



# Mit Computerintelligenz Lehr-Lern-Prozesse gestalten: Anwendung von *ChatGPT* zur Unterrichtsplanung

Paul P. Martin<sup>1</sup>, Nicole Graulich<sup>1,\*</sup>

<sup>1</sup> Justus-Liebig-Universität, Institut für Didaktik der Chemie, Heinrich-Buff-Ring 17, 35392 Gießen

\*Email Korrespondenzautor:in: [Paul.Martin@didaktik.chemie.uni-giessen.de](mailto:Paul.Martin@didaktik.chemie.uni-giessen.de)

**Abstract:** Aufgrund der allgemein wachsenden Popularität künstlicher Intelligenzen in Wissenschaft und Gesellschaft sowie der zunehmenden Vielfalt an technologischen Optionen gewinnt Computerintelligenz im Chemieunterricht eine immer größere Bedeutung. Neben der Diagnose von Schüler- bzw. Studierendenvorstellungen und der damit einhergehenden Möglichkeit der automatisierten Bewertung von offenen Textantworten können Methoden der Computerintelligenz auch zur Unterrichtsplanung eingesetzt werden. Die Anwendung all dieser Methoden wurde durch die Veröffentlichung von ChatGPT, einem textgenerierenden Sprachmodell, deutlich erleichtert, sodass nun auch Laien im Bereich der künstlichen Intelligenz mit gezielten Anfragen menschenähnliche Antworten aus komplexen Algorithmen gewinnen können. Ziel dieses Konferenzband-Beitrages ist es, die zugrundeliegende Methodik von ChatGPT einzuführen, Anwendungsmöglichkeiten für die Planung von Chemieunterricht zu veranschaulichen und diese zu reflektieren. Insbesondere wird anhand eines Praxisbeispiels aufgezeigt, wie Lehrende bei der Anwendung von ChatGPT zur Unterrichtsplanung vorgehen können, welche didaktischen Potentiale ChatGPT bietet und wie die Güte algorithmischer Antworten zu beurteilen ist. Die Einblicke in die Anwendung von ChatGPT können langfristig zu einer gewinnbringenden Integration von Methoden der künstlichen Intelligenz in Lehr-Lern-Prozesse führen.

**Keywords:** Künstliche Intelligenz, Computerlinguistik, ChatGPT, computergestütztes Lehren, Unterrichtsplanung

## 1. Einleitung

Methoden der künstlichen Intelligenz, genauer gesagt des maschinellen Lernens, haben in den vergangenen fünfzehn Jahren Einzug in die chemiedidaktische Forschung gehalten [1]. Forschende haben maschinelles Lernen dabei überwiegend genutzt, um Schüler- bzw. Studierendenvorstellungen detailliert zu diagnostizieren sowie um formative, also unbenotete Lernstanderhebungen automatisiert auszuwerten [1-3]. Zudem wurden erste Lehrkonzepte entwickelt, die die Anwendung von maschinellem Lernen in die Hochschullehre des Faches Chemie tragen [4-6]. Die Vielzahl an Anwendungsmöglichkeiten zeigt, dass maschinelles Lernen das Potential hat, Lehre in den naturwissenschaftlichen Unterrichtsfächern zu transformieren [2,7].

Bis zuletzt war die Nutzung all jener Technologien größtenteils auf Expertinnen und Experten beschränkt, die über umfangreiche Kenntnisse im Fach Informatik verfügen mussten. Die Entwicklung von ChatGPT (Chat Generative Pretrained Transformer) – eine Technologie aus dem Bereich der künstlichen Intelligenz, die auf Benutzerfragen menschenähnliche Konversationen eingeht – hat in den vergangenen Monaten allerdings erheblich zur Demokratisierung künstlicher Intelligenz beigetragen. Laien in diesem Bereich können nun von den vielschichtigen Potentialen künstlicher Intelligenz profitieren, was Lehrenden die Planung von Lehr-Lern-Prozessen erleichtern kann. Mit gezielten Anfragen gewinnen Lehrende beispielsweise einen Diskussionspartner bei der Planung von Unterricht und der Abstimmung einzelner Unterrichtsphasen auf die individuellen Bedürfnisse der Lernenden. In diesem Artikel werden anhand eines Praxisbeispiels erste Ansätze der Integration von ChatGPT in den Planungsprozess von Unterricht aufgezeigt und reflektiert. Dies ermöglicht zukünftig die Vernetzung der Vorteile menschlicher Planungskompetenz und maschinellen Outputs.

