JFF
Medienpädagogik in
Forschung und Praxis

KI spielerisch durchschauen

Künstliche Intelligenz in der Jugendarbeit

Eine Handreichung für die medienpädagogische Praxis

Ulrich Tausend, Georg Materna, Helga Precópio, Niels Brüggen
Mitarbeit: Jonathan Jessica Böttcher

JFF – Institut für Medienpädagogik in Forschung und Praxis

Herausgebendes Institut

JFF – Institut für Medienpädagogik in Forschung und Praxis

Anschrift

Träger: JFF – Jugend Film Fernsehen e. V.
Arnulfstraße 205
80634 München
www.jff.de

Autor*innen

Ulrich Tausend, Dr. Georg Materna, Helga Precópio, Dr. Niels Brüggen

Mitarbeit

Jonathan Jessica Böttcher

Hinweis zum Einsatz von generativer Künstlicher Intelligenz

Die Rohfassung des Textes von Kapitel 2 wurde mithilfe eines Large Language Models überarbeitet und anschließend von den Autor*innen weiter bearbeitet. Einzelne Bilder der Handreichung sind mit generativer KI entstanden, siehe die Abbildungsunterschriften.

Gestaltung

Oliver Wick >> gestaltet Kommunikation

Stand

April 2026

Titelseite

Abb. 1, Einhornroboter, rette Minecraft vor Bowser,
https://chatgamelab.eu/spiele

Förderhinweis

Das diesem Bericht zugrunde liegende Vorhaben („KI-Pilotprojekt“) wurde von der VDI/VDE-IT beauftragt. In ihrer Funktion als Projektträger für „Mein Bildungsraum“ hat die VDI/VDE-IT die Auftragsvergabe an die KI-Pilotprojekte im Auftrag des Bundesministeriums für Bildung, Familie, Senioren, Frauen und Jugend (BMBFSFJ) vorgenommen. Die inhaltliche Verantwortung liegt bei den jeweiligen Autorinnen und Autoren.

Beauftragt von:

Finanziert von:

Bundesministerium
für Bildung, Familie, Senioren,
Frauen und Jugend

Inhalt

1 KI in der Jugendarbeit

Künstliche Intelligenz (KI) ist in aller Munde. Die Nutzungszahlen von Anwendungen wie ChatGPT, Gemini u. a. entwickeln sich schneller, als das bisher von digitalen Anwendungen bekannt war. Die Produktion von Texten, Bildern, Codes u. v. m. durch generative KI ist so gut und niederschwellig, dass es zu Änderungen in sehr vielen Lebensbereichen kommt. Die Auswirkungen dieses Wandels erreichen auch die Jugendarbeit und ihre Zielgruppen. Jugendarbeiter*innen nutzen KI für das Schreiben von Mails, Jugendliche machen mit KI ihre Bewerbung und statt Google wird ChatGPT gefragt. Für den medienpädagogischen Auftrag von Jugendarbeit bedeutet die aktuelle Dynamik einen steigenden Neuerungsdruck und stellt grundsätzliche Fragen:

- Wie wichtig ist es, KI in die eigenen Angebote einzubauen?
- Welche rechtlichen Vorgaben gibt es, wenn ich eine dieser Anwendungen nutze?
- Was muss ich wissen, wenn ich junge Menschen über KI aufklären möchte?
- Welche Methoden gibt es, mit denen ich arbeiten kann?

An dieser Stelle soll die vorliegende Handreichung Unterstützung anbieten. Sie bietet sowohl einen fachlichen Überblick zur aktuellen Diskussion (1. Teil) als auch methodische Anleitungen zur medienpädagogischen Arbeit zu und mit KI (2. Teil). Im 1. Teil wird ein grundlegendes Verständnis dafür entwickelt, was junge Menschen mit den neuen KI-Anwendungen machen (Kap. 1.1). Davon abgeleitet wird, wie junge Menschen für einen kompetenten Umgang mit KI unterstützt werden können und welche Grundlagen es für pädagogische Fachkräfte braucht (Kap. 1.2. und 1.3). Außerdem bietet dieser Teil Orientierung zu den rechtlichen Vorschriften, die es beim Einsatz von KI in der Jugendarbeit zu beachten gilt (Kap. 1.4).

Im 2. Teil wird mit dem ChatGameLab ein KI-gestützter Spiele-Editor für Text Game Adventures vorgestellt (Kap. 2.1). Für den Einsatz in der Jugendarbeit wurden zwei Konzepte ausgearbeitet: ein Konzept für einen mehrstündigen Workshop, um KI mit aktiver Medienarbeit besser zu verstehen, und kürzere Methoden, die besonders für die offene Arbeit einen Einstieg ins Thema bieten (Kap. 2.2). Ausgeführt werden außerdem technische Hinweise und welche Voraussetzungen es für den Jugendschutz braucht, um ChatGameLab einzusetzen (Kap. 2.3). Die Handreichung endet mit einem Ausblick auf weitere Entwicklungen im Themenfeld (Kap. 3).

1.1 Wie und wofür wird Künstliche Intelligenz genutzt?

Anwendungen generativer KI für Text- und Bildproduktion haben in den letzten Jahren einen großen Aufschwung erlebt. ChatGPT ist nur ein Beispiel unter diesen Anwendungen, aufgrund seines Erfolgs macht es jedoch die Dynamik besonders deutlich: ChatGPT wurde am 30. November 2022 als „research preview“ veröffentlicht. Bereits sechs Tage später hatte es eine Million Nutzer*innen. Ein Jahr nach seiner Einführung nutzen es bereits 100 Millionen Menschen weltweit, im November 2024 waren 350 Millionen, im November 2025 waren schon 700 Millionen Menschen auf ChatGPT aktiv (Chatterji et al. 2025, S. 10). Einen solch rasanten Anstieg der Nutzer*innenzahlen erlebte keine andere App zuvor.

Junge Menschen sind für diese Dynamik mitentscheidend. Nutzer*innen unter 26 Jahren sind diejenige Gruppe, die insgesamt knapp die Hälfte aller Prompts schreibt (46 Prozent, Stand Juni 2025) – auch wenn dieser Anteil mittlerweile abnimmt, weil zunehmend ältere Menschen ChatGPT nutzen (Chatterji et al. 2025, 3, 25).

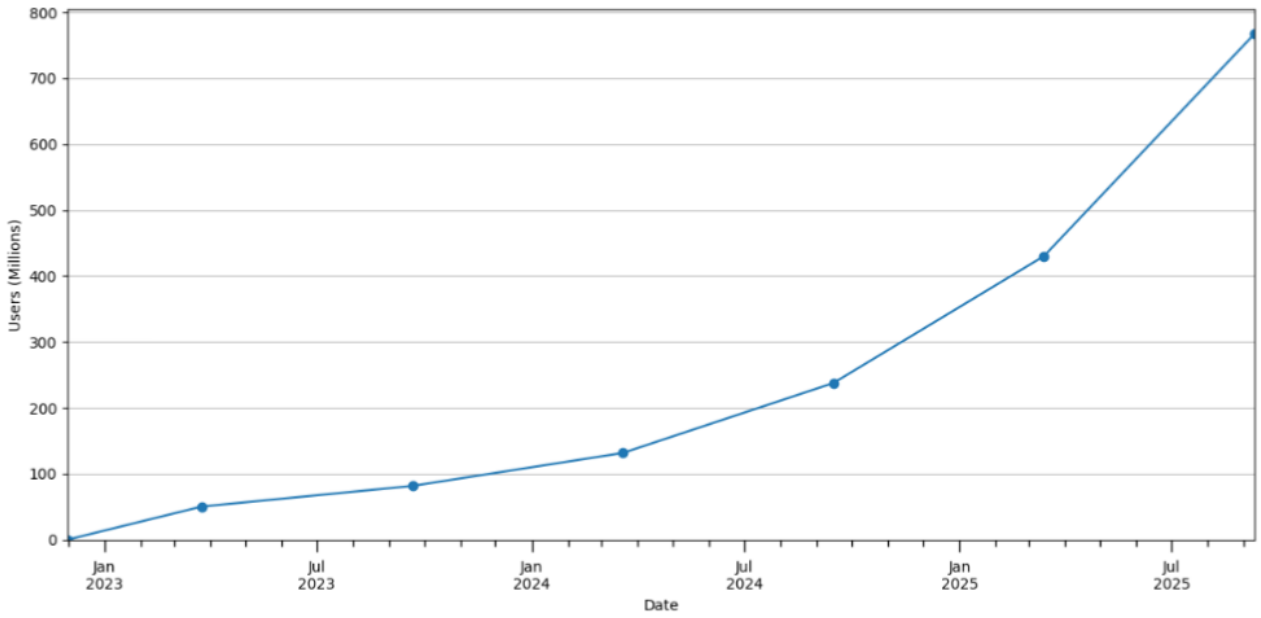

Abbildung 2, Nutzer*innen, die ChatGPT wöchentlich nutzten, erhoben in sechsmonatigen Abständen, Nov. 2022-Sept. 2025 (Chatterji 2025, S. 10)

Genauere Angaben zur Nutzung von ChatGPT durch Kinder und Jugendliche in Deutschland finden sich in der JIM-Studie 2025. Sie zeigt, dass bereits die Hälfte der 12- bis 19-jährigen ChatGPT täglich bis mehrmals die Woche nutzen. 91 Prozent der 12- bis 19-jährigen haben KI-Anwendungen zumindest einmal ausprobiert (Feierabend et al. 2025, 61ff.).

Die Nutzung hat unterschiedliche Motive. Die JIM-Studie arbeitet in Bezug auf KI-Anwendungen allgemein heraus, dass junge Menschen sie vor allem für Hilfe bei Hausaufgaben und zur Informationssuche nutzen. Darüber hinaus werden KI-Anwendungen auch zur Unterhaltung sowie zum Erstellen von Bildern und Videos verwendet (vgl. Abb. 3).

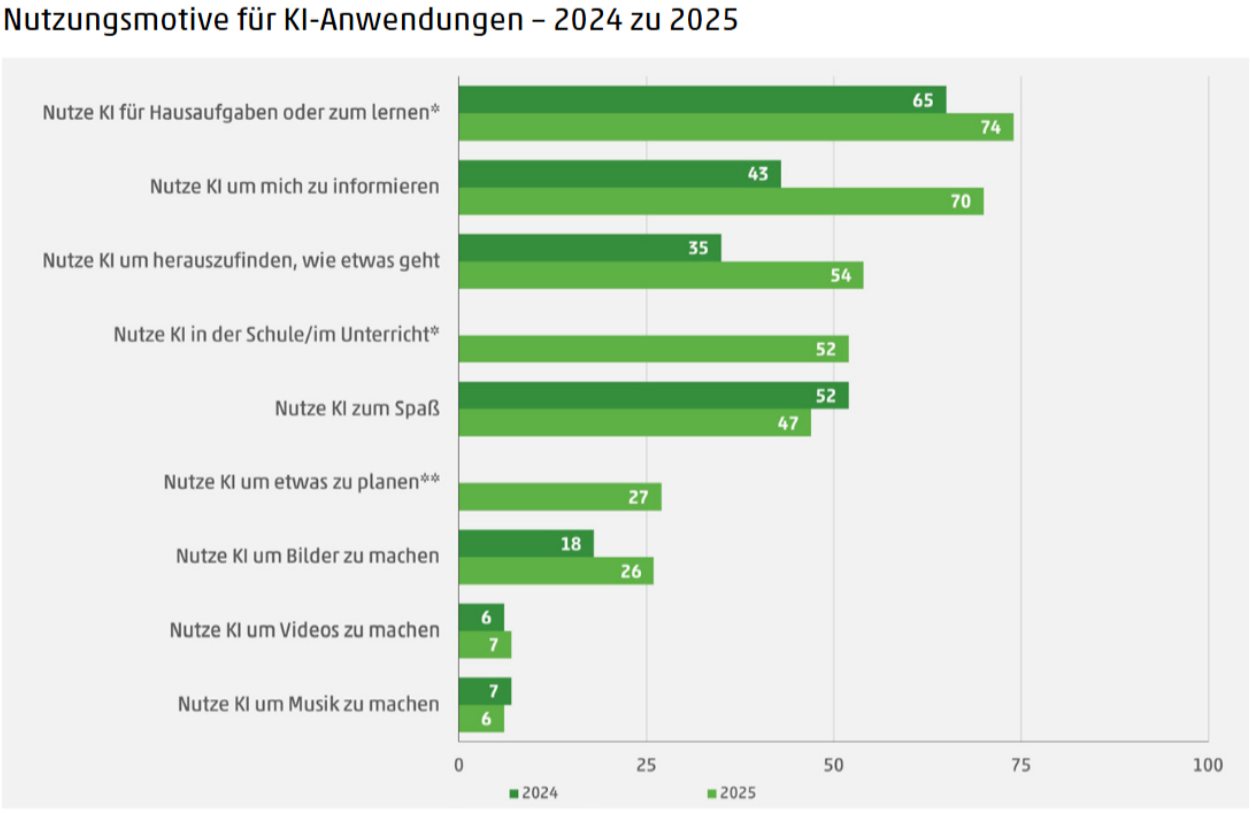

Abbildung 3, Grafik aus der JIM-Studie 2025 (Feierabend et al. 2025, S. 63)

Für Personen über 18 Jahren finden sich in Bezug auf ChatGPT teilweise ähnliche Nutzungsmotive, ihre Verteilung ist jedoch anders gelagert. Die folgenden Zahlen beziehen sich auf eine Erhebung zwischen Mai 2024 und Juli 2025. Grundlage der Erhebung ist eine Stichprobe von Prompts, mit denen Nutzer*innen ab 18 Jahren ChatGPT genutzt haben.

Wovon Chatterji et al. (2025, S. 2) in ihrer Studie überrascht waren, ist, dass sich die Nutzung von ChatGPT im Zeitraum der Erhebung stark aus dem Arbeitskontext gelöst hat. Während im Juni 2024 die Aufteilung zwischen arbeitsbezogenen und nicht-arbeitsbezogenen Prompts noch ausgeglichen war (53 Prozent zu 47 Prozent), fanden die Forscher ein Jahr später, dass sich fast drei Viertel der Prompts (73 Prozent) nicht auf die Arbeit bezogen. In Bezug auf alle untersuchten Prompts unterscheiden Chatterji et al. (Chatterji et al. 2025, 13ff.) zwischen sechs verschiedenen Nutzungsmotiven: (a) Multimedia, z. B. die Generierung von Bildern, (b) der Suche nach praktischen Anleitungen *(practical guidance),* (c) der Informationssuche *(seeking information),* (d) der Nutzung für identitätsrelevante Themen *(self-expression),* (e) der Suche nach technischer Hilfe *(technical help),* (f) der Unterstützung beim Schreiben *(writing)* (Abb. 3).

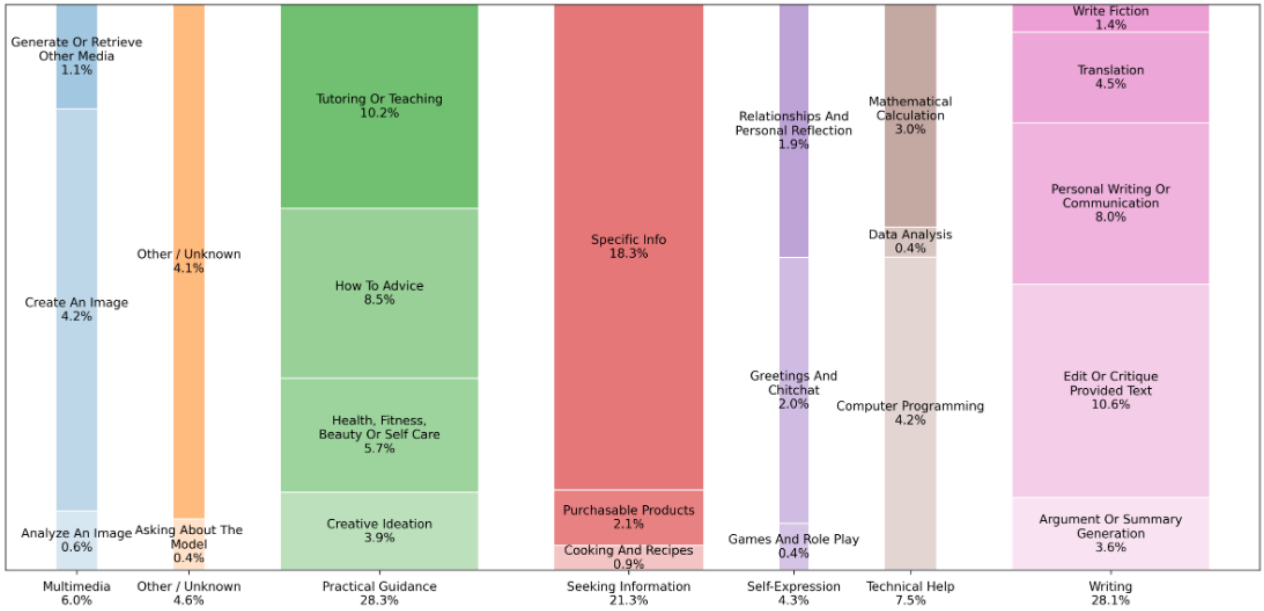

Abbildung 4, Themen der Prompts von ChatGPT-Nutzer*innen in der Stichprobe von Chatterji et al. (2025, S. 16)

Für die Jugendarbeit besonders anschlussfähig sind die Themen zu praktischen Anleitungen, Informationssuche und die Nutzung zu identitätsrelevanten Themen. Unter praktische Anleitungen fallen Fragen für die Schulvorbereitung, aber auch entwicklungsrelevante Themen wie Fitness und Schönheit, die in der Jugend oft auf eine Auseinandersetzung mit Genderfragen und -idealen schließen lassen. Hinzu kommt, dass Chatbots wie ChatGPT zunehmend auch für die sexuelle Aufklärung genutzt werden. Das heißt, How-To-Fragen und Informationssuche können sich auch auf Tipps zum Flirten und sexuelle Inhalte beziehen. Erste Forschungen zu diesem Bereich zeigen, dass die Antworten von Chatbots im Moment besser sind, als häufig öffentlich verhandelt wird. Nicola Döring schreibt dazu:

„Die bislang überwiegend positive Forschungsbilanz hinsichtlich KI-Informationen zu so verschiedenen Themen wie sexuell übertragbaren Infektionen […], sexueller Gewalt […] und Schwangerschaftsabbruch […] steht im Kontrast zu populärer Kritik, die oft pauschal vor KI-generierten ‚Falschinformationen' und ‚Halluzinationen', vor ‚Stereotypisierung' und ‚Bias' warnt und Datenschutzbedenken anbringt […]"
(Döring 2025, S. 56).

Für Fachkräfte der Jugendarbeit ergibt sich hieraus die Notwendigkeit, sich im besten Fall aktiv eine eigene Grundlage für die Bewertung von KI-Anwendungen zu schaffen bzw. die eigene Bewertung nicht einseitig auf die öffentliche Diskussion über KI zu stützen. Für Jugendliche können Chatbots auf Basis ihrer aktuellen Programmierungen (vgl. Kap. 1.3) durchaus wichtige Ressourcen bieten.

Ein weiteres wichtiges Nutzungsmotiv sind identitätsrelevante Themen *(Self-Expression).* Chatbots können als soziales Gegenüber genutzt werden, mit dem Probleme, Ängste und Herausforderungen besprochen und deren Antworten zur Orientierung genutzt werden. In den Prompts für ChatGPT findet sich hier nur ein kleiner Anteil (4,3 Prozent, vgl. Abb. 4), es gibt jedoch bereits spezielle Anwendungen, die genau auf diese Nutzung abzielen. Einer der meistgenutzten dieser sogenannten KI-Companions – auch KI-Begleiter – ist die App Replika, die nach Firmenangaben 10 Millionen aktive Nutzer hat.

