

Pilotprojekt zum Thema KI und Musikschulunterricht

Erkenntnisse aus der Erprobung von drei (Custom) ChatGPT-Versionen

Inhalt

Kontext	1
Key Findings / Ergebnisse	2
Fazit	5
Auswertung der Gruppendiskussion (2.10.2025)	6
Einblicke in die Custom-ChatGPT-Dokumentationen	12

Kontext

Das bundeslandübergreifende Pilotprojekt zur Nutzung von Künstlicher Intelligenz (KI) im Musikschulbereich wurde vor dem Hintergrund wachsender Diskussionen über den Einsatz von KI (KI und LLMs wird hier synonym genutzt) in Bildungskontexten initiiert. Ausgangspunkt waren Impulse auf dem Musikschulkongress in Dresden (2025), bei dem der damalige Bundesgeschäftsführer Holger Denckmann die Möglichkeit von speziell konfigurierten sogenannten Custom GPTs für musikpädagogische Fragestellungen thematisierte. Aufbauend darauf übernahm Dr. Rhea Richter, Geschäftsführerin des LVdM.SH, die Initiative, mit Projektmitteln aus DigiMuve.SH (gefördert durch das Bundesprogramm Ländliche Entwicklung und Regionale Wertschöpfung BULEplus), eine praxisnahe Erprobung zu realisieren.

Ziel des Pilotprojekts war es, erfahrungsbasierte Einschätzungen von Musikschullehrkräften zur Nutzung von Custom GPTs zu gewinnen. Custom GPTs sind individuell konfigurierte Varianten von ChatGPT, bei denen festgelegt werden kann, wie die KI reagieren soll und auf welche spezifischen Dokumente oder Wissensquellen sie zugreift, um Antworten gezielt auf einen bestimmten Anwendungsbereich – hier die Instrumentalpädagogik – auszurichten. Insbesondere sollte erprobt werden, ob und wie mit Hilfe von Custom GPTs ein praxisrelevanter Zugang zu vorhandenen Lehrplänen und Literaturlisten geschaffen werden kann.

Das Projekt wurde als landesverbandsübergreifendes Vorhaben umgesetzt. Beteiligt waren Musikschullehrkräfte aus Schleswig-Holstein, Nordrhein-Westfalen und Baden-Württemberg, die im Rahmen einer Testphase unterschiedliche KI-Varianten erprobten. Begleitet wurde diese Erprobung durch Iris Meyer als Projektreferentin beim LVdMSH; die wissenschaftliche Betreuung sowie die Moderation und Auswertung der abschließenden Gruppendiskussion übernahm Matthias Krebs. Unterstützt wurde das Projekt zudem

durch die Digitalisierungsbeauftragten der beteiligten Landesverbände, Thomas Hanz (NRW) und Rainer Kropf (Baden-Württemberg).

Zentrale Bestandteile des Projekts waren: 1.) eine umfangreiche Testung in einem Zeitraum von 3 Monaten, in dem die Musikschullehrkräfte zwei unterschiedlich konfigurierte Custom GPTs sowie das normale ChatGPT mit verschiedenen Fragen aus der instrumentalpädagogischen Praxis konfrontierten. Die Chat-Verläufe wurden von ihnen mittels eines Dokumentationstools festgehalten. 2.) wurde eine Gruppendiskussion geführt, die nicht als Evaluationsinstrument, sondern als exploratives, reflexives Erkenntnisformat angelegt war. Sie diente als Resonanzraum, in dem individuelle Nutzungserfahrungen mit KI nicht nur berichtet, sondern gemeinsam eingeordnet, relativiert und konturiert wurden.

Der vorliegende Bericht fasst die zentralen Erkenntnisse dieses Austauschs zusammen und leitet daraus Perspektiven für die zukünftige Ausgestaltung musikschulischer Praxis sowie verbandlicher Unterstützungsstrukturen ab.

Key Findings / Ergebnisse

In dieser Übersicht werden zentrale Aspekte aus der Auswertung der Nutzungsdokumentationen sowie der Gruppendiskussion zusammengefasst und Empfehlungen entwickelt.

Kurzfazit aus der Erprobung von drei Custom ChatGPT-Versionen

Lehrkräfte bewerten Custom-GPT-Versionen danach, wie nachvollziehbar, kontextsensibel und verlässlich ihre Antworten auf sie wirken. Die Custom-GPT-Version mit zusätzlichem Internetzugang (als „Plus“-Variante bezeichnet) wurde am anschlussfähigsten eingeschätzt, sie

- lieferte nachvollziehbarere Antworten
- ermöglichte sowohl Struktur als auch kreative Impulse
- bot Impulse zur pädagogischen (Selbst-)Reflexion

Empfehlung: Für den VdM liegt der Hebel weniger in der Kontrolle von KI-Antworten als in der Qualität und Transparenz der zugrunde liegenden Materialien sowie in der bewussten Gestaltung von Custom GPTs als reflexive, kontextoffene Werkzeuge für professionelle Praxis.

Weitere Empfehlungen:

1. KI wird von Lehrkräften als Assistenz verstanden

Die Diskussion um KI ist im Musikschulbereich selbstverständlich auch angekommen – jedoch mit hoher Sensibilität gegenüber pädagogischen Werten. Musikschullehrkräfte

sehen KI überwiegend als unterstützendes Werkzeug zur Reflexion, Inspiration und Strukturierung – nicht als Ersatz pädagogischer Entscheidungskompetenz.

- Keine Akzeptanz für Automatisierungslogiken im Unterricht
- Hohe Sensibilität für professionelle Verantwortung
- Vertrauen entsteht dort, wo KI unterstützt, nicht vorgibt

Empfehlung: Der VdM sollte KI nicht als Effizienzprogramm, sondern als professionelles Assistenzsystem rahmen.

