

Lernräume unter Verwendung von generativen Sprachmodellen

Ein Erfahrungsbericht zum kreativen Einsatz von Tools wie ChatGPT in der Software-Engineering-Hochschullehre

Victoria Geisel¹, Christian Schindler², Nils Stein³, Stefan Bente⁴

Abstract: University education may cultivate independent study by design of supportive environments for students. Recent proliferation of generative AI tools has had a significant impact on this. The extent to which successful scholarship is influenced by these tools is the subject of current research and the focus of this study. Tasks in the field of databases with different competence levels and didactic goals were solved by students with the help of GPT-3.5. Intellectual performance and success were then evaluated. The results show that integration of generative language models has the potential to improve academic attainment. Success proved highly dependent upon the nature of task and the strategy employed. The paper emphasises the need to integrate the effective use of AI tools into SE teaching, in order to use generative language models to their full potential in independent academic development.

Keywords: Künstliche Intelligenz; Hochschullehre; Bildungspraktiken; Lernprozesse; ChatGPT; Lernraum; generative Sprachmodelle; Software Engineering

1 Einleitung

Der kostenlose Zugang zu generativen Sprachmodellen hat die Wahrnehmung der Hochschul- und Berufswelt verändert. In einer Umfrage des Pew Research Center vom Juli 2023 in den USA [PG23] gaben 75% derjenigen Teilnehmer*innen, die bereits von einem Chatbot wie beispielsweise ChatGPT gehört haben, an, dass ChatGPT in den nächsten 20 Jahren einen Einfluss auf den Beruf “Softwareentwickler*in” haben wird. Die Studie zeigt ebenfalls auf, dass je höher der Bildungsgrad der Teilnehmenden ist, desto höher der relative Anteil an Personen, die daran glauben, dass ChatGPT einen Einfluss auf ihren eigenen Beruf haben wird. Als Mentoren und Ausbilder der Fachkräfte von morgen müssen Lehrende an Hochschulen und Universitäten die Aufklärung sowie Auseinandersetzung mit generativen Sprachmodellen in den Fokus rücken. Sind Sprachmodelle wie ChatGPT nur eine Gefahr

¹ Technische Hochschule Köln, Fakultät für Informatik und Ingenieurwissenschaften, victoria.geisel@th-koeln.de

² Technische Hochschule Köln, Fakultät für Informatik und Ingenieurwissenschaften, schindler.chris@icloud.com

³ Technische Hochschule Köln, Fakultät für Informatik und Ingenieurwissenschaften, nils.stein@smail.th-koeln.de

⁴ Technische Hochschule Köln, Fakultät für Informatik und Ingenieurwissenschaften, stefan.bente@th-koeln.de

für die bisherige Form der akademischen Lehre – oder ergeben sich durch einen kreativen Einsatz als Werkzeug auch neue Möglichkeiten des Kompetenzerwerbs?

Diese Frage wurde im Rahmen eines Lehrforschungsprojekts im Masterstudiengang *Digital Sciences* an der Fakultät für Informatik und Ingenieurwissenschaften der TH Köln untersucht [AL23]. Anhand des Moduls *Datenbanksysteme (DBS)* [HT23] im 4. Semester des Bachelorstudiengangs Informatik wurden verschiedene Lernräume mit individuellen didaktischen Zielen unter Einschluss von ChatGPT gestaltet. Diese Lernräume werden in den nachfolgenden Kapiteln dargestellt und in ihrer Wirksamkeit evaluiert. Dabei wird ausführlich auf die jeweiligen Vor- und Nachteile der Integration von ChatGPT eingegangen. Die Konzepte und Ergebnisse werden zwar anhand von Aufgabenstellungen aus dem relationalen Datenbankdesign diskutiert, lassen sich aber auch auf andere Bereiche des Software Engineerings übertragen.

Beim Forschungsdesign wurde der Umstand genutzt, dass einer der Autoren gleichzeitig wissenschaftlicher Mitarbeitender mit Verantwortung für die genannte Veranstaltung ist, und dass die verantwortlichen Professoren (Dr. Birgit Bertelsmeier und Dr. Johann Schaible) die Studie aktiv unterstützten. In diesem Rahmen wurden zwei Aufgabenblätter für diese Studie mit insgesamt sieben verschiedenen Aufgaben zusammengestellt, die unter Einschluss von ChatGPT zu bearbeiten waren. Semesterbegleitend haben dies im Sommersemester 2023 insgesamt 43 Studierende getan. Die Ergebnisse und Lösungswege (z.B. ChatGPT-Prompts) mussten dabei von den Studierenden dokumentiert werden. Diese boten, zusammen mit vier geführten Interviews und zahlreichen informell protokollierten Diskussionen in den Übungsgruppen, detailliertere Einblicke in die Vorgehensweisen und Prozesse bei der Bearbeitung dieser Lernräume. Die Ergebnisse sind in den nachfolgenden Kapiteln beschrieben. Zum Abschluss dieses Papers wird ein Fazit gezogen mit dem Ziel, eine Diskussionsgrundlage zu schaffen sowie zu weiterer Forschung anzuregen, wie der künftige Umgang mit KI-Tools an akademischen Lehrinrichtungen aussehen könnte.