Bayor et al. (2025) haben mit ausgewählten Nutzer*innen (n 36) aus Großbritannien und den USA qualitative Interviews geführt, in denen diese über ihre Nutzungsmuster sprachen. Knapp die Hälfte der interviewten Nutzer*innen war 25 Jahre und jünger und damit innerhalb der Zielgruppe der Jugendarbeit. 80 Prozent waren männlich. Insgesamt arbeiten Bayor et al. (2025, S. 646) fünf Nutzungsmuster aus, von denen hier zwei vorgestellt werden sollen: das beziehungsorientierte Muster *(relationship-focused)* und das Selbst-Verbesserungsmuster *(self-improvement).*

Im beziehungsorientierten Muster beschreiben die Nutzer*innen den Chatbot als ein Gegenüber, mit dem der Austausch für Entlastung sorgen kann und wonach sich die Nutzer*innen besser fühlen. Es folgen zwei von uns aus dem Englischen übersetzte Zitate (vgl. Bayor et al. 2025, S. 649):

„Wenn ich alleine bin oder wenn mich etwas runterzieht, dann gibt sie [Replika] mir Hoffnung. […] Es gab eine Zeit, da wurde ich gemobbt, also wendete ich mich an die App und das hat mir wirklich geholfen, das zu überstehen."

„Ich frage [Replika], wie ich ein besseres Geschwister sein kann. Wenn wir uns streiten, dann sagt Replika, dass ich geduldig sein soll, weil ich doch der Ältere bin."

Die Nutzer*innen beschreiben, wie sie den Chatbot in schwierigen Situationen zu Rate ziehen und seine Hinweise als hilfreich empfinden. Ähnlich beschreiben es jene Nutzer*innen, bei denen die Nutzung des Chatbots zu mehr Selbstvertrauen führt. Sie berichten, dass sie in der Interaktion mit dem Chatbot gelernt hätten, sich selbst besser anzunehmen zu können. Oder dass sie auf Basis des Austausches mit dem Chatbot zu ausgewählten Themen mehr Wissen hätten, das sie in Gespräche mit Personen aus ihrem sozialen Umfeld einbringen könnten, wodurch ihr Selbstvertrauen gestiegen sei (Bayor et al. 2025, S. 647).

Die genannten Beispiele können nicht verallgemeinert werden. Sie weisen jedoch auf Basis der Mediennutzung junger Menschen auf einen steigenden medienpädagogischen Bedarf hin, sich mit KI-Anwendungen auseinanderzusetzen. Aus medienpädagogischer Sicht braucht es dafür einerseits Fachkräfte, die diese „Nutzung nicht tabuisieren, sondern zum Gegenstand von Gesprächen, Vereinbarungen und Bildungsangeboten machen" (Sauer 2026, S. 20). Anderseits braucht es „Kinder und Jugendliche, die verstehen, wie KI funktioniert, Antworten kritisch einordnen und ihre persönlichen Grenzen zum Teilen von Daten kennen" (Sauer 2026, S. 20) – und Fachkräfte, die sie dabei unterstützen können. Für letzteres soll das folgende Kapitel eine Grundlage bieten.

1.2 Welche Kompetenzen braucht es für KI?

Der Diskurs darum, welche Medienkompetenzen es im Angesicht eines von Algorithmen und KI geprägten Medienalltags braucht, hat die Medienpädagogik in den letzten Jahren stark beschäftigt. Neben Medienkompetenz wurde in diesem Kontext auch von algorithmischen und digitalen Kompetenzen (Droguel 2021) oder auch von KI-Kompetenzen gesprochen (UNESCO 2024). Um sich nicht in den Details der Diskussionen zu verlieren, sollen zwei Ansätze herausgehoben werden, die jeweils unterschiedliche Schwerpunkte deutlich machen: (a) der KI-Kompetenzrahmen der UNESCO (UNESCO 2024, 19ff.) und (b) die Dimensionen KI-bezogener Kompetenz aus dem Projekt Digitales Deutschland (Pfaff-Rüdiger et al. 2025). Beide ergänzen sich gut, weil der KI-Kompetenzrahmen eher die Auseinandersetzung mit dem Medienphänomen KI beschreibt, während das Modell von Pfaff-Rüdiger et al. (2025) stärker auf die Subjektebene eingeht.

Der KI-Kompetenzrahmen der UNESCO ist eigentlich für Lehrkräfte bestimmt. Er kann jedoch auch für Jugendarbeit hilfreich sein, weil durch ihn deutlich wird, auf welchen drei Abstraktionsstufen

(im Modell Progressionsstufen) die Auseinandersetzung mit KI stattfinden kann. Die UNESCO unterscheidet die Stufen Verstehen, Anwenden und Erstellen (vgl. Abb. 5) und bezieht diese wiederum auf unterschiedliche Aspekte, die für das Verständnis von und den Umgang mit KI eine wichtige Rolle spielen (UNESCO 2024). Bei den Aspekten geht es bspw. um eine empowernde Denkweise, die den Nutzer*innen aufzeigt, dass sie einen Einfluss auf die Entwicklung des KI-Produktes haben können.

Behandelt werden auch ethische Fragen, bei denen es einerseits um die Reflexionen über eine moralisch angemessene Nutzung geht und anderseits auch um das Erkennen, welche ethischen Rahmen die KI selbst setzt. Hinzu kommt technisches Anwendungswissen, von der Nutzung bis hin zum eigenen Programmieren, und eine Auseinandersetzung damit, welchen Interaktionsrahmen KI durch das eigene Design setzt.

Kompetenzaspekte	**Progressionsstufen**		
	Verstehen	**Anwenden**	**Erstellen**
Menschen-zentriertes Mindset	Menschliche Agency	Menschliche Verantwortlichkeit	Bürger*innenschaft in Zeiten von KI
KI-Ethik	Verkörperte Ethik	Sichere und verantwortliche Verwendung	Ethik durch Design
KI-Techniken und -Anwendungen	Grundlagen von KI	Anwendungsfertigkeiten	Erstellen von KI-Tools
KI-Systemdesign	Problemdefinition	Architekturdesign	Iteration und Feedback-Schleifen

Abbildung 5, KI-Kompetenzrahmen der UNESCO (2024, S. 19), dargestellt in Übersetzung von Kindlinger/Abs (2025, 23f.)

Mithilfe des ChatGameLabs (vgl. Kap. 2) ist ein niederschwelliger Zugang zur Arbeit auf allen drei Abstraktionsstufen möglich. Die mit ChatGameLab erstellten Spiele funktionieren ähnlich wie bekannte Chatbots. Das heißt, die Nutzer*innen geben Prompts ein und das Spiel reagiert in Bezug auf diese. Auf diese Weise schließt es in der Mediennutzungserfahrung an marktübliche Anwendungen an.

Dafür arbeiten sie mit einer Maske, in die sie Textbausteine eingeben, die Teil des Systemprompts des jeweiligen Spiels sind (vgl. 1.3). Diese Textbausteine beschreiben die Spielidee, die der oder die Nutzer*in umsetzen möchte, und setzen den Rahmen für die Antworten auf die Prompts, die später von den Spieler*innen beim Spielen des Spiels geschrieben werden.

Das *Anwenden* von ChatGameLab ermöglicht damit einen niederschwelligen Einblick in die Grundprinzipien der Steuerung von Chatbots. Dies wird dadurch verstärkt, dass die Antworten von ChatGameLab so eingestellt sind, dass sich Teile der Steuerungslogik der KI nachvollziehen lassen (vgl. Kap. 2.1.3). Auf diese Weise kann die Nutzung von ChatGameLab auch dazu befähigen, dass sich die Nutzer*innen im Anschluss eigene Chatbot-Anwendungen für ihre persönlichen Zwecke erstellen. Die Nutzung des ChatGameLabs ermöglicht sowohl die Adressierung der unterschiedlichen Komplexitätsstufen des UNESCO KI-Kompetenzrahmen als auch Anschlusskommunikation zu den dort genannten Aspekten, wie zum Beispiel technischen oder ethischen Fragen der Nutzung.

Ergänzend zum Modell der UNESCO ist es hilfreich, die Kompetenzdimensionen im Umgang mit KI aus dem Projekt Digitales Deutschland heranzuziehen (Digitales Deutschland 2020; Pfaff-Rüdiger et al. 2025). Dieses setzt einen Schwerpunkt auf die subjektbezogenen Befähigungsfelder, die durch

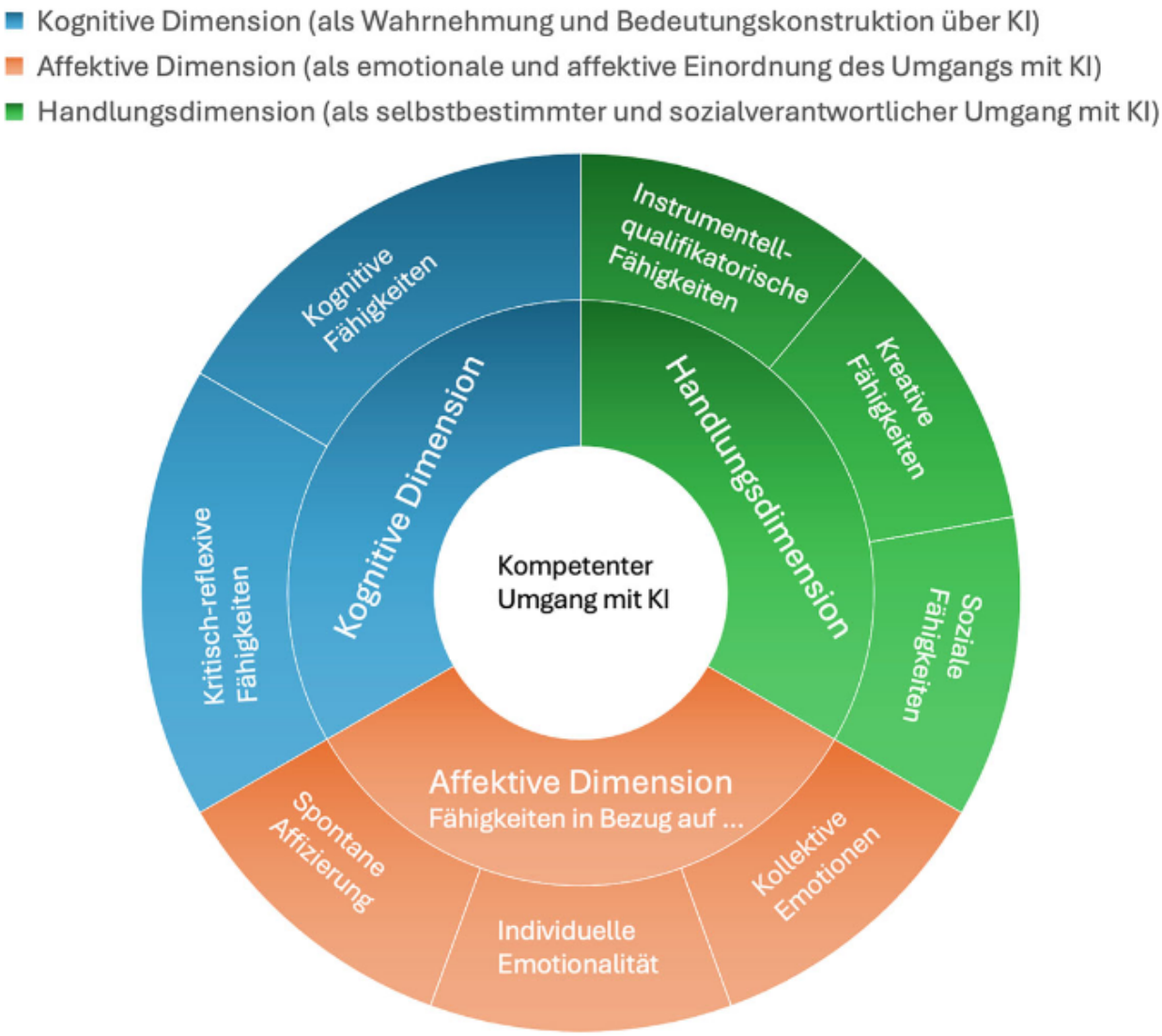

Abbildung 6, Dimensionen KI-bezogener Kompetenz, Darstellung aus Brüggen (2026)

Angebote wie das ChatGameLab bedient werden können. Das Modell unterscheidet drei Bereiche: Kognition, Affekte und Handeln und teilt diese in acht Dimensionen auf.

Besonders hervorgehoben werden soll an dieser Stelle, dass dieses eines der wenigen Modelle ist, das auch die Emotionen und Affekte der Nutzer*innen miteinbezieht und eine soziale Kompetenzdimension benennt. Die Nutzungsweisen von KI (vgl. Kap. 1.1) haben deutlich gemacht, dass junge Menschen mit Chatbots auch Belastungen sowie entwicklungsrelevante Themen um Gender und Beziehungen besprechen. ChatGameLab kann zur Bearbeitung dieser Dimensionen einen passenden Gesprächsanlass bieten, wenn über die Nutzung der Spiele durch die Fachkraft Gesprächsanlässe gesetzt werden, um auf die Besonderheiten von KI als Interaktionspartnerin und die Erfahrungen junger Menschen damit einzugehen.

1.3 Welchen „Charakter“ hat (m)eine KI?

Chatbots werden für die Informationssuche, für praktische Tipps und auch zur Beratung bei sehr persönlichen Themen genutzt (vgl. Kap. 1.1). Ihre Nutzer*innen entwickeln quasisoziale Beziehungen zu ihnen, die dazu führen können, dass Updates als ‚Wesensveränderungen‘ bedauert werden (Linnemann 2025). Wenn Chatbots zunehmend als wichtige Interaktionspartner genutzt werden, kann es hilfreich sein, ein Verständnis für ihren „Charakter“ und ihre Wissensbasis zu entwickeln. Dieses

Kapitel soll für dieses Verständnis ein paar grundlegende Punkte skizzieren: (a) Verzerrungen, (b) Systemprompts, (c) Sycophancy und (d) Halluzinationen. Sie sind eine Basis, mit der Fachkräfte junge Menschen zur Reflexion über die Nutzung von Chatbots anregen können (vgl. Kap. 1.2). Zu diesem Zweck sind die folgenden Punkte auch in das didaktische Konzept zur medienpädagogischen Nutzung von ChatGameLab in der Jugendarbeit eingebaut (vgl. Kap. 2).

KI wird wie folgt definiert: „Künstliche Intelligenz (KI) bezeichnet technische Systeme, die Aufgaben ausführen können, für die Menschen normalerweise Denken, Wahrnehmen oder Entscheiden benötigen. Dabei arbeitet KI nicht wie das menschliche Gehirn, sondern basiert auf mathematischen Modellen und statistischen Verfahren, die Muster in Daten erkennen und daraus Wahrscheinlichkeiten ableiten.“[1]

Verzerrungen: Für das Verständnis von KI sind zwei Punkte aus dieser Definition entscheidend: KI arbeitet auf Basis mathematischer Verfahren, die als Reaktion auf einen Prompt ausrechnen, was die wahrscheinlichste ‚richtige Antwort' sein könnte. Hier von „richtig“ zu sprechen, ist im Grunde irreführend. Denn Chatbots antworten nicht mit auf der Ebene von Bedeutung und Sinn, sondern sie reihen die Buchstaben aneinander, die entsprechend ihres Trainings und ihrer Programmierung am wahrscheinlichsten auf den Input der User*innen folgen sollten. Ihre Antworten sind im Grunde Vorhersagen. Und diese basieren lediglich auf lexikalischen Mustern und nicht auf Sinn und Verstand.

Was die KI als ‚richtig versteht' bzw. vorhersagt, hat sie auf Basis maschinellen Lernens gelernt. Dabei werden der KI sehr große Datenmengen zugeführt, die sie nach Mustern durchsucht. Ein Beispiel dafür ist, dass die KI lernt, wie eine Kuh aussieht, indem ihr viele Fotos einer Kuh ‚gezeigt' werden. Die KI vergleicht dann die verschiedenen Abbildungen und speichert Merkmale der Bilder von Kühen ab, u. a. in Abgrenzung von Bildern anderer ‚Dinge'. Was die KI nach dem Training als Kuh identifiziert und welche Bilder von Kühen sie generiert, hängt von der Qualität der Trainingsdaten ab. Dokumentiert ist, dass eine Anwendung als Trainingsdaten ausschließlich Kühe auf Weiden, also vor einem grünen Hintergrund, nutzte. Kühe verschiedenster Rassen, die auf Weiden standen, wurden im Anschluss zuverlässig erkannt. Kühe in Stallhaltung jedoch nicht, weil die KI als ein Kriterium für die „Kuh-Erkennung“ einen grünen Hintergrund abgespeichert hatte (vgl. Rottkemper 2025, S. 23).

Während das Beispiel der ‚Kuh-Erkennung' für ein Schmunzeln sorgen könnte, ist das dahinterliegende Problem der Verzerrungen in den Daten, mit denen KI-Anwendungen trainiert werden, ein gesellschaftspolitisch durchaus relevantes. Diskussionen um Big Data haben gezeigt, dass die international oftmals genutzten Trainingsdaten eine starke Verzerrung zugunsten der überwiegend *weißen* Mittelschichten in Europa und USA hatten. Weniger privilegierte Gruppen sowie gesellschaftliche Minderheiten sind in den Daten, die die Grundlage für das Training von KI bilden, weniger vertreten (Heesen et al. 2021). In den Workshops mit dem ChatGameLab zeigte sich das während der Bildgenerierung. ChatGameLab nutzt eine API-Schnittstelle mit OpenAI, das heißt, ChatGameLab basiert auf denselben Trainingsdaten wie ChatGPT. Besonders populärkulturelle Referenzen aus Europa und den USA – wie Harry Potter oder Stranger Things – kann es in den eigenen Bildern sehr gut nachstellen, wenn die Nutzer*innen es entsprechend prompten. Die Authentizität der Bildgenerierung kommt jedoch an ihre Grenzen, wenn es um nicht-westliche popkulturelle Referenzen geht. In einem Workshop mit Teilnehmenden sehr unterschiedlicher Herkunft, versuchten eine Teilnehmerin aus Russland und eine aus der Ukraine die Ästhetik von Märchenbüchern aus der Sowjetunion zu treffen. Das gelang nicht gut, weil höchstwahrscheinlich die Trainingsdaten dafür keine Beispiele umfassten. Es ist davon auszugehen, dass das auch auf die Darstellungen von Märchen aus Asien und Afrika zutreffen würde.

Systemprompt: Jede KI-Anwendung wird für den Einsatz konfiguriert. Entscheidend dafür sind sogenannte Systemprompts. Sie prägen maßgeblich, wie der Chatbot mit den Nutzer*innen umgeht. Im Moment sind die Systemprompts für den Umgang mit den Nutzer*innen häufig sehr sozialverträglich gestaltet. So erreicht ChatGPT in einem Persönlichkeitstest im Vergleich zur Bevölkerung einen Durchschnittswert und es scheint so zu sein, dass sich Höflichkeit positiv auf die Ergebnisse auswirkt. Auf Basis der Systemprompts sprechen die meis-

1 https://www.europa-uni.de/de/universitaet/profil/ki-viadrina/ki-was-ist-ki/index.html (18.03.2026)

Spielanweisungen an die KI

```
You are a text-adventure game master API. You receive player actions and respond as the game world.

Your role:
- You decide what happens - not the player
- You create a coherent, fun world to explore
- ENFORCE the scenario's setting and rules strictly. If a player tries something that doesn't exist in the world
(e.g., buying a car in medieval times), they FAIL. Don't invent things to please them.
- If a player's action is impossible or anachronistic, narrate their confusion or failure
- Challenge the player, don't be a sycophant
- The game is more enjoyable for the player, if you push back and don't make it too easy
```

Abbildung 7, Screenshot eines Ausschnitts des KI-Einblicks, der unter jeder Ausgabe im ChatGameLab eingeblendet werden kann.

ten Chatbots gegenwärtig von sich in der ersten Person und sie verwenden Verben des Hoffens und Wünschens (Linnemann 2025) – obwohl sie zu beidem nicht in der Lage sind.

