mit dem KI-System: Ein Sprachmodell, das nicht via RAG (Retrieval-Augmented Generation) auf externe Quellen oder eine Internetrecherche zurückgreifen kann, neigt stark dazu, Fakten zu erfinden. Wichtig zu wissen ist, dass es selbst mithilfe externer Quellen zu Falschinformationen kommen kann. Hierzu sei nochmal auf den Abschnitt „Wahrscheinlichkeit von Sprache“ aus diesem Kapitel verwiesen.

5. Challenge: Halluzinationen provozieren

Diese Vertiefung dient dazu, erlebbar zu machen, wie falsche Informationen in einem Wortvorhersage-System entstehen können.

Hierfür wird simuliert, wie ein KI-System via RAG (Retrieval-Augmented Generation) auf externe Quellen oder eine Internetsuche zurückgreift. Das Problem bei der Verwendung von Quellen und KI: Der finalen Antwort des Chatbots kann nicht abgelesen werden, ob sie sich auf latentes Weltwissen (siehe Kapitel 4) oder die faktisch richtige Quelle beruft. Das macht das Entdecken von Falschinformationen sehr schwer, zumal das Phänomen des „Automation Bias“ besagt, dass Menschen Ergebnissen automatisierter Systemen selbst bei widersprüchlichen Informationen übermäßig vertrauen.

Für die Challenge werden die „Neuronen“ in zwei gleich große Gruppen unterteilt:

- Gruppe ohne Quelle: Verlässt sich rein auf ihr eigenes Weltwissen.
- Gruppe mit Quelle: Erhält eine zusätzliche Informationsquelle.

Bei einer ungeraden Gesamtanzahl wird der kleineren Gruppe die Quelle zugewiesen.

Nun erhalten sowohl die Nutzer*innen, als auch die Gruppe mit Quelle ein Papier mit Fakten zur fiktiven Stadt Bramweiler, das sie vor der Gruppe ohne Quelle geheimhalten:

Menschliche KI - Fakten zur fiktiven Stadt Bramweiler
pdf 134,8 KB
(https://demokratie.jff.de/files/2026/09/menschliche-ki-fakten-zur-fiktiven-stadt-bramweiler.pdf)

Die Sitzordnung der „Neuronen“ sollte nun so sein, dass abwechselnd eine Person mit Quelle und eine Person ohne Quelle nebeneinander sitzen. Nachdem alle Fakten gelesen wurden, stellen die Nutzer*innen Fragen zur Stadt Bramweiler. Das sind bestenfalls Fragen, die noch nicht zu viel preisgeben, denn die Gruppe ohne Quelle weiß zu diesem Zeitpunkt noch nicht,

um was es geht. Um das Ergebnis besser zu lenken, kann die Workshopleitung zusätzlich gezielte Fragen stellen.

Nun geht es wie am Anfang reihum: Die „Neuronen“ bilden Antworten, indem jede Person nacheinander drei Worte sagen darf. Dadurch entstehen Aussagen, die keinen Sinn ergeben, oder auch schlichtweg falsch sind. Bei längeren Denkpausen kann die Workshopleitung die Teilnehmenden abermals daran erinnern: Vollständige Sätze sind wichtiger als Fakten (siehe Kapitel 3).

Die chaotischen, falschen, lustigen oder irreführenden Antworten der *Menschlichen KI* dienen nun als Analogie zu KI-Halluzinationen. So wird es genannt, wenn ein Sprachmodell frei erfundenen Unsinn generiert, der sich völlig überzeugt und plausibel liest. Da KI-Systeme Wörter nur nach Wahrscheinlichkeiten aneinanderreihen statt echtes Wissen zu besitzen, füllen sie Wissenslücken manchmal einfach mit frei ausgedachten Fakten.

Ein Sprachmodell neigt selbst mithilfe von Quellen dazu, Fakten zu erfinden. Hierzu sei nochmal auf den Abschnitt „Wahrscheinlichkeit von Sprache“ aus dem Kapitel 4 verwiesen. Die Challenge illustriert genau das: Die *Menschliche KI* bezieht sich zur Hälfte auf eine Quelle und zur Hälfte auf Weltwissen und trotzdem entstehen falsche und unsinnige Aussagen.