## 2. Grundlagen zur Funktionsweise von künstlicher Intelligenz

Spätestens die Einführung von ChatGPT durch OpenAI hat gezeigt, dass Methoden der künstlichen Intelligenz Einzug in die Gesellschaft gehalten haben. Algorithmen der künstlichen Intelligenz übernehmen Tätigkeiten, die wir bislang für spezifisch menschlich gehalten haben. Aus diesem Grund ist im Zusammenhang mit künstlicher Intelligenz häufig auch von der Automatisierung intelligenten Verhaltens die Rede, weshalb dieser Begriff im weitesten Sinne eine Software beschreibt, die kognitive Tätigkeiten ausführt [8].

Diese allgemeine Definition künstlich-intelligenten Verhaltens beinhaltet jedoch zwei Unklarheiten: Erstens verlagert es die Definition von künstlicher Intelligenz auf die Frage, was kognitive Tätigkeiten genau sind. Zweitens verändert sich die Definition, wenn das Ziel erreicht wurde. Sobald ein Computer das gewünschte Verhalten automatisiert ausführen kann, wird diese Tätigkeit als weniger intelligent wahrgenommen, gerade weil sie ein Computer ausführen kann [8]. Aus diesem Grund wird bei künstlichen Intelligenzen zwischen unterschiedlichen Teilbereichen unterschieden, die im Folgenden kurz vorgestellt werden.

### 2.1 Künstliche Intelligenz, maschinelles Lernen und tiefgreifendes maschinelles Lernen

Künstliche Intelligenz umfasst die Anwendung von verschiedenen Techniken wie die softwarebasierte Planung, Problemlösung und Entscheidungsfindung in häufig komplexen Prozessen. Im Gegensatz zum maschinellen Lernen, bei dem Algorithmen auf vorhandenen Daten lernen, beinhaltet künstliche Intelligenz auch all jene Systeme, die von Menschen vorgegebene Regeln ausführen. Der Begriff der künstlichen Intelligenz beschäftigt sich somit allgemein mit softwarebasierten Programmen, die erfassen, schlussfolgern, handeln und sich anpassen können.



Maschinelles Lernen ist hingegen ein Teilbereich der künstlichen Intelligenz, der sich mit der Fähigkeit von Computersystemen beschäftigt, menschenähnliche Leistungen durch das Lernen aus Erfahrung zu erbringen. Algorithmen des maschinellen Lernens führen also aus Erfahrungen heraus menschenähnliche Leistungen aus, ohne dafür explizit programmiert werden zu müssen. Bei den Erfahrungen, aus denen ein solcher Algorithmus lernt, handelt es sich um Zahlen, die in einem Datensatz gespeichert sind. Mithilfe dieser kodierten Datenmengen versuchen Algorithmen des maschinellen Lernens, Muster und Trends zu erkennen und Vorhersagen zu treffen.

Tiefgreifendes maschinelles Lernen ist wiederum eine spezifische Technik, die sich auf das Training von künstlichen neuronalen Netzen konzentriert, die mehrere Schichten von verknüpften künstlichen Neuronen enthalten – inspiriert vom menschlichen Nervensystem. Tiefgreifendes maschinelles Lernen hat in den letzten Jahren enorm an Popularität gewonnen, da es sehr erfolgreich bei der Lösung komplexer Probleme in Bereichen wie der Sprachverarbeitung und Bilderkennung eingesetzt werden konnte.

## 2.2 Große Sprachmodelle

ChatGPT funktioniert auf der Grundlage des großen Sprachmodells GPT 3 bzw. GPT 4, Modellen des tiefgreifenden maschinellen Lernens. Diese Sprachmodelle werden mit großen Korpora an Textdaten vortrainiert, um menschliche Sprache zu analysieren und menschenähnliche Antworten zu generieren. Nutzerinnen und Nutzer können durch ChatGPT auf diese großen Sprachmodelle zugreifen und Aufgaben im Zusammenhang mit Sprachverarbeitung, beispielsweise das Zusammenfassen, Übersetzen oder Ergänzen von Texten, an das Modell auslagern.

