Kann eine KI besser Unterricht planen als Lehramtsstudierende? – ein Bericht aus dem innovativen Lehrprojekt CUKI (Chemieunterricht geplant durch künstliche Intelligenz)

Michele Brott & Constantin Egerer

Im vergangenen Jahr war der Begriff „künstliche Intelligenz“ in aller Munde. Besonders der Ende 2022 veröffentlichte Chatbot „ChatGPT“ des Unternehmens OpenAI machte die KI-Technologie für jedermann zugänglich. Mit Sorge wurde diese Entwicklung besonders im Bildungssektor verfolgt. Doch anstatt dieses Tool aus der universitären Lehre zu verbannen, stellte das innovative Lehrprojekt „CUKI – Chemieunterricht geplant durch künstliche Intelligenz“ die KI ins Zentrum eines Seminars des Masterstudiengangs Lehramt Chemie und verglich insbesondere die Qualität von Unterrichtsplanungen von Studierenden mit denen des Chatbots.

Für Lehramtsstudierende stellt die Unterrichtsplanung erfahrungsgemäß eine große zeitliche und auch inhaltliche Herausforderung dar. Besonders bezüglich der Qualität der Unterrichtsplanung herrscht Heterogenität zwischen den Studierenden. Die Schwierigkeit besteht vor allem darin, allgemeine Konzepte wie beispielsweise die Merkmale von gutem Unterricht (z.B. Helmke, 2012) oder fachdidaktische Prinzipien in die eigene Planung einzubeziehen. Um die angehenden Chemielehrkräfte bei diesen Problemen zu unterstützen, wurde ein Seminarkonzept entwickelt, in dem Sequenzplanungen, Verlaufsplanungen und Unterrichtsmaterialien mithilfe des Tools ChatGPT erstellt und im Anschluss kritisch reflektiert wurden. Die Wirksamkeit des Seminars sowie die Ergebnisse der KI generierten Unterrichts–entwürfe wurden mit einer Begleitforschung evaluiert. Dazu stellte die Analyse von Verlaufsplanungen das zentrale Forschungsinstrument dar (Vgl. Abb.1). In der ersten Seminarsitzung erstellten die Studierenden zu einem vorgegebenen Thema eine eigene Stundenplanung ohne die Hilfe von KI, um einen Status Quo zu erheben. Im Verlauf des Seminars wurden mehrere Planungsaufgaben mit ChatGPT bearbeitet und schließlich eine Verlaufsplanung mit der KI erstellt. Zum letzten Seminartermin wurde dann wieder eine Stundenverlaufsplanung ohne die Hilfe der KI erstellt, um so einen Pre-Post-Vergleich in Bezug auf die Entwicklungen der Planungsqualität durch das Seminar ziehen zu können.

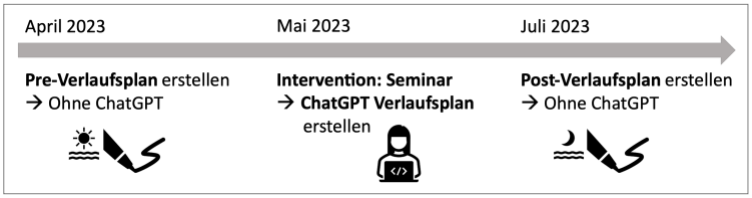

Abb. 1 Schematischer Ablauf der Begleitstudie

Die folgenden Forschungsfragen sollten mit dem vorstellten Design untersucht werden und erste Ansätze zur Beantwortung werden im Anschluss diskutiert.

1. Können die Studierenden durch die Beschäftigung mit ChatGPT ihre eigene Planungskompetenz steigern?
2. Sind die von ChatGPT generierten Planungen brauchbar, unterliegen sie ausgewählten Qualitätskriterien?
3. Hängt die Qualität dieser Stundenplanungen von der Planungskompetenz der jeweiligen Anwender:innen ab?
4. Welche Haltungen haben die Studierenden zu ChatGPT, wie bewerten sie den Einsatz im Seminar und wie bewerten sie selbst ihre Planungskompetenz?

