KI-unterstützte Bild- und Videogenerierung in der Wissensvermittlung

Executive Summary

In einer Welt, die zunehmend von digitalen Inhalten geprägt ist, hat sich die Art und Weise, wie Wissen erworben und weitergegeben wird, grundlegend gewandelt. Multimediales Lernen, also die Kombination von Text, Bild, Video oder Audio, hat stark an Bedeutung gewonnen. Der rasante Fortschritt der Künstlichen Intelligenz (KI) erhöht mithilfe von Bild- und Videogenerierung die Zahl der Möglichkeiten. Bild- und Videogeneratoren ermöglichen es, komplexe Konzepte auf der Basis einer Textanweisung in sehr kurzer Zeit in realistische Bilder oder animierte Erklärvideos zu verwandeln. Anstatt auf kostenintensive Produktionen oder umfangreiche Bilddatenbanken angewiesen zu sein, können Lehrende und Lernende nun maßgeschneiderte visuelle Hilfsmittel erstellen, die exakt auf ihre jeweiligen Lernziele zugeschnitten sind. Zudem generieren Lernende Bilder und Videos als Teil des Lernprozesses, beispielsweise im Zuge von Projektarbeiten.

Multimediale Inhalte können die Aufmerksamkeit der Lernenden binden und das Verständnis für Inhalte vertiefen. Visuelle Inhalte spielen hier eine besondere Rolle. Die KI-gestützte Generierung von Bildern und Videos ermöglicht grundsätzlich eine breite Verfügbarkeit von anschaulichen, skalierbaren und – je nach Kanal – auch leicht zugänglichen Lernressourcen in der Wissensvermittlung. Zudem wird die Personalisierung gefördert: Die Inhalte der Bilder und Videos können in verschiedenen Sprachen, Stilen und Schwierigkeitsgraden ausgegeben oder an etwaige Beeinträchtigungen angepasst werden, um den individuellen Bedürfnissen der Nutzer:innen gerecht zu werden.

So vielversprechend diese KI-Tools, wie beispielsweise der europäische Bildgenerator FLUX, integriert in Le Chat von Mistral, Synthesia als Videogenerator oder DALL-E 3 auch sind, bringen sie auch erhebliche Herausforderungen und Risiken mit sich. Zu den größten Gefahren zählen die Verbreitung visueller Falschinformationen, inhaltlicher Fehler, Datenschutzverletzungen, urheberrechtliche Probleme und die Entstehung sowie Nutzung sogenannter „Deepfakes". Dabei verschwimmt die Grenze zwischen realen und synthetisch erzeugten Bildern und Videos zunehmend, wodurch das Vertrauen der Gesellschaft in digitale Medien und KI-Technologien weiter sinkt. Zudem besteht das Risiko, dass KI-Bild- und Videogenerierung bestehende gesellschaftliche Stereotype und Verzerrungen (KI-Bias) reproduziert oder verstärkt, da sie unter anderem mit nicht repräsentativen Daten aus dem Internet trainiert wurden. Dadurch kann in der Wissensvermittlung ungewollt ein verzerrtes Bild der Wirklichkeit geprägt werden.

Gemäß *EU AI-Act* werden bild- und videogenerierende KI-Systeme mit „begrenztem" Risiko eingestuft. Aufgrund potenzieller Schäden im Hinblick auf Deepfakes, Datenschutzverletzungen und Desinformation im spezifischen Anwendungsbereich der Wissensvermittlung stuft das ALAIT-Risikoradar das Risiko dennoch als „hoch" ein (Stufe 4 im Anwendungsbereich, x-Achse in der Grafik) ein.

ALAIT Risikoradar für die KI-unterstützte Bild- und Videogenerierung in der Wissensvermittlung



BVG Bild- und Videogenerierung

Das ALAIT Risikoradar ist ein wissenschaftlich entwickeltes Risikoanalysetool für Künstliche Intelligenz (KI), das KI-Anwendungen kontextbezogen und unter Berücksichtigung ihres technischen Autonomiegrades einstuft und so die Risikosphäre für Anwender:innen auf einen Blick sichtbar macht. Dabei gilt: Je höher das Einsatz-Risiko aus dem Anwendungskontext und je größer der Autonomiegrad des KI-Systems in Bezug auf Entscheidungen, desto riskanter ist der Einsatz einzustufen. Eine erweiterte Klammer weist auf eine Bandbreite in der Risikoeinstufung hin. Ein geringer Autonomiegrad eines KI-Systems bedeutet nicht, dass man sich zurücklehnen kann. Es erfordert eine starke Rolle der Menschen, die es anwenden. (Details zum Stufenmodell s. S. 9f).

**ALAIT ist ein vom Bundesministerium für Innovation, Mobilität und Infrastruktur (BMIMI) initiiertes Forschungsprojekt zur Entwicklung einer neuartigen Form von Technologiebewertung für KI-Technologien mit wissenschaftlichen Methoden.*

Der Autonomiegrad ist hingegen gering: KI in dieser Anwendung ist ein passives Werkzeug, Menschen leiten den Prozess, vom Prompting bis zur Entscheidung über die Verwendung der Ergebnisse (Stufe 1). Insgesamt ergibt die Bewertung von KI-unterstützte Bild- und Videogenerierung im ALAIT-Risikoradar ein mittleres Gesamtrisiko (gelb, siehe Grafik) bei der Nutzung in der Wissensvermittlung.
Für einen verantwortungsvollen Umgang ist es daher entscheidend, neben der technologischen Kompetenz auch das Wissen darüber, wie die KI arbeitet und wie man KI-generierte Inhalte erkennt, bei Lehrenden und bei Lernenden zu stärken. Weiters sind eine transparente Kennzeichnung für KI-generierte Bilder und Videos sowie eine kritische Prüfung und Darlegung der den generierten Inhalten zugrundeliegenden Quellen durch die erstellende Person unerlässlich. Nur durch die Kombination von KI-Technologie und menschlicher Aufsicht kann sichergestellt werden, dass KI-gestützte Bild- und Videogenerierung die Lernprozesse von Kindern und Jugendlichen und den Bildungssektor insgesamt bereichern, ohne den Wahrheitsgehalt und die Vielfalt und Differenziertheit bei der Darstellung von Bildungsinhalten zu schmälern.