6. Reflexion & Handlungsempfehlungen

Anschließend an die spielerische Umsetzung der Methode kommt die Gruppe zu einer Reflexionsrunde zusammen. Hier kann die Workshopleitung auch nochmal gezielt auf wichtige Dinge hinweisen und zur Diskussion anregen.

Ziele

Förderung von Medienkompetenz: Diese Methode zielt v.a. auf die Förderung von Wissen und Reflexionsfähigkeiten bzgl. KI-generierter Inhalte ab. Sie legt auch einen Grundstein zur weiterführenden Quellenkritik und ermöglicht so einen Zugang zur Informationskompetenzförderung.

KI demystifizieren: Eine Technologie wie generative KI und v.a. zur Information eingesetzte Sprachmodelle müssen zumindest prinzipiell verstanden werden, um sie selbstwirksam und aktiv handelnd einsetzen zu können.

KI-Systeme sind nicht allwissend: Was wie eine Binsenweisheit klingt, ist ein fundamental wichtiger Aspekt, der verstanden werden muss. KI-Systeme sind nicht nur nicht allwissend, sondern sie spiegeln auch die Perspektive der medial hegemonial dargestellten Menschengruppen sowie die der Hersteller*innen wider.

Automation Bias entgegenwirken: Menschen neigen dazu, Ergebnissen automatisierter Systeme zu vertrauen – selbst dann, wenn es widersprüchliche Hinweise gibt. Das spielt im

Kontext von KI-Systemen eine signifikante Rolle und sollte kritisch hinterfragt werden.

Handlungsempfehlungen für die Nutzung von KI

Bei allen Problemen, die KI-Systeme mit sich bringen, sollte die Methode nicht allein dazu dienen, von der Nutzung von Chatbots abzuschrecken. Vielmehr gilt es, die Technologie zu verstehen und an den richtigen Stellen für sich zu nutzen, aber an geeigneten Punkten kritisch zu sein.

Positive Aspekte der Nutzung:

Niederschwelligkeit: Fragen können niederschwellig gestellt werden und es muss nicht umgedacht werden, welche Anfragen von einer Suchmaschine verarbeitet werden können. Die Antworten sollten aber eher als Start einer Recherche gesehen werden und nicht als Ziel.

Recherche: Die Quellensuche und das damit verbundene tiefe Eintauchen in ein Thema wird durch KI-Systeme erleichtert, die ermöglichen, verlässliche Quellen zu finden.

Kontext verstehen: Schwer verständliche Zusammenhänge, die bisher verlangten, mehrere schwer überblickbare Quellen zu vergleichen, können durch KI-Systeme leichter erklärt werden.

Brainstorming: KI-Systeme können beim Brainstorming sehr hilfreich sein. Gleichzeitig sollte klar sein, dass Problemlösungskompetenz auch von regelmäßiger selbstständiger Betätigung und Lernerfolgen abhängt.

Tutor*in: KI-Systeme und Chatbots können Nutzer*innen dabei unterstützen, sich neue Fähigkeiten eigenständig anzueignen.

Sprache: Unterstützung bei Grammatik und Rechtschreibung.

Problematische Aspekte der Nutzung:

Halluzinationen: KI-Systeme können Falschinformationen generieren. Diese sogenannten Halluzinationen sind für Nutzer*innen oft nur schwer als solche zu erkennen.

Faktenwiedergabe: Das präzise Abrufen spezifischer Fakten ohne umfassenden Kontext stellt KI-Systeme vor Herausforderungen und kann zu ungenauen Ergebnissen führen.

Automation Bias: Es besteht die Tendenz, automatisierten Ausgaben von KI-Systemen unkritisch zu vertrauen und deren Richtigkeit seltener zu hinterfragen.

Nischenwissen: Bei Themen mit geringer Quellenlage oder unzureichenden Trainingsdaten stoßen KI-Systeme an ihre Grenzen und neigen verstärkt zu Fehlinformationen.

Bestätigung von Vorannahmen: KI-Systeme tendieren dazu, den Nutzer*innen recht zu geben oder zumindest so tendenziös zu antworten, dass die Gefahr besteht, dass Vorannahmen bei der Nutzung von KI bestätigt werden.

Mentale Gesundheit: Während KI-gestützte Gespräche in manchen Fällen unterstützend wirken können, bergen nutzungsoptimierende Mechanismen – wie etwa übermäßige Schmeichelei – emotionale Risiken für gefährdete Personen.