Besonders interessant bei der Nutzung von ChatGPT ist, dass das zugrundeliegende Sprachmodell umfangreiche Zusammenhänge in natürlicher Sprache erkennen kann. So erfassen die Sprachmodelle der GPT-Familie nämlich die implizite Bedeutung einzelner Wörter sowie deren semantische Zusammenhänge, was die Generierung menschenähnlicher Antworten ermöglicht. Zur Erzeugung dieser menschenähnlichen Antworten greift ChatGPT auf große Datensätze bereits existierender Texte zurück, erkennt Muster in diesen Texten und verwendet diese Muster schließlich, um mit einem wahrscheinlichkeitsbasierten Ansatz für gegebene Eingaben passende Antworten zu formulieren. Das Formulieren der Antworten funktioniert dabei mithilfe der sogenannten autoregressiven Generation, bei der die Wahrscheinlichkeit für das Auftauchen jedes automatisch generierten Wortes vom vorherigen Wort abhängt. Diese beschriebene Funktionsweise zeigt, dass ChatGPT nicht den Zweck eines objektiven Recherchewerkzeugs erfüllen kann. Vielmehr wird das zugrundeliegende Sprachmodell auf riesigen Textkorpora trainiert, um möglichst kohärente, nicht notwendigerweise jedoch evidenzbasierte Antworten zu produzieren.

## 3. Leitfragen

Mit künstlicher Intelligenz funktionierende Tools wie ChatGPT werden im Bildungsbereich zunehmend genutzt sowie kontrovers diskutiert [9-11]. Bislang beschränkte sich die Anwendung von ChatGPT überwiegend auf die automatisierte Beantwortung von Fragen aus Prüfungen [12,13] oder auf die Diskussion von Ergebnissen in Protokollen [14]. Dieser Beitrag fokussiert hingegen die Anwendungsmöglichkeiten von ChatGPT als ergänzendes Werkzeug für die Unterrichtsplanung. Dabei werden die folgenden Leitfragen adressiert, um die Potentiale und Limitationen von ChatGPT besser beurteilen zu können:

- 1) An welcher Schrittfolge können sich Lehrkräfte orientieren, um mithilfe von ChatGPT eine Unterrichtsstunde im Fach Chemie mit experimentell-praktischem Schwerpunkt zu planen?
- 2) Welche Planungsideen äußert ChatGPT zur Einführung des Oxidationsbegriffs?
- 3) Sind die von ChatGPT geäußerten Gestaltungsmöglichkeiten didaktisch sinnvoll?

## 4. Ergebnisse und Diskussion

Die im Folgenden präsentierten, von ChatGPT generierten Antworten wurden mit einem kostenfreien Konto vom Erstautor erstellt. ChatGPT arbeitete zu diesem Zeitpunkt mit dem Sprachmodell GPT 3.5 sowie dem Update von 14. März 2023. Alle Antworten wurden innerhalb eines einzigen Chats generiert. Um die Exaktheit des algorithmischen Outputs zu erhöhen, wurde im Chat immer wieder gezielt nach weiteren Optionen gefragt. Die beigefügte Supporting Information bietet Möglichkeiten, den Output von ChatGPT im Detail nachzuvollziehen.

### 4.1 Anwendung von ChatGPT zur Unterrichtsplanung

Bei der Anwendung von ChatGPT zur Unterrichtsplanung gibt es einige allgemeine Aspekte, die Lehrende beachten können: Zunächst sollten Lehrende sich klarmachen, in welchen Unterrichtsphasen sie Inspiration und Diskussion von ChatGPT wünschen, um anschließend möglichst präzise Fragestellungen zu formulieren [1]. Die Zerlegung einer Anfrage in mehrere Teilschritte oder die Beschreibung des Kontextes kann ChatGPT helfen, die gewünschte Antwort zu verfeinern (Supporting Information, Auszug 1). Dabei ist es sinnvoll, den Detailgrad der automatisch erzeugten Antworten durch gezielte Nachfragen zu erhöhen. Des Weiteren ist zu beachten, dass ChatGPT alle Anfragen innerhalb eines Chats berücksichtigt. Ein zusammenhängendes Thema, beispielsweise die Planung einer Unterrichtsstunde, sollte somit in einem einzigen Chat bearbeitet werden; verschiedene Themen sind hingegen in verschiedenen Chats zu diskutieren.

Eine mögliche Vorgehensweise bei der Planung von Chemieunterricht mit experimentell-praktischer Arbeit unter Zuhilfenahme von ChatGPT ist in Abb. 1 dargestellt. Hierbei handelt es sich sicherlich nicht um die einzige Möglichkeit der Planung von Unterricht mit ChatGPT. Vielmehr hat sich diese Vorgehensweise durch zahlreiche Erprobungen als sinnvoll erwiesen.

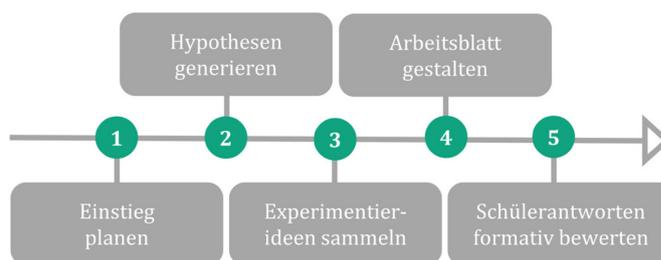


Abb. 1: Möglicher Arbeitsablauf für die Unterrichtsplanung mit ChatGPT.