Einleitung

Der Einsatz generativer KI in verschiedenen Bildungsbereichen, wie beispielsweise im schulischen Sprach- oder Kunstunterricht, hat sich rasant entwickelt.[1,30] KI-Systeme zur Bild- und Videogenerierung können aus einfachen Textvorgaben äußerst realistische und komplexe Bilder generieren, sodass auch Personen ohne Programmierkenntnisse niederschwellig und mit geringem Vorwissen Tools wie FLUX in Le Chat von Mistral oder DALL-E nutzen können.[1] Darüber hinaus sind visuelle Lernmethoden von Vorteil, da Bilder eine geringere kognitive Leistung verursachen als Texte und somit effektiver sind.[2]

Einige empirische Studien haben gezeigt, dass der Einsatz solcher KI-Systeme zu Bildungszwecken von Vorteil sein kann, da Lernende Informationen effektiver verarbeiten, wenn verbale und visuelle Darstellungen (unabhängig davon, ob diese mit oder ohne KI erstellt wurden) miteinander verknüpft werden und so ein tieferes Verständnis und eine langfristige Beibehaltung des Gelernten gefördert werden.[30] Zum Beispiel können künstlerische und handwerkliche Aktivitäten von Kindern durch eine KI-generierte Visualisierung eines Fantasietiers unterstützt werden.[1] Ein Einsatz von KI-Systemen, bei dem die Schüler:innen selbstständig Lernmaterialien generieren, kann (zumindest kurzzeitig) ein höheres Maß an Selbstwirksamkeit und eine stärkere emotionale Beteiligung fördern, indem die Aufmerksamkeit der Schüler:innen gefesselt, ihr Interesse geweckt und so das Lernerlebnis insgesamt verbessert wird. Der interaktive Charakter der Arbeit mit solchen Systemen fördert die aktive Teilnahme und das praktische Lernen und verbessert das Engagement der Schüler:innen. Für Lehrer:innen können KI-Systeme die Arbeitsbelastung reduzieren und interaktivere und interessantere Unterrichtsstunden schaffen, die motivierend und produktiv sind. [1,5]

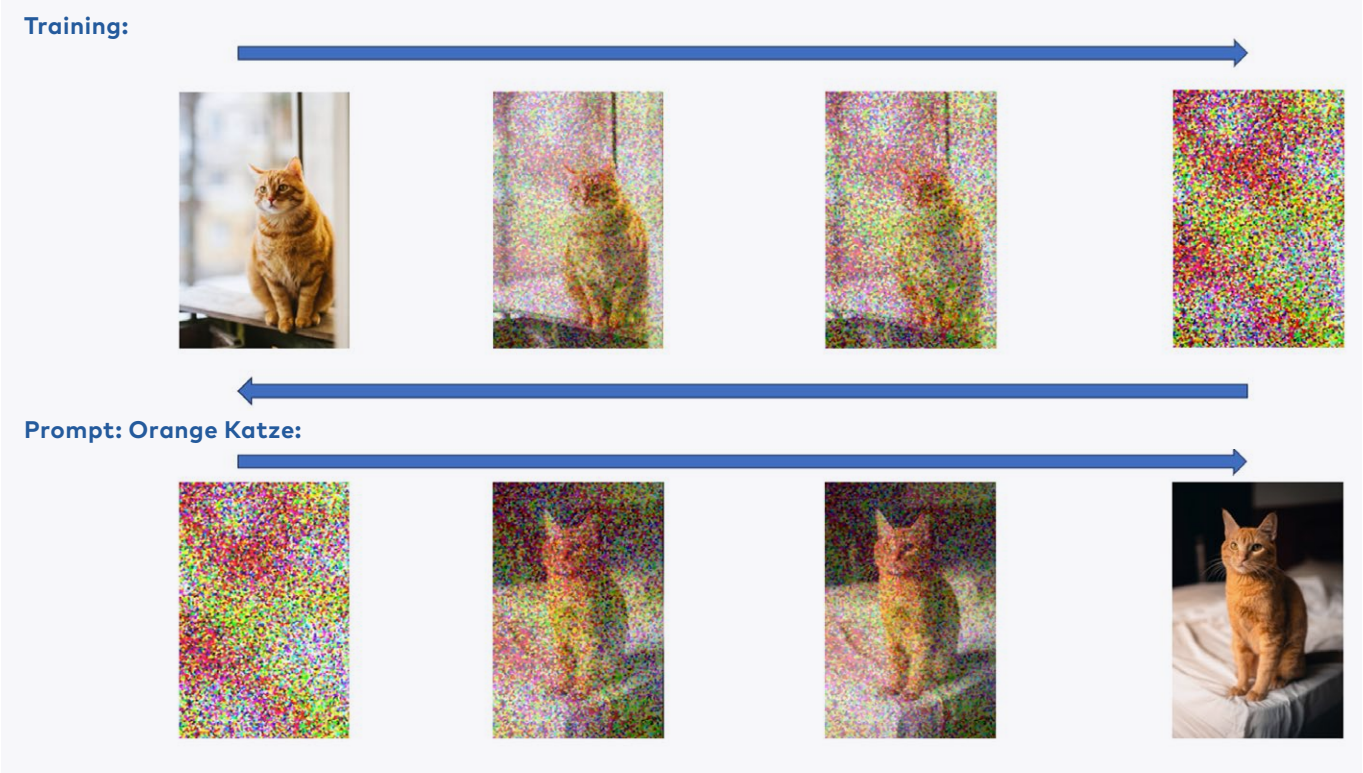
Technologiebeschreibung

Bei der KI-Bild- oder Videogenerierung handelt es sich um eine Technologie, bei der aus Textvorgaben oder vorhandenem Bildmaterial neue Bilder oder Videos erstellt werden. Basierend auf den textlichen Vorgaben und Beschreibungen des gewünschten Bildes oder Videos (=Prompt) generiert das System den gewünschten Output. Das generierte Bild oder Video spiegelt im Prompt beschriebene Aktionen, Bewegungen und Erzählabläufe wider und ist daher die Grundlage des generierten Materials.[31] Dahinter liegen Modelle, die anhand sehr großer Datensätze mit Millionen von Bildern und zugehörigen Texten trainiert wurden. Durch das Lernen aus diesen Beispielen entwickeln die Modelle ein Verständnis für Muster wie Formen, Farben, Stile und Kontexte. Dieses Wissen nutzen sie, um neue Bilder und Videos zu generieren.[3]

Im Laufe der Zeit wurden verschiedene Architekturen entwickelt, darunter GANs (Generative Adversarial Networks), VAEs (Variational Autoencoders) und die neuesten und erfolgreichsten Diffusionsmodelle. Viele der stark verwendeten KI-Tools basieren alle auf Diffusionsmodellen. Dabei handelt es sich um Bildgeneratoren, die durch „Entrauschen" die Generierung von Bildern lernen. Das heißt, dass sie lernen, hinzugefügtes Rauschen zu einem Bild (Störung in digitalen Bildern, die sich als zufällig verteilte, bunte oder helle Punkte zeigt) rückgängig zu machen. Während des Trainings werden ihnen zunächst reale Bilder gezeigt, denen nach und nach zufälliges Rauschen hinzugefügt wird, bis sie schließlich zu reinem Rauschen werden. Bei jedem Schritt lernt das Modell, wie es das Bild leicht entrauschen kann, d.h. es lernt, wie die

sauberere Version aussehen sollte. Nachdem es diesen Prozess viele Male durchlaufen hat, wird das Modell in der umgekehrten Aufgabe trainiert. Dabei geht es von reinem Rauschen aus und entfernt dieses nach und nach, um ein realistisches Bild zu erstellen. Wird eine Textvorlage bereitgestellt, wird diese in Zahlen umgewandelt. Diese Zahlen steuern den Entrauschungsprozess und formen das zufällige Rauschen langsam zu einem Bild um, das der Beschreibung entspricht.[3,4]

Das folgende Diagramm zeigt das Grundprinzip von Diffusionsmodellen:



(© Bild der Katze: EVG Kowalievska auf pexels.com, eigene Erstellung der Grafik)

Videos – im Gegensatz zu statischen Bildern – enthalten Bewegung und Kontinuität in sequenzieller Abfolge. Dies bedeutet größere Herausforderungen für eine narrative Konsistenz und chronologische Kohärenz. Insbesondere für die Videogenerierung wurden daher Diffusionsmodelle zu sogenannten flowbasierten Modellen weiterentwickelt, welche zusätzlich zu den Entrauschungsprozessen auch noch eine deterministische und verlustfreie Umwandlung zwischen Daten- und Merkmalsraum ermöglichen und sehr präzise Bild- und Audio-Output zusammenführen können.[31]

Das zugrunde liegende Problem dieser Technologie ist nicht das Offensichtlichste: Die für das Training von KI-Systemen verwendeten Daten sind oft voreingenommen (***Biased***) oder nicht ausreichend diversifiziert. Dies führt zur Erzeugung von Bildern und Videos, die systematische Verzerrungen aufweisen und Stereotype weiter verschärfen. Dies wird später im Detail erläutert.

Chancen

Für die Lehrkräfte:

- **Unterstützung der Lehrer:innen:** Lehrer:innen benötigen oft viel Zeit für die Recherche und Aufbereitung von Inhalten. KI-Tools können hier Abhilfe schaffen und Lehrer:innen bei sich wiederholenden Tätigkeiten entlasten.[1,5] Zu beachten ist, dass es jedoch auch zu einer Verschiebung der Arbeit kommt, weil KI-generierte Outputs kontrolliert werden müssen. Weiters gibt es zusätzlichen Koordinations-, Prüf- und Reflexionsaufwand für Lehrpersonen darüber, welche KI-Tools eine zielführende Ergänzung zum Unterricht darstellen.

- **Verbesserung des Lernens (Lernprozesse):** Forschungsergebnisse haben gezeigt, dass innovative Unterrichtsmaterialien das emotionale Engagement steigern können, indem sie Neugier und Interesse wecken.[6] Dies lässt sich auf den Einsatz von KI-Werkzeugen durch Lehrkräfte im Bildungsbereich übertragen. Studien zeigen, dass KI-generierte Videos im gleichen Ausmaß wie traditionelle Lehrvideos das anfängliche Verständnis und die Anwendung von Wissen durch die Lernenden wirksam verbessern können.[30] Lehrer:innen können mithilfe von KI-Werkzeugen mit einem geringeren Zeitaufwand einführende Lehrvideos erstellen, die beispielsweise mehrere Schwierigkeitsgrade aufweisen, in unterschiedlichen Sprachen verfügbar sind oder Untertiteln haben. Weiters hat ein Experiment einer Forschungsgruppe ergeben, dass der Einsatz von KI-Tools im Kunstunterricht die aktive Beteiligung und das praktische Lernen fördern und somit das Engagement der Schüler:innen verbessern können.[1] Naturgemäß hängt die jeweilige Wirkung stark von der Lernsituation und den verwendeten Methoden ab. So sind KI-generierte Videos im Bildungskontext an Hochschulen effektiver als in der frühkindlichen Bildung.[30]

Für die Jugendlichen:

- **Niederschwellige Lernmöglichkeiten:** Ein weiterer möglicher Vorteil der Nutzung von Bild- und Videogenerierung für den Wissenstransfer sowie von KI im Allgemeinen ist, dass Wissen durch den Zugang zu KI-Tools potenziell niederschwelliger möglich ist. So können KI-Tools wie DALL-E oder Sora Videolernmaterialien für Menschen mit Hörbeeinträchtigungen zugänglicher machen oder die lernenden Jugendlichen können (auch außerhalb des Unterrichts) ihr Verständnis und ihre Aussprache üben, indem sie Szenen beschreiben und Videogeneratoren entsprechende Videos erstellen lassen.[32] Ein aktuell breiter, kostenloser Zugang zu solchen Tools kann Jugendlichen und Familien mit begrenzten finanziellen Mitteln den gleichen Zugang zu Informationen und Bildungsqualität ermöglichen wie jenen mit mehr finanziellen Ressourcen.[2,7] Lerninhalte können auf diesem Weg an das jeweilige Wissensniveau der Jugendlichen angepasst, von diesen selbst angefragt, erstellt und konsumiert werden.

- **Selbstwirksamkeit:** Einzelne Studien deuten darauf hin, dass Schüler:innen, die mit einer KI-gestützten Lehrmethode für Bildende Kunst unterrichtet wurden, im Vergleich zu Schüler:innen, die mit der herkömmlichen Lehrmethode unterrichtet wurden, kurzfristig ein deutlich höheres Maß an Selbstwirksamkeit aufweisen.[1,31] Dies bedeutet, dass sie stärker davon überzeugt sind, ihre Ziele durch eigene Fähigkeiten und Mittel erreichen und etwaige Hindernisse auf dem Weg dorthin erfolgreich überwinden können. KI kann darüber hinaus die Innovationsfähigkeit und Kreativität von Menschen unterstützen, indem einfacher neue Bereiche erschlossen werden können. In gewissem Maße können KI-Tools auch dazu verwendet werden, erworbenes Wissen zu üben. In der Medizin gibt es beispielsweise Versuche zur Generierung von Röntgen- und EKG-Bildern, mit denen Medizinstudierende ihre Interpretationsfähigkeiten trainieren und verbessern können.[8] Insgesamt fehlen jedoch Studien, welche eine langfristig positive Wirkung auf die Selbstwirksamkeit belegen können.

Risiken und Herausforderungen

Für die Jugendlichen und Lehrkräfte:

1. **Stereotypisierung und KI-Bias (*Systematische Verzerrungen*):** Die häufigsten Verzerrungen, die in KI-Tools zur Umwandlung von Text in Bildern zu finden sind, sind Fehlrepräsentation (z.B. schädliche Stereotypisierung von Frauen oder Minderheiten), Unterrepräsentation (z.B. Ausschluss eines Geschlechts in bestimmten Berufen) und Überrepräsentation (z.B. anglozentrische Perspektiven als Standard).[2,41] Dies führt in weiterer Folge zur Verstärkung bestimmter bestehender Stereotypen von Personengruppen, da beispielsweise Frauen eher lächelnd dargestellt werden während Männer ernst und professionell gezeigt werden. Eine aktuelle Studie zu KI-generierten Bildern von Berufstätigen ergab beispielsweise, dass Männer in 76% der Bilder vertreten waren, Frauen hingegen nur in 8%, also im Vergleich zu ihrer tatsächlichen Vertretung in den Berufen wesentlich unterrepräsentiert wurden.[24] Durch den Konsum KI-generierter Inhalte können somit Vorurteile und Stereotypen auf gesellschaftlicher Ebene weiter verschärft werden.[9,41]