2. Der größte Mehrwert von KI liegt außerhalb der Unterrichtssituation

KI wird vor allem in der Vor- und Nachbereitung sowie zur individuellen Reflexion genutzt, nicht jedoch im Instrumentalunterricht selbst.

Beobachtungen:

- Zeitdruck im Unterricht
- Unsicherheit über Validität in Echtzeit
- Wunsch nach pädagogischer Kontrolle

Fokus auf Einsatzfelder:

- Planung
- Projektkonzeption
- Reflexion
- Dokumentation
- Keine Erwartung an „KI im Unterrichtsgeschehen“

Empfehlung: Förder- und Entwicklungsmaßnahmen sollten genau diese Einsatzfelder adressieren.

3. KI macht strukturelle Schwächen bestehender Materialien sichtbar

Viele der Kritikpunkte in der Arbeit mit KI legen weniger Defizite der Lehrkräfte offen als vielmehr strukturelle Grenzen der eingepflegten Lehrpläne und Materialien:

- veraltete Inhalte (Lehrpläne)
- mangelnde Konkretisierung des „Wie“ (Lehrpläne)
- geringe Maschinenlesbarkeit (Lehrpläne und Referenzlisten)
- fehlende Aktualisierungslogik bei den „alten“ Lehrplänen

Empfehlung: KI kann Impulsgeber für eine zeitgemäße Weiterentwicklung von Lehrplänen sein, ohne deren pädagogischen Kern infrage zu stellen.

- KI wirkt als Diagnoseinstrument für Strukturfragen

- Lehrpläne werden neu gelesen, nicht ersetzt

4. Vertrauen in KI entsteht durch Transparenz, nicht durch Einschränkung

Lehrkräfte akzeptieren KI dort, wo Herkunft, Begrenzung und Unsicherheit von Ergebnissen nachvollziehbar sind.

- Skepsis gegenüber „Halluzinationen“
- Ablehnung intransparenter Empfehlungen
- positives Feedback bei klaren Quellenangaben

Empfehlung: Offenheit und Quellenklarheit sind wirksamer als starke Restriktion.

5. KI verstärkt bestehende professionelle Unterschiede – sie nivelliert sie nicht

Lehrkräfte nutzen KI unterschiedlich: abhängig von Erfahrung, Fach, Haltung und Arbeitsweise. KI gleicht diese Unterschiede nicht aus, sondern macht sie sichtbar.

- Einheitliche Vorgaben wären kontraproduktiv
- Bedarf an differenzierten Angeboten

Empfehlung: Der VdM sollte Orientierungsräume schaffen, keine Standardlösungen.

6. Der Wunsch der Lehrkräfte richtet sich auf Orientierung, nicht auf Vorgaben

Lehrkräfte wünschen sich vom Verband keine Anweisung zur Nutzung von KI, sondern Orientierung, Austausch und professionelle Einordnung. Konkret gewünscht werden:

- Erfahrungsräume
- Fortbildung
- Austauschformate
- konzeptionelle Klarheit

Empfehlung: Solche Formen stärken Akzeptanz und Beteiligung. Vorstandsrolle als moderierender Ermöglicher, nicht als normierende Instanz.

Fazit

Das Pilotprojekt zeigt, dass Musikschullehrkräfte KI weder euphorisch noch ablehnend begegnen. Die Custom GPTs werden von ihnen vor allem als potenzielles Assistenzinstrument in einem komplexen pädagogischen Gefüge verortet. Für den VdM eröffnet sich damit weniger die Aufgabe, KI einzuführen, als vielmehr, Orientierung, Transparenz und strukturelle Anschlussfähigkeit zu schaffen.

Custom GPTs können eine sinnvolle Unterstützung für Musikschullehrkräfte darstellen, indem sie die Entwicklung und Stabilisierung einer reflektierten pädagogischen Haltung begleiten. Sie lassen sich mit vergleichsweise geringem Aufwand (auch individuell) einrichten und eignen sich daher für eine regelmäßige Nutzung im pädagogischen Arbeitsalltag. Neben Bildungsmaßnahmen zu Prompting und einer erweiterten pädagogischen Reflexion sollten auch die Erschließung von aktuellem pädagogischem Material unterstützt werden.

Darüber hinaus zeigt das hohe Interesse unter Musikpädagog:innen, dass KI zunehmend als möglicher Zugang zu kollektiver bzw. „Crowd“-Intelligenz verstanden werden kann, mit der sich Erfahrungswissen vieler Lehrkräfte bündeln und nutzbar machen ließe. In Verbindung mit Austausch- und Vernetzungsstrukturen verschiebt sich damit der Blick von der Frage „Was kann KI im Unterricht leisten?“ hin zu der weiterführenden Perspektive, dass KI vor allem Impulse für die Weiterentwicklung von Lehrplänen, digitalen Infrastrukturen sowie verbandlicher Moderations- und Ermöglichungsrollen geben kann.

Auswertung der Gruppendiskussion (2.10.2025)

Eine Expertendiskussion über das bundesländerübergreifende Pilotprojekt (August bis Oktober 2025): Lehrkräfte verschiedener Instrumentalfächer teilten ihre Nutzungserfahrungen mit verschiedenen Custom ChatGPT-Versionen zu unterrichtspraktischen Themen, mit Lehrplänen als Datenbasis. Die Teilnehmenden erörterten dabei kritisch die Qualität von Literaturempfehlungen sowie das Problem der Halluzinationen, bei denen ChatGPT fiktive Quellen erfindet. Während das Potenzial als Inspirationsquelle für didaktische Fragen betont wurde, erfolgte der Einsatz zunächst primär in der Unterrichtsvor- und Nachbereitung, eine Nutzung im direkten Kontakt mit Schüler:innen wurde zunächst nicht in Erwägung gezogen. Zentral thematisierte die Runde die Notwendigkeit einer besseren digitalen Datenbasis für zukünftige Tests. Die Diskussion unterstreicht, dass die Custom ChatGPTs aktuell eher als unterstützendes Werkzeug und weniger als eigenständiger Problemlöser wahrgenommen wird. Als Präferenz zeichnete sich die Variante Custom GPT mit Internetzugriff aus.