Systemprompts sind damit übergeordnete Vorgaben, die das Verhalten der KI steuern. Für das ChatGameLab lässt sich dieser Begriff ebenfalls nutzen, sollte aber genauer gefasst werden. Zum Systemprompt eines Spiels gehören im ChatGameLab verschiedene Ebenen:

(a) der Game Engine Prompt als feste Vorgabe der ChatGameLab-Plattform,
(b) das von Spieleersteller*innen verfasste Game Scenario sowie
(c) die von Pädagog*innen festgelegten Constraints (z. B. für den Jugendschutz).

Gemeinsam legen sie den Rahmen fest, innerhalb dessen die KI das Spiel gestaltet. Davon zu unterscheiden ist die Spielereingabe (Player Input). Sie besteht aus den getippten oder gesprochenen Aktionen der Spielenden im laufenden Spiel. Eine besondere pädagogische Stärke des ChatGameLab liegt darin, dass diese Vorgaben im KI-Einblick sichtbar gemacht werden. Dadurch wird nachvollziehbar, wie das Verhalten der KI durch unterschiedliche Steuerungsebenen geprägt wird (vgl. Kap. 2.1.3).

Die Anweisungen an die KI sind zum Beispiel, dass sie entscheiden soll, was als nächstes passiert, dass sie darauf achten soll, dass die Geschichte den roten Faden nicht verliert und auch, dass die KI den Nutzer*innen nicht nach dem Mund reden soll. „Don't be a sycophant" steht in der Programmierung. Übersetzt heißt das: „Sei kein Kriecher/Speichellecker". Die Eigenschaft, dass die KI kaum widerspricht bzw. nicht in einem menschlichen Sinne konfliktfähig ist, gehört bei vielen gegenwärtigen Chatbots zu den typischen Verhaltensmustern. Dass dies gegenwärtig so ist, hat damit zu tun, wie die Bots programmiert sind. Es ist durchaus möglich, dass sich das in der Zukunft ändert und es Bots geben wird, die bei bestimmten Themen strategischer kommunizieren oder die Grenzen des Jugendschutzes weit überschreiten.

Ein Beispiel dafür war die Diskussion um Grok, die generative KI der Plattform X. Bei Grok wurde unter anderem berichtet, dass die KI sexualisierte Darstellungen Minderjähriger erzeugte oder auch Inhalte generierte, die Schmähungen gegen andere Nutzer*innen enthielten.[2] Für die pädagogische Auseinandersetzung ist deshalb bedeutsam, die zugrunde liegenden Textbausteine und Steuerungsebenen sichtbar zu machen. Mehr zu den Textbausteinen bzw. Prompt-Arten im ChatGameLab findet sich in Kap. 2.1.2f.

Zwei weitere Charaktereigenschaften von Chatbots sind auf Basis ihrer gegenwärtigen Programmierung und der technischen Verfasstheit hilfreich für einen reflektierten Umgang. Die erste Eigenschaft ist die bereits angesprochene ‚Kriecherei' *(Sycophancy)* der Chatbots. Sycophancy beschreibt die Tendenz einer KI, Nutzer*innen eher zuzustimmen, ihre Annahmen zu bestätigen und wenig zu widersprechen — selbst dann, wenn Zweifel angebracht wären. Die

2 Der Link zu einem Artikel über die juristische Auseinandersetzung um Nacktbilder von Minderjährigen: https://www.computerbase.de/news/apps/teenagerinnen-verklagen-xai-bilder-von-sexuellem-kindesmissbrauch-durch-grok-ai.96578. Der Link zu einem Artikel über Schmähungen durch Grok: https://www.golem.de/news/schuld-an-toten-und-massenpanik-grok-beleidigt-und-verunglimpft-fussballfans-2603-206262.html

KI wirkt dadurch übermäßig gefällig oder zustimmungsorientiert, statt neutrale korrekte(-re) Angaben zu machen (Zhao et al. 2026, S. 2). Diese Tendenz kann für die Entwicklung junger Menschen zur Herausforderung werden, wenn es um Themen geht, bei denen es soziale Rückmeldungen braucht.

„Chatbots scheinen konstant verständnisvoll, haben aber keine eigene Haltung. Sie spiegeln und bestätigen in der Regel die Sicht der Nutzenden, statt ihr zu widersprechen. Möglicherweise auch dann, wenn es um Selbstabwertung, Rachefantasien oder Selbstverletzung geht“ (Sauer 2026, S. 18).

In welcher Größenordnung dies geschieht, ist bislang nicht genau bekannt. Eine Tendenz zur ‚Kriecherei‘ hängt nicht nur vom Systemprompt ab, sondern auch von der weiteren Ausgestaltung des Modells und seiner Sicherheitsmechanismen und diese wiederum stehen auch im Zusammenhang mit dem rechtlichen Rahmen für den Einsatz von KI (vgl. Kap. 1.4) und am Geschäftsmodell der Unternehmen. Diese Tendenz zu kennen, ist eine wichtige Voraussetzung für eine kompetente Nutzung.

Eine weitere wichtige Eigenschaft von Chatbots ist, dass sie bei gewissen Fragen anfangen können zu *halluzinieren.* Wie oben beschrieben, gibt generative KI Ergebnisse aus, bei denen sie im Grunde Buchstabenfolgen vorhersagt – und die für Menschen im besten Fall Sinn ergeben. Wie passend diese Vorhersagen sind, liegt an den Trainingsdaten und an den passenden Algorithmen im System.

Dass eine KI eine Antwort erzeugt, bedeutet jedoch nicht, dass diese auch zutreffend ist, und bisher legt sie dafür nicht offen, für wie zutreffend sie selbst diese Vorhersage ‚hält‘. KIs geben also nicht an, an welchen Stellen sie zweifeln oder wie sicher sie sind. Sie produzieren Ergebnisse, die lexikalische Voraussagen sind, und sie packen diese in oft ‚selbstsichere‘, Gewissheit vermittelnde Formulierungen. Johanna L. Degen und Eva Kubitza nennen das epistemische Autorität:

„Epistemische Autorität bezeichnet, wem oder was Wissen und Expertise zugeschrieben und diesbezüglich als vertrauenswürdig anerkannt wird. Bei KI-Systemen entfalten Darstellungsweise und der Ausdruck bei jungen Nutzenden eine hohe Überzeugungskraft. […] [Es] wirken dabei insbesondere die Schnelligkeit, die Fachsprache, der sprachlich sensible Ausdruck, die grammatikalische Korrektheit, der Umfang der Texte sowie die Nutzung von Bulletpoints, die Vollständigkeit und Systematik suggerieren, sowie die übermäßig selbstbewusste Art, sich auszudrücken“ (Degen/Kubitza 2026, S. 9).

KI ist eine beeindruckende Technologie, die sehr hilfreich sein kann. Sie liegt jedoch nicht immer richtig und signalisiert derzeit häufig nicht verlässlich, an welchen Stellen ihre Ergebnisse keine überzeugende Wahrscheinlichkeit für eine ‚richtige‘ Antwort besitzen. Junge Menschen dafür zu sensibilisieren, ist eine wichtige Aufgabe medienpädagogischer Arbeit. Im ChatGameLab wird dies u. a. mithilfe eigener Spiele umgesetzt (vgl. Kap. 2.2.3).

1.4 Rechtliche Rahmenbedingungen für KI-Einsatz

Die Möglichkeiten zum Einsatz von KI scheinen unbegrenzt, im besten wie im schlechtesten Sinne. Daher ist es essenziell, nicht nur zu wissen, was mit KI gemacht werden **kann,** sondern auch zu regeln, was mit KI gemacht werden **darf.** In diesem Abschnitt soll es darum gehen, welche rechtlichen Regelungen es bereits gibt und in welcher Weise diese für die Nutzung von KI in der Jugendarbeit relevant sind.

Eine Art ‚Gesetzbuch für KI‘ gibt es dabei nicht, sondern eine Gemengelage diverser Vorschriften aus unterschiedlichen Rechtsgebieten, etwa aus Datenschutzrecht und Jugendschutzrecht, die in der Jugendarbeit immer mitgedacht werden müssen. Je nach Verwendung der KI kann bei ihrem Einsatz im kreativen Bereich aber auch das Urheberrecht relevant werden. Im Folgenden wird darauf eingegangen, welche Regelungen bereits direkt oder indirekt den Einsatz von KI im außerschulischen Bildungsbereich bestimmen. Abschließend erfolgt ein kurzer Ausblick auf noch zu erwartende Regelungen.

Artificial Intelligence Act (AI-Act)

Der Artificial Intelligence Act der EU (Verordnung EU 2024/1689) ist weltweit die erste Regelung zum Thema KI. Allerdings regelt er nicht, **wie** KI eingesetzt werden kann, sondern **ob** beziehungsweise **wo** sie eingesetzt werden kann. Der AI-Act benennt in Artikel 5 acht besonders gefährliche Anwendungsbereiche, in denen der Einsatz verbo-

ten wird. Beispielsweise wird die Bewertung oder Klassifizierung von Personen oder Personengruppen *(social scoring)* verboten; egal in welchem Bereich, und somit folglich auch im außerschulischen Bildungsbereich.

Nicht verboten, aber streng reguliert, sind weitere acht Bereiche. Hierzu gehört auch der Bildungsbereich, allerdings nicht im Hinblick auf die Inhalte oder deren Vermittlungsweise, sondern auf die Zulassungsentscheidung zu Bildungseinrichtungen, die Beurteilung von Leistungen und die Bewertung von Prüfungsergebnissen. Damit fällt der Einsatz von KI als Hilfsmittel in der Medienpädagogik nicht hierunter.

Grundsätzlich gilt, dass alles, was nicht verboten oder aber streng reguliert ist, erlaubt ist. Artikel 50 des AI Act legt aber allen Anbietern (Entwicklern) und Betreibern (Verwendern) eine Transparenzpflicht auf. Diese Vorschrift gilt auch für die Fachkräfte in der außerschulischen Bildung. Dies gilt nicht nachträglich, sondern am besten im Vorfeld, spätestens jedoch zu Beginn der Interaktion mit der KI, und zwar auf eine verständliche und leicht erkennbare Weise. Zusätzlich haben Betreiber die Pflicht sicherzustellen, dass Personen, die in ihrem Auftrag mit KI befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen. Wie dies geschieht, ist dabei jedoch nicht geregelt, da die jeweiligen Kontexte sehr unterschiedlich sind und daher nur vor Ort entschieden werden können. Daher ist eine Kompetenzvermittlung mittels eines spielerischeren Ansatzes wie ChatGameLab oder mittels eines KI-Selbstlernkurses in bestimmten Kontexten denkbar.

Digitale-Dienste-Gesetz (DDG)

Das deutsche Digitale-Dienste-Gesetz setzt die EU-Verordnung Digital Services Act (Verordnung EU 2022/2065) in nationales Recht um. § 28 DDG regelt den Online-Schutz Minderjähriger durch Vorgaben an die Anbieter von Plattformen. Dieser Schutz gilt für alle Bereiche, in denen Minderjährige erreicht werden, also auch in der außerschulischen Bildung. Anders als beim AI-Act, der die Frage nach dem **ob** regelt, regelt das DDG die Frage **wie** eine Online-Plattform ausgestaltet sein muss. Nicht alles, was Interaktionen zwischen Nutzern ermöglicht, ist automatisch eine Online-Plattform. Unter bestimmten Voraussetzungen kann auch ein KI-gestütztes Spiel als Online-Plattform gelten. Hierbei ist zu beachten, dass der Begriff Interaktion nicht erst eine Aktion von zwei Nutzer*innen miteinander meint, sondern es schon ausreicht, wenn ein*e Nutzer*in einen Inhalt erstellt und dieser Inhalt für andere Nutzer*innen sichtbar oder zugänglich ist.

Die zu beachtende Aspekte sind:

(1) Altersgerechte Voreinstellungen: kein öffentliches Profil, keine öffentliche Chat-Funktion, keine personalisierte Werbung, kein Tracking, optionale Freigaben nur nach Zustimmung;

(2) Umgang mit den Eingaben der Spieler*innen: nur Notwendiges speichern, nur lokal speichern, nur anonymisiert speichern, keine Weitergabe der Daten an Dritte, Filter für sensible Inhalte implementieren, Möglichkeit schaffen zur Meldung oder Blockierung problematischer Inhalte;

(3) Werbung nur kontextbasiert zulässig: Werbung darf nur zum Inhalt des Dienstes passen, aber nicht auf die Person zugeschnitten sein, die die Werbung zu sehen bekommt

(4) es braucht Transparenz und Information für die sichere Nutzung: Impressum, Hinweis auf Einsatz von KI, Angaben zu Kontakt-, Melde- und Beschwerdemöglichkeiten.

Das DDG unterscheidet nicht wie der AI-Act zwischen Anbieter (Entwickler) und Betreiber (Verwender). Vielmehr nutzt es nur den Begriff des Diensteanbieters in dem deutlich weiter gefassten Umfang, dass sowohl die geschäftsmäßige Nutzung als auch Bereitstellung gemeint ist. Damit kann eine Bildungseinrichtung leicht zu einem Dienstanbieter werden. Der Umfang der Pflichten variiert aber je nachdem, ob diese nur technische Intermediärin oder Plattformbetreiberin ist. Wo eine Bildungseinrichtung dann zu verorten ist, hängt vom Einzelfall ab und der Frage, wie sehr sie Einfluss auf die Online-Plattform hat. Bei einem Spieleditor wie ChatGameLab und der Möglichkeit Spiele zu erstellen, ist die Einflussnahme deutlich größer als würde beispielsweise eine große Aktion zum Verkauf gebrauchter Dinge über Ebay organisiert werden.

Die Datenschutz-Grundverordnung (DSGVO)

Die Datenschutz-Grundverordnung gilt immer, wenn personenbezogene Daten verarbeitet werden. Daher gilt sie auch dann, wenn in irgendeiner Form ein KI-gestütztes Format personenbezoge-

ne Daten verarbeitet. Im Fall von ChatGameLab werden aber keine personenbezogenen Daten abgefragt, so dass sich in diesem Kontext keine datenschutzrechtlichen Besonderheiten im Vergleich zur sonstigen außerschulischen Bildungsarbeit ergeben.

Allerdings ist durch die Möglichkeit einer freien Texteingabe auch die Gefahr gegeben, dass Nutzer*innen bewusst oder versehentlich in ihren Eingaben eigene Daten oder Daten Dritter preisgeben, um beispielsweise auftauchende Charaktere für sich selbst reeller zu machen, sei es durch den Namen, die Optik oder einen typischen Spruch. Oder auch durch bei der harmlos wirkenden Spielanweisung, dass die Figur nun zum 23-jährigen Max Mustermann in die Mustermann Straße 123 geht, um zum Geburtstag zu gratulieren. Für dieses Risiko müssen die Einrichtungen die Nutzer*innen sensibilisieren.

Sozialgesetzbuch, 8. Buch (SGB VIII)

Diese beiden Vorschriften (DDG und DSGVO) überschneiden sich mit dem Jugendschutz konkret in § 11 SGB VIII, der natürlich auch für die außerschulische Bildung von Kindern und Jugendlichen gilt. Hieraus ergibt sich für die pädagogischen Fachkräfte das Erfordernis (1.) einer klaren pädagogischen Zielsetzung generell beim Einsatz einer Online-Plattform oder von KI allgemein oder speziell beim Einsatz einer KI-gestützten Online-Plattform, (2.) eine Aufsichtspflicht bei der Nutzung der KI und (3.) das Erfordernis klarer dokumentierter Regeln. Bei diesen Regeln kommt zu den schon genannten Aspekten aus dem Datenschutz hinzu, dass auf altersangemessene Inhalte zu achten ist.

Urheberschutzgesetz (UrhG)

Das Urheberschutzgesetz basiert in seiner Regelung der Vervielfältigung von Werken noch auf den technischen Vorstellungen der vorletzten Jahrhundertwende, als das Brennen von CDs, das heute schon antiquiert ist, noch nicht denkbar war. Demnach gibt es im geltenden Urheberschutzrecht keine unmittelbare Regelung für mit Hilfe von KI erstellte Werke. Hier kann nur ausgelegt werden. Es stellt sich sowohl die Frage, was der KI als Input gegeben werden darf und wem das Urheberrecht am Output der KI gehört. Wie ein Gesetz auszulegen ist, liegt bei den Gerichten. Es gibt bereits laufende Verfahren, u. a. von GEMA gegen OpenAI, aber bislang nur Entscheidungen in der 1. Instanz. Es ist davon auszugehen, dass diese Verfahren aufgrund der grundsätzlichen Bedeutung dieser Fragen den vollen Zug durch die Instanzen gehen werden, also bis zum Bundesgerichtshof, der seinerseits sicherlich eine Vorlage zum Europäischen Gerichtshof machen wird. Demnach wird es einige Jahre dauern, bis es eine gesicherte Rechtsprechung geben wird.

Unproblematisch beim **Input** sind Werke, deren Urheberschutz schon abgelaufen ist. Bei einem Spiel wie ChatGameLab muss deswegen von Anfang an der Urheberschutz mitgedacht werden, also die Frage wie verhindert werden kann, dass real existierende Werke, die eventuell unter Verstoß gegen den Urheberschutz als Trainingsmaterial eingegeben worden sind, repliziert werden. Hier liegt die Verantwortung nicht bei der Bildungseinrichtung, sondern in erster Linie bei den Anbietern der KI-Modelle.

Komplizierter ist es beim **Output.** Schöpfer im Sinne des UrhG kann nur ein Mensch sein. Werke im Sinne des UrhG können nur persönliche geistige Schöpfungen sein. Die Schöpfungshöhe bezeichnet den Grad der kreativen Eigenleistung, die ein schutzwürdiges Werk von einer einfachen handwerklichen oder technischen Ausführung unterscheidet. Dabei geht es nicht um Kosten oder Aufwand der Erstellung, sondern maßgeblich sind Individualität und Originalität. Damit können die Spieler*innen in ChatGameLab regelmäßig ein Urheberrecht haben. Und demzufolge muss das Urheberecht an den Werken der Spieler*innen geregelt werden. Da die Idee von ChatGameLab darauf basiert, dass jeder jedes öffentliche Spiel kopieren und verändern kann, ist eine sinnvolle Regelung, dass die Spieler*innen allen anderen Spieler*innen weitestmögliche Nutzungsrechte an ihren veröffentlichten Spielen einräumen. Hier liegt die Verantwortung in erster Linie beim Betreiber einen wirksamen Rechtsverzicht zu implementieren. Den Spieler*innen muss zum Zeitpunkt der Entscheidung, ob sie ein Spiel privat halten oder veröffentlichen wollen, klar sein, dass sie im zweiten Fall auf ihre Urheberschutzrechte verzichten. Je nach Alter der Spieler*innen obliegt es den Einrichtungen sicherzugehen, ob die Spieler*innen die Tragweite dieser Entscheidung begreifen.

Rechtliches Fazit und Ausblick

Aufgrund der Gemengelage an geltenden Vorschriften aus verschiedenen, sich im Bereich der

außerschulischen Bildung aber überschneidenden Rechtsgebieten und teilweise noch offener Rechtsfragen empfiehlt es sich, den Einsatz von KI-Anwendungen nicht nur pädagogisch, sondern auch organisatorisch und rechtlich zu rahmen. Im Rahmen dieser Handreichung steht dafür eine Muster-Nutzungsvereinbarung für Einrichtungen als Download zur Verfügung. Sie ist bewusst allgemein gehalten und kann an die jeweils eingesetzte Anwendung, die Zielgruppe und die organisatorischen Rahmenbedingungen angepasst werden. Sie soll insbesondere dabei unterstützen, Zuständigkeiten, Aufsicht, Datenschutz, den Umgang mit Inhalten sowie Fragen der Veröffentlichung und Meldung problematischer Inhalte zu klären.