2. **Halluzinationen und Desinformation:** Ein Problem, das bei allen LLM-basierten Systemen auftritt – und LLMs sind Teil von Bild- und Videogeneratoren –, sind Halluzinationen, d. h. ungenaue oder falsche Informationen, die authentisch erscheinen, aber faktisch falsch und als unabsichtliche KI-Fehler einzuordnen sind. Hinzu kommt das große Potenzial für absichtliche Desinformation, also absichtliche Verbreitung falscher oder irreführender Information. Im Gegensatz zu früheren Formen von menschlicher Manipulation können KI-Systeme mit wesentlich geringerem Aufwand überzeugende Bilder, Videos und Texte generieren, die die Grenze zwischen Fakt und Fiktion verwischen.[36] Halluzinationen und Desinformation stellen ein großes Risiko dar, insbesondere für technologisch benachteiligte Gruppen wie ältere Menschen und Kinder.[34,35] Falsche, irreführende oder auch qualitativ minderwertige Informationen, die von einem KI-Tool glaubhaft vermittelt werden, können zur Polarisierung der Gesellschaft beitragen und damit allgemeingültige gesellschaftliche Werte in Frage stellen.[37]

3. **Datenschutz & Urheberrechte:** Generative KI-Systeme benötigen für das Training große Datenmengen, in denen sensible Informationen, wie beispielsweise persönliche Identitätsdaten, enthalten sein können. Werden solche Daten gestohlen oder missbraucht, verlieren Betroffene die Kontrolle über Verwendung, Kontext und Weiterverbreitung ihrer Daten und Bilder, mit potenziell gravierenden Folgen für Privatsphäre und Sicherheit, etwa durch Identitätsdiebstahl.[10] Besonders relevant ist dies, weil neue KI-Modelle mit tausenden Bildern aus dem Internet trainiert werden, unter denen sich auch Fotos von realen Personen befinden können. Nach der Datenschutz-Grundverordnung (DSGVO), Art.4, zählen Fotos von erkennbaren Personen als personenbezogene Daten.[11] Angesichts der Tatsache, dass diese Daten zur Erstellung von ***Deepfakes*** genutzt werden können, ist das Datenschutzproblem äußerst besorgniserregend. Lehrpersonen dürfen daher keine KI-generierten Videos oder Bilder auf Basis von realen Personen (inklusive Schüler:innen) verwenden. Zugleich müssen sie Schüler:innen und Jugendliche über Datenschutz- und Deepfake-Risiken aufklären, insbesondere, weil es Kindern schwerer fällt, Deepfakes zu erkennen und sie darüber hinaus anfälliger für schädliche Online-Praktiken wie Grooming oder Cybermobbing sind.[42] Auch urheberrechtliche Fragen gehören im Unterricht thematisiert, da Sprach- und Bildmodelle regelmäßig mit einer großen Menge an urheberrechtlich geschützten Daten trainiert werden. Inwieweit die Nutzung dieser Inhalte zum Training von KI zulässig ist, wird aktuell kontrovers diskutiert und wird in den kommenden Jahren die Gerichte beschäftigen.[13] Besonders bei KI-Bildern muss man aber aufpassen: KI-Bilder, die real existierende Personen darstellen oder nachahmen, können das Recht am eigenen Bild (§ 78 UrhG) oder datenschutzrechtliche Persönlichkeitsrechte verletzen (Artikel 8 EMRK, DSGVO).[12] Dies gilt insbesondere bei der Erstellung und Veröffentlichung von Deepfakes oder realitätsnahen Porträts.[14]

4. **Vereinheitlichung von Stilen:** Da KI-generierte Bilder oder Videos immer basierend auf der Wahrscheinlichkeit bereits bestehender Kunstwerke oder Bilder erstellt werden, besteht insbesondere in Künstler:innenkreisen und im Bereich der kunsthandwerklichen Ausbildung die Befürchtung, dass eine zu starke Abhängigkeit von bestehenden Ideen und Stilen zu einem Teufelskreis führen könnte, in dem sich dieselben Ideen wiederholen und eine Tendenz zur Vereinheitlichung entsteht.[2] Gerade der kreative Prozess des erfinderischen Lernens erfordert es, über das bereits Bekannte hinauszugehen und etwas zu erschaffen, was noch nicht existiert. Das kann von der KI nicht geleistet werden. Sowohl bei der Erstellung

von Lehrmaterialien durch Lehrer:innen als auch von Werken der Schüler:innen im Kunstunterricht sollte daher ein gewisses Maß an Authentizität und Persönlichkeit enthalten sein. Die reine Übernahme der generierten Bilder oder Videos als finale Produkte verschärft hingegen den Mangel an Vielfalt.[33]

5. **Übermäßiges Vertrauen in Technologien:** Mehrere Studien haben gezeigt, dass Menschen dazu neigen, KI-Systemen mehr zu vertrauen als Menschen (ein Phänomen, das als „Automatisierungsbias" bezeichnet wird).[19] Es besteht die Gefahr, dass Menschen die von KI generierten Bilder und Videos nicht mehr ausreichend hinterfragen, selbst dann, wenn sie widersprechende Informationen und Wissen über einen konkreten Fall haben.[18]

6. **Verringerung des Vertrauens in die eigenen Fähigkeiten:** Da Lernen ein biologischer Prozess ist, aktiviert das Überwinden von Denkwiderständen das Belohnungssystem des Gehirns und verankert damit auch nachhaltig Motivation. KI-Tools, die innerhalb von Sekunden Aufgaben „abarbeiten" können, untergraben den Belohnungsmechanismus. Da Kinder und Jugendliche das Gefühl haben, mit diesen Tools nicht mithalten zu können, verschiebt sich die Rolle des oder der Lernenden von einer aktiven Auseinandersetzung mit Inhalten zu einer passiven Konsumation. Die kognitionswissenschaftliche Forschung warnt dabei vor langfristigen Effekten: Wenn anspruchsvolle Denkaufgaben systematisch an KI delegiert werden, verliert das Gehirn die neuronalen Trainingsreize, die für die Entwicklung kritischen Denkens und kreativer Problemlösefähigkeit essenziell sind. Dadurch schwindet das Vertrauen in die eigenen intellektuellen Fähigkeiten.[25, 26]

7. **Versäumnis von wichtigen Entwicklungsschritten:** Es gibt die Befürchtung, dass die Nutzung von KI-Tools die gesunde Entwicklung von Kindern negativ beeinflussen kann. Neurowissenschaftliche Forschungen haben gezeigt, dass handwerkliches Schaffen, wie beispielsweise das Berühren und Formen verschiedener Materialien mit den Händen, Bereiche des Gehirns stimuliert, die für die Entwicklung von Kindern von entscheidender Bedeutung sind. Dies deutet darauf hin, dass solche Tätigkeiten nicht durch KI ersetzt werden sollten.[2,7] Generieren Schüler:innen mithilfe von KI-Tools Bilder, anstatt selbst zu zeichnen oder zu fotografieren, entfällt nicht nur die feinmotorische Tätigkeit des Zeichnens, Malens oder Fotografierens, sondern auch die haptische Auseinandersetzung mit unterschiedlichen Materialien wie Papier, Stiften, Farben oder der Kamera. Dabei zeigen Studien, dass nicht nur die eigene handschriftliche Auseinandersetzung mit Material die besten Lerneffekte erzielt, sondern auch die motorischen Fähigkeiten von Schulkindern ab 6 Jahren gefördert werden sollen, indem händisch kreativ gestaltet wird.[39] Die Forschung zeigt, dass Kinder ihre eigenen Erwartungen an die Fähigkeiten der KI anpassen, wenn KI-Tools ihre kreativen Ideen nicht sofort umsetzen können.[40] Auch deshalb ist es wichtig, handwerkliche Aktivitäten von Kindern zu fördern und auszubauen.