2 Gestaltung von Lernräumen unter Einschluss von ChatGPT

Das Modul DBS befasst sich mit einfachen bis semi-komplexen Code-Strukturen im Datenbank-Bereich, größtenteils anhand der Programmiersprache PL/SQL. Kategorisiert in *Basisaufgaben*, *Vertiefende Aufgaben* und *Experimentelle Aufgaben* wurden die nachfolgend vorgestellten Lernräume wie folgt evaluiert:

- Die Autoren protokollierten die Ergebnisse der unsystematischen Selbstbeobachtungen, die sie beim eigenen Erstellen der Aufgabenblätter inklusive Musterlösungen gewinnen konnten, um vorab die Ergebnisqualität, den Lernerfolg und mögliche Stolpersteine identifizieren zu können. Die Ergebnisse der Beobachtungen mündeten in dem Forschungsdesign entsprechenden Anpassungen der Aufgabenstellungen.

- Die Studienteilnehmenden mussten nicht nur die eigene Lösung für die jeweiligen Aufgaben einreichen, sondern auch den gesamten in diesem Zusammenhang geführten Dialog mit ChatGPT. Teilweise wurde dieser noch von den Studierenden im Nachgang kommentiert.
- Im Verlauf der Studie gab es mehrere Online-Meetings mit den Studienteilnehmern, in denen die Aufgaben erläutert sowie über die Lösungen gesprochen wurde. Teilweise mündete dies in sehr detaillierten Diskussionen. Diese in Zoom durchgeführten Interviews folgten keinem definierten Leitfaden, um zum einen den persönlichen Austausch mit den Studierenden zu fördern, und zum anderen, um die Gesprächsthemen des noch unerforschten Bereichs nicht durch definierte Fragen einzuschränken. Die Interviews wurden informell protokolliert.
- Am Ende der Studie nahmen alle studentischen Teilnehmer anonym an einer Online-Umfrage teil. Dadurch wurde ihnen die Möglichkeit für ehrliches, vielleicht auch negatives oder kritisches Feedback gegeben. Die Studierenden wurden nach Aufwandschätzungen, Qualitätsbeurteilungen, Problemen und persönlichen Nutzungsverhalten gefragt.

Das Ziel der Studierenden im Zusammenhang mit diesem Experiment war die Generierung von lauffähigen Code-Lösungen mithilfe von GPT-3.5, die der Semantik der jeweiligen Aufgabenstellungen entsprachen. Im Detail bedeutet dies, dass die Vorgehensweise den Studierenden freigestellt war und durch iterative Anfragen an das generative Sprachmodell Aufgabenstellungen gelöst werden sollten. Für Aufgabenstellungen, die keine Generierung von Code erforderten, sollte durch iterative Anfragen eine korrekte Lösung für das in der Aufgabenstellung beschriebene Problem erzeugt werden. Die Motivation der Studierenden, gute Lösungen zu erarbeiten, wurde durch die Verteilung von Klausurbonuspunkten erhöht. Die Bewertung, ob eine Aufgabe korrekt gelöst ist, erfolgte durch die Autoren. Die im Folgenden verwendeten Prozentangaben zur Korrektheit einzelner Aufgaben beziehen sich auf diese Bewertungen.

2.1 Lernerfolg

Lernerfolg beschreibt den nachweisbaren Erwerb von Wissen und Fähigkeiten, in diesem Paper konkret beim selbstständigen Lernen im Rahmen der Lehre. Dabei können sowohl persönliche als auch vorgegebene Lernziele erreicht werden. Werden Lernziele nicht erreicht und Kompetenzen nicht erworben, liegt **nahezu kein Lernerfolg** vor. **Geringer Lernerfolg** bedeutet ansatzweises Erreichen von Lernzielen, wobei Kompetenzen nicht im ausreichenden Maß erworben werden. Bei **partielltem Lernerfolg** werden die Ziele nur in Teilen erreicht. Erfolgt ein **dependenzieller Lernerfolg** können Lernziele erreicht, sowie Kompetenzen unter Berücksichtigung von klar definierten Vorgehensweisen erworben werden. Bei einem **guten Lernerfolg** werden alle Lernziele erreicht und Kompetenzen vollumfänglich entsprechend des Learning Outcomes erworben.

2.2 Lernräume

Ein Lernraum bezeichnet die Umgebung, in der Lernaktivitäten stattfinden [St18]. Diese können physischer oder virtueller Natur sein. Insbesondere zum selbstständigen Lernen eignen sich virtuelle Lernräume, da keine Interaktion mit anderen Personen stattfindet. In einem digitalen Umfeld kann dieser Raum stark durch generative Sprachmodelle gestaltet und personalisiert werden.