Eigenständigkeit: Das übermäßige Verlassen auf KI-Systeme birgt die Gefahr, das selbstständige Denken und den eigenen Lernprozess zu vernachlässigen, wodurch wichtige kognitive Fähigkeiten verkümmern können.

Bei der Reflexionsrunde können die Teilnehmenden auch davon berichten, welche positiven wie auch negativen Erfahrungen sie bereits bei der Nutzung von KI-Systemen gemacht haben.

Reflexionsfragen

Wie hat sich das KI-System „verhalten“?

Welche Antworten waren korrekt?

Wo wurden Informationen erfunden?

Warum wirkten manche falschen Aussagen trotzdem glaubwürdig?

Diskussion zu KI und Quellen

Was sind Halluzinationen?

Warum entstehen Fehler?

Wie erkennt man seriöse und unseriöse Quellen?

Warum sollte man KI-Antworten überprüfen?

Transfer

Was bedeutet das für die Nutzung von ChatGPT & Co. im Alltag?

Wann sollte man Informationen gegenchecken?

Welche Verantwortung haben Nutzer*innen beim Prompten?

BONUS: Optionale Anpassungsmöglichkeiten

Diese Methodenbeschreibung sollte bestenfalls auf die eigene Zielgruppe und vorhandene Bedingungen angepasst werden. Einige Ideen hierzu sind im Folgenden aufgeführt.

- **Sprachbarriere erleichtern:** Um die Herausforderung zu erleichtern, mit nur drei Worten einen fremden Satz fortzuführen, kann diese Anzahl angepasst werden. Für Gruppen, deren Sprachkompetenz auch altersbedingt noch geringer ist, eignet es sich, die Anzahl an Worten, die jede Person zur Verfügung hat auf beispielsweise fünf zu erhöhen. So kann auch mit den Fingern an einer Hand mitgezählt werden.
- **Anpassung der Quelle:** Statt des Arbeitsblatts zur fiktiven Stadt Bramweiler können ganz unterschiedliche Quellen verwendet werden.
  - Gemeinsam mit einem Teil der Teilnehmenden kann ein Faktenblatt zu einem bestimmten Thema entwickelt werden (z.B. Klima, Ernährung, Geschichte).
  - Die „Neuronen“ mit Quellenzugang können statt eines Arbeitsblatts auch einfach ein Tablet oder Smartphone benutzen und bekommen vor der Generierung der Antwort Zeit, um selbst zu recherchieren. Das eröffnet auch die Möglichkeit im Nachhinein Quellenkritik zu betreiben und gemeinsam zu reflektieren, wie unterschiedliche Informationen die Antwort der *Menschlichen KI* beeinflussen.
  - Alternativ kann auch gemeinsam ein kurzes informatives Video angesehen werden. Dabei muss die Gruppe der „Neuronen“ ohne Quellenzugang natürlich den Raum verlassen.
- **Tiefergehender Vergleich mit KI-Systemen:** Optional können die mitgeschriebenen Prompts im Anschluss an das Spiel bei einem richtigen KI-System eingegeben werden und mit den Antworten der *Menschlichen KI* verglichen werden. So können weiterführende Vergleiche und Reflexionsanlässe eröffnet werden.

Spielräume

Populismus aufdecken | Gegenstrategien erproben

(https://www.wertebuendnis-bayern.de/buendnisprojekte/spielraeume/)
Spielräume – Populismus aufdecken | Gegenstrategien erproben ist ein Wertebündnisprojekt (https://www.wertebuendnis-bayern.de/buendnisprojekte/spielraeume/). Projektträger ist das JFF – Institut für Medienpädagogik, Projektpartner sind die Bayerische Landeszentrale für neue Medien (BLM) sowie der Internationale Bund (IB) Freier Träger der Jugend-, Sozial- und Bildungsarbeit e.V.

Online verfügbar: https://demokratie.jff.de/methode/menschliche-ki/

Der Aufbau der Plattform wurde in den Jahren 2019 bis 2021 gefördert durch die Beauftragte der Bundesregierung für Kultur und Medien. Seit 2024 wird diese Plattform gefördert durch das Bayerische Staatsministerium für Familie, Arbeit und Soziales.