8. **Verringerung von Kreativität:** Wissenschaftlich wird Kreativität als eine Fähigkeit konzeptualisiert, die durch Lernen, Experimentieren und reflexives Engagement entwickelt wird.[27] Aus dieser Perspektive ist Kreativität mehr als die Erzeugung neuartiger Kombinationen, worauf generative KI beruht. Kreativität ist ein Prozess, bei dem Einzelpersonen und Teams Fähigkeiten zur Problemformulierung, Urteilsvermögen und eine kreative Identität entwickeln. Durch den Einsatz generativer KI entstehen daher seltener neue Ideen, wodurch langfristig auch die Innovationsfähigkeit der Menschheit gemindert sein kann.[28, 29]

9. **Niedrige Genauigkeit:** Trotz der Fortschritte in der KI-Technologie der letzten Jahre sind KI-Bilder in Bezug auf Details immer noch recht ungenau, insbesondere bei der Textgenerierung innerhalb des Bildes selbst. Diese Ungenauigkeiten können zu einer negativen Wahrnehmung und Verwirrung führen, wenn Menschen sie zu Bildungszwecken nutzen. Manche Studien, die die Anwendung der Text-zu-Bild-Generierung in der Radiologie untersuchten, ergaben beispielsweise, dass das medizinische Personal die anatomische Genauigkeit von KI-Tools wie DALL-E als nicht zufriedenstellend empfand.[8,15]

Für die Gesellschaft:

1. **Energie- und Ressourcenverbrauch:** Die Entwicklung, die Herstellung von Hardware, sowie Training und die tägliche Nutzung von KI-Werkzeugen erfordern enorme Mengen an Strom, Wasser und weiteren Ressourcen.[16] Die Internationale Energieagentur (IAEA) schätzt den Anteil der Rechenzentren am globalen Stromverbrauch zwar aktuell noch auf 1,5 %, erwartet jedoch, dass dieser bis 2030 auf 10 % steigen wird.[17] Eine derart drastische Veränderung des Energieverbrauchs dürfte aufgrund der hohen Nachfrage zu Instabilität im Energiesektor führen. Dies hätte höhere Preise für die Bevölkerung zur Folge. Darüber hinaus kann der gesteigerte Energiebedarf dazu führen, dass fossile Kraftwerke verstärkt eingesetzt werden, wodurch sich die bestehende Klimakrise weiter verschärfen würde.[16]

2. **Menschliche Anti-KI-Voreingenommenheit:** Die Kennzeichnung von Inhalten als KI-generiert kann zu einer abwertenden Reaktion bei den Nutzer:innen führen: Sie bewerten die vermittelten Inhalte mitunter als weniger wirksam, bringen häufiger Gegenargumente vor und zeigen eine geringere Konsistenz in den Reaktionen auf die von KI-generierten Inhalte. Wenn die Teilnehmenden die Verwendung von KI erkennen können, überwiegt daher möglicherweise ihre Skepsis gegenüber den potenziellen Vorteilen bei der Gestaltung der präsentierten Inhalte.[7] Dies sollte von Lehrenden berücksichtigt werden.

Empfehlungen für den Praxiseinsatz

Für Bildungseinrichtungen, Lehrkräfte und Bildungspolitik:

- **Lehrer:innen als Vorbilder für den Umgang mit KI-Tools:** Lehrpersonen stellen für die Schüler:innen ein Vorbild im Umgang mit KI-Tools dar. Es ist daher unumgänglich, dass Lehrpersonen einen kritischen und reflektierten Umgang mit KI-Tools in Hinblick auf ökonomische Abhängigkeiten und Souveränität, Datenschutz, Deepfakes, Falschinformationen, ökologische Herausforderungen und Kennzeichnungspflichten vorleben. Damit können sie die Bedeutung eines reflektierten KI-Einsatzes an ihre Schüler:innen vermitteln.

- **Human-in-the-Loop:** Insbesondere im Bildungsbereich, wo bestimmte Anwendungen gemäß dem EU-KI-Gesetz als Hochrisikokategorie gelten, sollten KI-Tools nie automatisiert eingesetzt werden, sondern möglichst immer unter menschlicher Kontrolle bleiben. Pädagog:innen könnten KI-generierte Vorlagen verwenden, um Schüler:innen zum Experimentieren anzuregen.[5] Anschließend ist es jedoch zentral, dass sie mit den Schüler:innen gemeinsam erarbeiten, nach welchen Prinzipien die künstliche Intelligenz arbeitet, um die Komposition der generierten Inhalte besser zu verstehen. Dieser Ansatz verbindet die Effizienz der KI mit der unersetzlichen Rolle menschlicher Mentor:innen bei der Förderung des Wissensaufbaus.[1]

- ***Transparenz*** **und Erklärbarkeit:** Transparenz ist einer der vertrauensbildenden Faktoren im Einsatz von KI. Viele Menschen sind besorgt darüber, die Kontrolle über Technologie und die Verwendung ihrer Daten zu verlieren. Das bedeutet, dass Lehrer:innen und Lernende in der Lage sein sollten, zu verstehen, wie ein KI-System funktioniert, welche Auswirkungen es haben kann, und ob es vertrauenswürdig arbeitet und eingesetzt wird.[22] Indem sie bewusst KI-Systeme mit ***Erklärbarer KI (XAI)*** zur Verfügung gestellt bekommen oder selbst auswählen, können Lehrkräfte und Lernende die zur Generierung von Informationen verwendeten Quellen und Logiken einsehen. Dadurch kann eine kritische Bewertung der Ergebnisse gefördert werden.[10,22] Nutzer:innen sollten außerdem über den Einsatz von KI informiert werden, da die Kennzeichnungspflichten des EU AI-Act im August 2026 in Kraft getreten sind. Vor allem bei täuschend echt aussehenden KI-generierten Inhalten, wie Texten, Bildern oder Videos, welche von Menschen nicht als KI-generiert erkannt werden können, gilt eine Kennzeichnungspflicht.[23]

- **Risikobewertung und -monitoring von KI-Tools:** Vor der Einführung von Bild- und Videogeneratoren ist eine Risikobewertung zu empfehlen, um rechtliche, technische und kontextuelle Anforderungen und Auswirkungen des KI-Einsatzes abschätzen und frühzeitig entsprechende Vorkehrungen treffen zu können. Eine Methode dafür ist die ALAIT Ex-Ante Risikobewertung von KI-Systemen. Wesentlich ist dabei eine klare Definition der KI-Aufgaben: Für welche Aufgabenstellungen können KI-generierte Bilder oder Videos verwendet werden, für welche nicht?