Tab. 1: Übersichtstabelle Aufgaben

ID	Bloom Level	Aufgabe	Didaktisches Ziel	Ergebnisqualität	Lernerfolg
B1	3	Prozedur in PL/SQL programmieren	Abfrage und Anwendung von Grundlagen-Wissen in einem definierten Anwendungskontext	Nahezu perfekte Lösungen (98%)	Nahezu keiner
B2	3	Trigger in PL/SQL programmieren	Abfrage und Anwendung von Grundlagen-Wissen in einem definierten Anwendungskontext	Gute Lösungen (70%)	Gering
V1	4	Compound-Trigger programmieren	Erkennung von Zusammenhängen durch kritisches Auseinandersetzen mit dem Thema	Befriedigende Lösungen (60%)	Partiell
V2	4	Python-Skript für Oracle-Datenbank	Erkennung von Zusammenhängen durch kritisches Auseinandersetzen mit dem Thema	Gute bis sehr gute Lösungen (84%)	Dependenziell
E1	2	Fragen zu B-Bäumen beantworten	Nachhaltiges Lernen durch Erklären	Ausreichende Lösungen (51%)	Gering
E2.1	5	Fehlersuche in komplexer Prozedur, theoretische Fragen	Lernen durch Auffinden von Fehlern	Gute Lösungen (74%)	Gut
E2.2	2	Theoretische Fragen bekommen und selbst beantworten	Assistiertes Lernen durch selbstständiges Recherchieren in Literaturquellen	Gute Lösungen (74%)	Gering
E3	4	Trigger aus Grafik erstellen	Eigenständige Entwicklung sowie Interpretation einer nicht näher definierten Aufgabenstellung	Sehr gute Lösungen (86%)	Gut

Die Lernräume werden anhand von *Bloom's Revised Taxonomy* [AKB01] klassifiziert, mit deren Hilfe Lernziele in sechs Stufen (Wissen, Verstehen, Anwenden, Analysieren, Bewerten, Erschaffen) differenziert werden können. So kann sichergestellt werden, dass Lernende auf verschiedenen Ebenen des kognitiven Denkens gefordert werden. Tabelle 1 gibt einen Überblick über die gestellten Aufgaben, deren didaktisches Ziel sowie der Eignung der Aufgaben für das selbstständige Lernen von Studierenden gemessen an der Qualität der Ergebnisse und des Lernerfolgs.

3 Basisaufgaben (B)

Die in diesem Kapitel beschriebenen Aufgaben wenden Grundlagenwissen an bzw. fragen dieses ab. Die zu implementierenden Funktionalitäten waren einfach gehalten und sollen Aufgaben zum Einstieg in eine neue Programmiersprache repräsentieren. Beide gestellten Aufgaben dieses Lernraumes sind als Bloom-Level 3 einzuordnen. Die Studierenden sollen eine konkrete Technik in einem klar definiertem Umfeld anwenden. Das Lernziel ist dabei die Fähigkeit zu stärken, selbstständig theoretisches Wissen in einem Anwendungskontext praktisch umzusetzen.

Ziel der ersten Aufgabe (B1) war die Programmierung einer Prozedur ohne Rückgabewert. Die zweite Aufgabe (B2) fokussierte die Implementierung eines Triggers in PL/SQL mithilfe einer textuellen Beschreibung der Funktionalitäten und erforderte die Anwendung des Konzepts der Transition.

3.1 Bewertung

98% der eingereichten Lösungen zur Prozeduraufgabe (B1) waren lauffähig und semantisch korrekt gelöst. Die aus den Gesprächsverläufen beobachtete Strategie der Studierenden war mehrheitlich das Kopieren und Einfügen der Aufgabenstellung an ChatGPT und es konnte in den meisten Fällen initial eine korrekte Lösung erfolgreich generiert werden.

70% der eingereichten Lösungen bei der Triggeraufgabe (B2) waren korrekt. Copy and paste der Aufgabenstellung führte seltener zu einer initial korrekten Lösung. ChatGPT hatte Probleme bei der Unterscheidung zwischen einer Transitionsvariable (bspw. OLD) und der Referenz auf eine Transitionsvariable (bspw. :OLD).⁵ Es konnte ebenfalls beobachtet werden, dass der gewählte SQL-Dialekt sich im Verlauf des Gesprächs ohne separate Anweisung verändern kann (bspw. plötzlicher Wechsel zu MySQL als Antwort auf eine Anfrage).

Die vorgestellten Basisaufgaben unterstützen nach Evaluation der Studie nicht das selbstständige Lernen. Studierende empfanden den plötzlichen Wechsel des gewählten SQL-Dialekts als frustrierend. Wenn in Gesprächsverläufen ein Wechsel des Dialekts erkenntlich war, haben nur die wenigsten der betroffenen Studierenden dies korrekterweise erkannt und das generative Sprachmodell auf diesen Fehler aufmerksam gemacht. Mehrheitlich wurde dieser Fehler nicht erkannt und die Lösungen in einem falschen Dialekt eingereicht. Zusätzlich, durch den hohen Anteil an initial korrekt generierten Lösungen, konnte im Rahmen der Basisaufgaben kein hoher Lerneffekt beobachtet werden. Dieser wäre aber gegeben, wenn Studierende mehrheitlich den Wechsel des Dialekts verstanden sowie evaluiert hätten oder durch eigenständige Weiterentwicklung des Codes (bspw. Abfangen von möglichen Fehlerquellen) eine tiefgreifendere sowie qualitative Auseinandersetzung erfolgt wäre.