Zu Beginn des Planungsprozesses können Lehrende ChatGPT um einen möglichen Unterrichtseinstieg bitten. Um ChatGPT einen Überblick über die Rahmenbedingungen zu verschaffen, sollten neben dem Thema der Unterrichtsstunde und der zu planenden Phase auch die Jahrgangsstufe und Schulform sowie Ziel und Zweck genannt werden (Tab. 1). Durch präzise Angaben erhöht sich der Genauigkeitsgrad der von ChatGPT generierten Antworten (Supporting Information, Auszug 1).



Tab. 1: Beispiel-Anfrage, die für die Planung einer Unterrichtsstunde mit ChatGPT genutzt werden kann. Die Anfrage kann beginnen mit: „Generiere einen problemorientierten Einstieg zu folgender Unterrichtsstunde...“

Rahmenbedingung	Konkretisierung
Thema	Oxidationsbegriff als Sauerstoffaufnahme
Unterrichtsphase	Einstieg
Jahrgangsstufe	9
Schulform	Gymnasium
Ziel	Oxidationsbegriff an Alltagsbeispielen entwickeln
Zweck	Motivierung und Angabe der Zielstellung

ChatGPT eignet sich im Planungsprozess nicht nur, um Lehrendenverhalten zu planen, sondern auch um Schülerantworten zu antizipieren. Dabei kann das textgenerierende Sprachmodell valide und alternative Hypothesen zu Phänomenen erstellen. Die generierten Hypothesen helfen, den Lernenden bei der Planung von Unterricht stärker in den Blick zu nehmen. Lehrkräfte können auf Basis der fiktiven Vorstellungen zielgerichtet planen, wie sie auf die jeweiligen Hypothesen reagieren würden und gegebenenfalls Experimente zu deren Überprüfung vorbereiten. Die fiktiven Schülervorstellungen können darüber hinaus in einen Concept Cartoon eingebettet werden, um wissenschaftlich korrekte und alternative Schülervorstellungen explizit zu kontrastieren [15]. Ein solcher Concept Cartoon könnte zu Beginn einer geplanten Unterrichtsstunde zur Ermittlung von Präkonzepten oder am Ende der Stunde zur Festigung valider Vorstellungen eingesetzt werden.

Anschließend kann ChatGPT unter Berücksichtigung der automatisch generierten Schülerhypothesen Experimente zu deren Überprüfung vorschlagen. Durch eine Vielzahl an Alternativen können sich Lehrkräfte Inspirationen einholen, woraufhin eine reflektierte Entscheidung für oder gegen die Durchführung einzelner Experimente zu treffen ist.

Zudem kann ChatGPT Aufgaben für ein Arbeitsblatt zur Ergebnissicherung konzipieren (Supporting Information, Auszug 2). Hierbei ist ebenfalls zu spezifizieren, welche Elemente das Arbeitsblatt enthalten soll, beispielsweise einen Erklärungstext, Multiple-Choice-Aufgaben, Zuordnungsaufgaben, Lückentexte oder Kurzantworten. Das Künstliche-Intelligenz-Tool Dall.E, bereitgestellt von OpenAI, kann zudem automatisch Bilder erzeugen, wobei aktuelle Urheberrechtsbestimmungen bei der Nutzung dieser Bilder stets zu berücksichtigen sind.

Für formative Zwecke kann ChatGPT darüber hinaus Schülerantworten bewerten, insofern die Lehrkraft dem Modell einen klar definierten Erwartungshorizont zur Verfügung stellt. Dies kann insbesondere hilfreich sein, wenn – ergänzend zur herkömmlichen Bewertung der Lehrkraft – Freitexte und Essays formativ zu beurteilen sind und Lehrende kontinuierlich Feedback erteilen möchten. Hierbei ist jedoch zu betonen, dass ChatGPT keine perfekte Bewertung von Freitext-Antworten ausstellt. Aus diesem Grund sollte es keineswegs zur Notenvergabe verwendet werden.