- **Bias-Testing und Debiasing:** Da Verzerrungen und die Förderung von Stereotypen erwiesenermaßen Schwachstellen von generativen KI-Werkzeugen darstellen, ist es unerlässlich, gegenzusteuern und KI-Systeme im Bildungsbereich auf Bias zu testen und gegebenenfalls Debiasing-Maßnahmen durchzuführen, um Bias und Stereotype gezielt zu reduzieren (Auswahl von Tools mit weniger Bias-Risiken, Schulungen für Nutzende, etc.).[9] Um der Komplexität und Sensibilität dieses Themas gerecht zu werden, ist außerdem eingehendere Forschung zum Verständnis von Vorurteilen und Diversität nötig.

- **Aufbau und Stärkung des KI-Wissens:** Es ist wichtig, sich über KI-Technologien zu informieren, die richtigen Kompetenzen aufzubauen und die eigenen Rechte zu kennen.[18] In Österreich besteht mit der Rundfunk- und Telekom Regulierungs-GmbH (RTR) eine Informations- und Anlaufstelle für die breite Öffentlichkeit. Die RTR-KI-Servicestelle unterstützt Bürger:innen bei Fragen zur Transparenz, zu Rechten im Umgang mit KI-Systemen oder zur Überprüfung von KI-Entscheidungen.[20,21]

Für KI-Anbieter:innen:

- **Privacy by Design:** Bei der Verwendung generativer KI im Bildungsbereich ist der Schutz der Privatsphäre und die Sicherheit von Personen besonders zu berücksichtigen. Wichtige Maßnahmen betreffen beispielsweise die Verwendung sicherer Methoden zur Datenspeicherung und -übertragung (u.a. Verschlüsselung, Anonymisierung, De-Identifizierung), die Stärkung der Datenzugriffskontrolle sowie Maßnahmen zur Berechtigungsverwaltung (Zugriff nur für autorisierte Personen).[10]

- **Präzise, hochwertige und diversifizierte Daten:** Die Zuverlässigkeit von KI-Systemen hängt zu einem wesentlichen Teil von der Qualität der Datensätze ab, mit denen diese trainiert bzw. im laufenden Betrieb gespeist werden. Um ***KI-Bias*** zu vermeiden und die ***Genauigkeit*** zu verbessern, sollen die Datensätze repräsentativ sein sowie eindeutig und strukturiert beschriftet werden.

- **Souveräne KI-Tools:** Nicht alle KI-Tools gehen sorgsam mit Nutzerdaten um und erlauben eine Kontrolle des Datenflusses. Zudem sind Lock-in-Effekte und steigende Preise zu berücksichtigen. Ziel sollte es daher sein, insbesondere für den Unterricht DSGVO-konforme, geopolitisch unabhängige, und erklärbare KI-Systeme zu schaffen, die den Unterricht punktuell und gezielt bereichern können.

Wichtige Begriffe

KI-Genauigkeit (AI-Accuracy): Bezieht sich auf die Fähigkeit eines KI-Systems, korrekte Vorhersagen oder Entscheidungen zu treffen. Sie ist ein wichtiger Maßstab für ihre Leistung und entscheidend für die Bestimmung ihrer Wirksamkeit und Zuverlässigkeit.

KI-Autonomie: Die Fähigkeit eines KI-Systems, eine Reihe von Zielen unter einer Reihe von Unsicherheiten in ihrer Umgebung selbstständig und ohne externe Eingriffe zu erreichen.

Gesetz der Europäischen Union über künstliche Intelligenz (EU AI-Act): Eine europäische Verordnung über künstliche Intelligenz (KI) – die erste umfassende Verordnung über KI von einer großen Regulierungsbehörde. Sie konzentriert sich insbesondere auf KI-Systeme mit hohem Risiko.

Datenschutz in der KI: Die Gesamtheit der Praktiken und Bedenken im Zusammenhang mit der ethischen Erfassung, Speicherung und Nutzung personenbezogener Daten durch Systeme der künstlichen Intelligenz.

Transparenz: Bedeutet, dass die Funktionsweise, Entscheidungsprozesse und Einsatzbereiche eines KI-Systems nachvollziehbar, erklärbar und offen zugänglich sind – für Entwickler:innen, Nutzer:innen und andere Stakeholder.

Erklärbare KI (XAI): Bezieht sich auf Methoden und Techniken, die es ermöglichen, die Entscheidungen und Ergebnisse von KI-Systemen besser zu verstehen und nachzuvollziehen.

KI-Bias, systematische Verzerrungen: Bias ist eine systematisch unterschiedliche Behandlung bestimmter Objekte, Personen oder Gruppen im Vergleich zu anderen. Behandlung ist jede Art von Handlung, einschließlich Wahrnehmung, Beobachtung, Darstellung, Vorhersage oder Entscheidung.

Deepfakes: Sind realistisch wirkende Medieninhalte (Fotos, Audiodateien oder Videos), die durch Techniken der KI verändert, erstellt bzw. verfälscht wurden.

Erklärung Stufenmodell des ALAIT Risikoradars

Im ALAIT KI-Risikoradar wird die Beziehung zwischen Anwendungsrisiko und Autonomie eines KI-Systems dargestellt. Die Risikostufen stützen sich auf das EU KI-Gesetz (EU AI-Act), insbesondere auf Artikel 6 und Annex III, die sich mit risikoreichen Anwendungsbereichen von KI befassen. Geringere System-Autonomie und Anwendungsrisiken werden durch kältere Farben (blau) und höhere System-Autonomie und Anwendungsrisiken durch wärmere Farben (rot) dargestellt.

Der Farbwechsel vermittelt das erhöhte Risiko solcher Entscheidungen. Mithilfe dieser Farbskala lässt sich das Gesamtrisiko erkennen: Violett und dunkelrot – sehr hoch, rot und dunkelorange – hoch, hellorange und Gelbtöne – mittel, Blautöne – geringes Gesamtrisiko. Im Idealfall sollten hohe Anwendungsrisiken und System-Autonomie vermieden oder nur nach sehr sorgfältiger Abwägung eingesetzt werden.

Autonomiegrad des KI-Systems

Stufe 1: Keine Autonomie

KI ist ein passives Werkzeug. Menschen treffen alle Entscheidungen und leiten die Maßnahmen ein.
Empfohlene Anwendungsfälle: Szenarien mit hohen Risiken oder bei denen ethische Entscheidungen bedeutend sind (z. B. medizinische Diagnostik, Justizsystem).

Stufe 2: Geringer Autonomiegrad (Human-in-the-Loop)

Die KI gibt Empfehlungen oder Optionen, aber die Benutzer:innen bleiben für die Auswahl und Genehmigung von Maßnahmen verantwortlich.
Empfohlene Anwendungsfälle: Aufgaben mittlerer Komplexität mit mäßigen Risiken.

Stufe 3: Mittlerer Autonomiegrad (Human-on-the-Loop)

Die KI führt bestimmte Aufgaben autonom aus, wobei Menschen in Ausnahmefällen eingreifen.
Empfohlene Anwendungsfälle: Szenarien, in denen eine kontinuierliche menschliche Beteiligung nicht erforderlich ist, kritische Risiken jedoch eine menschliche Überwachung erfordern.