⁵ Dieses Problem trat bei der verwendeten Version GPT-3.5 auf. Inwiefern es andere generative KI Modelle betrifft, etwa GPT-4.0, könnte in weiterführender Forschung untersucht werden.

4 Vertiefende Aufgaben (V)

Diese Aufgaben betrachteten erweiterte Datenbank-Konzepte. Zur Lösung dieser Aufgaben ist ein gewisses Vorwissen notwendig. Auf der Bloom-Skala sind beide Aufgaben als Level 4 einzuordnen. Die Studierenden analysierten ein konkretes Problem und lösten es mit einer passenden Technik. Sie mussten tiefer in das Material eintauchen und kritisch über die Anwendung und die Auswirkungen der Technik nachdenken, und lernten so, Strukturen und Zusammenhänge zu erkennen.

Eine Aufgabe befasste sich mit der Implementierung eines Compound-Triggers (V1), der die beiden in PL/SQL verfügbaren Arten (Statement- und Row-Trigger) verknüpft. Die zweite Aufgabe befasste sich mit der anwendungsseitigen Datenbankanbindung (V2) und verknüpfte zwei Programmiersprachen, indem ein Python-Skript entworfen werden musste, das sich mit einer Oracle-Datenbank verbindet und eine semantisch vorgegebene Query an die Datenbank abschickt.

4.1 Bewertung

Bei beiden Aufgaben konnte ChatGPT 3.5 im Rahmen der Studie keine initial korrekte Lösung generieren. 60% der eingereichten Lösungen zur Compound-Trigger Aufgabe (V1) waren korrekt und lauffähig. Bei dieser Aufgabe konnte mehrheitlich ein guter Lernprozess beobachtet werden, da das Kopieren und Einfügen der Aufgabenstellung nicht zum erwünschten Ziel führte und Studierende sich mit semantischem Fachwissen sowie mit dem generierten Code befassen mussten, um das Problem erkennen und beheben zu können. Es wurde auch beobachtet, dass ChatGPT bei optionalen Code-Strukturen leere Code-Blöcke erzeugen kann, die keine Funktionalitäten haben und somit nicht erforderlich wären, weshalb qualitativ minderwertiger Code erzeugt wird.

Die Aufgabe zur Anbindung einer Datenbank an eine Anwendung (V2) konnte zu 84% korrekt gelöst werden. Die Verknüpfung von zwei Programmiersprachen in einem Skript scheint für ChatGPT kein Problem darzustellen, die vorgeschlagenen Lösungen hatten aber kleine Mängel, die korrigiert werden mussten. Das Kopieren und Einfügen von Fehlermeldungen, die beim Testen des Codes ausgegeben werden, führte bei ChatGPT zu einer korrekten Lösung, was auch die mehrheitlich von den Studierenden verwendete Strategie darstellte. Dies könnte dazu führen, dass der Lernerfolg auf Seiten der Studierenden ausbleibt.

Die vorgestellten vertiefenden Aufgaben unterstützen nach Evaluation der Studie das selbstständige Lernen bedingt. Der Lernerfolg bei der Programmierung eines Compound-Triggers (V1) zeigte sich deutlich in den eingereichten Chatverläufen der Teilnehmenden. Es konnte mehrfach beobachtet werden, dass die Strategie, die Aufgabenstellung oder die Oracle-Fehlermeldungen zu übergeben, nicht mehr funktioniert hat. In diesem Fall mussten die Fehler eigenhändig in dem generierten Codeausschnitt gesucht werden, und somit war

eine tiefgreifendere Auseinandersetzung mit dem Code erforderlich. Es konnte ebenfalls beobachtet werden, dass nach Scheitern dieser Strategie vermehrt Verständnisfragen zur Semantik des Codes gestellt wurden, die zum Lernerfolg beitrugen. Wie bereits bei der Trigger-Aufgabe (B1) konnte in der Compound-Trigger Aufgabe (V1) beobachtet werden, dass ChatGPT Schwierigkeiten mit Transitionen hatte, zusätzlich verstärkt durch Verwendung von Transitionen in Codeblöcken, die keine Transitionen erlaubten.

In der zweiten Aufgabe (V2) hing der Lernerfolg stark von der Strategie der Studierenden ab, da das Kopieren und Einfügen der Fehlermeldungen von Oracle zu einem Erfolg führen konnte. Bei Eintragung benutzerspezifischer Daten (sowie Serverdaten) zeigten Studierende Kreativität und modifizierten durch Folgeanfragen an ChatGPT die generierte Lösung so, dass das Auslesen der Daten aus einer Datei erfolgte oder über einen Input-Wert bei der Ausführung des Skripts eingetragen werden musste. Vereinzelt gab es auch Studierende, die persönliche Login-Daten der zur Verfügung gestellten Datenbankinstanz an ChatGPT zur Generierung einer Lösung übergaben.