Übergeordnete Fragen: Wie wird die Nutzung von verschiedenen Custom GPTs bewertet? Wie erleben Musikschullehrkräfte Anwendungen von KI im musikschulischen Alltag und welche Chancen, Grenzen und Unterstützungsbedarfe ergeben sich daraus für Musikschulen und den Verband?

Konkret gestellte Diskussionsfragen (als Impulse):

- Hat sich deine Sicht auf ChatGPT in den letzten Monaten verändert? Mit welchen Erwartungen bist du in die Testrunde gegangen – und wie haben sich diese im Verlauf bestätigt oder verändert? Schlüsselerlebnisse?
- In welchen Situationen nutzt du im Alltag ChatGPT? Mit welchem Gerät? Wie häufig? Mit welchem Anliegen?
- Was waren die heraushebenswerten Unterschiede zwischen diesen 3 Varianten?
- Fragen zur Anwendung von ChatGPT im pädagogischen Alltag: Es wurden konkrete Fallbeispiele (aus der Test-Dokumentation) zur Diskussion gestellt. Auch: Was hindert dich daran, ChatGPT direkt im Musik- oder im Instrumentalunterricht einzusetzen?

Nutzungserfahrungen und Einsatzbereiche

In welchen Situationen und für welche Aufgaben nutzen Lehrkräfte KI aktuell – und wofür bewusst nicht?

Die Gruppendiskussion zeigt, dass Lehrkräfte KI derzeit vor allem außerhalb der eigentlichen Unterrichtssituation nutzen, insbesondere in der Vor- und Nachbereitung ihres Unterrichts sowie zur individuellen Reflexion. KI wird eingesetzt, um Anregungen für Stundenplanung, Repertoire-Empfehlung und methodische Alternativen zu erhalten. Als hilfreich wird KI dort erlebt, wo sie Inspiration bietet, Denkprozesse anstößt oder Übersicht schafft, ohne unmittelbaren Entscheidungsdruck zu erzeugen. Bewusst nicht genutzt wird KI hingegen im direkten Instrumentalunterricht, insbesondere in Anwesenheit der Schüler:innen (möglicherweise auch, da Testzeitraum teilweise in den Ferien). Gründe hierfür sind knappe Unterrichtszeit, Zweifel an der inhaltlichen Verlässlichkeit sowie der Anspruch, pädagogische Entscheidungen selbst zu verantworten und situativ zu treffen. Bei konkreten Literaturentscheidungen oder sensiblen pädagogischen Situationen (z. B. Leistungsbewertung, individuelle Förderung) werden keine guten Erfahrungen geschildert. Insgesamt zeigt sich ein klar abgegrenztes Nutzungsverständnis: KI gilt als unterstützendes Werkzeug zur Reflexion und Vorbereitung, nicht als operative Instanz im Unterrichtsgeschehen.

- Typisch ist die Beschreibung von KI als „parallele Denkfolie“, gegen die eigene Ideen gespiegelt werden
- KI dient als Anstoß, nicht als Autorität
- Mehrere Lehrkräfte nutzen KI gezielt zur Reflexion schwieriger pädagogischer Situationen, z. B.: Motivationsprobleme, Schüchternheit oder Verweigerung, Eltern-Schüler:in-Lehrer:in-Konstellationen
- Eine einzelne Lehrkraft berichtet von der Nutzung von KI zur Zusammenfassung umfangreicher Unterrichtsnotizen (anonymisiert), Rückschau auf Lernverläufe

Wahrgenommene Chancen für die musikpädagogische Praxis

Welche Potenziale sehen Lehrkräfte in KI-Anwendungen für ihre pädagogische Arbeit?

Im Rahmen der Custom-ChatGPT-Testung lag der Fokus der Betrachtung vor allem auf dem Potenzial, ihre musikpädagogische Arbeit anzuregen, zu erweitern und zu entlasten, ohne dabei ihre professionelle Verantwortung abzugeben: die Nutzung von KI wird als Inspirations- und Reflexionsinstrument beschrieben. Einige Lehrkräfte betonen zudem das Potenzial, über KI alternative methodische Zugänge oder kreative Lösungsansätze kennenzulernen, die sie mit ihrer eigenen Erfahrung weiterentwickeln können.

Insgesamt wird ChatGPT weniger als fachlicher Experte verstanden, sondern als Werkzeug, das individuelle Professionalität stärkt, indem es Denkräume öffnet, Reflexion ermöglicht oder pädagogische Entscheidungsprozesse vorbereitet.

Wahrgenommene Grenzen, Risiken und Irritationen

Wo erleben Lehrkräfte KI als problematisch, unzuverlässig oder nicht anschlussfähig an ihre Praxis?

Lehrkräfte erleben KI insbesondere dort als problematisch, wo sie Unzuverlässigkeit, Intransparenz oder zusätzlichen Prüfaufwand erzeugt. Als zentrales Risiko werden inhaltliche Fehler und Halluzinationen genannt, insbesondere erfundene oder falsch zugeordnete Literaturangaben, unpassende Schwierigkeitsgrade oder scheinbar plausible, aber fachlich nicht sinnvolle Empfehlungen. Irritationen entstehen zudem durch die geringe Tiefe vieler Antworten, die häufig bei allgemeinen Auflistungen stehen bleiben und komplexe pädagogische Fragestellungen nur oberflächlich behandeln. Bei sensiblen Entscheidungen – etwa individueller Förderung, Leistungsbewertung oder konkreter Literaturauswahl – wird KI bewusst begrenzt oder ausgeschlossen.