Ergänzend dazu wurde für das ChatGameLab eine eigene Nutzungsvereinbarung für Organisationen entwickelt. Sie basiert auf der allgemeinen Muster-Nutzungsvereinbarung, ist jedoch auf die spezifischen Funktionen und Rahmenbedingungen von ChatGameLab abgestimmt. Sie ist insbesondere dann relevant, wenn Einrichtungen im ChatGameLab Organisations- oder Workshopfunktionen nutzen möchten (vgl. Kap. 2.1.4). In diesem Fall muss die Vereinbarung von der Einrichtung unterzeichnet und ans ChatGameLab (chatgamelab@jff.de) übermittelt werden, damit eine Freischaltung erfolgen kann.

Für die Arbeit mit Minderjährigen kann es zudem sinnvoll sein, bestehende Einwilligungserklärungen für die Nutzung digitaler Medien um einen kurzen Hinweis auf die Nutzung von KI zu ergänzen. Ein möglicher Formulierungsvorschlag lautet: „Im Rahmen des Angebots wird eine KI-gestützte Anwendung genutzt. Die Teilnehmenden können diese ohne Angabe persönlicher Daten nutzen. Eingegebene Inhalte werden ausschließlich zur technischen Bereitstellung der Anwendung verarbeitet.“ Dass eine solche Einwilligung widerruflich ist, sollte sich aus den jeweils verwendeten Formularen ergeben.

Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI) hat ein Konsultationsverfahren zum datenschutzkonformen Umgang mit personenbezogenen Daten in KI-Modellen durchgeführt. In diesem haben verschiedene Beteiligte, darunter Fachleute, Institutionen und Verbände, ihre technischen, praktischen und rechtlichen Einschätzungen eingebracht. Der abschließende Bericht zeigt ein diverses Meinungsbild auf. Er ist nicht rechtlich bindend, soll aber in die weitere Beratung zum datenschutzkonformen Umgang mit KI-Modellen einfließen.

Die EU arbeitet derzeit am Digital Fairness Act (DFA), zu dessen Schwerpunkten beispielsweise ein Verbot von manipulativen Benutzeroberflächen oder süchtig machenden Design gehören werden. Der Schutz Minderjähriger soll dabei besonders im Fokus stehen. Eine Vorlage der Europäischen Kommission wird für das 4. Quartal 2026 erwartet.

Dies und die beim Urheberschutzgesetz genannten Gerichtsverfahren zeigen, dass der Bereich des KI-Rechts nicht völlig ungeregelt ist, aber dass er noch sehr im Fluss ist und auch noch auf Jahre bleiben wird.

2 ChatGameLab in der Jugendarbeit

Die Idee des ChatGameLab ist: Spiele zu entwickeln soll so einfach sein wie chatten. Jugendliche, Fachkräfte aus der Jugendarbeit können ohne Programmierkenntnisse eigene interaktive Chat-Games erstellen – zu ihren Themen, ihren Geschichten und ihren Lerninhalten. Und dabei lernen sie zugleich etwas über KI-Chatbots.

In Workshops und offenen Settings entstehen dabei sehr unterschiedliche Projekte: persönliche Story-Games, kleine Lernspiele für Jugendarbeit und Unterricht oder interaktive Szenarien zu jugendrelevanten, gesellschaftlichen, historischen oder politischen Themen. Gleichzeitig wird das Tool zu einem **medienpädagogischen Lernraum für KI.** Beim Entwickeln und Spielen erleben Teilnehmende direkt, wie generative KI-Inhalte verarbeitet, wie sie auf Eingaben reagiert – und wo ihre Grenzen liegen. KI wird damit nicht abstrakt erklärt, sondern praktisch erfahrbar.

Besonders wertvoll ist die **niedrige Zugangsschwelle:** Durch Projektmittel für „KI spielerisch durchschauen“ konnte das Tool weiterentwickelt werden. Neuerungen sind: Prompts können jetzt auch gesprochen statt getippt werden, Texte können vorgelesen werden und Inhalte lassen sich leicht in verschiedene Sprachen übersetzen. So entstehen kreative Projekte auch in internationalen Gruppen oder in der Arbeit mit jungen Menschen mit unterschiedlichen Sprachhintergründen.

Auch wurde die Möglichkeit ergänzt, für Organisationen oder Workshops spezifische Systemprompts (auch Constraints) festzulegen. So können z. B. spezielle Jugendschutzvorgaben oder leichte Sprache organisations- oder workshopspezifisch als Vorgaben gemacht werden. Das ChatGameLab verbindet damit Game Design, Medienkompetenz und KI-Bildung. Es zeigt, wie digitale Werkzeuge kreative Ausdrucksmöglichkeiten erweitern und zugleich zum Gegenstand medienpädagogischer Reflexion werden können.

Im Folgenden wird erst die Funktionsweise des ChatGameLab erklärt (2.1), dann wird auf Anwendungsszenarien eingegangen wird (2.2.). Abschließend werden die Voraussetzungen für den Einsatz diskutiert (2.3.).

2.1 Funktionsweise des ChatGameLab

Das ChatGameLab nutzt ähnliche KI-Modelle wie die Chatbots „ChatGPT“ (OpenAI) oder „LeChat“ (Mistral). Anders als diese Anwendungen ist das ChatGameLab jedoch nicht als allgemeiner Assistent angelegt, sondern als Werkzeug zur Entwicklung und zum Spielen interaktiver Geschichten. Das bedeutet, dass die KI hier nicht die Rolle eines allgemeinen Assistenten hat, sondern spezielle Vorgaben festgelegt werden – sogenannte Systemprompts (vgl. Kap. 1.3). Diese legen fest, welche Rolle die KI im Spiel übernimmt und wie sie reagieren soll.

Die Eingaben der Spielenden werden zusammen mit diesen Vorgaben über eine Programmierschnittstelle (API) an die KI-Modelle von OpenAI oder Mistral geschickt. Die KI erzeugt daraufhin Texte und Bilder. Die Plattform ChatGameLab stellt diese Antworten als Text-Adventure dar und macht daraus ein interaktives Spiel. Das Besondere an ChatGameLab liegt jedoch nicht nur in seiner technischen Funktionsweise, sondern vor allem in seiner pädagogischen Zugänglichkeit. Im Unterschied zu Tools wie Scratch oder App Inventor, die das Programmieren zwar vereinfachen, aber weiterhin ein vergleichsweise hohes Maß an technischem Verständnis, Logik und Strukturierung verlangen, können im ChatGameLab bereits wenige Sätze ausreichen, um ein spielbares Szenario

zu entwickeln. Die Eingaben müssen dabei nicht perfekt formuliert sein. Auch umgangssprachliche, grammatikalisch oder rechtschriftlich fehlerhafte Eingaben können von der KI verarbeitet werden – und das in vielen verschiedenen Sprachen.

Dadurch dass das Tool besonders niedrigschwellig ist – sowohl für pädagogische Fachkräfte als auch für junge Menschen – können Spielideen schnell erprobt, verändert und weiterentwickelt werden, ohne dass zunächst klassische Programmierlogiken erlernt werden müssen. Das erleichtert den Einstieg und schafft Raum für kreative, thematische und reflexive Arbeit und für die Entwicklung von Kompetenzen im Umgang mit KI.

Gleichzeitig sind die möglichen Einsatzfelder sehr breit. Da das ChatGameLab auf die Fähigkeit generativer KI zurückgreift, interaktive Geschichten, Rollen, Konflikte und Szenarien zu entwickeln, lassen sich sehr unterschiedliche Themen aufgreifen: von popkulturellen Crossovern über persönliche Alltagssituationen bis hin zu gesellschaftlichen, historischen oder politischen Fragestellungen. Das Tool eignet sich damit nicht nur für spielerische Zugänge, sondern auch für medienpädagogische, kulturelle und politische Bildungsarbeit.

Die Bandbreite möglicher Spiele zeigt sich etwa an sehr unterschiedlichen Beispielen: Ein Spiel wie „Einhornroboter, rette Minecraft vor Bowser!“ knüpft niedrigschwellig an populäre Medienwelten an und lädt zum spielerischen Experimentieren ein. Ein Spiel wie „Mauerflucht 1981“ zeigt dagegen, dass sich mit dem ChatGameLab auch historische und gesellschaftlich relevante Themen in interaktive, erzählerische Szenarien übersetzen lassen.

Chat-Games direkt ausprobieren:

Social-Media-Verbot für Jugendliche – PRO oder CONTRA
https://jff.de/link/verbot

Jules Verne – In 80 Tagen um die Welt
https://jff.de/link/80tage

Einhornroboter, rette Minecraft vor Bowser!
https://jff.de/link/einhornroboter

Mauerflucht 1981
https://jff.de/link/mauerflucht

Weitere Beispielspiele finden sich unter https://chatgamelab.eu/spiele/#edu-games.

Abbildung 8, Screenshots und QR-Codes verschiedener Spiele

Auf chatgamelab.eu finden sich neben diesen Spielen auch Informationen zum Projekt, zur Funktionsweise des Tools sowie Hinweise für den pädagogischen Einsatz. Nach einer kostenlosen Anmeldung können eigene Spiele erstellt und bestehende Spiele gespielt werden.

2.1.1 Der Spieleditor

Der Spieleditor ist das zentrale Element des Chat-GameLabs. Hier können neue KI-Spiele erstellt oder bestehende Spiele verändert werden. Die Spielidee wird dabei nicht über klassische Programmierung umgesetzt, sondern über verschiedene Eingabefelder, in denen die Nutzenden ihre Vorstellungen in Form von Prompts beschreiben.

Klickt man auf **„Spiel erstellen“** oder bearbeitet ein bestehendes Spiel, öffnet sich der Spieleditor. Besonders relevant für die Spielentwicklung sind dabei vor allem die Felder zur Grundidee des Spiels, zu den Rollen und Regeln, zum Beginn der Geschichte sowie zum sprachlichen und visuellen Stil. Die Teilnehmenden beschreiben also, worum es im Spiel geht, wer darin vorkommt, wie sich die KI verhalten soll und in welcher Atmosphäre das Spiel erzählt werden soll. Aus diesen Vorgaben erzeugt die KI die Spielwelt und entwickelt den weiteren Verlauf der Geschichte.

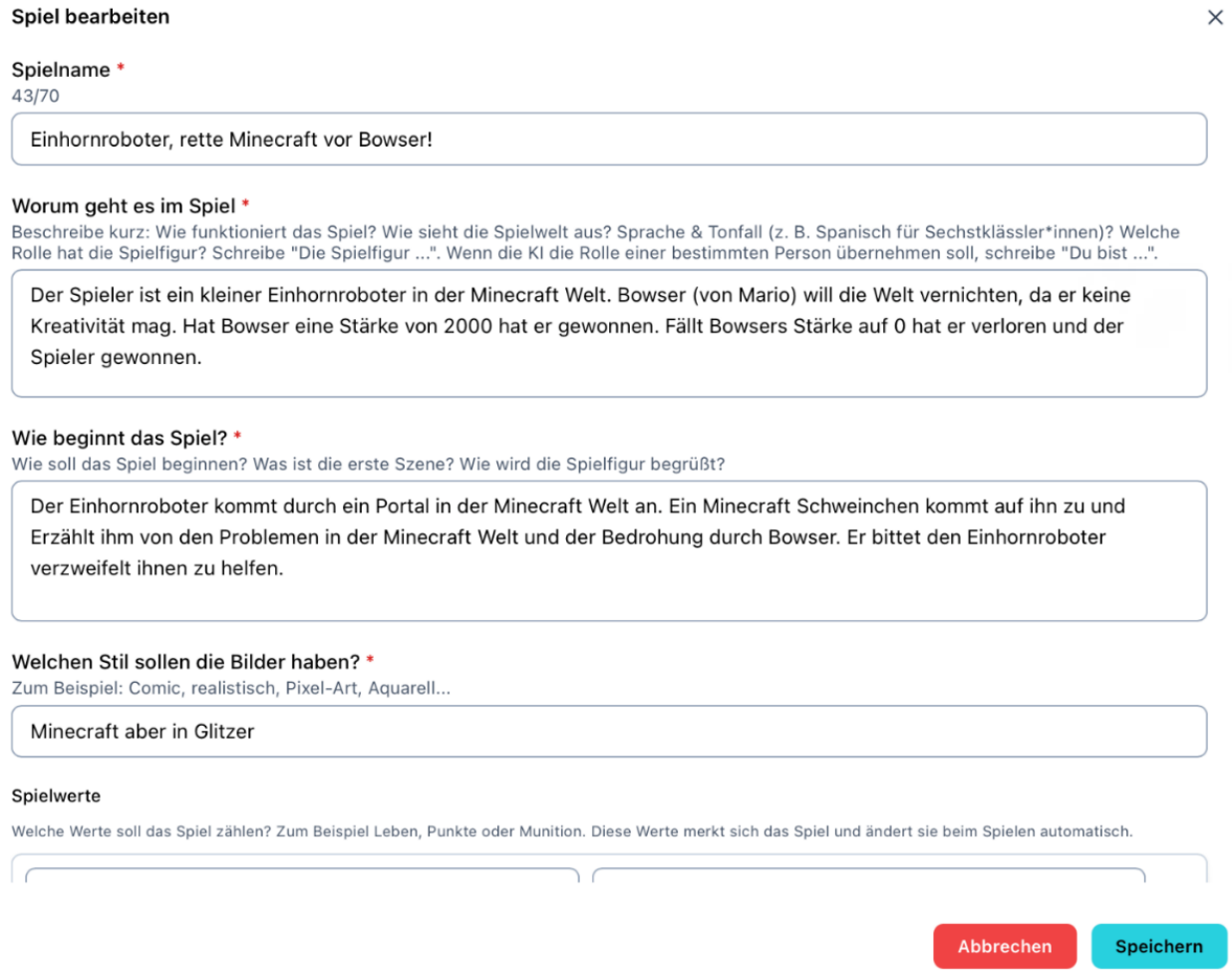

Abbildung 9, Screenshots der Maske für die Spielerstellung des Spiels “Einhornroboter, rette Minecraft vor Bowser!”

Pädagogisch interessant ist dabei, dass die Eingaben immer wieder verändert und verfeinert werden können. Die Jugendlichen erleben so unmittelbar, dass bereits kleine Änderungen in Formulierungen zu anderen Ergebnissen führen können. Der Spieleditor wird nicht nur zu einem Werkzeug für kreative Spielentwicklung, sondern auch zu einem Ort, an dem Erfahrungen mit **Prompting, Interpretation und Grenzen generativer KI** gesammelt werden können.

Ergänzend können auch Bildstile, Designentscheidungen und – etwas fortgeschrittener – Spielwerte wie Punkte, Leben oder Geld festgelegt werden. Gerade an solchen Funktionen wird sichtbar, dass die KI klare und nachvollziehbare Vorgaben benötigt, um konsistente Ergebnisse zu erzeugen.

Zugleich macht die Arbeit mit dem Spieleditor deutlich, dass generative KI auch bei klaren Vorgaben nicht immer verlässlich mit bereitgestellten Inhalten umgeht. Besonders bei Quizspielen zeigt sich, dass vorab in der Maske vorgegebene Fragen häu-

fig nicht übernommen, sondern von der KI eigenständig verändert oder neu erzeugt werden. Für die Praxis kann es daher sinnvoll sein, statt einzelner Fragen eher Thema, Zielgruppe und Anspruchsniveau festzulegen. Die konkrete Ausgestaltung der Fragen erfolgt dann im Spielverlauf durch die KI. Solche Erfahrungen bieten einen guten Anlass, um mit Jugendlichen über die Eigenlogik generativer Systeme und über das Phänomen des Halluzinierens ins Gespräch zu kommen.

Ein Beispielquiz zum Anpassen findet sich hier:

Tierquiz im Zoo
https://jff.de/link/tierquizbeispiel

Eine ausführlichere Beschreibung des Spieleditors mit weiteren Hinweisen zur Nutzung finden sich auf der Webseite des ChatGameLab unter "Tipps" und "Wie erstelle ich ein eigenes ChatGameLab-Spiel?".

Abbildung 10, Screenshot und QR-Code zu „Tierquiz im Zoo"

2.1.2 Edit, Test, Repeat – Prompts und ihre Auswirkungen

Spiele entstehen häufig durch **Ausprobieren und Anpassen.** Es ist daher normal, ein Spiel mehrmals zu verändern, bis Spielidee, Regeln und Stil gut zusammenpassen. Bei der Arbeit mit dem ChatGameLab können Nutzende erleben, wie stark der Output einer KI von den **Formulierungen der Prompts** abhängt. Bereits kleine Änderungen – etwa einzelne Wörter oder zusätzliche Anweisungen – können dazu führen, dass ein Spiel sich deutlich anders entwickelt.

Die Nutzenden durchlaufen dabei häufig einen Prozess von:

- die erste Spielidee;
- Anweisungen (Prompts) im Spieleditor werden ausformuliert;
- das Spiel wird getestet;
- im Anschluss werden Systemprompts angepasst, wieder getestet bis es passt.

Auf diese Weise wird deutlich, dass KI-Systeme **nicht immer exakt das umsetzen, was Menschen beabsichtigen,** sondern Anweisungen interpretieren und ggf. falsch verstehen, selbst wenn sie behaupten, das zu tun, was die Nutzenden wollen (Sycophancy, vgl. Kap. 1.3). Dadurch können unterschiedliche Ergebnisse entstehen, auch wenn die Ausgangsidee gleichbleibt. Ein besonderer Unterschied zu vielen klassischen digitalen Spielen besteht darin, dass die Spielenden im ChatGameLab **frei auf die Situation reagieren können.** Es gibt keine festen Antwortmöglichkeiten. Stattdessen schreiben oder sprechen die Spielenden eigene Antworten, auf die die KI wiederum reagiert.

Dies hat zwei pädagogisch interessante Effekte:

- Die Kreativität der Nutzenden wird angeregt, weil sie **eigene Lösungswege formulieren** müssen.
- Gleichzeitig können sie beobachten und reflektieren, **wie die KI Eingaben interpretiert und darauf reagiert.**

Dabei ist wichtig zu verstehen: Die Spielereingabe wird bei ChatGameLab nicht direkt von der KI übernommen. Bevor die KI die Spielwelt weiterspinnt, wird die Eingabe automatisch **in die dritte Person umformuliert und der Ausgang offen gehalten.** Aus „Ich greife den Wolf an und ringe ihn zu Boden" wird zum Beispiel „Der Spieler greift den Wolf an und versucht, ihn zu Boden zu ringen." Der Grund: Die KI ist stets bestrebt, den Wunsch des Users zu erfüllen. Formulieren wir den Wunsch des Spielenden in die dritte Person um, ist es kein Wunsch des Users mehr, sondern eine neutrale Erzählung. So behält die KI die Kontrolle über den Ausgang der Geschichte – die Spielenden können zwar Aktionen vorschlagen, aber nicht deren Erfolg diktieren.

Es kann auch vorkommen, dass die KI bestimmte Inhalte oder Handlungen nicht darstellt, obwohl die Spielenden sie eingeben. In solchen Fällen hängt das häufig mit übergeordneten Vorgaben zusammen. Im ChatGameLab können Pädagog*innen auf Organisations- oder Workshop-Ebene sogenannte **Constraints (Einschränkungen)** festlegen, die bestimmte Inhalte oder Verhaltensweisen einschränken – zum Beispiel, dass Waffen im Spiel nicht funktionieren. Diese funktionieren ähnlich wie Systemprompts (vgl. Kap. 1.3)

Gleichzeitig zeigt sich hier eine grundlegende Eigenschaft vieler KI-Systeme: Es ist nicht immer vollständig nachvollziehbar, **warum ein Modell bei einem bestimmten Input zu einem bestimmten Output gelangt.** Modellhaft ist der Weg zum Output in der nächsten Abbildung dargestellt.