1. Planungskompetenz der Studierenden im Pre-/Postvergleich

In einer Literaturrecherche wurde deutlich, dass bereits eine gute Forschungslage im Bereich der kriteriengestützen Analyse von Unterrichtsqualität vorhanden ist. Kriterien für die Planungsqualität von Unterricht sind eher rar. Aus den vorhandenen Arbeiten von Rothland (2022), König et al. (2021) und Großmann & Krüger (2022) wurden Redundanzen ausgeschlossen und ein Katalog von 17 Gütekriterien für eine qualitativ hochwertige Unterrichtsplanung aufgestellt, um Verlaufsplanungen zu bewerten. Dabei wurde für jedes Kriterium eine Intervallskala von 1 – 5 (nicht erfüllt – voll erfüllt) angewendet. Wir nehmen an, dass der Mittelwert der Skalenwerte Rückschlüsse auf die Planungskompetenz zulassen.
Auf dieser Grundlage wurden die Entwürfe, die die Studierenden (N=7) zu Beginn und am Ende des Seminars ohne ChatGPT entworfen haben, analysiert. Dabei fiel auf, dass alle Studierenden besser bewertete Unterrichtsentwürfe nach der Auseinandersetzung mit dem Seminarthema „Unterrichtsplanung mit ChatGPT“ hatten als zuvor. Insgesamt gab es einen Anstieg der gemittelten Skalenwerte von 3,26 auf 3,63. Das entspricht einer Zunahme um 11%. Besonders in den Bereichen „didaktische Reduktion“, „Lehrplanorientierung“ und „Anpassung an die Lerngruppe“ waren besonders große Zuwächse zu messen. Somit kann angenommen werden, dass die Beschäftigung mit ChatGPT in diesem Seminar die Planungskompetenz der Studierenden verbesserte.

2. Planungsqualität der mit ChatGPT generierten Unterrichtsplanungen

Besonderes Interesse bestand darin, die Qualität der Unterrichtsplanungen von ChatGPT zu analysieren. Dabei wurden die gleichen Kriterien angewendet, wie für die Bewertung der von den Studierenden selbst entwickelten Planungen. Grundlegend fiel auf, dass die KI-Planungen deutlich allgemeiner und weniger konkret formuliert waren, wodurch zwei Kriterien nicht valide bewertet werden konnten. Die Stärken des Chatbots lagen vor allem im Bereich der Stimmigkeit mit dem Lehrplan und des Vorhandenseins eines roten Fadens. In den Kriterien Lernziele, Differenzierungsmaßnahmen und Handlungsalternativen schnitten die Ergebnisse der KI eher schlecht ab. Die Planungen von ChatGPT erreichten durchschnittlich einen Skalenwert von 2,46 (um 32% niedriger im Vergleich zu den Post-Unterrichtsplanungen der Studierenden). Jedoch zeigt dies, dass die KI trotz des niedrigeren Skalenwerts eine basale Planungsgrundlage liefert und die beste Planung von ChatGPT sogar besser war als die schlechteste Planung der Studierenden.

3. Zusammenhang zwischen Planungskompetenz der Studierenden und Qualität der Unterrichtsentwürfe von ChatGPT

ChatGPT basiert auf einem sogenannten Large Language Model (LLM), wodurch der Output, den die KI generiert, maßgeblich von den Benutzereingaben (Prompts) beeinflusst wird. Je detailreicher und genauer ein Prompt formuliert ist, desto maßgeschneiderter ist die Antwort des Chatbots. Somit liegt die Vermutung nah, dass Studierende, die über eine größere Planungskompetenz verfügen, besser dazu in der Lage sind, eine vorgegebene Aufgabe zu analysieren und die wichtigsten Informationen an ChatGPT in Form von Prompts weiterzugeben als Studierende mit weniger Kompetenzen in der Unterrichtsplanung.
Um den vermuteten Zusammenhang zwischen Planungskompetenz und Qualität der KI-Planung zu überprüfen, wurde eine Korrelationsanalyse zwischen den Pre-Unterrichtsplanungen der Studierenden und den KI-Entwürfen durchgeführt. Dabei wurde