5 Experimentelle Aufgaben (E)

Experimentelle Aufgaben hatten den Zweck, kreative Aufgabenstellungen und das Potenzial von generativen Sprachmodellen zu untersuchen. Zwei Aufgaben fokussierten ein Rollenspiel zwischen Sprachmodell und Studierenden, bei dem die Rollen der Studierenden und Lehrenden eingenommen wurden. Die beiden Aufgaben verfolgten das didaktische Ziel, einen Dialog zwischen Lernenden und Lehrenden zu schaffen, wie er außerhalb des digitalen Lernumfelds zu finden ist. Durch das Einnehmen der Rolle als Lehrende durch die Studierenden sollte der bewährte Ansatz erprobt werden, bei dem sich bei Lernenden durch das Erklären von Inhalten an Dritte ein nachhaltiger Lernerfolg einstellt [Ma02]. In einem entgegengesetzten Ansatz sollte ChatGPT den Lernenden eine falsche Antwort liefern, sodass sich aus dem Auffinden von Fehlern ein Lernerfolg einstellen konnte. Des Weiteren sollte ChatGPT den Lernenden Fragen zu theoretischen Grundlagen stellen. Nach dem Prinzip „Assistiertes Lernen“ sollte ChatGPT den Lernenden einen thematisch relevanten Quellverweis liefern, der diese bei der Beantwortung der Frage unterstützt. Die Komplexität der Lernziele befindet sich dabei auf unterschiedlichen Stufen (einfachem Verstehen (2), Analysieren (4) und Beurteilen (5)). In der letzten Aufgabe sollten Studierende auf Basis eines visuellen Inputs selbstständig kreativ agieren und eine Aufgabenstellung samt Lösung entwickeln. Die Aufgabe wurde auf Grundlage der Inquiry-Based Learning Methode entwickelt [SAC23]. Dabei soll durch Erfahrungswerte und tiefgründiges Hinterfragen der verfügbaren Informationen eigenständig Wissen ermittelt werden.

Die erste Aufgabe verfolgte den Ansatz, die Rollen von Studierenden und ChatGPT zu vertauschen (E1). Es wurde die zum Zeitpunkt der Studie vorhandene mathematische Schwäche [Zh23] von ChatGPT verwendet, damit Studierende die Perspektive eines Lehrenden einnehmen und versuchen konnten, anhand von Erklärungen mathematischer Formeln das Sprachmodell zu einer korrekten Lösung zu führen. Die Aufgabenstellung

setzte sich mit binären Bäumen, bspw. mit der Fragestellung nach der minimalen Höhe eines binären Baumes unter bestimmten Bedingungen, auseinander. Studierenden war Zugriff auf eine Musterlösung gewährt. Diese sollte nicht an ChatGPT übergeben werden, sondern durch iterative Anfragen auf die vorgeschlagene Lösung von ChatGPT Stellung zu nehmen und durch Hinweise zu unterstützen.

In der zweiten Aufgabe sollten Studierende einen vorgegebenen Prompt an ChatGPT versenden, der eine entsprechende zufällige Aufgabe generiert. Diese sollten die Studierenden anschließend lösen (E2). In der ersten Teilaufgabe (E2.1) sollte eine komplexe Prozedur mit bewusst eingebauten Fehlern generiert werden und Studierende hatten die Aufgabe diese Fehler zu finden. In der zweiten Teilaufgabe (E2.2) sollten fünf theoretische Fragen zu PL/SQL, mit einem Hinweislink versehen, von ChatGPT gestellt und von den Studierenden beantwortet werden. Die gestellten Fragen waren mehrheitlich nicht spezifisch, sondern allgemein gehalten und beschäftigten sich z.B. mit dem Unterschied zwischen einer Prozedur und einer Funktion in PL/SQL, oder wofür ein Cursor benötigt wird.

In der dritten Aufgabe sollten Studierende selbst eine Aufgabe konstruieren und lösen, ohne kopierfähige Informationen an ChatGPT weitergeben zu können (E3). Hierfür wurden den Studierenden visuelle Informationen in Form eines Bildes zur Verfügung gestellt, damit sie einen Trigger mithilfe der verfügbaren Informationen implementieren.⁶

5.1 Bewertung

51% der eingereichten Lösungen der B-Baum-Aufgabe (E1) wurden als korrekt deklariert. Der niedrige Anteil daran lässt sich damit erklären, dass die Aufgabe als sehr frustrierend empfunden wurde, da das generative Sprachmodell bspw. die Vorschläge zur Berechnung der korrekten Höhe teilweise ignorierte oder nicht umsetzte. Diese Aufgabe erforderte auch eine andere Herangehensweise von den Studierenden, da das Tool die richtige Lösung nicht kannte und andere Hilfsmittel als Werkzeuge mit eingebunden werden mussten. Nach Gesprächen mit Studierenden konnte festgestellt werden, dass die intensiven Trial-and-Error Versuche zu einer tiefgreifenden Auseinandersetzung mit der Thematik führten, aber der Nutzen den Frust nicht überwog.