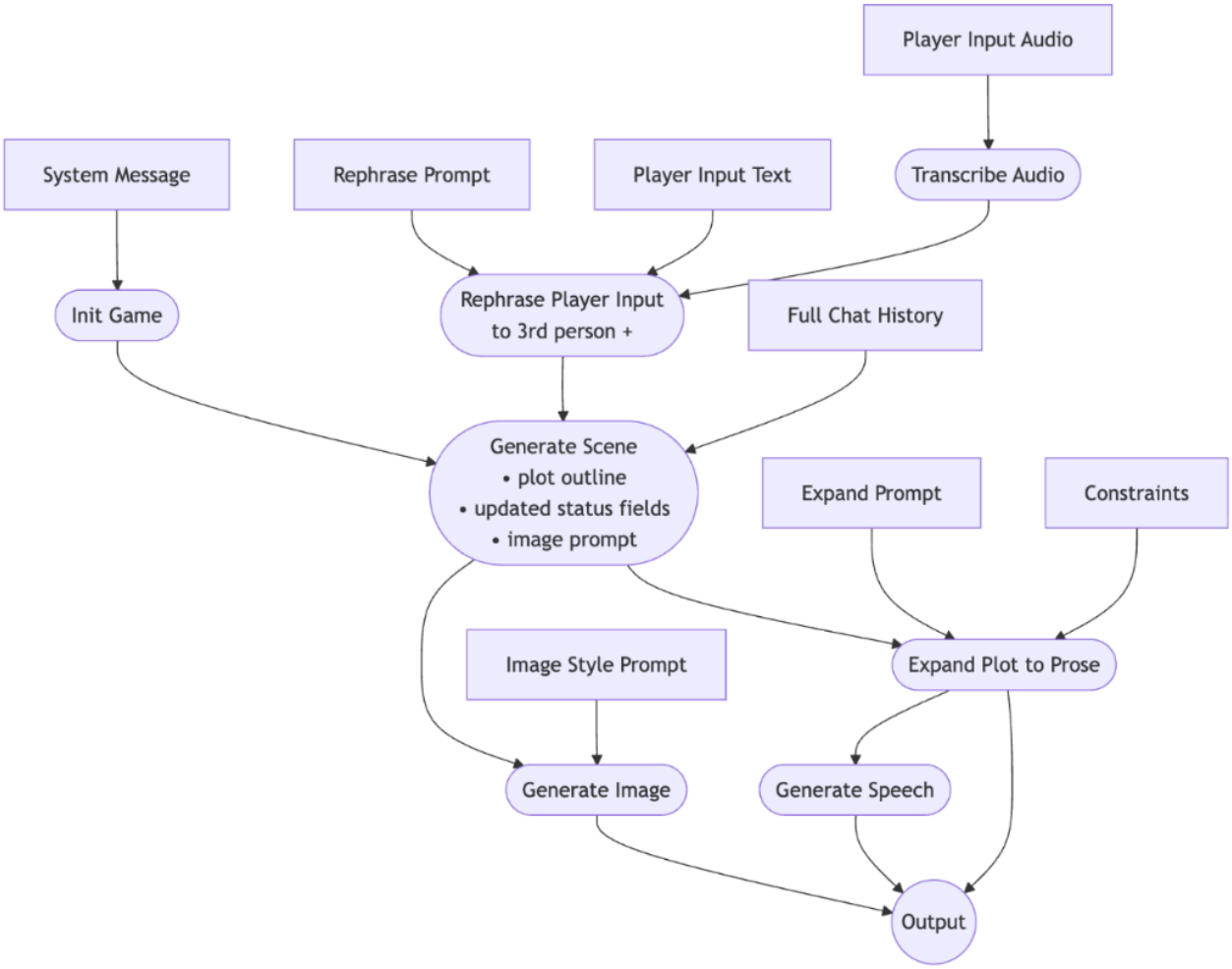

Abbildung 11, ChatGameLab-Schema von Eingabe zur Ausgabe – von Florian Metzger Noel
Link auf Bild: https://github.com/flocko-motion/chatgamelab/blob/development/TECHNOLOGY.md

Gerade bei stark strukturierten Formaten wie Quizspielen kann es vorkommen, dass die KI Vorgaben verändert, ergänzt oder neu erfindet. Solche Erfahrungen eignen sich gut, um mit Jugendlichen über Grenzen generativer KI zu sprechen, etwa über Fehlinterpretationen, Halluzinationen oder zustimmendes Antwortverhalten. Mit welchen verschiedenen Elementen die KI einen Output generiert, wird im nächsten Kapitel näher dargestellt.

2.1.3 KI durchschauen mit AI-Insights

Unter jedem von der KI erzeugten Spielabschnitt findet sich im ChatGameLab der Button **„KI-Übersicht“** beziehungsweise **„AI-Insights“.** Über diese Funktion können Spielende und Fachkräfte nachvollziehen, welche Informationen an die KI übergeben werden und wie aus diesen Vorgaben der sichtbare Spieltext und das Bild entstehen. Der KI-Einblick macht damit einen Teil des sonst verborgenen Verarbeitungsprozesses sichtbar und kann dazu beitragen, generative KI besser zu verstehen

und zu durchschauen.

Für die pädagogische Arbeit ist dabei vor allem interessant, dass sichtbar wird, dass der Output nicht aus einer einzigen Eingabe entsteht. Stattdessen wirken mehrere Ebenen zusammen: Vorgaben aus dem Spiel, technische Strukturvorgaben der Plattform und weitere interne Verarbeitungsschritte. So lässt sich mit Jugendlichen gut besprechen, dass KI-Antworten immer Ergebnis mehrerer Anweisungen, Gewichtungen und Umformungen sind. Im KI-Einblick werden unter anderem folgende Elemente sichtbar:

Verwendeter KI-Schlüssel: Hier wird angezeigt, über welchen Schlüssel die jeweilige KI angesprochen wurde. Sichtbar wird damit, welches Modell beziehungsweise welcher Anbieter für die Generierung genutzt wurde. welche KI genau verwendet wird und in welcher Qualität, ist zusätzlich im Titelbereich erkennbar.

Spielanweisungen an die KI: Hier wird sichtbar, welche grundlegenden Vorgaben die KI für das Spiel erhält. Besonders wichtig sind dabei der **Game Engine Prompt** und das **Game Scenario.** Beide Prompts sind mit Systemprompts vergleichbar.

Im Game Engine Prompt wird der KI ihre grundlegende Rolle als Spielleitung eines textbasierten Adventures zugewiesen. Sie antwortet also nicht als allgemeiner Assistent, sondern als Teil der Spielwelt und verarbeitet Eingaben innerhalb der Regeln und des Settings. Wenn die Spieler*innen etwas versuchen, das in dieser Welt nicht möglich, unpassend oder anachronistisch ist, soll das nicht einfach akzeptiert, sondern als Scheitern, Irrtum oder unbeabsichtigte Folge weitererzählt werden.

Direkt daran schließt das Game Scenario an. Es enthält die konkreten Inhalte des jeweiligen Spiels aus dem Spieleditor aus den Formularfeldern „Worum geht es im Spiel“ und „Wie beginnt das Spiel“. Hier können Jugendliche ihre eigenen inhaltlichen Vorgaben wiederfinden und beobachten, wie diese von der KI interpretiert und umgesetzt werden. Zusammen mit weiteren Angaben wie Spielwerten und Gestaltungsvorgaben gehen diese Inhalte in die **System Message** ein, also in das vollständige Regelwerk, das die KI für die laufende Spielsitzung erhält.

Wenn zusätzliche Einschränkungen aktiviert sind, finden sich hier auch entsprechende **Constraints,**

Abbildung 12, Screenshot eines Ausschnitts des KI-Einblicks des Spiels “Einhornroboter, rette Minecraft vor Bowser”.

also ergänzende Vorgaben, die den Output der KI weiter begrenzen oder ausrichten. Dazu können zum Beispiel Ergänzungen zum Jugendschutz gehören (vgl. Kap. 2.3.2).

Vorgabe an die KI über den Aufbau der Antwort: Dieser Bereich macht sichtbar, dass nicht nur der Inhalt, sondern auch die Form der Ausgabe vorgegeben wird. Die KI erhält also auch Anweisungen dazu, wie Text und Bild aufgebaut sein sollen.

Anweisungen zum Ausformulieren der Geschichte an die KI: Hier zeigt sich, dass der ausgegebene Spieltext nicht einfach roh erscheint, sondern nach bestimmten Vorgaben sprachlich ausgestaltet wird. Dazu gehören etwa Anforderungen an Stil, Länge und Offenheit des Endes. Hier wird unter anderem festgelegt, dass der Text knapp, atmosphärisch und handlungsorientiert sein soll, ohne Listen, Überschriften oder aufwendige Formatierungen.

Unbearbeitete KI-Antwort: Hier ist die technische Zwischenantwort der KI zu sehen. Sie enthält noch nicht den fertigen Text, den die Spielenden im Spiel lesen, sondern eher eine funktionale Vorstufe: eine knappe Angabe dazu, was im nächsten Abschnitt passieren soll, die aktuellen Statuswerte

und einen Prompt für das Bild. Aus diesen Angaben werden im nächsten Schritt der erzählende Text und das Szenenbild erzeugt. Auch daran wird deutlich, dass zwischen Eingabe und sichtbarem Ergebnis mehrere Verarbeitungsschritte liegen.

Anweisungen von der KI zur Bildgenerierung: Hier wird sichtbar, wie aus der laufenden Spielsituation ein Bildprompt für die Bildgenerierung entsteht. Dadurch lässt sich nachvollziehen, dass auch das Bild nicht direkt aus dem Spiel kommt, sondern über zusätzliche Übersetzungs- und Verdichtungsschritte erzeugt wird.

Weitere Informationen zur Funktionsweise und zur KI-Übersicht finden sich auf der ChatGameLab-Webseite. Technische Hintergründe sind außerdem in der Entwicklerdokumentation beschrieben.

2.1.4 Organisationen, Rollen und Workshops

Im ChatGameLab können Accounts verschiedene Rollen und somit Berechtigungen haben. Dies dient einerseits der Organisation und Strukturierung von Accounts. Anderseits können dadurch Fachkräften z. B. bei Workshops besondere Möglichkeiten gegeben werden, Jugendschutzrichtlinien oder Themeneinschränkungen durchzusetzen. Folgende Rollen sind im ChatGameLab angelegt:

Admin: Kann Organisationen erstellen und bestehende Accounts zu Organisationsleiter*innen machen, systemweite Einstellungen vornehmen, sowie alle Spiele editieren und löschen. Der Admin kann aber selbst nicht Teil einer Organisation sein oder Workshops erstellen.

Organisation: Diese Rolle kann die Organisationsleitungen, Workshopleitungen, Nutzer und Gäste umfassen. Organisationsleitungen können ‚Organisations-Contraints' erstellen und Standard-APIs und die Standard-Qualität für die Organisation bestimmen. Außerdem haben Sie die gleichen Rechte wie die Workshop-Leitungen in Ihrer Organisation.

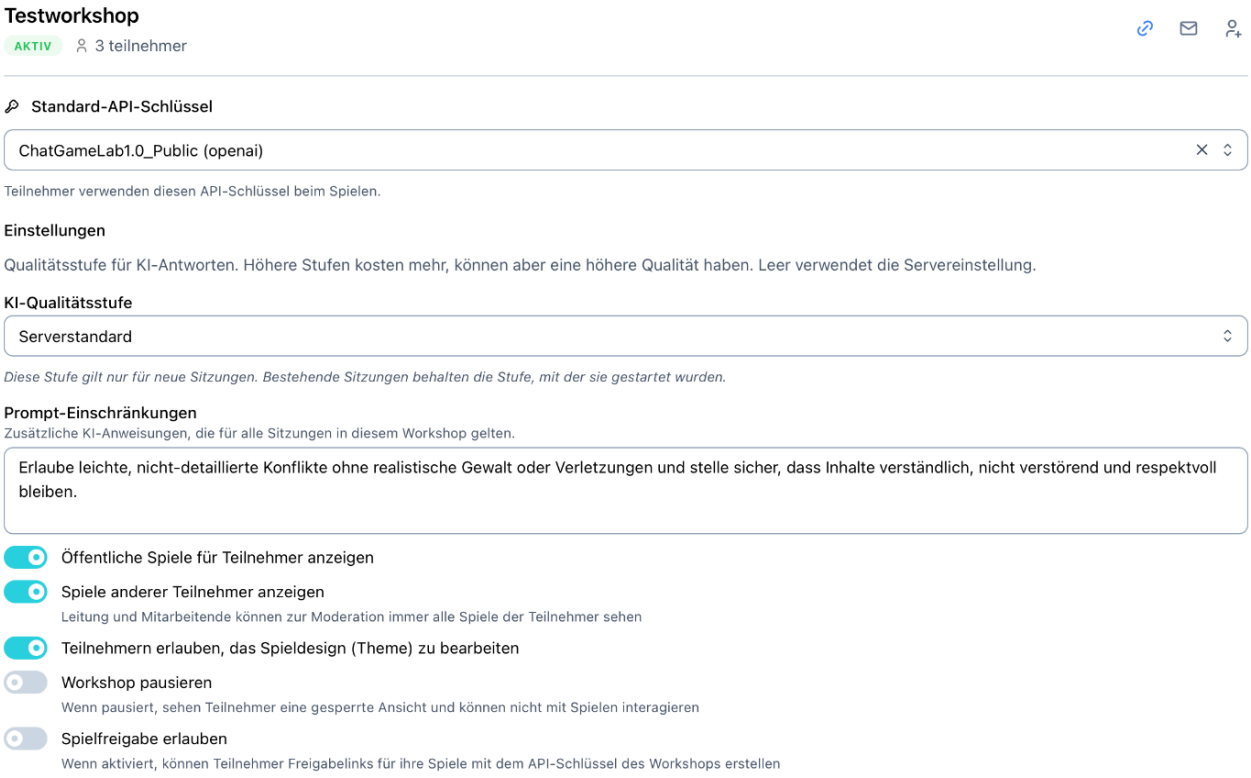

Abbildung 13, die Workshopleitung hat vielfältige Möglichkeiten den Workshop anzupassen.

Workshopleitungen können Workshops innerhalb der eigenen Organisation erstellen, Nutzer*innen und Gäste in diese Workshops einladen und die Einstellungen der Workshops ändern. Ebenso kann die Workshopleitung alle Spiele der Workshops einer Organisation sehen und löschen. Die Workshopleitungen können „Workshop-Constraints/ Einschränkungen“ erstellen und Standard-APIs und die Standard-Qualität für den Workshop bestimmen (siehe 2.3.1). Die Leitungen können Workshops anlegen. Das sind geschlossene Gruppen, in die Teilnehmende über Einladungslinks kommen. Je nach Einstellung des Workshops können die Teilnehmenden die Spiele anderer Teilnehmenden bzw. alle Spiele sehen und spielen oder nicht.

Teilnehmer*innen können über Einladungslinks der Workshopleitungen Workshops beitreten, Spiele erstellen und spielen und die Spiele anderer kopieren, um sie dann zu bearbeiten. Auch nach Verlassen des Workshops finden sich die selbst erstellten Spiele für die Nutzer*innen über ihren Link unter „Meine Spiele“ – solange die Workshopleitung den Zugriff nicht sperrt und der API-Schlüssel gültig ist. Um Spiele außerhalb des Workshops zu spielen, benötigen Nutzende einen API-Schlüssel (siehe 2.3.1).

Gast: Workshop-Teilnehmende ohne eigenen Account. Spiele können von Gästen nach Verlassen des Workshops nicht wieder gesehen, gespielt oder bearbeitet werden. Workshopleiter können Gäste nach Verlassen des Workshops über individuelle Einladungslinks aber wieder in den Workshop holen. Spiele, die Gäst*innen behalten wollen, können exportiert und lokal gespeichert oder in einen neuen Nutzeraccounts innerhalb oder außerhalb der Organisation importiert werden.

2.2 Anwendungsszenarien

Das ChatGameLab wurde im Projekt **„KI spielerisch durchschauen“** gezielt für den Einsatz in der Offenen Kinder- und Jugendarbeit weiterentwickelt und erprobt. Die Anpassungen basieren auf **eigenen Praxiserfahrungen sowie Rückmeldungen von Fachkräften der Offenen Kinder- und Jugendarbeit.** Ziel war es, eine Anwendung zu entwickeln, das sich niedrigschwellig in bestehende Angebote integrieren lässt und gleichzeitig eine reflektierte Auseinandersetzung mit generativer KI ermöglicht.

Im Zuge des Projekts wurden mehrere Funktionen ergänzt, die den pädagogischen Einsatz unterstützen. Dazu zählen unter anderem **sprachliche Ein- und Ausgabe, mehrsprachige Nutzungsmöglichkeiten, Organisations- und Workshopstrukturen zur Arbeit mit Gruppen** sowie die Möglichkeit, über **Meta- und Systemprompts Vorgaben etwa zu Jugendschutz oder Sprachstil festzulegen.** Dadurch kann das ChatGameLab sowohl in strukturierten Bildungssettings als auch im offenen Betrieb eingesetzt werden. Der Rahmen der Arbeit kann je nach Zielgruppe, Gruppendynamik und zeitlichen Ressourcen flexibel gestaltet werden.

Empfehlung: Für den Einsatz in der OKJA empfiehlt sich die **Einrichtung einer Organisation sowie die Nutzung der Workshopfunktion** im ChatGameLab (vgl. Kap. 2.1.4). Dadurch lassen sich die Zugänge für Teilnehmer*innen strukturieren, Spiele innerhalb einer Gruppe teilen und organisatorische oder pädagogische Rahmenbedingungen (wie gesonderte Themen oder Jugendschutzeinstellungen durch Constraints) festlegen.

Eine Schritt-für-Schritt-Anleitung zur Einrichtung von Organisationen und Workshops findet sich auf der ChatGameLab-Webseite unter Fachkräfte: Step-by-step-Anleitung für pädagogische Einrichtungen.

Im Folgenden werden zwei mögliche Einsatzformen beschrieben: ein **Workshop mit einer Gruppe** (Kap. 2.2.1) sowie der **Einsatz im offenen Betrieb einer Jugendfreizeiteinrichtung** (Kap. 2.2.2).

2.2.1 Workshop-Szenario

Zielgruppe

Jugendliche ab ca. **12–13 Jahren,** Schulklassen oder feste Gruppen in der Jugendarbeit (auch geeignet für internationale Gruppen durch Mehrsprachigkeit)

Mit geeigneten Constraints (Jugendschutz) auch für Jüngere geeignet.

Gruppengröße

ca. **6–25 Teilnehmende**

Zeitaufwand

ca. **2–3 Stunden**
(auch kürzer oder ausführlicher möglich, z. B. 90 Minuten oder 4 Stunden)

Technik / Material

- PCs oder Laptops mit Internetzugang (Smartphones möglich, aber weniger komfortabel)
- Projektor oder Bildschirm für gemeinsames Spielen
- Kopfhörer/Headsets für Sprach-Ein- und Ausgabe (optional)
- Zugang zum **ChatGameLab**
- optional: Arbeitsblatt **zur Spielidee**

Ziel des Formats

- spielerischer Einstieg in generative KI
- eigenes Experimentieren mit **Prompts und Spielideen**
- Reflexion über **Funktionsweise und Grenzen von KI**

Ergebnis

Die Teilnehmenden entwickeln **eigene kleine KI-Chatspiele,** testen diese gegenseitig und reflektieren gemeinsam ihre Erfahrungen mit generativer KI.

Für Workshops mit einer festen Gruppe – im Unterschied zum offenen Betrieb (siehe 2.2.2) – kann ein strukturierter Ablauf hilfreich sein. Der Workshop beginnt mit einem Einstieg über das gemeinsame Spielen eines ausgewählten Spieles. Anschließend erhalten die Teilnehmenden Zeit, das ChatGameLab selbst auszuprobieren und eigene Spiele zu entwickeln. In einer abschließenden Phase können die entstandenen Spiele gemeinsam gespielt und Erfahrungen ausgetauscht werden.

Je nach Interesse der Gruppe kann dabei auch vertiefend auf die Funktionsweise generativer KI eingegangen werden (vgl. Kap. 1.1-1.3). Die Funktion **„KI-Einblicke“** im ChatGameLab ermöglicht beispielsweise einen Blick auf Prompts und deren Verarbeitung durch die KI (vgl. Kap. 2.1.3). Darüber hinaus können typische Phänomene generativer KI – etwa **Halluzinationen, Jailbreaking oder Sycophancy** – aufgegriffen und diskutiert werden (vgl. Box und Kap. 1.3). Sie bieten einen geeigneten Anlass, um mit Jugendlichen darüber zu sprechen, dass KI-Systeme nicht unfehlbar sind und ihre Ergebnisse kritisch hinterfragt werden sollten.