eine hohe Effektstärke (r=0,917) festgestellt, wodurch die Hypothese bestätigt wurde. Das bedeutet, dass die Studierenden mit hoher Planungskompetenz auch die besten KI-Entwürfe und die mit niedriger Planungskompetenz die schlechtesten Ergebnisse mit ChatGPT generierten.
In der Analyse der Prompts der Studierenden, die besonders gute Ergebnisse erzielten, ergaben sich einige Elemente, die zu einem guten Prompting zur Unterrichtsplanung beigetragen haben. Dazu gehören die Nennung zu implementierender Kontexte und Experimente, wichtiger didaktischer Rahmenbedingungen und Anforderungen wie Phasierung, Impulse, Vorwissen, Bundesland, Klassenstufe, Schulart und Zeitrahmen sowie formaler Vorgaben an die Unterrichtsplanung (z.B. Ausgabe als Tabelle).

4. Haltungen zu ChatGPT und dessen Planungskompetenz sowie Selbsteinschätzung

Eine Pre-/Postbefragung der Studierenden (N=8) im Sommersemester 2023 ergab, dass alle ChatGPT kannten und über 60% die KI bereits im privaten und universitären Bereich in Gebrauch hatten. Eine Person hatte den Chatbot auch bereits zur Unterrichtsplanung eingesetzt. Zu Beginn des Seminars war die Haltung bzgl. ChatGPT für den Einsatz zur Planung von Chemieunterricht noch eher skeptisch. Nach der Beschäftigung mit dem Tool im Seminar stiegen die Zustimmungswerte zur gewinnbringenden Anwendung. Außerdem schrieben die Studierenden der KI eine zufriedenstellende Erfüllung einiger ausgewählter Kriterien qualitätsvoller Unterrichtsplanung zu, wie z.B. roter Faden oder Handlungsorientierung. Die Studierenden wählten außerdem aus dem Kriterienkatalog, die Kriterien aus, die sie ihrer Ansicht nach selbst zufriedenstellend in ihren Planungen erfüllen. Dabei sich zeigte, dass sie selbst einen Zuwachs ihrer Planungskompetenz wahrnahmen.

Zusammenfassend lässt sich feststellen, dass ChatGPT die Planung von Lehrkräften (noch) nicht ersetzen kann. Jedoch überrascht, dass das Tool bereits nach weniger als einem Jahr nach Markteinführung bereits brauchbare Planungsgrundlagen liefert, die die Arbeit von Lehrkräften unterstützen können. Dabei sind die Ergebnisse der KI umso besser, je höher die Planungskompetenz der Bediener:innen ist.
Die Befragung und die Analyse der Unterrichtsentwürfe vor und nach dem Seminar zeigten, dass die Beschäftigung mit ChatGPT in einem Seminar des Masterstudiengangs des Lehramtes Chemie sowohl subjektiv als auch objektiv einen messbaren Kompetenzzuwachs bezüglich der Planung von Unterricht bei den Studierenden bewirkt hat. Aktuell werden in dem Modul „Forschung & Entwicklung in der Chemiedidaktik“ KI-generierte Unterrichtsentwürfe in realen Unterrichtssituationen erprobt und ausgewertet.

Großmann, L., & Krüger, D. (2022). Welche Rolle spielt das fachdidaktische Wissen von Biologie-Referendar*innen für die Qualität ihrer Unterrichtsentwürfe? *Zeitschrift für Didaktik der Naturwissenschaften*, *28*(1), 4. https://doi.org/10.1007/s40573-022-00141-w

Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Franz Emanuel Weinert gewidmet. 4. Aufl.* (F. E. Weinert, Hrsg.). Kallmeyer.

König, J., Krepf, M., Bremerich-Vos, A., & Buchholtz, C. (2021). Meeting Cognitive Demands of Lesson Planning: Introducing the CODE-PLAN Model to Describe and Analyze Teachers' Planning Competence. *The Teacher Educator*, *56*(4), 466–487. https://doi.org/10.1080/08878730.2021.1938324

Rothland, M. (2022). Anmerkungen zur Modellierung und Operationalisierung (allgemeindidaktischer) Unterrichtsplanungskompetenz. *Unterrichtswissenschaft*, *50*(3), 347–372. https://doi.org/10.1007/s42010-021-00111-0