## 4.2 Planungsideen von ChatGPT

Die in Tab. 1 formulierte Anfrage für die Planung einer Unterrichtsstunde zur Einführung des Oxidationsbegriffs an Alltagsbeispielen wurde an ChatGPT gestellt. Das textgenerierende Modell antwortete wie folgt, redundante Passagen des Outputs wurden nachträglich gekürzt:

„Stell dir vor, du bist spät dran und musst schnell aus dem Haus, um zur Schule zu kommen. Du öffnest den Kühlschrank und nimmst einen geschnittenen Apfel heraus [...]. Doch du bemerkst, dass der Apfel seltsam schmeckt und eine braune Farbe hat. Was könnte pas-

siert sein? [...] In dieser Unterrichtsstunde werdet ihr lernen, welcher chemische Prozess für das Braunwerden des Apfels sorgt [...]. Ziel ist es, anhand dieses Alltagsbeispiels den zugrundeliegenden chemischen Prozess zu identifizieren und zu definieren.“

Auf Nachfrage nannte ChatGPT anschließend mögliche Hypothesen, die Lernende zur Erklärung des Phänomens äußern könnten. Zu diesen fiktiven Schülerhypothesen gehörte, dass der Apfel braun wurde, weil er alt, schmutzig oder von Insekten befallen sei. Auch könne es sein, dass der Apfel zu viel Sonne abbekommen habe, zu lange gelagert wurde oder in Berührung mit Lebensmitteln gekommen sei, die das Braunwerden verursachen.

Zur Überprüfung dieser Hypothesen schlug ChatGPT eine zielgerichtete Variablenkontrolle vor. Zum Beispiel solle ein früh und ein spät geernteter Apfel oder ein beschmutzter und unbeschmutzter Apfel vergleichend beobachtet werden, nachdem diese aufgeschnitten wurden. Auch sei es sinnvoll, einen Apfel mit und ohne Wurmbefall sowie einen Apfel mit und ohne Lichteinwirkung zu vergleichen. Zudem schlug ChatGPT vor, einen offenliegenden, aufgeschnittenen Apfel weiteren Äpfeln gegenüberzustellen, die entweder mit Frischhaltefolie bedeckt, mit Gummiband wieder in die ursprüngliche Position fixiert, unter Wasser gelegt oder mit Zitronensäure benetzt wurden.

Zur Ergebnissicherung erstellte ChatGPT ein Arbeitsblatt mit Informationstext, Multiple-Choice-Aufgabe, Lückentext und Freitext-Antworten (Supporting Information, Auszug 2). Die Freitext-Antworten wurden abschließend von ChatGPT bewertet. Die Höchstpunktzahl von zwei Punkten wurde vergeben, wenn der Begriff Oxidation in einer Antwort als Aufnahme von Sauerstoff erklärt wurde.

## 4.3 Reflexion der Planungsideen

ChatGPT lieferte eine Vielzahl an Gestaltungsmöglichkeiten für die Planung einer Unterrichtsstunde zur Einführung der Oxidation als Sauerstoffaufnahme. Positiv ist hierbei hervorzuheben, dass ChatGPT einen alltagsnahen, praxisbezogenen und problemorientierten Unterrichtseinstieg vorschlagen konnte, der im Schulunterricht tatsächlich umgesetzt werden kann. Weiterhin konnte ChatGPT eine Vielzahl fiktiver Schülervorstellungen erzeugen, die Lehrkräfte zur Testung von Hypothesen im naturwissenschaftlichen Gang der Erkenntnisgewinnung überprüfen können. Zudem erscheint der Ansatz der Variablenkontrolle als naturwissenschaftliche Denk- und Arbeitsweise didaktisch sinnvoll.

Beim Einsatz von ChatGPT ist hervorzuheben, dass das zugrundeliegende große Sprachmodell keine evidenzbasierten Unterrichtspraktiken, sondern vielmehr wahrscheinlichkeitsbasierte Ansätze zur Planung von Unterricht nutzt. Der Vorschlag zur Einführung des Oxidationsbegriffs am Beispiel des Braunwerdens von Apfelanschnitten wird demnach genannt, weil diese Unterrichtsidee häufig im Trainingsdatensatz des zugrundeliegenden großen Sprachmodells auftritt. Zur Integration evidenzbasierter Unterrichtsideen in Modelle des maschinellen Lernens bedarf es weiterhin die menschliche Interpretation der genannten Ideen mithilfe von fachdidaktischen Erkenntnissen [7].