Stufe 4: Hoher Autonomiegrad (Human in Control)

Das KI-System arbeitet weitgehend autonom, erlaubt es den Benutzer:innen jedoch, es selbst zu übersteuern, um unerwünschte Ergebnisse zu vermeiden.
Empfohlene Anwendungsfälle: Umgebungen mit geringem bis mittlerem Risiko.

Stufe 5: Vollständige Autonomie mit minimaler Aufsicht

Das KI-System arbeitet unabhängig und erfordert nur minimale oder gar keine menschliche Intervention. Die Beteiligung des Menschen beschränkt sich auf die langfristige Aufsicht (Audits).
Empfohlene Anwendungsfälle: Umgebungen mit geringen Sicherheits- oder ethischen Risiken und hoher Zuverlässigkeit des KI-Systems (z. B. sich wiederholende Aufgaben in kontrollierten Umgebungen).

Anwendungsbereich-Risiko

Stufe 1: Minimales Risiko

Das KI-System hat keine Auswirkungen auf die Benutzer:innen oder die Entscheidungsfindung.
Kriterien: Keine direkte Auswirkung auf die Hochrisikobereichen des EU AI-Acts.

Stufe 2: Begrenztes Risiko

KI-Systeme, die mit Benutzer:innen interagieren, aber keine Entscheidungen mit hohen Risiken treffen. Das Risiko steigt, wenn es an Transparenz über die Beteiligung von KI mangelt.
Kriterien: Bereiche, die nicht in der Liste der „hohen Risiken" des EU AI-Acts enthalten sind.

Stufe 3: Mittleres Risiko

KI-Systeme haben keine besonderen Auswirkungen auf einzelne Personen, aber sie entfalten Wirkung auf kollektiver oder gesellschaftlicher Ebene.
Kriterien: KI-Systeme, die für die öffentliche Nutzung verfügbar sind und das Potenzial haben, bestehende Gepflogenheiten zu beeinflussen und langfristig zu verändern.

Stufe 4: Hohes Risiko

Jeder Algorithmus, der in den laut EU AI-Act „Hochrisikobereichen" angewendet wird oder direkte Auswirkungen auf einzelne Personen hat.
Kriterien: Zugehörigkeit zum „Hochrisikobereich" des EU AI-Acts, nur wenn die Regeln für Transparenz und Datenqualität eingehalten werden.

Stufe 5: Extremes Risiko

Jeder Algorithmus, der in den laut EU AI-Act „Hochrisikobereichen" angewendet wird.
Kriterien: Zugehörigkeit zum „Hochrisikobereich", wenn die Regeln für Transparenz und Datenqualität NICHT eingehalten werden (zum Beispiel aufgrund von Übergangsfristen).

Quellen

1 Bian, C., Wang, X., Huang, Y., Zhou, S., & Lu, W. (2025). Effects of AI-generated images in visual art education on students' classroom engagement, self-efficacy and cognitive load. Humanities and Social Sciences Communications, 12(1), 1548. https://doi.org/10.1057/s41599-025-05860-2

2 Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: Crafting with text-to-image generative models. Digital Creativity, 34(1), 1–21. https://doi.org/10.1080/14626268.2023.2174557

3 Understanding AI Image Generation: Models, Tools, and Techniques | DigitalOcean. (n.d.). Retrieved 27 November 2025, from https://www.digitalocean.com/community/tutorials/understanding-ai-image-generation-models-tools-and-techniques

4 Improving Diffusion Models as an Alternative To GANs, Part 2. (2022, April 26). NVIDIA Technical Blog. https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-2/

5 Leiker, D., Finnigan, S., Gyllen, A. R., & Cukurova, M. (2023). Prototyping the use of Large Language Models (LLMs) for adult learning content creation at scale (arXiv:2306.01815). arXiv. https://doi.org/10.48550/arXiv.2306.01815

6 Reed, J. (2023). AI Image-Generation as a Teaching Strategy in Nursing Education.

7 Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. Computers & Education, 224, 105164. https://doi.org/10.1016/j.compedu.2024.105164

8 Miftahul Amri, M., & Khairatun Hisan, U. (2023). Incorporating AI Tools into Medical Education: Harnessing the Benefits of ChatGPT and Dall-E. Journal of Novel Engineering Science and Technology, 2(02), 34–39. https://doi.org/10.56741/jnest.v2i02.315

9 Barve, S., Mao, A., Shi, J. M., Juneja, P., & Saha, K. (2025). Can we Debias Social Stereotypes in AI-Generated Images? Examining Text-to-Image Outputs and User Perceptions (arXiv:2505.20692). arXiv. https://doi.org/10.48550/arXiv.2505.20692

10 Yu, H., & Guo, Y. (2023). Generative artificial intelligence empowers educational reform: Current status, issues, and prospects. Frontiers in Education, 8. https://doi.org/10.3389/feduc.2023.1183162

11 Recht am eigenen Bild | datenschutz.rlp.de. (n.d.). Retrieved 23 January 2026, from https://www.datenschutz.rlp.de/themen/recht-am-eigenen-bild

12 Geuer, D. K. G. und D. E. (2025, July 28). Das Recht am eigenen Bild im Kontext von KI-Tools. GEUER Rechtsanwälte OG. https://www.geuer.at/2025/07/28/ki-bildrechte/

13 Künstliche Intelligenz: Eine Herausforderung für das Urheberrecht. (n.d.). Retrieved 15 October 2025, from https://www.wu.ac.at/forschung/forschungsportal/news/archiv-news/detail/kuenstliche-intelligenz-eine-herausforderung-fuer-das-urheberrecht

14 Die Haftungsfrage bei KI-generierten Bildern | KOMMUNAL. (n.d.). Retrieved 26 January 2026, from https://kommunal.at/die-haftungsfrage-bei-ki-generierten-bildern

15 Temsah, M.-H., Alhuzaimi, A. N., Almansour, M., Aljamaan, F., Alhasan, K., Batarfi, M. A., Altamimi, I., Alharbi, A., Alsuhaibani, A. A., Alwakeel, L., Alzahrani, A. A., Alsulaim, K. B., Jamal, A., Khayat, A., Alghamdi, M. H., Halwani, R., Khan, M. K., Al-Eyadhy, A., & Nazer, R. (2024). Art or Artifact: Evaluating the Accuracy, Appeal, and Educational Value of AI-Generated Imagery in DALL·E 3 for Illustrating Congenital Heart Diseases. Journal of Medical Systems, 48(1), 54. https://doi.org/10.1007/s10916-024-02072-0

16 Ressourcenverbrauch von KI: Die Nimmersatt-Industrie und ihre Kosten. (n.d.). AlgorithmWatch. Retrieved 26 January 2026, from https://algorithmwatch.org/de/ressourcenverbrauch-von-ki/

17 Executive summary – Energy and AI – Analysis. (n.d.). IEA. Retrieved 26 January 2026, from https://www.iea.org/reports/energy-and-ai/executive-summary

18 Thapa, B. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society. The European Liberal Forum. https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-data-driven-government-in-a-free-society/

19 Wihlborg, E., Larsson, H., & Hedström, K. (2016). 'The Computer Says No!' – A Case Study on Automated Decision-Making in Public Authorities. 2016 49th Hawaii International Conference on System Sciences (HICSS), 2903–2912. https://doi.org/10.1109/HICSS.2016.364

