

Szenarien für den KI-Einsatz in der Demokratiebildung

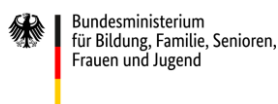
Szenario 7: KI als Chance: Gegenrede – Argumentationstraining gegen Hassrede

Die Schüler:innen setzen sich mit Hassrede auseinander. Mithilfe eines KI-Chatbots trainieren sie das Argumentieren gegen Ungleichwertigkeitsvorstellungen und reagieren auf Hassrede.

Hinweis: Dieses Szenario dient als Diskussionsgrundlage für Lehrkräfte. Es muss an die jeweiligen schulischen Rahmenbedingungen, Ressourcen und Kompetenzen der Lernenden angepasst und weiterentwickelt werden.

Lernziel(e):	<ul style="list-style-type: none"> Die Schüler:innen nutzen einen KI-Chatbot, um ihre praktische Handlungskompetenz (Argumentationskompetenz) im Umgang mit Hassrede zu fördern.
Zielgruppe:	Sekundarstufe I und II
Arbeitsform:	Partnerarbeit, Kleingruppen und Plenum
geförderte Kompetenzen:	<ul style="list-style-type: none"> Sprachliche Kompetenzen Empathie und Perspektivwechsel Soziale Kompetenzen Medienkompetenz
Voraussetzungen für die Durchführung:	<ul style="list-style-type: none"> Technische Ausstattung: Endgeräte mit Internetzugang, Möglichkeit zum Speichern/Exportieren von KI-Chats Emotionale Sicherheit & Gesprächskultur: klare Regeln, Schutzraum, sensibler Umgang mit persönlichen Erfahrungen Grundlagenwissen zu Hassrede: z. B. Definitionen, rechtliche Aspekte, gesellschaftliche Wirkung Selbstorganisation & Metakognition: eigenständige Recherche, Reflexion der emotionalen Wirkung von Hassrede und Gegenrede Kritische Bewertungskompetenz: Fähigkeit, KI-Aussagen auf Angemessenheit, Fehlinformationen und Sprachstil zu prüfen Datenschutz & Selbstschutz: anonymisierte Eingaben, keine persönlichen Betroffenheitserfahrungen als Promptinhalt Sozialstruktur: Partnerarbeit, moderierter Austausch, wechselnde Präsentations- und Reflexionsformen optional: Moderationskarten, Pinnwand, Online-Pinnwand

Gefördert vom:



Dieses Produkt ist unter der [Lizenz CC BY-SA 4.0](#) veröffentlicht. Ausgenommene Inhalte sind an den einzelnen Inhalten angegeben. Die Urheber:innen sollen bei der Weiterverwendung wie folgt angegeben werden: Achenbach-Carret, Auerswald, Kindlinger & Hahn-Laudenberg entstanden im Projekt ProKIS (Prozesse KI-bezogenen Wandels in der Schule) im Projektverbund SchuDiDe (Schulentwicklung: digital-demokratisch), lernen:digital Kompetenzzentrum Schulentwicklung.

ProKIS wird finanziert durch die Europäische Union – NextGenerationEU und gefördert durch das Bundesministerium für Bildung, Familien, Senioren, Frauen und Jugend. Die geäußerten Ansichten und Meinungen sind ausschließlich die des Autors/der Autorin und spiegeln nicht unbedingt die Ansichten der Europäischen Union, Europäischen Kommission oder des Bundesministeriums für Bildung, Familien, Senioren, Frauen und Jugend wider. Weder Europäische Union, Europäische Kommission noch Bundesministerium für Bildung, Familien, Senioren, Frauen und Jugend können für sie verantwortlich gemacht werden. Förderkennzeichen 01JA23E03B.

Schrittweises Vorgehen:

Schritt 0: Informationen durch Videoinput: Die Schüler:innen sehen ein kurzes Erklärvideo zu Hate Speech (z.B. MrWissen2go¹), das einen groben Überblick über das Thema gibt. Sie notieren dazu, was sie bereits wussten und was für sie neu war.

Schritt 1: Erfahrungen austauschen: Die Lehrkraft moderiert einen Austausch über persönliche Erfahrungen der Schüler:innen. Die Schüler:innen werden gefragt, wem schon einmal Hassrede begegnet ist und in welchen Situationen dies stattfand. Die Klasse führt ein Gespräch darüber, worin die Unterschiede (on- und offline) bestehen (Beispielfragen: Wie hast du dich in den Situationen gefühlt? Wie hast du reagiert? Wie würdest du gern reagieren? Welche Unterschiede gibt es on- und offline?).