Zudem ist bei der beschriebenen Unterrichtsplanung zu beachten, dass ChatGPT nicht immer in der Lage war, die Fachinhalte didaktisch zu reduzieren. Um das Braunwerden eines Apfelanschnittes Lernenden im Chemie-Eingangsunterricht verständlich zu erklären, müssen die Inhalte vereinfacht werden. Die von ChatGPT als Erklärung genannte enzymatische Katalyse müsste beispielsweise je nach Jahrgangsstufe vernachlässigt werden. Die im Apfel enthaltenen Polyphenole, die ChatGPT richtigerweise für das Braunwerden der Äpfel verantwortlich machte, könnten im Unterricht durch Begriffe wie Pflanzenstoffe, Geschmacksbildner oder Abwehrstoffe ersetzt werden. Eine stark vereinfachte Wortgleichung des Phänomens könnte im Schulunterricht wie folgt lauten: Der farblose



Pflanzenstoff reagiert mit Luftsauerstoff zu einem braunen Farbstoff, der für die Farbveränderung der Apfelschnitte sorgt.

Zu dieser didaktischen Reduktion war ChatGPT nicht immer fähig (Supporting Information, Auszug 3). Erste von ChatGPT erstellte Erklärungen für das Braunwerden von Apfelschnitten enthielten zahlreiche Fachbegriffe wie Enzyme, Polyphenolverbindungen, Polyphenoloxidase, Melanin und Pigment, was das Verständnis der Erklärung für Lernende deutlich erschweren sollte. Auf konkrete Anfrage nach einer didaktischen Reduktion ersetzte ChatGPT den Begriff Polyphenole durch Pflanzenstoffe oder sekundäre Pflanzenmetabolite. Nach Bitte um weitere didaktische Reduktion tendierte ChatGPT zu animistischen Sprechweisen [15] und teleologischen Erklärungsansätzen [16], wodurch der eigentliche Kern des chemischen Phänomens vernachlässigt wurde. Insgesamt hatte ChatGPT Probleme, sich bei Erklärungen nicht in Details zu verlieren, das Verständnisniveau von Lernenden einer bestimmten Jahrgangsstufe zu antizipieren, zwischen relevanten und irrelevanten Informationen zu unterscheiden und die Sprache an das Wissensrepertoire von Lernenden anzupassen. Für eine individuelle didaktische Reduktion braucht es somit weiterhin fachdidaktische Expertise. Dies zeigt, dass ChatGPT Lehrkräfte in der Planung von Unterricht unterstützen, diese jedoch nicht ersetzen kann.

## 5. Limitationen

Die in diesem Beitrag vorgestellten Ergebnisse sind teilweise limitiert. Vorab ist zu betonen, dass die Analyse der Antworten von ChatGPT stets den aktuellen Forschungsstand widerspiegelt. Durch Updates der Sprachmodelle und Fortschritte im Bereich des maschinellen Lernens verändern sich die von ChatGPT generierten Antworten. Aus diesem Grund sind die Ergebnisse unserer Analyse vom Stand der Technologie bestimmt.

Antworten von ChatGPT sind des Weiteren, analog zu den Ergebnissen zahlreicher Anwendungen des tiefgreifenden maschinellen Lernens, nicht vollständig reproduzierbar, weil sie auf einer probabilistischen Methode basieren. Selbst bei identischen Anfragen kann es daher dazu kommen, dass aufgrund unterschiedlicher Zufallsvariationen verschiedene Ergebnisse produziert werden. Neben der Eingabe beeinflussen zudem der Verlauf des vorangestellten Chats und die verwendete Version des GPT Sprachmodells die generierten Ergebnisse.

Bei den erstellten Ergebnissen kommt es darüber hinaus manchmal dazu, dass ChatGPT halluziniert. Dies bedeutet, dass Antworten unter Umständen Informationen enthalten, die nicht korrekt oder irrelevant sind. Dieses Erfinden von Informationen kann zu Verzerrungen im Antwortverhalten von ChatGPT führen. Für die Planung von Unterricht bedeutet dies konkret, dass ChatGPT Unterrichtsideen möglicherweise nicht begründet in Bezug auf Unterrichtsqualität auswählt, sondern diese wahrscheinlichsbasiert auf Grundlage des vorliegenden Trainingsdatensatzes ermittelt. Als Lehrkraft ist es daher wichtig, die Qualität der vorgeschlagenen Unterrichtsideen kritisch zu prüfen.

In Verbindung mit dem Phänomen des Halluzinierens ist zu erwähnen, dass große Sprachmodelle wie GPT 3 und GPT 4 mit nicht aufgereinigten Sprachdaten trainiert werden, die zum Beispiel aus dem Internet stammen. Aus diesem Grund können diese Sprachmodelle Stereotype ausbilden, die zur Aufrechterhaltung oder Verschärfung von Diskriminierungen gegen gesellschaftlich marginalisierte Gruppen führen können [17]. Zur Eindämmung dieser Diskriminierung sind gegebenenfalls fachspezifische Gegenmaßnahmen zu treffen [18,19] oder individuelle Algorithmen unter Beachtung von Fairness-Aspekten zu programmieren.