Alles in allem wird KI weniger als gefährlich, sondern als noch unreif und nicht verlässlich genug wahrgenommen, um zentrale pädagogische Entscheidungen zu tragen; ihre Nutzung bleibt daher klar auf unterstützende, reflexive Funktionen beschränkt.

→ Was waren die heraushebenswerten Unterschiede zwischen diesen 3 Varianten?

Wie beurteilt ihr insgesamt die Qualität der Antworten der drei KIs? Wie unterscheiden sich die drei KIs?

Im Rahmen des Pilotprojekts wurden drei verschiedene KI-Varianten getestet: ein auf Lehrpläne und Dokumente (VdM/KOMU) beschränktes Custom GPT, ein Custom GPT mit zusätzlichem Internetzugriff (als „Plus“-Variante bezeichnet) sowie das standardmäßige ChatGPT. Die Tester:innen hielten dabei folgende heraushebenswerte Unterschiede fest:

1. Das auf Lehrpläne beschränkte Custom GPT

Probleme bei der Datenextraktion: Teilnehmer:innen berichteten, dass diese Variante Schwierigkeiten hatte, Informationen zuverlässig aus den bereitgestellten PDFs (oft Scans) zu ziehen. Eine Lehrkraft stellte fest, dass die KI teilweise explizit darauf hingewiesen werden musste, in die Daten zu schauen, und dennoch oft halluzinierte, indem sie nicht existierende Autor:innen-Buch-Kombinationen erfand.

Eingeschränkte Antwortbasis: Eine andere Lehrkraft bemerkte, dass diese Version oft angab, keine Informationen im Lehrplan zu finden, oder lediglich Teile daraus rezitierte, was die kreative Nutzbarkeit einschränkte.

Fehlende Belege: Kritisiert wurde auch, dass diese Version konsequent keine Quellenangaben machte.

Repetitive Vorschläge: Es fiel auch auf, dass für den Klavierbereich fast unabhängig von der Fragestellung immer die gleichen Werke (z. B. Czerny) vorgeschlagen wurden.

2. Das Custom GPT mit Internetzugriff („Plus“-Variante)

Höhere Kohärenz und Nachvollziehbarkeit: Diese Variante wurde von verschiedenen Lehrkräften als die schlüssigste empfunden, da sie einige richtige Quellennachweise lieferte und teilweise sogar Seitenzahlen zum Nachschlagen in den Lehrplänen nannte.

Größere Informationstiefe: Durch den Internetzugriff konnte die KI Lücken in den VdM-Lehrplänen (z. B. im Bereich E-Gitarre oder bei spezifischen pädagogischen Spannungsfeldern) besser ausgleichen.

Nützlichkeit für die Inspiration: Eine Lehrkraft hob hervor, dass diese Variante sich auch gut dafür eignen würde, neue Impulse zu gewinnen. Zugleich seien die Vorschläge nachvollziehbarer gewesen als bei den anderen Versionen.

3. Das normale (unveränderte) ChatGPT

Kreativität vs. Oberflächlichkeit: Während eine Lehrkraft die Standard-KI als kreativer bei pädagogischen Fragestellungen empfand, bezeichnete eine andere Lehrkraft die Ergebnisse als eher oberflächlich.

Qualität der Quellen: Eine Lehrkraft beobachtete, dass die allgemeine KI zwar valide wirkende Quellen (wie Verlage) nannte, diese aber nur auf Basis von Internet-Rankings auswählte, ohne eine musikpädagogische Einordnung vorzunehmen.

Suchmaschinen-Ersatz: In der allgemeinen Nutzung wurde diese Variante oft als effizienter als Google empfunden, um Informationen aus verschiedenen Webseiten zusammenzufassen.

Zusammenfassend lässt sich sagen, dass die „Plus“-Variante mit Internetzugriff den besten Kompromiss aus fachspezifischer Ausrichtung und kreativer Unterstützung bot, während die rein dokumentenbasierte Custom GPT stark unter der Qualität der zugrunde liegenden PDF-Daten litt und das reine ChatGPT zu allgemein antwortete.

Wie müssten die Custom GPTs entwickelt werden, um nützlicher zu werden?

Zentral erscheint eine deutlich verbesserte Datenbasis: Lehrpläne, Literaturlisten und Materialien sollten aktuell, vollständig und vor allem maschinenlesbar vorliegen (nicht als gescannte PDFs), damit die KI tatsächlich aus diesen Quellen arbeiten kann. Ebenso wichtig erscheint Quellentransparenz: Lehrkräfte erwarten klare Hinweise darauf, worauf sich Aussagen stützen, idealerweise mit Seitenangaben oder eindeutiger Herkunft, um Ergebnisse schnell einordnen zu können. Darüber hinaus wünschen sie sich Custom GPTs, die weniger generisch und weniger normativ antworten, sondern stärker auf konkrete pädagogische Kontexte reagieren (z. B. Alter, Lernstand, Unterrichtsform), ohne dabei vorzugeben, „die richtige Lösung“ zu liefern.

Lehrpläne, Materialien und Datenstrukturen

Wie beeinflusst der Einsatz von KI den Blick auf bestehende Lehrpläne und Materialien?