Tipp: Da Schüler:innen in der Gruppe evtl. bereits selbst Opfer von Hassrede waren, eignet sich hier die offene und indirekte Frage nach der Begegnung mit dem Thema. Dies erlaubt es Schüler:innen, sich zunächst in der Beobachterrolle zu präsentieren und die eigene Betroffenheit nicht öffentlich darstellen zu müssen.

Schritt 2: Zu Hassrede recherchieren: Die Schüler:innen recherchieren in Kleingruppen vertieft zu Hassrede² und stellen diese Ergebnisse vor. Die Lehrkraft entscheidet, welche Fragen für ihre Klasse relevant sind und sucht die Texte dementsprechend heraus.

- Fragen zur Selbstreflexion: Was ist der Unterschied zwischen Beleidigung und Hassrede? Welche Menschen und Berufsgruppen sind am häufigsten davon betroffen? Welche Rolle spielen Medienkonzerne und deren Regularien? Welche Konsequenzen hat Hassrede für die Gesellschaft? Welche Rolle spielt KI zur Verbreitung von Hassrede? Welche juristischen Möglichkeiten gibt es? Welche Organisationen zur Unterstützung für Betroffene gibt es?*



Tipp: Die Ergebnisdarstellung kann durch Stichworte auf Moderationskarten oder auf einer Online-Pinnwand (z.B. Padlet oder Taskcards) erfolgen. Alternativ können auch Poster erstellt werden.

¹ ARD, „Hate Speech / Mirko Drotschmann“ in: „So geht Medien“, online unter: <https://t1p.de/o1v5p>, zuletzt abgerufen am 04.02.2025

² Beispielsweise diverse Kurztexte dazu: Amadeo-Antonio-Stiftung, „Hate Speech und Fake-News“, online unter: <https://t1p.de/xog0c>, zuletzt aufgerufen am 04.02.2025

Schritt 3: Gegenrede trainieren: Die Schüler:innen diskutieren mit dem KI-Chatbot zu Hassrede. Die Schüler:innen geben den unten stehenden Prompt in Partnerarbeit in einen KI-Chatbot ein. Sie erhalten von dem KI-Chatbot Beispiele für Hassrede und müssen Gegenargumente finden.

Der Prompt:

Du bist Trainer:in für Gegenrede. Ich habe in meinem Umfeld mit Aussagen gruppenbezogener Menschenfeindlichkeit zu tun und brauche ein Training gegen Hassrede. Bitte nenne mir ein Beispiel für Hassrede in Bezug auf das Thema XXX / die Themen XXX und warte auf mein Gegenargument. Gehe dann auf mein Gegenargument ein und führe ein weiteres Beispiel an. Achte darauf, dass meine Gegenargumente sachlich sind und nicht beleidigend. Gib mir ganz am Ende des Gesprächs bitte Feedback dazu. Nenne insgesamt 8-10 Argumente. Das Gespräch soll für Schüler:innen der XXX Klasse verständlich sein. Hast du das verstanden?³



Alternative Möglichkeiten für das Szenario

Anstatt allgemein von Hassrede zu sprechen, kann das Argumentationstraining auch spezifiziert werden (z.B. Antisemitismus). Dies bietet sich sogar an, da Hassrede in unterschiedlichsten Formen auftritt.



Achtung

Die Eingabe des gleichen Prompts kann immer wieder unterschiedliche Ergebnisse liefern. Daher sollte die Lehrkraft in den Kleingruppen prüfen, ob es funktioniert. Es kann zu Fehlern seitens des KI-Chatbots kommen.

Schritt 5: Training reflektieren: Nachdem die Schüler:innen Gegenargumente in Partnerarbeit formuliert haben, notieren sie stichpunktartig auf einer Online-Pinnwand (z.B. Padlet oder Taskcard), wie die Diskussion funktioniert hat sowie Auffälligkeiten oder Besonderheiten. Alternativ laden die Schüler:innen ihre Gespräche aus dem KI-Chatbot auf einer Online-Pinnwand hoch und können der Klasse bestimmte Stellen präsentieren.

Falls einzelne Schüler:innen Unterstützung benötigen, können diese einen KI-Chatbot um eine Zusammenfassung des Gesprächs bitten.

³ Die Nachfrage stellt sicher, dass der KI-Chatbot die Anfrage wiederholt und hier noch Missverständnisse und Fehler berichtigt werden können.

Es findet im Plenum nun ein Austausch über die geführte Diskussion statt.

Beispielfragen:

- Gab es Unterschiede und Gemeinsamkeiten bei den Diskussionen?
- Wie hat das Gespräch funktioniert?
- Wie habt ihr euch gefühlt?
- Gab es Argumente, die ihr nicht gut widerlegen konntet?
- Welche Argumente des Chatbots waren verharmlosend, vereinfachend oder schlichtweg unangemessen? Welche Kategorien lassen sich unterscheiden?
- In welchen Situationen würdet ihr Hassrede mit Argumenten widersprechen?
- Welche anderen Strategien gibt es, etwas gegen Hassrede zu tun?