Zudem kann nicht sichergestellt werden, dass ChatGPT alle Datenschutzbestimmungen und Urheberrechtsrichtlinien beachtet. Das textgenerierende Modell speichert beispielsweise die Eingaben sowie die Logbuchdaten der Nutzerinnen und Nutzer, um die zukünftige Genauigkeit der Ausgaben zu erhöhen. Aus diesem Grund sollten keine persönlichen Daten in das Sprachmodell eingegeben werden. Weiterhin wurde das Modell unter anderem auf großen Mengen an Benutzerdaten trainiert, deren Herkunft im Detail nicht geklärt ist.

Aus praktischer Perspektive ist letztlich anzumerken, dass ChatGPT mit einem kostenfreien Konto nicht immer verfügbar ist. Die Software-Anwendung wird auf Servern oder Cloud-Plattformen ausgeführt, die bei erhöhter Nachfrage überlastet sind. Dies führt dazu, dass ChatGPT bei erhöhter Auslastung unter Umständen nicht genutzt werden kann.

## 6. Fazit und Implikationen

Große Sprachmodelle stellen einen bedeutenden Fortschritt im Bereich der Technologien künstlicher Intelligenz dar. Trotz kritischer Ansichten, beispielsweise bezüglich der Gefahr des automatisierten Erstellens von Plagiaten [20], bietet die zugrundeliegende Technologie in vielzähligen Anwendungsbereichen das Potential für weitreichende Innovationen. Aus diesem Grund erscheint es sinnvoll, dass sich Lehrende mit der Anwendung von ChatGPT und den weitreichenden Folgen auseinandersetzen. Nutzerinnen und Nutzer sollten dabei mit einem kritischen Bewusstsein für gesellschaftliche und ethische Fragestellungen sowie einem Verständnis von algorithmischer Wahrscheinlichkeit und Unsicherheit die Antworten des Sprachmodells beurteilen. Die automatisch erzeugten Antworten sind hinsichtlich ihres Wahrheitsgehaltes und ihrer Passung zu hinterfragen. ChatGPT ist demnach keine objektive Informationsquelle im Sinne eines Recherchewerkzeugs, sondern vielmehr ein subjektiver Diskussionspartner.

Bei der Unterrichtsplanung können große Sprachmodelle Lehrkräfte unterstützen, indem sie Stundenentwürfe gestalten und Aufgaben konzipieren. Darüber hinaus können sie verwendet werden, um Alltagsbezüge herzustellen, wissenschaftlich korrekte sowie alternative Schülervorstellungen zu generieren und mögliche Experimente vorzuschlagen. All diese Planungsideen können helfen, den naturwissenschaftlichen Gang der Erkenntnisgewinnung im Chemieunterricht umfassend zu unterstützen.

Insgesamt bleibt zu betonen, dass ChatGPT keineswegs einen Ersatz für menschliche Lehrkräfte darstellt. Dies wird unter anderem daran deutlich, dass Sprachmodelle keine sozialen Entscheidungen treffen und Unterricht demnach nicht auf die individuellen Lernbedürfnisse der Schülerinnen und Schüler eingehen können. Im Planungsprozess kann ChatGPT als gewinnbringende Ergänzung wahrgenommen werden. Aussagen von ChatGPT müssen jedoch stets reflektiert und gegebenenfalls korrigiert werden. Besonders zu betonen ist, dass Lehrkräfte die Expertinnen und Experten bei der Planung und Durchführung von Unterricht bleiben. Werkzeuge künstlicher Intelligenz können den Planungsprozess zwar bereichern, ersetzen Lehrkräfte allerdings nicht.

## 7. Supporting Information

In der Supporting Information werden Auszüge aus dem Interview mit ChatGPT dargelegt. Diese Auszüge präzisieren die im Tagungsband-Beitrag getroffenen Aussagen zur Zerlegung einer Anfrage in mehrere Teilschritte, die Möglichkeiten der Nutzung von ChatGPT zur Generierung von Unterrichtsmaterial sowie die mangelnden Fähigkeiten von ChatGPT in der didaktischen Reduktion.