74% der Studierenden reichten eine korrekte Lösung bei der zweiten Aufgabe (E2) ein. Nach Evaluation der ersten Teilaufgabe (E2.1) konnte festgestellt werden, dass die Qualität der von ChatGPT generierten Aufgaben sehr stark schwankte. Es wurden sowohl Aufgaben generiert, die mehrere Fehler angemessener Schwierigkeit beinhalteten, als auch Aufgaben,

⁶ Die von den Autoren initial überlegte Aufgabenstellung lautete wie folgt: Eine Kundin tätigt eine Bestellung bei einem Händler und wartet auf ihr Paket. In Abhängigkeit von der benötigten Lieferzeit in Tagen ergeben sich unterschiedliche Szenarien. Wenn die Zustellung des Pakets an die Kundin in unter zehn Tagen erfolgt, gilt das als erfolgreiche Zustellung, die keine weiteren Handlungen auslöst. Wenn das Paket länger als zehn, aber weniger als 20 Tage benötigt, bekommt die Kundin einen Gutschein/Rückerstattung im Wert von 20 Euro. Sollte die Lieferung länger als 20 Tage in Anspruch nehmen, wird der Kundin ein Gutschein/Rückerstattung im Wert von 50 Euro gewährt.

die keine, unlogische oder nur einen Fehler beinhalteten, der sich lediglich auf Tippfehler bezog. Es kam ebenfalls vermehrt vor, dass Fehler bereits beim Stellen der Aufgabe an die Studierenden aufgelöst und erklärt wurden. Bei der zweiten Teilaufgabe (E2.2) kam es vermehrt vor, dass die Hinweislinks fehlerhaft waren. Verfügbare Links entstammten offiziellen Quelle (bspw. Handbuch von Oracle). Auch bei dieser Teilaufgabe erläutert ChatGPT gelegentlich die Lösung bereits in der Fragestellung. Die gestellten Fragen waren mehrheitlich nicht spezifisch, sondern allgemein gehalten.

86% der eingereichten Lösungen zu Aufgabe 3 (E3) waren korrekt. 30% der Teilnehmer missverstanden die initiale Aufgabenstellung und konstruierten teilweise Trigger, die keine sinnvolle Semantik aufwiesen. Der hohe Anteil an korrekten Lösungen lässt sich dadurch erklären, dass eine sinnvolle Semantik kein Teil der Aufgabenstellung war, sondern, wie auch bei den übrigen Aufgaben, ein lauffähiger Code als Endergebnis gefordert wurde. Studierende zeigten bei dieser Aufgabenstellung mehrheitlich Kreativität bei der Gestaltung der Aufgabe. Studierende tendierten vermehrt dazu, Verbesserungen an dem generierten Code vorzunehmen. Vermutlich war das Fehlen einer konkreten Aufgabenstellung sowie die gegebene Freiheit bei der Generierung einer Lösung ein Motivator dafür, sich mit dem eigenen Code intensiver auseinanderzusetzen. In der Aufgabenstellung war zudem absichtlich ein Logikfehler eingebaut, denn es gab keine Fälle für genau zehn bzw. genau 20 Tage. Es sollte überprüft werden, ob diese entgegen der Aufgabenstellung von ChatGPT, oder den Studierenden berücksichtigt werden – beides war nicht der Fall.

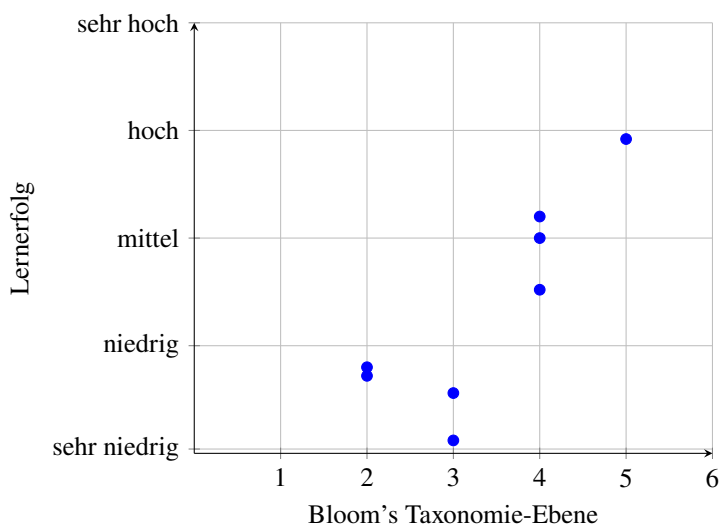
Die vorgestellten experimentellen Aufgaben unterstützen nach Evaluation der Studie das selbstständige Lernen bedingt. Das Ausnutzen der mathematischen Schwäche von ChatGPT gestaltete sich als schwierig, da die Studierende sich zwar einerseits tiefgreifend mit dem Thema mithilfe anderer Quellen auseinandersetzen mussten, um die Rolle eines Lehrenden einzunehmen. Andererseits aber löste dies bei den Studierenden eine hohe Frustration aus, da ChatGPT trotz richtiger Erklärungen und Vorschlägen zur Berechnung seitens der Studierenden Schwierigkeiten hatte, eine korrekte Lösung zu generieren. Diese Frustration scheint den Lernerfolg zu überwiegen, weshalb die Aufgabenstellung in dieser Form nicht zu Zwecken des selbstständigen Lernens empfohlen werden kann. Zudem ist es sehr wahrscheinlich, dass die mathematischen Fähigkeiten von ChatGPT sich mit der Einführung von Code Interpreter [Zh23] verbessern werden.