In den Diskussionsbeiträgen verdeutlicht sich ein Bedarf nach mehr Material in Form von aktuellen Lehrplänen oder auch anderen musizierpädagogischen Texten. Im Umgang mit KI wird deutlich, dass viele Lehrpläne als inhaltlich überholt, wenig konkret im Hinblick auf das „Wie“ des Unterrichtens und nur eingeschränkt anschlussfähig an aktuelle pädagogische Fragestellungen wahrgenommen werden. Statt mühsam in PDFs zu blättern, nutzen Lehrkräfte die KI, um gezielt nach Inhalten zu suchen.

Auch wird deutlich, dass sich im Rahmen des Projekts und die Auseinandersetzung mit den Custom GPTs ein „neues Interesse“ für die Lehrpläne entwickelt. Dies kann vorsichtig so interpretiert werden, dass ChatGPT die pädagogische Reflexion von Lehrkräften fördern kann. Interessanterweise kann die KI die Diskrepanz zwischen Lehrplanvorgaben (Idealnorm) und dem realen Unterrichtsgeschehen (Unterrichtsrealität) aufzeigen.

Lehrkräfte vergleichen KI-Antworten zunehmend mit Lehrplaninhalten und stellen fest, dass KI teils mehr Anregungen und Kontexte liefert als die Dokumente selbst, insbesondere bei methodischen oder sozialen Fragestellungen, die im Lehrplan kaum ausgeführt sind. Dadurch fungiert KI weniger als Ersatz, sondern als Spiegel struktureller Grenzen der bestehenden Materialien und deutet auf den Wunsch nach pädagogischen Materialien oder Angeboten, die nicht nur normativ festlegen, was gelernt werden soll, sondern auch Hinweise darauf geben, wie pädagogische Praxis gestaltet werden kann.

Rolle von Erfahrung, Urteilskraft und Verantwortung

Wie verorten Lehrkräfte ihre professionelle Verantwortung im Umgang mit KI?

Lehrkräfte in der Musikschulpädagogik verorten ihre professionelle Verantwortung im Umgang mit KI vor allem in der Rolle einer kritischen Kontrollinstanz. Sie sehen sich nicht als passive Nutzer:innen, sondern als Expert:innen, die die Ergebnisse der KI filtern, bewerten und an die individuellen Bedürfnisse ihrer Schüler:innen anpassen müssen. Lehrkräfte sehen es als Teil ihrer Professionalität an, eine skeptische Grundhaltung gegenüber der Technologie einzunehmen.

Lehrkräfte verorten ihre professionelle Verantwortung im Umgang mit KI eindeutig bei sich selbst und nicht bei der Technologie. KI wird durchgehend als unterstützendes Werkzeug verstanden, dessen Vorschläge nur dann Bedeutung erlangen, wenn sie durch eigene Erfahrung, fachliches Wissen und pädagogische Urteilskraft geprüft, eingeordnet und gegebenenfalls verworfen werden. Die Teilnehmenden heben hervor, dass pädagogische

Entscheidungen nicht delegierbar sind und stets kontextabhängig getroffen werden müssen. Gerade die wahrgenommenen Unzuverlässigkeiten von KI, etwa fehlerhafte oder zu allgemeine Antworten, verstärken dieses Verantwortungsbewusstsein und führen zu einer klaren Grenzziehung gegenüber automatisierten Empfehlungen: Die Lehrkräfte sind sich einig, dass die KI zwar Inspiration liefern kann, die didaktische Entscheidung aber bei der Lehrperson verbleiben muss.

Unterstützungsbedarfe und nächste Schritte

Welche Rahmenbedingungen wünschen sich Lehrkräfte für einen reflektierten Einsatz von KI im Musikschulbereich?

Lehrkräfte wünschen sich für einen reflektierten Einsatz von KI im Musikschulbereich vor allem klare Orientierungs- und Unterstützungsstrukturen, nicht jedoch verbindliche Vorgaben oder standardisierte Nutzungsszenarien. Als größte Hürde wird lange die derzeitige Qualität der bereitgestellten Lehrpläne diskutiert. Da einige davon veraltet sind (z. B. Violine von 1992), wird die KI als Chance gesehen, diese Lücken durch aktuelle Impulse zu füllen. Der VdM plant bereits, künftig auf Printausgaben zu verzichten und die KI in die digitale Veröffentlichung und Weiterentwicklung von Lehrplänen einzubinden.

Darüber hinaus äußern Lehrkräfte den Wunsch nach Fortbildungs- und Austauschformaten, in denen Erfahrungen mit KI gemeinsam reflektiert, Grenzen diskutiert und gelingende Einsatzformen sichtbar gemacht werden. Wichtig ist ihnen dabei eine Haltung des Verbandes, die Ermöglichung und Orientierung in den Vordergrund stellt, statt normative Erwartungen zu formulieren.

Auch rechtliche und ethische Klarheit – etwa im Hinblick auf Datenschutz, Urheberrecht und Verantwortlichkeiten – wird als Voraussetzung genannt, um KI mit größerer Sicherheit einsetzen zu können. Ein interessanter Ansatz ist die Bereitstellung von KI auf eigenen Servern (z. B. durch „Smart Musikschule“), um unabhängig von großen „Hyperscalern“ (=Technologieunternehmen wie OpenAI und Google) agieren zu können und die Datenhoheit zu behalten.

Zusammenfassend wird deutlich, dass Lehrkräfte den VdM vor allem in einer moderierenden, strukturierenden und qualitätssichernden Rolle sehen, die professionelle Autonomie respektiert und zugleich Räume für eine verantwortungsvolle Weiterentwicklung des KI-Einsatzes eröffnet.

Einblicke in die Custom-ChatGPT-Dokumentationen

Im Rahmen des bundeslandübergreifenden Pilotprojekts erprobten insgesamt acht Musikschullehrkräfte aus Schleswig-Holstein, Nordrhein-Westfalen und Baden-Württemberg im Zeitraum von Anfang August bis Anfang Oktober unterschiedliche Custom GPTs für den Instrumentalunterricht in den Fächern Gitarre, Klavier und Violine. Frage war: Wie unterscheiden sich die drei Custom GPTs?