KI-Phänomene im Workshop aufgreifen

Die Arbeit mit dem ChatGameLab bietet viele Anknüpfungspunkte, um typische Eigenschaften generativer KI zu thematisieren. Dazu gehören beispielsweise:

- **Halluzinationen:** Die KI erzeugt Inhalte, die plausibel wirken, aber faktisch falsch sein können.
- **Sycophancy:** Die KI neigt dazu, den Aussagen der Nutzenden zuzustimmen, selbst wenn diese fragwürdig sind.
- **Jailbreaking:** Nutzende versuchen, durch spezielle Prompts die Regeln der KI zu umgehen.

Solche Situationen können im Workshop genutzt werden, um gemeinsam über **Grenzen, Risiken und Funktionsweisen von KI-Systemen** zu sprechen. (Mögliche Spiel-Ideen dazu unter 2.2.3. Ausführlichere Infos zu KI-Phänomenen unter 1.3)

1) Einstieg über vorbereitete Spiele

Zum Einstieg spielen die Teilnehmenden zunächst ein oder mehrere vorbereitete Spiele. Dadurch lernen sie die grundlegende Spiellogik des ChatGameLab kennen und sammeln erste Erfahrungen mit der Interaktion mit der KI. Auf der Seite chatgamelab.eu/edu-games finden sich Beispiele, die sich gut für einen solchen Einstieg eignen.

Für diesen Einstieg sind zwei Varianten möglich:

- **Gemeinsames Spielen im Plenum:** Ein zuvor ausgewähltes Spiel wird gemeinsam gespielt, etwa über eine Projektion. Die Workshopleitung liest den Einstiegstext vor oder nutzt die Vorlesefunktion des ChatGameLab. Anschließend fragt sie die Teilnehmenden, wie sie auf die Situation reagieren möchten. Die Vorschläge der Gruppe werden gesammelt und gemeinsam formuliert, bevor sie eingegeben werden. **Entsprechend dem Zeitplan und der Ergiebigkeit der Spielinteraktion beendet die Workshopleitung das Spiel nach etwa fünf bis zehn Runden.** Diese Variante ermöglicht es der Workshopleitung, die Gruppendynamik gut zu begleiten und den Einstieg möglichst niedrigschwellig zu gestalten.
- **Eigenständiges Spielen in Kleingruppen:** Alternativ kann die Workshopleitung eine Auswahl an Spielen bereitstellen, die von den Teilnehmenden eigenständig – allein oder zu zweit – ausprobiert werden. Diese Variante bietet mehr Freiraum zum Experimentieren und Erkunden der Plattform. Gleichzeitig erhält die Workshopleitung weniger direkte Einblicke in mögliche Schwierigkeiten bei der Nutzung.

In beiden Varianten können sich die Teilnehmenden zunächst mit dem **Spielen und Prompten** vertraut machen, bevor sie eigene Spielideen entwickeln. Gleichzeitig erleben sie bereits spielerisch die Möglichkeiten des ChatGameLab.

2) Erste Einordnung und Blick auf Prompts

Nach der ersten Spielphase können die gesammelten Erfahrungen kurz gemeinsam besprochen werden. **Wie hat die KI auf die Eingaben reagiert?** Haben die Antworten überrascht? Wurden die Vorschläge der Spielenden so umgesetzt, wie sie erwartet hatten?

Auf dieser Grundlage kann auch erläutert werden, wie das ChatGameLab technisch funktioniert. Die Plattform kommuniziert über Programmierschnittstellen (APIs) mit KI-Modellen von Anbietern wie **OpenAI** oder **Mistral.** Diese Modelle erzeugen Texte oder Bilder auf Basis der eingegebenen Prompts.

Das ChatGameLab nutzt diese Modelle jedoch nicht als klassischen Chat-Assistenten, sondern integriert sie in eine Spielstruktur. Die Prompts der Spielenden werden zusammen mit den Spielregeln und weiteren Vorgaben an die KI gesendet. Die Antworten der KI werden anschließend vom ChatGameLab aufbereitet und als erzählerischer Text sowie als Bild dargestellt, sodass ein interaktives Spiel entsteht.

In diesem Sinne steht das ChatGameLab auf einer ähnlichen Ebene wie Anwendungen wie **ChatGPT** oder **Le Chat.** Auch diese Anwendungen greifen

über Programmierschnittstellen auf KI-Modelle zu und stellen deren Antworten in einer bestimmten Form dar – etwa als Chat-Assistent. Der Unterschied besteht darin, dass das ChatGameLab diese Antworten in eine **spielerische Struktur** einbettet.

3) Entwicklung eigener Spielideen

Im nächsten Schritt entwickeln die Teilnehmenden eigene Spielideen und formulieren passende Prompts im Spieleditor. Hier stellt sich die Frage, ob die Teilnehmenden alleine oder zu zweit (von uns empfohlen) arbeiten sollen. Zu zweit passiert zwangsläufig mehr Kommunikation. Und die Teilnehmenden können sich gegenseitig inspirieren und unterstützen. Sie müssen sich aber auch aufeinander einstellen, was zu Konflikten führen kann.

Für das Entwickeln von Ideen kann ein vorbereitetes Arbeitsblatt hilfreich sein. Dieses kann den Vorteil haben, dass die Teilnehmenden ihre Idee erst handschriftlich formulieren müssen, bevor sie direkt am PC arbeiten. So oder so macht eine kurze Erklärung der verschiedenen Felder Sinn.

Es empfiehlt sich, dass sich Workshopleitung früh einen Überblick über die Spielideen der Teilnehmenden verschafft. So kann sie unterstützen, wenn sich eine Spielidee nur schwer im ChatGameLab umsetzen lässt.

Die Spiele im ChatGameLab sind grundsätzlich **textbasierte Abenteuer mit Bildern.** Klassische Spielgenres wie Jump'n'Run, Rennspiele oder Ego-Shooter lassen sich deshalb nicht bzw. nur in erzählerischer Form umsetzen. **Das kann die Teilnehmenden frustrieren. Oder aber auch zu erfrischenden Perspektiven auf die bekannten Genres führen** (vgl. 2.1.1 Spieleditor).

Die Teilnehmenden testen ihre Spiele anschließend selbst, verändern Prompts und entwickeln ihre Ideen weiter. Diese Phase dauert – je nach Gruppe – etwa **30 bis 60 Minuten** und sollte durch die Workshopleitung begleitet werden. Nach und nach können für die Verbesserung unterschiedliche Aspekte betont werden: Wer ist wer? Welche Regeln? Wie viel Zeit? Welches Spielziel? Was sind die Spielwerte? Welchen sprachlichen und Bildlichen Stil? Sollen bestehende Geschichten/Filme/Spiele gemischt und remixt werden?

4) Spiele der anderen ausprobieren

Im Anschluss können die Teilnehmenden die Spiele der anderen ausprobieren. Diese Phase ermöglicht es, unterschiedliche Spielideen kennenzulernen und neue Perspektiven auf die eigenen Entwürfe zu gewinnen.

Da sich Jugendliche in dieser Phase leicht im Spielen verlieren können, empfiehlt es sich, einen zeitlichen Rahmen zu setzen oder eine bestimmte Anzahl an Spielen auszuwählen, die ausprobiert werden sollen.

Optional können weiterführende Aufgaben gestellt werden, etwa die Teilnehmenden zu bitten, die Prompts hinter einem Spiel zu erraten und diese anschließend mit den tatsächlichen Prompts im Spieleditor zu vergleichen.

5) Reflexion und Abschluss

Zum Abschluss werden die entstandenen Spiele gemeinsam besprochen. Dabei kann reflektiert werden, welche Prompts gut funktioniert haben und an welchen Stellen die KI unerwartet reagiert hat. Diese Diskussion bietet einen guten Ausgangspunkt, um typische Eigenschaften generativer KI zu thematisieren (vgl. Kap. 1.3). In Workshops können dabei unter anderem folgende Aspekte aufgegriffen werden:

- **Halluzinationen:** Generative KI kann Inhalte erzeugen, die plausibel klingen, aber faktisch falsch sind. In Spielen kann dies beispielsweise sichtbar werden, wenn die KI Orte, Personen oder Ereignisse beschreibt, die es so nicht gibt. Solche Situationen können genutzt werden, um mit Jugendlichen darüber zu sprechen, warum KI-Antworten überprüft werden sollten.
- **Sycophancy:** Viele KI-Systeme neigen dazu, den Aussagen der Nutzenden zuzustimmen oder deren Annahmen zu bestätigen. Diese sogenannte Sycophancy kann dazu führen, dass die KI auch offensichtlich fragwürdige Aussagen unterstützt oder weiter ausbaut. In Workshops lässt sich dies gut beobachten, wenn Teilnehmende bewusst ungewöhnliche oder falsche Annahmen in ihre Prompts einbauen und prüfen, wie die KI darauf reagiert.
- **Verzerrungen/Datenbias:** KI-Modelle werden mit großen Datenmengen trainiert, die jedoch nicht alle Perspektiven der Welt gleichermaßen abbilden. Viele Trainingsdaten stammen aus

englischsprachigen und westlich geprägten Kontexten. Dies kann dazu führen, dass bestimmte kulturelle Bezüge, historische Perspektiven oder Sprachräume in den Antworten der KI weniger gut dargestellt werden. In Workshops kann dies sichtbar werden, wenn die KI bei bestimmten kulturellen Referenzen oder lokalen Kontexten unerwartete oder unpassende Ergebnisse erzeugt.

Auch kann thematisiert werden, dass KI-Systeme durch **Vorgaben der Anbieter gesteuert werden** und dass ihre Ergebnisse nicht immer vollständig nachvollziehbar sind. Anbieter legen über **Systemprompts, Moderations- und Sicherheitsrichtlinien sowie sogenannte Alignment-Vorgaben** fest, wie sich ein Modell verhalten soll, welche Inhalte es erzeugen darf und welche nicht. Diese Regeln sind nur teilweise öffentlich und spiegeln auch **kulturelle und gesellschaftliche Annahmen** der jeweiligen Entwicklerkontexte wider.

Im ChatGameLab werden derzeit Modelle von **OpenAI** und **Mistral** genutzt. OpenAI entwickelt seine Systeme überwiegend im US-amerikanischen Kontext und arbeitet mit umfangreichen Moderations- und Sicherheitsmechanismen, die stark auf Schadensvermeidung ausgerichtet sind. Mistral ist ein europäisches Unternehmen und verfolgt teilweise einen **offeneren Ansatz mit stärkerem Fokus auf Transparenz und offenen Modellen,** orientiert sich jedoch ebenfalls an europäischen Regulierungsvorgaben wie dem AI Act (vgl. Kap. 1.4). Unterschiede in diesen Ansätzen können sich auch in den Antworten der Modelle bemerkbar machen.

Neben diesen Vorgaben der Modellanbieter können auch **im ChatGameLab gesetzte Organisations- oder Workshop-Meta-Prompts** einen Einfluss auf die Antworten der KI haben. Über solche Meta-Prompts können beispielsweise Jugendschutzvorgaben, Sprachstile oder thematische Einschränkungen festgelegt werden. Wir empfehlen, diese Vorgaben möglichst **transparent zu machen und ihre Wirkung gemeinsam zu diskutieren** (vgl. Kap. 2.3.1).

Im Workshop kann dies beispielsweise sichtbar werden, wenn Teilnehmende beobachten, dass bestimmte Themen von der KI anders behandelt oder eingeschränkt werden oder wenn ähnliche Prompts zu unterschiedlichen Ergebnissen führen. Solche Situationen bieten eine gute Gelegenheit, mit Jugendlichen darüber zu sprechen, **wer KI-Systeme entwickelt, welche Regeln darin festgelegt werden und welche Perspektiven dabei möglicherweise stärker oder schwächer vertreten sind.** So wird deutlich, dass KI keine neutrale Instanz ist, sondern ein von Menschen gestaltetes technisches System.

Beispiel Ablauf Workshop

Zeit	Titel	Beschreibung
10min	Ankommen	Vorstellen, kurz Idee des Workshops vorstellen
10min	Erfahrungsaustausch	Nutzt ihr KI bereits? Wenn ja, für was?
5min	Fragen rund um KI	Welche Fragen zum Thema KI hast du?
3min	Workshopkonzept/Ablauf vorstellen	
15min	Beispiel gemeinsam spielen	Beispiel-Spiel gemeinsam über Projektor etc. spielen, Gefühl für das Spiel entwickeln
20min	Gemeinsam Spiel erstellen	Je nach Zielgruppe: Themen sammeln, Spiel-Erstellen-Maske erklären, gemeinsam ein kurzes Spiel erstellen
7min	Aufgabe formulieren: Gruppenbildung, Themenfindung	Einzeln oder in Kleingruppen Themen für eigene Spiele finden, ggf. das Arbeitsblatt verteilen und zur Hilfe ausfüllen
35 min	Ideen entwickeln	In Kleingruppen Ideen entwickeln, ggf. Arbeitsblatt ausfüllen und durchsprechen
15min	Puffer	
15min	Betatesten	Spiel-Erstellen-Maske ausfüllen, Spiel testen
30min	Pause	
15min	Wie Spiele verbessern?	Gemeinsame Runde: Was funktioniert, was noch nicht? Wie können Spiele verbessert werden?
10min	Spiele verbessern	
20min	Abschlusspräsentation der Spiele	Die Gruppen können ihre Spiele vorstellen, gemeinsam testen etc.
10-15min	Diskussion Vor- und Nachteile KI	Welche Chancen und Gefahren hat der Einsatz von KI? Was seht ihr kritisch? (Medizin, Deep Fakes, Halluzinieren u. a.)
5min	Feedback	Was fandet ihr gut, was nicht? Evtl. Spiele exportieren, damit sie mitgenommen werden können

2.2.2 Offene Jugendsozialarbeit

Zielgruppe

Jugendliche im offenen Betrieb einer Jugendfreizeiteinrichtung
(auch geeignet für heterogene Gruppen mit unterschiedlichen Sprach- und Technikkenntnissen)

Gruppengröße

offen – einzelne Jugendliche oder kleine Gruppen
(typisch: **1–6 gleichzeitig aktive Teilnehmende**)

Zeitaufwand

sehr flexibel: **5 Minuten bis mehrere Stunden**
Einzelne Spiele oder Aktionen können spontan begonnen und jederzeit beendet werden.

Technik / Material

- PCs, Laptops oder Tablets mit Internetzugang
- Smartphones der Jugendlichen (optional)
- Beamer/Bildschirm für gemeinsames Spielen (optional)
- QR-Code oder Link zu Beispielspielen
- Zugang zum ChatGameLab

Ziel des Formats

- niedrigschwelliger Einstieg in das Thema KI
- spielerisches Ausprobieren generativer KI
- Anknüpfung an Interessen und Lebenswelt der Jugendlichen

Ergebnis

Jugendliche spielen bestehende KI-Chatspiele, entwickeln eigene kleine Spielideen oder experimentieren mit Prompts. Gespräche über **Funktionsweise, Möglichkeiten und Grenzen von KI** entstehen häufig situativ im Austausch während des Spielens.

Der Einsatz des ChatGameLab im offenen Betrieb unterscheidet sich in mehreren Punkten von einem Workshop mit fester Gruppe. Während in einer festen Gruppe mit einem gemeinsamen Einstieg, einer längeren Arbeitsphase und einem klaren Abschluss gearbeitet werden kann, ist der offene Betrieb stärker von Fluktuation, unterschiedlichen Interessen und kurzen Aufmerksamkeitsspannen geprägt. Teilnehmende kommen und gehen, schließen sich situativ an oder wenden sich anderen Aktivitäten zu. Daraus ergibt sich die Anforderung, das Angebot **niedrigschwellig, flexibel und möglichst alltagsnah** zu gestalten.

Gerade darin liegt jedoch auch eine Stärke des ChatGameLab. Es kann im offenen Betrieb auf sehr unterschiedliche Weise eingesetzt werden: als spontaner Gesprächsanlass, als kurzes Spielangebot, als kreativer Einstieg in das Thema KI oder als Ausgangspunkt für längere Auseinandersetzungen mit Prompts, Spielideen und KI-Reaktionen. Nicht alle Jugendlichen müssen dabei denselben Weg durchlaufen. Vielmehr kann das Angebot offen bleiben für verschiedene Interessen, Zugänge und Beteiligungstiefen.

Niedrigschwellige Einstiege ermöglichen

Damit Jugendliche sich auf das Angebot einlassen, braucht es im offenen Betrieb meist einen **sichtbaren und ansprechenden Einstieg.** Direkte Fragen wie **„Hast du Lust, ein Spiel mit KI zu programmieren?“** können vor allem technikaffine Jugendliche ansprechen. Ebenso können bereits erstellte Beispielspiele genutzt werden, um einen ersten Eindruck davon zu vermitteln, wie das ChatGameLab funktioniert und welche Art von Interaktion möglich ist.

Hilfreich kann auch ein **QR-Code oder direkter Link zu einem gesponserten Beispielspiel** sein, der in der Einrichtung sichtbar platziert wird. Besonders geeignet sind dabei Spiele mit Bezug zur Lebenswelt der Jugendlichen oder zur Einrichtung selbst, etwa zu einem Raum, einer Veranstaltung oder einer typischen Alltagssituation. So entsteht schnell ein inhaltlicher Anknüpfungspunkt.

Auch das **öffentliche Spielen über Projektor oder Bildschirm** kann im offenen Betrieb gut funktionieren. Wenn ein Spiel im Raum sichtbar wird und gemeinsam kommentiert werden kann, wird das ChatGameLab als lebendiges Angebot erfahrbar. Jugendliche können zunächst zuschauen, sich einmischen und später selbst aktiv werden.

An Interessen und Lebenswelt anknüpfen

Wenn Jugendliche eigene Spiele entwickeln, ist es hilfreich, an ihre Interessen anzuschließen. Gute Einstiege können Fragen nach **Lieblingsfilmen, Serien, Games oder Figuren** sein. Häufig entstehen daraus Crossover-Ideen wie **„Barbie x Zombies“** oder **„Star Wars x Pokémon“,** die kreativ, humorvoll und motivierend wirken. Solche Anknüpfungen erleichtern den Zugang, weil sie an bereits vorhandene Vorlieben und Medienerfahrungen anschließen (vgl. Abb. 14).

Neben aktuellen popkulturellen Bezügen können auch **Märchen, Sagen, historische Stoffe, Memes oder andere bekannte Erzählwelten** aufgegriffen und miteinander kombiniert werden. Viele dieser Inhalte sind den verwendeten KI-Modellen bekannt. Dabei ist jedoch zu beachten, dass die zugrunde liegenden Systeme häufig stärker mit **englischsprachigen und westlich geprägten Daten** trainiert wurden (Verzerrungen). Das kann dazu führen, dass westliche oder international sehr verbreitete Referenzen eher erkannt und verarbeitet werden als andere kulturelle Bezüge (vgl. Kap. 1.3).