Als Verbesserung und Idee für weiterführende Studien könnte ChatGPT dazu aufgefordert werden, sich wie ein Studierender zu verhalten und Rückfragen zu einem bestimmten Sachverhalt zu stellen, der dann von den Studierenden erläutert werden muss. Die Bearbeitung von zufällig generiertem Content, wie in der zweiten Aufgabenstellung, wird als eine gute Möglichkeit für das selbstständige Lernen angesehen. Es ist zwar möglich, dass nicht gelehrt Inhalte in diesem Kontext abgefragt werden können, was jedoch durch eine Einschränkung in einer detaillierten Anfrage an das generative Sprachmodell umgangen werden kann. Obwohl beobachtet werden konnte, dass Lösungen teils direkt bei der Fragestellung übermittelt wurden, wird der Einfluss auf den Lernerfolg eher als gering betrachtet. Einerseits können die Informationen nach wie vor entnommen und verinnerlicht werden. Andererseits sollten die

Lösungen von ChatGPT nach wie vor evaluiert werden, da initial nicht von einer korrekten Antwort ausgegangen werden konnte. Ein Lernerfolg war erkennbar, wenn Studierende die Plausibilität der Lösungen eigenständig beurteilen mussten.

Die dritte Aufgabe, die sich mit der Entnahme visueller Informationen beschäftigte, kann auch als eine gute Möglichkeit für das selbstständige Lernen gelten. Studierende mussten die gegebenen Informationen verschriftlichen und eine Anfrage an das generative Sprachmodell stellen. Dabei mussten sie die Semantik der Aufgabe verstehen, um plausible Lösungen generieren zu lassen. Auch die Kreativität sowie selbstständig durchgeführte Verbesserungen an der Aufgabe sprechen für eine positive Einstellung der Studierenden bei der Bearbeitung. Allerdings sollte beachtet werden, dass die Interpretation der Aufgabenstellung von der initial intendierten Aufgabe stark abweichen kann, und es nicht die *eine* korrekte Lösung gibt. Zudem kann man davon ausgehen, dass das Auslesen eines Bildes mit ChatGPT zukünftig möglich sein wird, was den Lernerfolg durch die Möglichkeit des Kopierens und Einfügens der Aufgabenstellung in den Prompt stark einschränken würde.

6 Abhängigkeit Taxonomie – Lernerfolg



Die Ergebnisse der Fallstudie zeigen, dass höhere Ebenen der Bloom'schen Taxonomie tendenziell mit einem höheren Lernerfolg verbunden sind.⁷ Dies deutet darauf hin, dass

⁷ Bei einem unterdurchschnittlichen Lernerfolg überwiegen die Nachteile des Einsatzes von generativer KI, da eigene Kompetenzen durch gut generierte Inhalte nicht ausreichend abgefragt werden. Im Gegensatz dazu konnten bei der Einstufung in der oberen Diagrammhälfte Vorteile festgestellt werden, da die Integration des eigenen Fachwissens eine entscheidende Rolle spielt. Die dargestellten Werte sind eine Einschätzung der Autoren, basierend auf den im Laufe des Projekts gesammelten Daten.

Aufgaben, die komplexere kognitive Fähigkeiten erfordern und mit generativer KI gelöst werden, zu einem größeren Lernerfolg führen können. Dennoch ist der Lernerfolg auch stark abhängig von verschiedenen Faktoren wie der Art der Aufgabe, der Lernumgebung und den individuellen Lernstrategien der Studierenden, wie die gewisse Variabilität zeigt. Andere Faktoren, wie die spezifischen Lernziele der Aufgabe, die Motivation und das Engagement der Studierenden, sowie die Unterstützung und Ressourcen, die den Studierenden zur Verfügung stehen, spielen ebenfalls eine wichtige Rolle. Auch die Größe der betrachteten Aufgabenstellung wäre noch interessant, was in einem weiteren Paper aus diesem Lehrforschungsprojekt [BRW23] vertieft wird. Weitere Forschung ist hier notwendig, um einen möglichen Zusammenhang nachzuweisen. Weiteres Analysepotential liegt zum einen in einer möglichen Korrelation der Aufgabenart und dem Lernerfolg und zum anderen, inwiefern die Strategie der Studierenden den Lernerfolg beeinflusst.

7 Fazit

Die Studie konnte in einem Modul an einer deutschen Hochschule am Beispiel von ChatGPT zeigen, dass die Integration von generativen Sprachmodellen in den akademischen Lehrkontext das Potenzial hat, den Lernerfolg von Studierenden zu verbessern. Der Erfolg hängt jedoch stark von der Strategie und Motivation der Studierenden ab. Wenn Studierende sich dafür intrinsisch motiviert entscheiden, eine vorgeschlagene Lösung des generativen Sprachmodells zu evaluieren und nicht ohne Weiteres zu übernehmen, können Lernerfolge auch bei einfachen Fragestellungen gemessen werden.