Zu Beginn des Projekts erhielten die teilnehmenden Lehrkräfte („Tester:innen“) eine Einführung per Zoom, in der sowohl das Projektziel als auch das Konzept der Custom GPTs erläutert wurden. Ergänzend wurde ihnen eine strukturierte Anleitung zum Testprozess und zur Dokumentation ihrer Erfahrungen zur Verfügung gestellt, die den systematischen Vergleich der GPT-Varianten sowie eine reflektierte Rückmeldung unterstützen sollte.

Insgesamt wurden im Formular 113 Eintragungen im Zeitraum vom Anfang August bis Anfang Oktober erfasst. Das Formular umfasste insgesamt 13 Eingabefelder zur Dokumentation und Bewertung. Im Folgenden werden die Daten weniger inhaltlich zusammengefasst, sondern in geclusterter Form dargestellt, um einen Überblick zu den Dimensionen zu vermitteln, die bei der Testung bedeutsam wurden.

1. Clusterung der ChatGPT-Anfragen (Einstiegsfrage)

Qualitative Clusterung mit kurzer Charakterisierung und typischen Fragebeispielen (paraphrasiert, nicht wörtlich zitiert)

Cluster 1: Literatur- und Repertoirefragen (größter Cluster)

Charakteristik

Fragen zur Auswahl, Einordnung und Progression von Spiel- und Unterrichtsliteratur, häufig mit Bezug auf Alter, Leistungsstand oder Wettbewerbe.

Typische Fragen

- Welche Stücke eignen sich für Schüler:innen mit bestimmtem Lernstand?
- Welche Literatur passt nach einem bestimmten Werk?
- Welche Stücke sind für Jugend musiziert (U-Stufen) geeignet?
- Welche Lehrwerke werden im Lehrplan empfohlen?

Bedeutung

- Hoher Orientierungsbedarf
- Hohe Erwartungen an Korrektheit → zugleich hohe Frustration bei Halluzinationen

Cluster 2: Methodisch-didaktische Fragen („Wie unterrichten?“)

Charakteristik

Fragen nach konkreten Vorgehensweisen im Unterricht, oft jenseits dessen, was Lehrpläne explizit ausformulieren (Interpretation im Anschluss an Gruppendiskussion).

Typische Fragen

- Wie kann ich Technik X vermitteln?
- Wie gestalte ich den Übergang von Einzel- zu Gruppenunterricht?
- Wie kann ich Übestrategien altersgerecht vermitteln?
- Wie kann ich musikalische Gestaltung fördern?

Bedeutung

- Zentrales Praxisinteresse
- Macht sichtbar, dass Lehrpläne hier oft als zu abstrakt erlebt werden

Cluster 3: Pädagogische und psychosoziale Situationen

Charakteristik

Fragen zu Motivation, Verhalten, Schüchternheit, Verweigerung oder besonderen Lernbedingungen.

Typische Fragen

- Wie gehe ich mit sehr zurückhaltenden Schüler:innen um?
- Was tun, wenn Motivation fehlt?
- Wie kann ich mit pubertären Blockaden umgehen?
- Wie reagiere ich auf Störungen oder Überforderung?

Bedeutung

- KI wird hier als Reflexionspartner genutzt
- Hohe Sensibilität, klare Grenzziehung gegenüber automatisierten Lösungen

Cluster 4: Planung, Strukturierung und Organisation von Unterricht

Charakteristik

Fragen zur mittel- und längerfristigen Planung von Unterricht, Projekten oder Reihen.

Typische Fragen

- Wie kann eine Unterrichtsreihe strukturiert werden?
- Wie plane ich ein mehrwöchiges Band- oder Ensembleprojekt?

- Welche Schritte sind sinnvoll zur Vorbereitung eines Vorspiels?

Bedeutung

- KI als Ordnungs- und Strukturhilfe

Cluster 5: Reflexion, Dokumentation und Analyse der eigenen Praxis

Charakteristik

Fragen, die auf die eigene professionelle Reflexion zielen, oft retrospektiv.

Typische Fragen

- Kannst du meine Unterrichtsnotizen zusammenfassen?
- Welche Muster erkennst du im Lernverlauf?
- Welche nächsten Schritte könnten sinnvoll sein?

Bedeutung

- Stark reflektive Nutzung

Cluster 6: Lehrplan- und Strukturfragen

Charakteristik

Fragen, die Lehrpläne selbst zum Gegenstand machen – weniger als normatives Dokument, mehr als Ressource.

Typische Fragen

- Was sagt der Lehrplan zu Thema X?
- Wo finde ich im Lehrplan Hinweise zu Methode Y?
- Welche Progression ist vorgesehen?

Bedeutung

- KI fungiert als „Index“ zu Lehrplänen
- Intensive Auseinandersetzung mit bestehenden Materialien

Cluster 7: Vergleichs- und Kontrollfragen (KI-kritisch, Custom-GPTs)

Charakteristik

Bewusst testende Fragen zur Überprüfung von Grenzen, Halluzinationen und Regelkonformität.

Typische Fragen

- Nenne Literatur, die nicht im Lehrplan steht.
- Beantworte Fragen außerhalb der Instrumentalpädagogik.
- Nenne Quellen explizit.