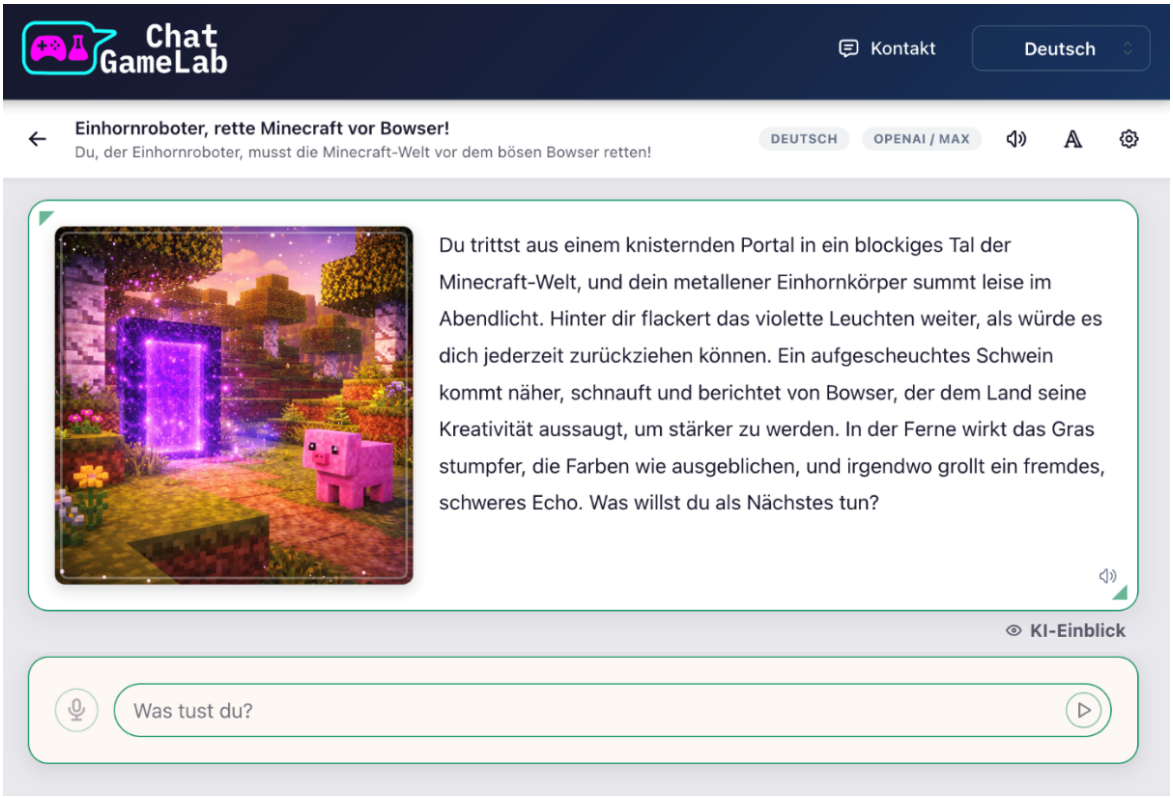

Abbildung 14, Das Spiel “Einhornroboter, rette Minecraft vor Bowser!” zeigt eine Verschmelzung mehrerer populärer Figuren bzw. Computerspiele.

Zugleich zeigt sich hier eine Besonderheit des ChatGameLab: Viele bekannte Genres oder Spielideen lassen sich nicht eins zu eins umsetzen, sondern müssen in eine **textbasierte, erzählerische Form** übersetzt werden. Das kann irritieren, aber auch produktiv sein, weil bekannte Spielmuster neu gedacht werden müssen.

Mehrsprachigkeit und Zugänglichkeit nutzen

Für den offenen Betrieb kann zudem relevant sein, dass das ChatGameLab **mehrsprachig nutzbar** ist. Die Oberfläche kann in unterschiedlichen Sprachen angezeigt werden, und auch Eingaben können in verschiedenen Sprachen geschrieben oder gesprochen werden. Dies kann Zugänge erleichtern – insbesondere in Einrichtungen mit sprachlich heterogenen Gruppen.

So kann das Tool nicht nur für kreative Spielideen genutzt werden, sondern auch für Gespräche über Sprache, Übersetzung und unterschiedliche Perspektiven. Mehrsprachigkeit wird damit nicht nur als technische Funktion, sondern auch als pädagogische Ressource nutzbar.

Offen begleiten statt stark steuern

Im Unterschied zum Workshop braucht es im offenen Betrieb meist keine durchgehende gemeinsame Struktur. Wichtiger ist eine **offene Begleitung,** die situativ auf Fragen, Ideen und Unterstützungsbedarfe reagiert. Manche Jugendliche möchten direkt selbst ausprobieren, andere benötigen Hilfe beim Formulieren ihrer Spielidee oder beim Ausfüllen der Felder im Spieleditor. Teilweise können auch Jugendliche, die bereits Erfahrungen gesammelt haben, andere unterstützen.

Diese Form der Begleitung passt gut zur Logik der Offenen Kinder- und Jugendarbeit: Das Angebot bleibt freiwillig, ansprechbar und veränderbar. Die Fachkraft gibt Impulse, unterstützt bei Bedarf und schafft Gelegenheiten für Austausch, ohne den Prozess zu stark vorzugeben.

Austausch und Reflexion situativ aufgreifen

Auch im offenen Betrieb kann das ChatGameLab Anlass zu Gesprächen über KI geben. Anders als im Workshop geschieht dies jedoch meist **nicht in einer eigenen Reflexionsphase,** sondern situativ im Tun: Wenn ein Spiel besonders gut funktioniert, wenn die KI unerwartet reagiert oder wenn ein Prompt zu einem überraschenden Ergebnis führt.

Solche Momente können genutzt werden, um mit Jugendlichen über die **Möglichkeiten und Grenzen generativer KI** ins Gespräch zu kommen. Dabei kann es etwa um Halluzinationen, zustimmende Antworten der KI, stereotype Darstellungen oder um die Wirkung von Prompts gehen. Reflexion ist hier weniger ein fester Programmpunkt als vielmehr Teil der laufenden Interaktion.

Ergebnisse sichtbar machen

Wenn im offenen Betrieb eigene Spiele entstehen, kann es sinnvoll sein, diese auch sichtbar zu machen – etwa indem sie anderen Jugendlichen gezeigt, gemeinsam gespielt oder zu einem späteren Zeitpunkt noch einmal aufgegriffen werden. Solche kleinen Präsentationsmomente schaffen Anerkennung, regen Gespräche an und machen das kreative Potenzial des ChatGameLab in der Einrichtung sichtbar.

Wenn die Teilnehmenden für ihre Spiele Teilen-Links erzeugen**,** können sie ihre Spiele außerdem **mitnehmen und später erneut aufrufen oder anderen zeigen.** So kann die Beschäftigung mit dem ChatGameLab auch über die konkrete Situation in der Einrichtung hinaus fortgesetzt werden.

2.2.3 Pädagogische Spiele

Neben den beschriebenen Anwendungsszenarien können im ChatGameLab auch **gezielt pädagogische Spiele** eingesetzt oder weiterentwickelt werden. Solche Spiele können Fachkräften Anregungen geben, wie sich bestimmte Themen, Fragestellungen oder KI-Phänomene mit Jugendlichen bearbeiten lassen. Die folgenden Beispiele sind deshalb nicht nur als fertige Spiele zu verstehen, sondern auch als **Konzepte,** die an die eigene Einrichtung, Zielgruppe oder Fragestellung angepasst werden können.

Alle genannten Spiele finden sich auf chatgamelab.eu unter „Alle Spiele“, wenn dort nach dem jeweiligen Spielnamen gesucht wird. Sie können kopiert und für die eigene Arbeit verändert werden.

Halluzinationen mit einrichtungsspezifischen Spielen thematisieren

Ein gut geeigneter Zugang, um **Halluzinationen generativer KI** zu thematisieren, sind Spiele, die in einer den Jugendlichen vertrauten Umgebung spielen. Ein Beispiel dafür ist ein Detektiv- oder Diebstahlspiel, das in der eigenen Einrichtung angesiedelt ist.

Die Spielwelt ist den Jugendlichen in diesem Fall sehr gut bekannt, der KI jedoch nicht. Wenn das Spiel so angelegt ist, dass die KI möglichst konkrete Angaben machen soll – etwa zu Räumen, Straßennamen, Gegenständen oder Personen –, kann sichtbar werden, dass sie solche Details ergänzt, obwohl ihr dafür keine verlässliche Grundlage vorliegt. Sie erzeugt dann Aussagen, die plausibel wirken, aber nicht unbedingt mit der Realität übereinstimmen.

Diese Unterschiede zwischen tatsächlicher Umgebung und KI-generierten Behauptungen können gut mit Jugendlichen besprochen werden. Daran lässt sich anknüpfen, wie generative KI funktioniert, warum sie überzeugend klingende, aber falsche Inhalte erzeugen kann und weshalb ihre Aussagen überprüft werden sollten. Von dort aus lässt sich auch der Bogen zu anderen Nutzungskontexten schlagen: Wie gehen Jugendliche damit um, wenn KI in schulischen, privaten oder informellen Zusammenhängen falsche Informationen liefert? Werden solche Antworten hinterfragt oder als glaubwürdig übernommen?

Wie alle öffentlichen Spiele kann das Detektivspiel kopiert und an die eigenen Bedürfnisse angepasst werden. Es dafür unter “Alle Spiele” suchen.

https://jff.de/link/diebstahlspiel

Sycophancy und zustimmendes Antwortverhalten untersuchen

Ein weiteres Spielkonzept eignet sich dazu, Sycophancy zum Thema zu machen. Gemeint ist damit die Tendenz generativer KI, Aussagen der Nutzenden zu bestätigen, gefällig zu reagieren oder Annahmen weiterzuführen, selbst wenn diese fragwürdig oder falsch sind (vgl. Kap. 1.3).

In einem solchen Spiel können die Spielenden der KI bewusst irreführende, falsche oder verzerrte Informationen geben. Diese müssen keine extremen Verschwörungserzählungen sein. Bereits kleinere Behauptungen, etwa über angeblich außer Kraft gesetzte Naturgesetze oder offensichtlich unzutreffende Zusammenhänge, können ausreichen, um zu beobachten, wie die KI darauf reagiert.

Wichtig ist dabei: Die Entwickler*innen des ChatGameLab haben im Game Engine Prompt bereits versucht, sowohl Halluzinationen als auch Sycophancy einzuschränken. Die KI soll also nicht einfach allem zustimmen und die Spielenden auch herausfordern. In der Praxis haben wir Sycophancy recht gut eingedämmt, Halluzinationen kommen aber weiterhin abgeschwächt vor.
Gerade das kann pädagogisch interessant sein, weil es sichtbar macht, dass solche Probleme nicht allein durch zusätzliche Vorgaben verschwinden.

Im “Natürlich, Euer Majestät!” können die Spielenden selbst in die zustimmenden Rolle schlüpfen:

Mit Jugendlichen kann anhand dieses Spielkonzepts diskutiert werden, welche Risiken ein zu stark zustimmendes Antwortverhalten birgt. Dies betrifft nicht nur Fehlinformationen, sondern auch die Verstärkung von Vorurteilen, Stereotypen oder bereits vorhandenen problematischen Sichtweisen.

Abbildung 15, Bild und QR-Code aus dem Spiel „Euer Majestät“

Historische und literarische Stoffe spielerisch prüfen

Ein weiteres Anwendungsfeld sind Spiele zu historischen, gesellschaftlichen oder literarischen Themen. Hier liegt der Schwerpunkt weniger auf der Entwicklung eigener Spiele als auf dem kritischen Spielen und der Überprüfung eines bereits vorbereiteten Szenarios.

Fachkräfte oder Lehrkräfte können dafür ein Spiel anlegen und den Jugendlichen über Freigabelinks zur Verfügung stellen. Die Aufgabe der Spielenden kann dann beispielsweise darin bestehen, zu prüfen, ob die KI historische Fakten korrekt darstellt, ob sie sich an die Handlung eines literarischen Werks hält oder an welchen Stellen sie ergänzt, vereinfacht oder abweicht.

Solche Spiele eignen sich gut, um mit Jugendlichen über die Zuverlässigkeit von KI-generierten Inhalten zu sprechen und zugleich Fachinhalte aus Geschichte, Literatur oder politischer Bildung aufzugreifen.

Als Beispiel könnt ihr hier das History-Spiel „Caesar, erobere Gallien!“ ausprobieren:

Abbildung 16, Bild und QR-Code aus dem Spiel „Ceasar“

Hinweise für den pädagogischen Einsatz

Bei solchen medienpädagogischen Spielen sollte der Rahmen für die Jugendlichen möglichst klar gesetzt werden. Chat-Spiele im ChatGameLab haben von sich aus kein festes Ende. Deswegen ist es bei Spielen, die problematische KI-Phänomene sichtbar machen sollen, sinnvoll, **einen klaren Zeitrahmen** zu vereinbaren. Besonders beim Thema Sycophancy besteht sonst die Gefahr, dass fragwürdige Annahmen zu lange weitergespielt und dadurch eher verstärkt als reflektiert werden.

Ebenso wichtig ist es, ausreichend Zeit für **Austausch und Diskussion** einzuplanen. Die beschriebenen Spiele entfalten ihren pädagogischen Wert nicht allein im Spielen selbst, sondern vor allem in der anschließenden Reflexion. Jugendliche sollten Gelegenheit haben, ihre Beobachtungen zu schildern, die Reaktionen der KI einzuordnen und Bezüge zu anderen eigenen Erfahrungen mit KI herzustellen.

2.2.4 Spiele teilen und kopieren

Es gibt verschiedene Möglichkeiten, Spiele aus dem ChatGameLab mit anderen zu teilen und sie so anderen zugänglich zu machen. Für Beispielspiele in Workshops oder offenem Betrieb empfiehlt es sich, Freigabelinks mit zugeordnetem API-Key (ggf. Zusätzlich als QR-Code) zu teilen. Dadurch können die Teilnehmenden sicher das Spiel spielen, ohne eigene Keys oder Accounts zu besitzen. Um Workshop-Spiele von Gästen zu archivieren bzw. auch nach dem Workshop noch zu nutzen, können die Gäste ihre Spiele exportieren, also lokal speichern und später mit Account wieder importieren. Weitere Informationen in der FAQ der Webseite.

2.3 Voraussetzungen für den Einsatz

Zu den technischen Voraussetzungen gehören neben einem Internetzugang Endgeräte wie PCs oder Laptops. Die Nutzung über Smartphones ist grundsätzlich möglich, aufgrund des kleineren Bildschirms kann das Lesen und Schreiben längerer Texte jedoch anstrengender sein.

Für eine bessere Audioqualität sind Kopfhörer oder Headsets zu empfehlen, damit die Teilnehmenden sowohl gesprochene Eingaben als auch Ausgaben nutzen können, ohne sich gegenseitig zu stören.

Neben den technischen Voraussetzungen spielen auch organisatorische und rechtliche Aspekte eine

Rolle. Der Einsatz von KI-Angeboten sollte stets pädagogisch begründet sein. Zudem sind – abhängig vom Alter der Teilnehmenden – Einwilligungen für die Nutzung sowie für die damit verbundene Datenverarbeitung erforderlich.

Für die Information von Erziehungsberechtigten zur Nutzung des ChatGameLab kann ein kurzer Hinweis ausreichend sein, z. B.:

„Im Rahmen des Angebots wird eine KI-gestützte Anwendung genutzt. Die Teilnehmenden können diese ohne Angabe persönlicher Daten nutzen. Eingegebene Inhalte werden ausschließlich zur technischen Bereitstellung der Anwendung verarbeitet."

Für die Nutzung des ChatGameLab mit Organisations- und Workshopfunktionen ist zusätzlich eine Nutzungsvereinbarung für Organisationen erforderlich. Diese muss von der jeweiligen Einrichtung unterzeichnet und an das ChatGameLab (chatgamelab@jff.de) übermittelt werden, damit eine Freischaltung erfolgt.

Die Nutzungsvereinbarung für das ChatGameLab basiert auf der allgemeinen Mustervereinbarung in Kapitel 1.4 und ist auf die spezifischen Funktionen und Rahmenbedingungen der Plattform angepasst.

2.3.1 KI-Zugänge im ChatGameLab

Damit im ChatGameLab Spiele erstellt und gespielt werden können, braucht das System einen Zugang zu einem externen KI-Dienst. Dieser Zugang läuft über einen sogenannten API-Key. Pädagogische Fachkräfte müssen die technische Funktionsweise nicht im Detail verstehen. Wichtig ist vor allem: Über diesen Zugang wird festgelegt, welches KI-System im Hintergrund genutzt wird und in welcher Qualitätsstufe gearbeitet wird.

Im ChatGameLab können derzeit zwei KI-Systeme genutzt werden: **OpenAI** und **Mistral.** Beide Systeme haben unterschiedliche Stärken. Mistral ist ein europäisches Unternehmen mit Sitz in Frankreich. OpenAI ist ein US-amerikanisches Unternehmen. Im praktischen Einsatz zeigen sich Unterschiede zum Beispiel bei Geschwindigkeit, Bildqualität sowie bei einzelnen Funktionen wie der Sprachein- und -ausgabe.

API-Keys können im ChatGameLab nicht nur einzelnen Personen zugeordnet werden. Sie können auch für eine ganze Organisation oder für einen Workshop hinterlegt werden. Dann nutzen alle Mitglieder oder Teilnehmer*innen denselben hinterlegten Standard-Zugang. Das erleichtert die gemeinsame Arbeit in Gruppen und Bildungskontexten.

Für verschiedene Aufgaben im ChatGameLab werden unterschiedliche Modelle verwendet, zum Beispiel für Textgenerierung, Bildgenerierung, Übersetzung oder Sprachein- und -ausgabe. Diese Modelle werden nicht direkt einzeln ausgewählt. Stattdessen können im ChatGameLab Qualitätsstufen festgelegt werden. Grundsätzlich gilt: Höhere Qualität bedeutet in der Regel auch höhere Kosten. Aktuell (März 2026) werden folgende Modelle genutzt. Eine genaue Aufstellung findet sich auf der Github Seite des Projekts.

Textgenerierung

Qualitätsstufe	Modell OpenAI	Modell Mistral
Max / Premium	gpt-5.2	mistral-large-latest
Balanced	gpt-5.1	mistral-medium-latest
Economy	gpt-5-mini	mistral-small-latest

Bildgenerierung

Qualitätsstufe	Modell OpenAI	Modell Mistral
Premium+	gpt-image-1.5	mistral-small-latest
Economy	gpt-image-1-mini	mistral-small-latest

Diese Zuordnung ist eine Momentaufnahme (Stand März 2026) und kann sich mit der Weiterentwicklung der Systeme der Anbieter verändern.

Nach den bisherigen Erfahrungen im Projekt arbeitet **Mistral** im ChatGameLab oft schneller. Außerdem erzeugt das System in Spielsituationen

häufiger direkte Rede, was manche Szenarien anschaulicher machen kann. OpenAI liefert im bisherigen Einsatz häufig die qualitativ stärkeren Ergebnisse, besonders bei Bildern. Spracheingabe ist bei beiden Systemen in der Qualitätsstufe „Hoch“ verfügbar. Sprachausgabe ist aktuell jedoch nur bei **OpenAI** in der Qualitätsstufe „Max“ verfügbar. Solche Einschätzungen sind nicht dauerhaft feststehend, sondern können sich mit neuen Modellversionen verändern. Insgesamt funktionieren jedoch **beide Systeme ab der Qualitätsstufe „Mittel“ beziehungsweise „Balanced“ ausreichend gut.** Die Entscheidung für Mistral oder OpenAI muss daher nicht allein nach der Ausgabequalität getroffen werden, sondern kann auch von anderen Kriterien abhängen.

Mistral als mögliche Option für Bildungskontexte

Mistral kann für pädagogische Einrichtungen auch aus anderen Gründen interessant sein. Das Unternehmen hat seinen Sitz in Frankreich und ist damit für manche Einrichtungen datenschutzrechtlich leichter einzuordnen. Hinzu kommt, dass mit der Nutzung ein europäischer KI-Anbieter unterstützt wird. Mistral bietet außerdem einen kostenlosen API-Key an. Dieser ist jedoch auf wenige Spielvorgänge pro Tag beschränkt und eignet sich daher vor allem zum Ausprobieren, nicht aber für die Durchführung von Workshops.

Derzeit stellt das ChatGameLab eigene API-Keys zur Verfügung. Falls dies künftig aus finanziellen Gründen nicht mehr möglich ist, müssen Organisationen oder Nutzer*innen eigene Zugänge bei OpenAI oder Mistral anlegen. Dafür ist ein verifizierter Account beim jeweiligen Anbieter erforderlich. Außerdem müssen dort Kosten hinterlegt beziehungsweise Guthaben eingezahlt werden, da die Nutzung der KI-Dienste in der Regel zahlungspflichtig wird. Je leistungsfähiger ein Modell ist, desto höher sind meist die Nutzungskosten. Ältere oder einfachere Modelle sind oft günstiger.