20 KI-Servicestelle der RTR. (n.d.). RTR. Retrieved 16 September 2025, from https://www.rtr.at/rtr/service/ki-servicestelle/ki-servicestelle.de.html

21 European Centre for Algorithmic Transparency—European Centre for Algorithmic Transparency. (2025, October 24). https://algorithmic-transparency.ec.europa.eu/index_en

22 Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074. https://doi.org/10.1016/j.caeai.2022.100074

23 Kennzeichnungspflicht für bestimmte KI-Inhalte ab August 2026. (n.d.). wko.at. Retrieved 26 January 2026, from https://www.wko.at/oe/gewerbe-handwerk/kennzeichnungspflicht-fuer-ki-inhalte

24 Gorska, A. M., & Jemielniak, D. (2023). The invisible women: Uncovering gender bias in AI-generated images of professionals. Feminist Media Studies, 23(8), 4370–4375. https://doi.org/10.1080/14680777.2023.2263659

25 Assefa, E. A. (2024). The Double-Edged Sword of Artificial Intelligence (AI) in Education: Maximizing benefits while mitigating risks. The Journal of Quality in Education, 14(24), 154-176. https://doi.org/10.37870/joqie.v14i24.450

26 Hasan M. K. (2025). How AI quietly undermines the joy and effort of learning: a call for rebalancing education in the digital age. Annals of medicine and surgery (2012), 87(8), 4693–4694. https://doi.org/10.1097/MS9.0000000000003456

27 Kovvuri, K. R. & Kumar, S. (2026). The Creativity Offloading Paradox: Governing Generative AI, Human Creativity, and Innovation Capability in Organizations http://dx.doi.org/10.2139/ssrn.6100446

28 Doshi, A. R. & Hauser O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances 10. DOI: 10.1126/sciadv.adn5290

29 Moon, K., Kushlev, K., Bank, A., Luttges, B. L., Viskontas, I., Kaufman, J. C., Dan Richard Johnson, Duckworth, A., Green, A. (2026). The Creative Link Between Words and Ideas Is Weakening in the AI-Era. Retrieved from osf.io/preprints/psyarxiv/jsz58_v6.

30 Nguyen, H. V., & Duong, M. T. T. (2026). AI-Generated Video in Education: A Scoping Review of Technologies, Pedagogical Applications, Benefits, and Challenges. *IEEE Access,* 14, 111696–111714. https://doi.org/10.1109/ACCESS.2026.3715230

31 Riedl, L. (2025). Videos mit künstlicher Intelligenz gestalten: Kreative Tools und professionelle Workflows für die Videoproduktion. Springer Fachmedien. https://doi.org/10.1007/978-3-658-46663-3

32 Adetayo, A. J., Enamudu, A. I., Lawal, F. M., & Odunewu, A. O. (2024). From text to video with AI: The rise and potential of Sora in education and libraries. Library Hi Tech News. https://doi.org/10.1108/LHTN-02-2024-0028

33 Spengler, S. (2024). KI im Kontext von Kunst und Kunstunterricht: Auswirkungen von Bild- und Textgenerierung mit Künstlicher Intelligenz auf künstlerisches und (kunst-)pädagogisches Arbeiten. *Ludwigsburger Beiträge zur Medienpädagogik, 24,* 1–26. https://doi.org/10.21240/lbzm/24/04

34 Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S., & Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. *Computer Science Review, 52,* 100632. https://doi.org/10.1016/j.cosrev.2024.100632

35 European Parliament. (o. J.). *Children and generative AI | Think Tank | European Parliament*. Abgerufen 12. März 2026, von https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769494

36 Park, S., & Nan, X. (2026). Generative AI and misinformation: A scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. *AI & SOCIETY, 41*(2), 1501–1515. https://doi.org/10.1007/s00146-025-02620-3

37 Weidinger, L., Uesato, J., Rauh, M. et al. (2022). Taxonomy of Risks posed by Language Models. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22,* 214–229. https://doi.org/10.1145/3531146.3533088

38 Van der Weel, F. R. (Ruud), & Van der Meer, A. L. H. (2024). Handwriting but not typewriting leads to widespread brain connectivity: A high-density EEG study with implications for the classroom. *Frontiers in Psychology, 14,* 1219945. https://doi.org/10.3389/fpsyg.2023.1219945

39 Frikha, M., & Alharbi, R. S. (2023). Optimizing Fine Motor Coordination, Selective Attention and Reaction Time in Children: Effect of Combined Accuracy Exercises and Visual Art Activities. *Children, 10(5),* 786. https://doi.org/10.3390/children10050786

40 Newman, M., Sun, K., Dalla Gasperina, I. B., Shin, G. Y., Pedraja, M. K., Kanchi, R., Song, M. B., Li, R., Lee, J. H., & Yip, J. (2024). „I want it to talk like Darth Vader": Helping Children Construct Creative Self-Efficacy with Generative AI. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24,* 1–18. https://doi.org/10.1145/3613904.3642492

41 Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT '23, 1493–1504. https://doi.org/10.1145/3593013.3594095

42 European Parliament. (2025). Children and deepfakes. Retrieved 13 August 2026, from https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf

Projekt ALAIT

Das Austrian Lab for AI Trust (ALAIT) ist ein vom österreichischen Bundesministerium für Innovation, Mobilität und Infrastruktur (BMIMI) gefördertes Forschungs- und Entwicklungs-Projekt zur Schaffung von Vertrauen durch Wissen im Bereich Künstliche Intelligenz (KI). Das Projekt ALAIT zielt darauf ab, Interessierte und wichtige gesellschaftliche Gruppen zu befähigen, KI-Technologien verantwortungsvoll zu nutzen und ethische sowie qualitativ hochwertige Standards für den Einsatz von AI zu etablieren.

Das Projekt wird von **winnovation** geleitet (Gertraud Leimüller und Brigitte Ömer-Rieder) und im Konsortium mit **leiwand.ai** (Rania Wazir), **TU Wien** (Sabine Köszegi und Karoline Bohrn) und **Austria Presse Agentur – APA** (Verena Krawarik und Sophia Marecek) umgesetzt.

Die ALAIT-Dossiers sind auf der Projekthomepage abrufbar: https://science.apa.at/project/alait/

Die Inhalte des Dossiers entsprechen dem aktuellen Stand der Technik und wurden sorgfältig nach wissenschaftlichen Kriterien erstellt. Sie dienen jedoch nicht als rechtsverbindliche Auskunft oder Beratung.

Impressum

Medieninhaberin und Herausgeberin:
winnovation consulting gmbh
Linke Wienzeile 124/Top 5
1060 Vienna

Dieses Dossier steht unter der Creative Commons Lizenz CC BY-NC-ND 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de
(Namensnennung-Nicht kommerziell-Keine Bearbeitungen 4.0 International)

Veröffentlicht 2026, gefördert durch das Bundesministerium für Innovation, Mobilität und Infrastruktur.

Danksagung:

Wir danken folgendem Experten für sein hilfreiches Feedback zu Vorversionen dieses Dossiers:
Fares Kayali

Grafik Design: drahtwerk.at