Bedeutung

- Hohe professionelle Skepsis
- zur Bewertung von Custom GPTs

2. Clusterung der Zusammenfassenden Bewertung zum Chatverlauf

Die Aussagen lassen sich weniger entlang von „gut/schlecht“, sondern entlang von Haltungen, Erwartungen und professionellen Grenzziehungen clustern. (qualitative Clusterung; paraphrasiert, nicht wörtlich zitiert)

Cluster 1: KI als Inspirations- und Reflexionshilfe (sehr häufig, überwiegend positiv)

Charakteristik

Lehrkräfte bewerten KI als hilfreich, wenn sie Denkprozesse anstößt, neue Perspektiven eröffnet oder Routinen irritiert, ohne Entscheidungen vorzugeben.

Typische Bewertungen

- „Gibt neue Ideen, auf die ich selbst nicht gekommen wäre“
- „Hilfreich zum Nachdenken, nicht zum Übernehmen“
- „Gut als Sparringspartner“

Kernhaltung

KI unterstützt Reflexion, nicht Handeln.

Cluster 2: Begrenzte Verlässlichkeit bei fachlich-konkreten Fragen (häufig, kritisch-sachlich)

Charakteristik

Bewertungen, die die mangelnde Zuverlässigkeit bei konkreten fachlichen Anforderungen (v. a. Literatur) thematisieren.

Typische Bewertungen

- falsche oder erfundene Literatur
- unpassende Schwierigkeitsgrade

- fehlende Differenzierung
- „Am Ende genauso viel Zeit fürs Kontrollieren“
- „Nur sinnvoll, wenn die Antwort stimmt“

Kernhaltung

Hoher Prüfaufwand relativiert den Nutzen.

Zeitgewinn ist kontextabhängig, nicht garantiert.

Cluster 3: Qualität der Datengrundlage als Schlüsselbedingung (häufig, strukturell argumentierend)

Charakteristik

Bewertungen, die weniger die KI selbst, sondern die zugrunde liegenden Daten in den Fokus rücken.

Typische Bewertungen

- „Mit besseren Daten wäre das hilfreich“
- „Lehrpläne sind nicht gut aufbereitet“
- „Scans funktionieren nicht“

Kernhaltung

Nützlichkeit hängt von Materialqualität, nicht vom Tool allein ab.

Cluster 4: Unterschiedliche Bewertung der getesteten KI-Varianten (vergleichend)

Charakteristik

Explizite Gegenüberstellung der drei getesteten Varianten.

Typische Bewertungen

- Lehrplan-only: zu eng, wenig hilfreich
- Plus-Version: ausgewogen, nachvollziehbar
- Allgemeine KI: kreativ, aber unscharf

Cluster 5: KI als Anlass zur Auseinandersetzung mit Lehrplänen (indirekt, aber wiederkehrend)

Charakteristik

Bewertungen, die beschreiben, dass die KI-Nutzung den Blick auf Lehrpläne verändert hat.

Typische Bewertungen

- „Ich habe mich wieder mit Lehrplänen beschäftigt“
- „Als Such- oder Indexfunktion hilfreich“
- „Zeigt, was fehlt“

Kernhaltung

KI wirkt als Spiegel und Katalysator, nicht als Ersatz.

3. Sind die Antworten für die Unterrichtspraxis hilfreich/nutzbar?

8. Ist die Quellenlage nachvollziehbar? (z. B. nennt die KI einen Abschnitt oder zitiert sinngemäß?)

[Weitere Informationen](#)



Clusterung:

- Falsche / erfundene Quellenangaben
 - Behauptung, ein Werk stünde im VdM-Lehrplan, obwohl es nicht existiert.
- Fehlende oder unpräzise Quellen
 - Quellen werden gar nicht genannt oder nur vage.
 - Beispiel: „Es wird vom Lehrplan gesprochen, aber nicht genau angegeben, wo.“
 - Beispiel: „Es wird nur grob angegeben, dass es aus den VdM-Empfehlungen kommt.“
- Quellen vorhanden, aber ohne Detailtiefe
 - Quellenangaben sichtbar, aber ohne konkrete Stellen/Seiten.

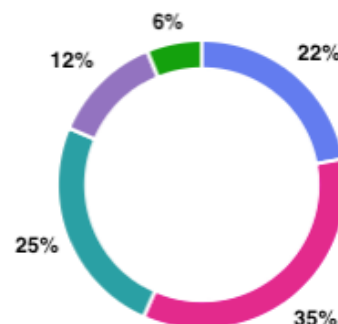
- Beispiel: „Es ist sichtbar, dass die Infos aus dem Lehrplan kommen, allerdings ohne genaue Seitenangaben.“
- Beispiel: „Es ist ersichtlich, welche Schulen genannt werden, aber nicht, woher genau die Zuschreibungen stammen.“
- Korrekte und detaillierte Quellenangaben
 - Lehrpläne zitiert, teils mit Seitenzahlen oder Verlinkungen.
 - Beispiel: „Bei den Inhalten werden die Quellen genau benannt.“
- Sonstige Beobachtungen / Abweichungen
 - Fehlinterpretationen oder terminologische Missverständnisse.
 - Beispiel: „Der Bär brummt‘ wird als Bild verstanden, nicht als Stücktitel.“

4. Sind die Antworten für die Unterrichtspraxis hilfreich/nutzbar?

10. Sind die Antworten für die Unterrichtspraxis hilfreich/nutzbar?

[Weitere Informationen](#)

sehr gut	25
gut	39
teilweise	28
wenig	14
gar nicht	7



Cluster 1: Hilfreich als Impuls- und Ideengeber (sehr häufig genannt)

Charakteristik

Antworten werden als nutzbar bewertet, wenn sie neue Ideen, Perspektiven oder Denkansätze liefern, die Lehrkräfte eigenständig weiterentwickeln können.