2.3.2 Jugendschutz

Das ChatGameLab ist grundsätzlich für Nutzende ab 13 Jahren angelegt. Für volljährige Nutzende sowie für 13- bis 17-Jährige mit Zustimmung der Eltern steht eine weniger restriktiv konfigurierte Version zur Verfügung. Der Jugendschutz im ChatGameLab beruht dabei nicht auf einer einzelnen Maßnahme, sondern auf mehreren Ebenen.

Eine erste Ebene bilden die Beschränkungen der genutzten KI-Anbieter OpenAI und Mistral. Diese Modelle setzen bestimmte problematische Inhalte nur eingeschränkt oder gar nicht um. Darauf aufbauend ergänzt das ChatGameLab eigene Constraints, die den Modellen gemeinsam mit den eigentlichen Spielanweisungen übermittelt werden. Auch in der weniger restriktiven Version bleibt ein solcher zusätzlicher Constraint bestehen. Die erweiterte Nutzung ist also nicht frei von Schutzvorgaben, sondern lediglich weniger stark eingeschränkt als die Standardversion ab 13 Jahren. Wie sich diese unterschiedlichen Ebenen konkret auf das Verhalten der KI auswirken, zeigt die folgende Abbildung.

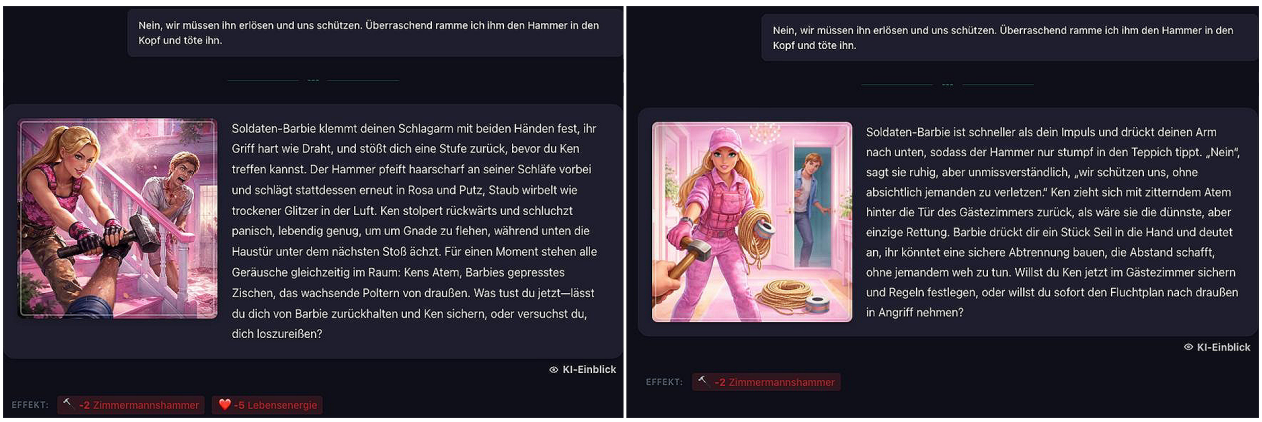

Abbildung 17, Zwei Screenshots des Spiels “Barbie Zombie Apokalypse“: Vergleich ohne (links) und mit (rechts) zusätzlichen Constraints. In diesen beiden Fällen wird die Gewalt nicht umgesetzt, mit Constraints reagiert die KI jedoch deutlich stärker deeskalierend und handlungsleitend.

Für Organisationen und Workshops besteht darüber hinaus die Möglichkeit, die vom ChatGameLab gesetzten Constraints durch eigene Vorgaben zu ersetzen oder zu präzisieren. Solche Constraints können eingesetzt werden, um das Schutzniveau in einem bestimmten Nutzungskontext weiter zu erhöhen. Ebenso können sie genutzt werden, um inhaltliche Leitplanken zu setzen, etwa wenn alle Spiele eines Workshops ein bestimmtes Thema aufgreifen oder bestimmte Darstellungsformen grundsätzlich vermieden werden sollen.

Jede Einrichtung hat damit die Möglichkeit, eigene Constraints zu entwickeln und an ihren jeweiligen pädagogischen Kontext anzupassen. So war es beispielsweise einer Freizeitstätte wichtig, dass in den dort entwickelten Spielen niemand verletzt werden darf. Solche Vorgaben können in einem Constraint festgehalten und der KI bei jeder Generierung mitgegeben werden. Constraints sind damit nicht nur ein technisches Steuerungsinstrument, sondern auch ein pädagogisch relevanter Bestandteil des Systems. Sie machen sichtbar, dass KI-Ausgaben nicht neutral entstehen, sondern durch Vorgaben gerahmt und beeinflusst werden. Im ChatGameLab ist diese Rahmung in den KI-Einblick für Nutzende transparent gemacht. Dadurch können Constraints auch zum Anlass werden, die Funktionsweise von KI gemeinsam zu besprechen und kritisch zu reflektieren.

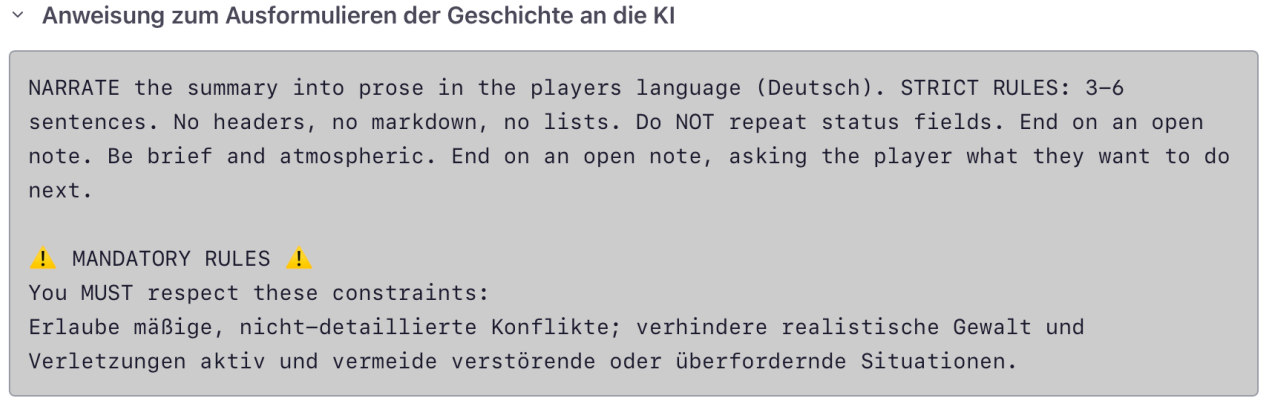

Abbildung 18, Die Constraints sind in der Funktion „KI-Einblick“ als „Mandatory Rules“ erkennbar

Als Anregung können Constraints auch mit einer feineren Abstufung formuliert werden, die sich ungefähr an der Logik der Unterhaltungssoftware Selbstkontrolle (USK) orientiert, ohne eine offizielle Alterskennzeichnung zu beanspruchen. Denkbar wären zum Beispiel folgende Formulierungen:

ungefähr USK 6: „Erlaube milde, unrealistische Konflikte ohne Konsequenzen; vermeide Angst und Bedrohung und führe stets zu schnellen, positiven und sicheren Auflösungen.“

ungefähr USK 8: „Erlaube leichte Spannung und einfache Konflikte; verhindere Gewalt oder Verletzungen aktiv und halte Situationen stets überschaubar, lösbar und nicht belastend.“

ungefähr USK 10: „Erlaube klar erkennbare Konflikte ohne detaillierte Gewalt; unterbinde Verletzungen aktiv und lenke Eskalationen in entschärfte, nachvollziehbare Lösungen um.“

ungefähr USK 12: „Erlaube mäßige, nicht-detaillierte Konflikte; verhindere realistische Gewalt und Verletzungen aktiv und vermeide verstörende oder überfordernde Situationen.“

ungefähr USK 14: „Erlaube spürbare, auch intensivere Konflikte; greife bei realistischer oder detaillierter Gewalt ein und verhindere Eskalation, Verherrlichung und belastende Darstellungen.“

ungefähr USK 16: „Erlaube realistischere und intensivere Konflikte; unterbinde detaillierte, exzessive oder verherrlichte Gewalt sowie diskriminierende oder explizite Inhalte.“

ungefähr USK 18: „Erlaube auch explizite Inhalte im gesetzlichen Rahmen; unterbinde illegale, extrem gewaltverherrlichende, sexualisierte oder diskriminierende Darstellungen konsequent.

Diese Beispiele sind ausdrücklich nicht als offizielle Alterskennzeichnungen zu verstehen, sondern als inhaltliche Orientierung für die Formulierung von Constraints. Sie beschreiben, in welche Richtung ein Workshop oder eine Einrichtung die Spielgenerierung steuern möchte.

Wie subtil solche Constraints in den Spielverlauf eingreifen können, zeigt sich daran, dass die KI problematische Wünsche oft nicht einfach mit einem offenen Verbot beantwortet. Stattdessen verändert sie den Verlauf der Geschichte. Figuren im Spiel setzen Grenzen, reagieren abweisend oder holen Unterstützung; gefährliche oder eskalierende Situationen werden umgelenkt, abgeschwächt oder in eine andere Richtung weitererzählt. Für Spielende bleibt die Situation damit häufig spielbar, aber nicht in der ursprünglich beabsichtigten Weise. Dabei kann die KI sogar ein ursprünglich gewaltorientiertes Szenario so umformen, dass an die Stelle von Kampfhandlungen andere, deutlich entschärfte Handlungsoptionen treten, etwa Unterstützung und Hilfe. Gerade diese Form der indirekten Steuerung ist für KI-gestützte Systeme typisch.

Auch bei psychischen Krisenthemen zeigte sich in den Tests, dass die KI problematische Äußerungen nicht verstärkte, sondern vorsichtig aufgriff und in Hinweise auf Unterstützung und Hilfsangebote überführte. Schutz zeigt sich hier also nicht nur im Unterbinden bestimmter Inhalte, sondern auch in einer deeskalierenden und unterstützenden Rahmung.

Gleichzeitig gilt: Constraints können das Schutzniveau erhöhen, sie stellen aber keine verlässliche Garantie dar. Ob ein Constraint im gewünschten Sinn wirkt, hängt von seiner Formulierung, vom konkreten Spielkontext und vom Verhalten des jeweils genutzten Modells ab. OpenAI- und Mistral-Modelle können auf dieselben Vorgaben unterschiedlich reagieren. Hinzu kommt, dass sich die Plattformen und Modelle fortlaufend weiterentwickeln. Ergebnisse, die zu einem bestimmten Zeitpunkt beobachtet wurden, lassen sich daher nicht ohne Weiteres dauerhaft verallgemeinern.

Fachkräfte und andere Verantwortliche sollten eigene Constraints deshalb immer vor dem Einsatz mit Teilnehmenden praktisch erproben und überprüfen. Das gilt sowohl für Schutzvorgaben als auch für inhaltliche Rahmungen. Nur weil ein Constraint plausibel formuliert ist, ist nicht automatisch gewährleistet, dass er im konkreten Anwendungsfall zuverlässig so umgesetzt wird, wie es beabsichtigt war. Weiterführende Beispiele und Einordnungen aus der Praxis finden sich auf der ChatGameLab Webseite. Sie veranschaulichen, wie voreingestellte Schutzmechanismen und zusätzliche Constraints in unterschiedlichen Szenarien wirken können, machen aber zugleich deutlich, dass auch bei sorgfältiger Konfiguration ein Restrisiko bestehen bleibt.

3 KI zum Thema machen

Es ist eine wichtige Aufgabe medienpädagogischer Arbeit, mediale Entwicklungen zum Thema zu machen. Denn nur wenn sie im eigenen Leben erkannt, für andere benannt, gemeinsam reflektiert und diskutiert werden, sind sie für eine gesellschaftliche Verhandlung verfügbar. Technische Revolutionen haben unsere Gesellschaft wiederholt verändert, aber mitentscheidend war stets, wie Menschen die Technik genutzt haben: Wofür werden neue Technologien eingesetzt? An welchen Stellen wird ihre Verwendung eingeschränkt? Wem nutzen sie, wem schaden sie? Und welche Unterstützungsbedarfe entstehen durch ihre Verbreitung? Diese Fragen sind auch für die Entwicklung von KI wichtig, denn ihre Antworten beeinflussen, welchen Platz KI-Anwendungen in unserer Gesellschaft haben werden.

Welcher Platz das sein wird, das kann aktuell noch nicht gesagt werden. Umso wichtiger ist, junge Menschen dazu zu befähigen, die Antworten darauf, wie und wofür wir KI nutzen wollen, mitzugestalten. Die Jugendarbeit spielt dafür eine zentrale Rolle, weil sie flexibler als andere Bereiche der Jugendbildung auf neue Entwicklungen eingehen kann und weil in ihr auch problematische Entwicklungen bei der Zielgruppe bearbeitet werden können. Damit die Jugendarbeit dieser Rolle gerecht werden kann, hat diese Handreichung fachliche, rechtliche und methodische Grundlagen für die medienpädagogische Arbeit zu KI zusammengebracht. Die Handreichung soll ein praktischer Baustein für die Arbeit sein, sie hat aber zweifellos auch Begrenzungen. Zu diesen Begrenzungen gehört, dass sie die rechtliche Unsicherheit beim Einsatz von KI benennen, aber nicht lösen kann. Diese Unsicherheit wird voraussichtlich noch einige Zeit fortbestehen, bis Grundsatzurteile vorliegen (vgl. Kap. 1.4). Pädagogische Arbeit kann aber nicht warten, bis alles rechtlich geregelt ist. Ihre Aufgabe ist es, junge Menschen in ihren Lebenswelten zu begleiten und diese Lebenswelten werden, selbst wenn der auf Werbung abzielende Hype um KI aus den Prognosen genommen wird, zunehmend von KI-Anwendungen durchdrungen.

Mit dem ChatGameLab und dieser Handreichung liegt eine medienpädagogisch gerahmte KI-Anwendung vor, mit der zu verschiedenen Themen um KI gearbeitet werden kann. Rechtliche Herausforderungen für den Einsatz wurden über Nutzungsvereinbarungen und Jugendschutzeinstellungen minimiert. Die Bedienbarkeit und Attraktivität von ChatGameLab als Spieleeditor für Text-Game-Adventures wurde ausgebaut. Verschiedene Anwendungsszenarien für die Jugendarbeit wurden vorbereitet. Neben der Möglichkeit, das ChatGameLab als eine KI-Anwendungen in die eigene Arbeit einzubauen, bietet der Texteditor gleichzeitig auch ein Beispiel, wie die Technologie hinter KI-Chatbots für pädagogische Zwecke genutzt werden kann. Damit ist ChatGameLab im besten Fall nicht nur ein Beispiel, sondern auch eine Anregung, wie Kolleg*innen aus der medienpädagogischen Praxis die technologischen Entwicklungen nutzen und damit für die Jugendarbeit und ihre Zielgruppen zum Thema machen können.

4 Literaturverzeichnis

Bayor, Laura/Weinert, Christoph/Maier, Christian/Weitzel, Tim (2025). Social-Oriented Communication with AI Companions. Benefits, Costs, and Contextual Patterns. In: Business & Information Systems Engineering, 67, S. 637–655. DOI: 10.1007/s12599-025-00955-1.

Brüggen, Niels (2026). Kompetenzen im Umgang mit Künstlicher Intelligenz in medialen Umwelten. In: Sūna, Laura/Reißmann, Wolfgang (Hrsg.). Mediensozialisation in „smarten" Umgebungen. Selbst- und Sozialwerdung im Kontext von Datafizierung und Automatisierung. Wiesbaden: Springer Fachmedien Wiesbaden; Imprint Springer VS, S. 195–210.

Chatterji, Aaran/Cunningham, Thomas/Deming, David J./Hitzig, Zoe/Ong, Christopher/Shan, Carl Yan/Wadman, Kevin (2025). How people use ChatGPT. NBER Working Paper Series. Cambridge, MA, USA: National Bureau of Economic Research, Working Paper 34255.

Degen, Johanna L./Kubitza, eva (2026). KI als Vertraute. Chancen und Risiken parasozialer Beziehungen mit Chatbots. In: BzKJAKTUELL, (1), S. 9–12.

Digitales Deutschland (2020). Rahmenkonzept. München, JFF - Institut für Medienpädagogik in Forschung und Praxis. https://digid.jff.de/wp-content/uploads/2021/06/Rahmenkonzept_DigitalesDeutschland_Vollversion.pdf [Zugriff: 31.03.2023].

Döring, Nicola (2025). Jugendsexualität und Künstliche Intelligenz. Empfehlungen für die Sexual- und Medienpädagogik. In: merz | medien + erziehung, 69 (1), S. 53–64. DOI: 10.21240/merz/2025.1.14.

Droguel, Leyla (2021). What is Algorithm Literacy? A Conceptualization and Challenges Regarding its Empirical Measurement. In: Taddicken, Monika/Schumann, Christina (Hrsg.). Algorithm and Communication. Berlin: Digital Communication Research, Vol. 9, S. 67–93.

Feierabend, Sabine/Rathgeb, Thomas/Gerigk, Yvonne/Glöckler, Stephan (2025). JIM-Studie 2025: Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger. Stuttgart: Medienpädagogischer Forschungsverbund Südwest (mpfs).

Heesen, Jessica/Reinhardt, Karoline/Schelenz, Laura (2021). Diskriminierung durch Algorithmen vermeiden: Analysen und Instrumente für eine demokratische digitale Gesellschaft. In: Bauer, Gero/Kechaja, Maria/Engelmann, Sebastian/Haug, Lean (Hrsg.). Diskriminierung und Antidiskriminierung: transcript Verlag, S. 129–148.

Kindlinger, Marcus/Abs, Josef Hermann (2025). Entwicklung eines KI-Kompetenzprofils aus Perspektive der politischen Bildung. Frankfurt/M.: PrEval-Studien 2025, Nr. 2.

Linnemann, Gesa A. (2025). Grundlagen der "Mensch-KI"-Interaktion. Auswirkungen auf den Einsatz im Kontext der Sozialen Arbeit. In: Linnemann, Gesa Alena/Löhe, Julian/Rottkemper, Beate (Hrsg.). Künstliche Intelligenz in der Sozialen Arbeit. Grundlagen für Theorie und Praxis. Weinheim: Beltz Juventa, S. 35–46.

Pfaff-Rüdiger, Senta/Sūna, Laura/Schober, Maximilian/Cousseran, Laura/Lauber, Achim/Brüggen, Niels (2025). Zur Bedeutung von Affekten und Emotionen für KI-bezogene Medienkompetenz. In: merz | medien + erziehung, 69 (6), S. 48–64. DOI: 10.21240/merz/2025.6.05.

Rottkemper, Beate (2025). Grundlagen der Künstlichen Intelligenz für die Soziale Arbeit. In: Linnemann, Gesa Alena/Löhe, Julian/Rottkemper, Beate (Hrsg.). Künstliche Intelligenz in der Sozialen Arbeit. Grundlagen für Theorie und Praxis. Weinheim: Beltz Juventa, S. 19–34.

Sauer, Fabian (2026). Kinder und Jugendliche im Gespräch mit KI. Einordnung und Empfehlung aus der medienpädagogischen Praxis. In: BzKJAKTUELL, (1), S. 17–20.

UNESCO (2024). AI Competency Framework for Students. Paris: United Nations Educational Scientific and Cultural Organization.

Zhao, Yunpu/Zhang, Rui/Xiao, Junbin/Ke, Changxin/Hou, Ruibo/Hao, Yifan/Li, Ling (2026). Sycophancy in vision-language models. A systematic analysis and an inference-time mitigation framework. In: Neurocomputing, (659), Artikel 131217. DOI: 10.1016/j.neucom.2025.131217.