Zusätzlich können die Schüler:innen aufgefordert werden, zwischen den selbst gefundenen/entwickelten und den auf Grundlage von Hinweisen den Chatbots gegebenen (Gegen-)Argumenten zu unterscheiden.

Diskussionsfragen für die ProKIS-Fortbildung:

- Wie kompetent geht die KI mit Hate-Speech-Inhalten um?
- Gab es Momente, in denen die KI unangemessen oder zu neutral reagiert hat?
- Wie kann man sicherstellen, dass die KI nicht selbst diskriminierende Muster reproduziert?
- In welchen realen Situationen wäre Gegenrede riskant – und welche anderen Optionen gibt es?
- Worin unterscheidet sich ein Dialog mit der KI von einem Dialog mit echten Menschen?
- Welche Verantwortung trägt die KI-Firma bei der Moderation von Hassrede?
- Wie kann man erkennen, ob die KI eine problematische Aussage lediglich „nachstellt“?

Begriffserklärungen:

Hassrede (übersetzt aus dem Englischen von „Hate Speech“): Hassrede bezeichnet abwertende und menschenverachtende Aussagen (verbal oder schriftlich), die zur Ausgrenzung von oder zum Aufhetzen gegen bestimmte Personen oder Gruppen führen soll.

Gruppenbezogene Menschenfeindlichkeit (GMF): Unter dem Begriff versteht man unterschiedliche Formen der Abwertung von Gruppen. Die Zuordnungen zu Gruppen kann aufgrund von Merkmalen wie Herkunft, Sexualität, Ethnie, Religion, sozialem Status etc. erfolgen. Dabei wird die Gleichwertigkeit aller Menschen zumindest implizit abgelehnt.

Literatur zu GMF und Hassrede:

Amadeu-Antonio-Stiftung, „Gruppenbezogene Menschenfeindlichkeit – Was ist das?“, online unter: <https://t1p.de/lvj0o>, zuletzt aufgerufen am 04.02.2025

Amadeu-Antonio-Stiftung, „Hate Speech und Fake-News“, online unter: <https://t1p.de/xog0c>, zuletzt aufgerufen am 04.02.2025

HateAid GmbH (2021). „Hass im Netz ist nicht Teil des Jobs. Für Kommunalpolitiker:innen und alle, die sich vor Ort engagieren, von Feuerwehr bis Flüchtlingshilfe.“, online unter: <https://t1p.de/r2sw6>, zuletzt aufgerufen am 04.02.2025

Weiterführende Lernangebote zu GMF und Hassrede:

Niedersächsische Landeszentrale für politische Bildung, „Die App KonterBUNT hilft, Stammtischparolen zu kontern“, online unter: <https://demokratie.niedersachsen.de/startseite/angebote/konterbunt/die-app-konterbunt-hilft-stammtischparolen-zu-kontern-177368.html>, zuletzt aufgerufen am 03.03.2025

Landeszentrale für politische Bildung Baden-Württemberg, „Hate Speech - Hass im Netz“, online unter: <https://www.lpb-bw.de/hatespeech>, zuletzt aufgerufen am 03.03.2025

Landeszentrale für politische Bildung Baden-Württemberg, „Mit Herz gegen hate speech! Online-Kurs zur Sensibilisierung für einen fairen Umgang miteinander“, online unter: <https://www.elearning-politik.de/mitherzgegenhatespeech>, zuletzt aufgerufen am 03.03.2025

Sponholz, Liriam (2024). Hate Speech und Inzivilität: Gegenmaßnahmen jenseits von Gegenrede. bpb. online unter: <https://www.bpb.de/themen/medien-journalismus/soziale-medien/545938/hate-speech-und-inzivilitaet-gegenmassnahmen-jenseits-von-gegenrede/#node-content-title-2>, zuletzt aufgerufen am 03.03.2025

Buchzik, Dana (2017). Strategien gegen den Hass. bpb. online unter: <https://www.bpb.de/themen/medien-journalismus/netzdebatte/260888/strategien-gegen-den-hass/#node-content-title-3>, zuletzt aufgerufen am 03.03.2025

Hofmann, Andy Alexander, „Hate Speech – Gruppenbezogene Menschenfeindlichkeit im Netz“, Themenblätter im Unterricht Nr. 118 (bpb), Download verfügbar unter: <https://www.bpb.de/shop/materialien/themenblaetter/270493/hate-speech-gruppenbezogene-menschenfeindlichkeit-im-netz/>, zuletzt aufgerufen am 03.03.2025