Typische Erläuterungen

- „Gibt Denkanstöße, auf die ich selbst nicht gekommen wäre“
- „Gut als Inspiration, nicht zur direkten Übernahme“
- „Regt zum Weiterdenken an“

Nutzungslogik

Indirekte Nutzbarkeit durch Anregung, nicht durch konkrete Umsetzung.

Cluster 2: Eingeschränkt nutzbar wegen fehlender Praxisnähe (häufig)

Charakteristik

Antworten werden als grundsätzlich interessant, aber zu allgemein oder zu wenig kontextbezogen für die konkrete Unterrichtssituation beschrieben.

Typische Erläuterungen

- „Zu allgemein für meine konkrete Unterrichtssituation“
- „Bleibt an der Oberfläche“
- „Keine Anpassung an Alter, Niveau oder Unterrichtsform“

Nutzungslogik

Potenzial vorhanden, aber Übersetzungsarbeit durch die Lehrkraft nötig.

Cluster 3: Nicht nutzbar aufgrund fachlicher Fehler oder Ungenauigkeiten (deutlich benannt, v. a. bei Literaturfragen)

Charakteristik

Antworten werden als nicht nutzbar eingeschätzt, wenn sie inhaltlich falsch, ungenau oder nicht überprüfbar sind.

Typische Erläuterungen

- falsche oder erfundene Literatur
- unpassende Schwierigkeitsgrade
- nicht existente Werke oder Autor:innen

Nutzungslogik

Hoher Prüfaufwand → keine direkte Praxistauglichkeit.

Cluster 4: Nutzbar zur Vor- und Nachbereitung, nicht im Unterricht selbst

Charakteristik

Antworten werden als hilfreich bewertet, sofern sie außerhalb der Unterrichtssituation genutzt werden.

Typische Erläuterungen

- „Gut für Planung und Reflexion“

- „Nicht geeignet für den Unterricht in Echtzeit“
- „Zu zeitaufwändig während der Stunde“

Nutzungslogik

Asynchrone Nutzbarkeit → kein Einsatz im direkten Unterrichtsgeschehen.

Cluster 5: Nutzbar als Strukturierungs- und Übersichtshilfe („mittelhäufig“)

Charakteristik

Antworten werden positiv bewertet, wenn sie helfen, komplexe Informationen zu ordnen oder einen Überblick zu gewinnen.

Typische Erläuterungen

- „Hilfreich, um Gedanken zu sortieren“
- „Gibt eine grobe Struktur“
- „Als Einstieg oder Überblick geeignet“

Nutzungslogik

KI als Ordnungsinstrument, nicht als Detailberater.

Cluster 6: Unterschiedliche Nutzbarkeit je nach KI-Variante (vergleichend)

Charakteristik

Die Nutzbarkeit wird explizit an die getestete KI-Version gekoppelt.

Typische Erläuterungen

- Lehrplan-only: „kaum nutzbar, zu eng“
- Plus-Version: „am ehesten praxisnah“
- Allgemeine KI: „kreativ, aber fachlich unscharf“

5. Das ist mir bei diesem Chatverlauf außerdem aufgefallen...

- 1. Halluziniertes / erfundenes Zusatzmaterial
 - „Viel anscheinend halluziniertes Beiwerk wie ‚Zielgruppe‘ und ‚Schwierigkeitsgrad‘.“
 - „Bei kurzer Nachfrage nach mehr Titeln hat er nicht mehr im Internet gesucht und wieder halluziniert.“
- 2. Quellen- und Dokumentennutzung

- „Höchstens, dass gar keine Quellen außerhalb des Lehrplans benutzt oder zumindest genannt wurden.“
- „Es wurden öfters Quellen genannt als bei den Custom ChatGPTs.“
- „Ich bin mir relativ sicher, dass er nicht die beigelegten Dokumente durchsucht hat, sondern frei verfügbare VdM-Listen im Internet.“
- 3. Dialogverhalten & Hilfsbereitschaft
 - „Ich war überrascht über die – fast immer – folgenden Angebote von ChatGPT, noch etwas zu ‚tun für mich‘.“
 - „Dass bei dieser Frage auch dieser Chat mich eigeninitiativ gefragt hat, ob er mehr in die Tiefe gehen soll.“
 - „ChatGPT ist sehr, sehr hilfswillig.“
- 4. Inhaltliche Grenzen / fehlende Methodik
 - „Es bleibt auf der informativen Ebene, es gibt keine konkreten methodischen Tipps.“
 - „Die Auswahl der Stücke habe ich der KI überlassen. Mich hätte die Antwort auf die Frage, wie seine Auswahl zustande gekommen ist, interessiert.“
 - „Dass bei näherer Nachfrage die Informationen teilweise redundant werden.“
- 5. Unterschiede zwischen Versionen
 - „Ich habe den (fast) selben Prompt mit ChatGPT ohne Lehrplan ausprobiert, die Antwort war doch ziemlich unterschiedlich.“
 - „Die KI Lehrplan VdM Plus geht mehr auf mich ein, als die KI Lehrplan VdM. Die andere hat ein bisschen den Lehrplan rezitiert, diese hat auch eine Reflexion geliefert.“
- 6. Selbstkorrektur & Lernbereitschaft
 - „Die KI hat zwei Anläufe gebraucht, um einzusehen, dass ihre Stücke zu schwer für meine Schülerin sind.“
 - „Die schnelle ‚Ehrlichkeit‘, dass keine Info zur 2. Frage da ist.“
- 7. Bewertung & Meta-Kommentare der KI
 - „Dass die KI eine Kategorie wie ‚empfehlenswert‘ auswählt ähnlich einer Bewertung.“

Das Projekt wurde gefördert durch das Bundesprogramm Ländliche Entwicklung und Regionale Wertschöpfung (BULEplus).

Gefördert durch:



**aufgrund eines Beschlusses
des Deutschen Bundestages**