Die Basisaufgaben zeigten, dass die Studierenden in der Lage waren, grundlegende Konzepte erfolgreich mit ChatGPT anzuwenden. Diese Aufgaben förderten jedoch nicht unbedingt das selbstständige Lernen, da ein einfaches Copy and paste möglich war und ein intensives Befassen mit der Aufgabenstellung bzw. der Thematik nicht nötig war.

Die vertiefenden Aufgaben erforderten trotz ChatGPT-Nutzung ein tieferes Verständnis der Thematik und förderten das selbstständige Lernen. Allerdings hing der Lernerfolg stark von der Strategie der Studierenden ab. Um einen wirklichen Mehrwert zu schaffen, ist es dafür in Zukunft sinnvoll, den Umgang mit LLMs sowie Prompt Engineering zu lehren.

Die experimentellen Aufgaben zeigten ebenfalls das Potenzial von generativen Sprachmodellen zur Förderung des selbstständigen Lernens auf. Auch wenn etwa das Rollenspiel zwischen ChatGPT und Studierenden zu Frustrationen führte, konnte durch das Generieren von zufälligem Inhalt und das Verarbeiten visueller Informationen ein Bereich gezeigt werden, in dem das selbstständige Lernen unterstützt werden kann.

Insgesamt zeigt die Studie an verschiedenen Beispielen, dass die Integration von generativen Sprachmodellen im akademischen Lehrkontext das Potenzial hat, den Lernerfolg zu verbessern. Allerdings ist weitere Forschung und Diskussion nötig, da es sich um einen auf den Hochschul- und Fachbereich beschränkten Erfahrungsbericht handelt. Weitere

Forschung ist notwendig, um praktische Leitlinien mit detaillierteren Empfehlungen zum Einsatz von KI-Werkzeugen in der Lehrpraxis zu erarbeiten, einschließlich möglicher Fallstricke und bewährter Verfahren.

Lehrende sollten Lernende dazu ermutigen, kritisch zu denken und selbstständig Lösungen zu finden, anstatt sich auf eine generierte Lösung zu verlassen. Erreicht werden könnte dies durch Aufgabenstellungen, die ein höheres Bloom-Level haben, ein tieferes Verständnis erfordern und durch das Lehren von Strategien, die Studierende dazu ermutigen, vorgeschlagene Lösungen kritisch zu hinterfragen, im Kontext einzuordnen und zu verbessern.

In Anbetracht der rasanten Fortschritte in der Künstlichen Intelligenz ist es unerlässlich, dass sich Lehrende nicht prinzipiell gegen diese Entwicklung stellen, sondern versuchen, sie zu beherrschen. Die Nutzung von Tools wie ChatGPT wird immer verbreiteter. Ein Verbot ist weder praktikabel noch wünschenswert. Die Lehre sollte sich darauf konzentrieren, Studierende darauf vorzubereiten, entsprechende Tools verantwortungsvoll zur Verbesserung ihres Lernens und zur Erweiterung ihrer Fähigkeiten nutzen zu können.

Literatur

- [AKB01] Anderson, L. W.; Kratwohl, D. R.; Bloom, B. S.: A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy. Longman, New York, 2001.
- [AL23] ArchiLab, 2023, URL: https://www.archi-lab.io/projects/ss23/gp_chatgpt_ss23.html, Stand: 17. 10. 2023.
- [BRW23] Bente, S.; Randall, N.; Wäckerle, D.: A Conceptual Framework to Transform Coding Education in Times of Generative AI, SEUH 2024, 2023.
- [HT23] Köln, H. T., 2023, URL: <https://hops.gm.th-koeln.de/hops/modules/modulelisting/module.php?mkz=1577>, Stand: 03. 11. 2023.
- [Ma02] Martin, J.-P.: Lernen durch Lehren (LdL). Die Schulleitung - Zeitschrift für pädagogische Führung und Fortbildung in Bayern/, 2002.
- [PG23] PARK, E.; GELLES-WATNICK, R.: Most Americans haven't used ChatGPT; few think it will have a major impact on their job. Pew Research Center/, 2023.
- [SAC23] College, S. A., 2023, URL: <https://sac.edu/AcademicAffairs/TracDat/Pages/Inquiry-Based-Learning-.aspx>, Stand: 12. 11. 2023.
- [St18] Stang, R.; Bernhard, C.; Kraus, K.; Schreiber-Barsch, S.: Lernräume in der Erwachsenenbildung. In (Tippelt, R.; von Hippel, A., Hrsg.): Handbuch Erwachsenenbildung/Weiterbildung. Springer Fachmedien Wiesbaden, Wiesbaden, S. 643–658, 2018.
- [Zh23] Zhou, A.; Wang, K.; Lu, Z.; Shi, W.; Luo, S.; Qin, Z.; Lu, S.; Jia, A.; Song, L.; Zhan, M.; Li, H.: Solving Challenging Math Word Problems Using GPT-4 Code Interpreter with Code-based Self-Verification. arXiv e-prints/, 2023.