## 8. Referenzen

- [1] P. P. Martin, N. Graulich (2023): When a machine detects student reasoning: a review of machine learning-based formative assessment of mechanistic reasoning, *Chem. Educ. Res. Pract.*, 24 (2), 407-427, DOI: 10.1039/D2RP00287F
- [2] X. Zhai, K. C. Haudek, L. Shi, R. H. Nehm, M. Urban-Lurain (2020): From substitution to redefinition: A framework of machine learning-based science assessment, *J. Res. Sci. Teach.*, 57 (9), 1430-1459, DOI: 10.1002/tea.21658
- [3] X. Zhai, Y. Yin, J. W. Pellegrino, K. C. Haudek, L. Shi (2020): Applying machine learning in science assessment: a systematic review, *Stud. Sci. Educ.*, 56 (1), 111-151, DOI: 10.1080/03057267.2020.1735757
- [4] A. Grant St James, L. Hand, T. Mills, L. Song, A. S. J. Brunt, P. E. Bergstrom Mann, A. F. Worrall, M. I. Stewart, C. Vallance (2023): Exploring Machine Learning in Chemistry through the Classification of Spectra: An Undergraduate Project, *J. Chem. Educ.*, 100 (3), 1343-1350, DOI: 10.1021/acs.jchemed.2c00682
- [5] D. Lafuente, B. Cohen, G. Fiorini, A. A. Garcia, M. Bringas, E. Morzan, D. Onna (2021): A Gentle Introduction to Machine Learning for Chemists: An Undergraduate Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling, *J. Chem. Educ.*, 98 (9), 2892-2898, DOI: 10.1021/acs.jchemed.1c00142
- [6] E. S. Thrall, S. E. Lee, J. Schrier, Y. Zhao (2021): Machine Learning for Functional Group Identification in Vibrational Spectroscopy: A Pedagogical Lab for Undergraduate Chemistry Students, *J. Chem. Educ.*, 98 (10), 3269-3276, DOI: 10.1021/acs.jchemed.1c00693
- [7] M. Kubsch, C. Krist, J. M. Rosenberg (2023): Distributing episodic functions and tasks – A framework for augmenting human analytic power with machine learning in science education research, *J. Res. Sci. Teach.*, 60 (2), 423-447, DOI: 10.1002/tea.21803
- [8] K. Zweig: Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können, 2019, Heyne, München.
- [9] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, G. Kasneci (2023): ChatGPT for good? On opportunities and challenges of large language models for education, *Learn. Individ. Differ.*, 103, 1-13, DOI: 10.1016/j.lindif.2023.102274
- [10] G. Lawrie (2023): Establishing a delicate balance in the relationship between artificial intelligence and authentic assessment in student learning, *Chem. Educ. Res. Pract.*, 24 (2), 392-393, DOI: 10.1039/D3RP90003G
- [11] X. Zhai (2022): ChatGPT User Experience: Implications for Education, *SSRN*, 1-18, DOI: 10.2139/ssrn.4312418
- [12] T. M. Clark (2023): Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions, *J. Chem. Educ.*, 100 (5), 1905-1916, DOI: 10.1021/acs.jchemed.3c00027
- [13] S. Fergus, M. Botha, M. Ostovar (2023): Evaluating Academic Answers Generated Using ChatGPT, *J. Chem. Educ.*, 100 (4), 1672-1675, DOI: 10.1021/acs.jchemed.3c00087
- [14] T. Humphry, A. L. Fuller (2023): Potential ChatGPT Use in Undergraduate Chemistry Laboratories, *J. Chem. Educ.*, 100 (4), 1434-1436, DOI: 10.1021/acs.jchemed.3c00006
- [15] H.-D. Barke: *Chemiedidaktik – Diagnose und Korrektur von Schülervorstellungen*, 2006, Springer, Berlin und Heidelberg.
- [16] V. Talanquer (2023): Interview with the Chatbot: How Does It Reason? *J. Chem. Educ.*, 100 (8), 2821-2824, DOI: 10.1021/acs.jchemed.3c00472
- [17] T. Cheuk (2021): Can AI be racist? Color-evasiveness in the application of machine learning to science assessments., *Sci. Educ.*, 105 (5), 825-836, DOI: 10.1002/sce.21671
- [18] A. Grimm, A. Steegh, J. Çolakoglu, M. Kubsch, K. Neumann (2023): Positioning responsible learning analytics in the context of STEM identities of under-served students, *Front. Educ.*, 7:1082748, 1-12, DOI: 10.3389/educ.2022.1082748
- [19] A. Grimm, A. Steegh, M. Kubsch, K. Neumann (2023): Learning Analytics in Physics Education: Equity-Focused Decision-Making Lacks Guidance! *J. Learn. Anal.*, 10 (1), 71-84, DOI: 10.18608/jla.2023.7793
- [20] H. H. Thorp (2023): ChatGPT is fun, but not an author, *Science*, 379(6630), 313, DOI: 10.1126/science.adg7879