

HERRMANN, Marc & DILLING, Frederik
Siegen

Beweisen mit ChatGPT in der Geometrie? Fallstudie zur selbstständigen Beweisentwicklung von Studierenden

Einleitung

Im Rahmen der Hochschulmathematik werden Beweise von Studierenden häufig nur in einer rezipierenden Rolle nachvollzogen, was dem Charakter vieler Mathematik-Vorlesungen geschuldet ist. Hierbei nehmen die Studierenden die Mathematik als ein fertiges Produkt wahr, statt diese aktiv zu entdecken und zu entwickeln (Winter, 2016). Bei der eigenständigen Beweisführung in Übungsaufgaben kommt es nach unserer Erfahrung häufig vor, dass anstelle einer eigenen Auseinandersetzung und möglicherweise nur teilweise korrekten Lösung, Beweise aus dem Internet oder von Mitstudierenden unreflektiert übernommen werden, um hier eine potenziell korrektere Lösung abgeben zu können, insbesondere wenn die Punkte für Übungsaufgaben an eine Klausurzulassung gekoppelt sind. Um das Beweisverständnis von Studierenden durch eine aktive Auseinandersetzung zu fördern, wurden in der Hochschuldidaktik verschiedene Ansätze wie zum Beispiel die Mid-Proof Peer Instruction entwickelt (Bauer et al., 2022). In diesem Beitrag wird der Einbezug eines digitalen Werkzeugs als Hilfsmittel für Studierende untersucht. Im Rahmen einer Studie wird der Frage nachgegangen, inwieweit die bewusste Verwendung der generativen künstlichen Intelligenz ChatGPT als Hilfsmittel, zu einer anderen, insb. aktiveren Auseinandersetzung mit den Beweisen führt.

Methodik

Zur Beantwortung dieser Frage wurde im Rahmen einer Vorlesung zur Geometrie für das Lehramt in der Primarstufe und der Sekundarstufe I an der Universität Siegen im Wintersemester 2023/24 eine Fallstudie durchgeführt. Die Studierenden mussten im Rahmen der Veranstaltung mehrere Übungsblätter bearbeiten und eine gewisse Anzahl von Punkten sammeln, um eine Zulassung zur Abschlussklausur zu erwerben. Im Rahmen des Themengebietes zu Dreiecken werden in der Veranstaltung klassischerweise mehrere Sätze zu Dreiecken in der Vorlesung bewiesen, zu welchen die Studierenden dann vertiefende Übungsaufgaben durchführen. Für diese Studie wurden stattdessen zwei der zentralen Beweise des Themengebietes, nämlich jene zur Innenwinkelsumme im Dreieck und zum Basiswinkelsatz in gleichschenkligen Dreiecken nicht in der Vorlesung bewiesen, sondern sollten von den Studierenden als ein Übungsblatt selbstständig bewiesen werden.

Hierbei wurde den Studierenden kommuniziert, dass sie die Beweise nicht

In: P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht & P. Scherer (Hrsg.),

Beiträge zum Mathematikunterricht 2024.

57. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.

<https://doi.org/10.37626/GA9783959872782.0>

allein herleiten müssen, sondern aktiv ChatGPT als Werkzeug zur Erstellung der Beweise nutzen sollen. Hierbei würden für eine vollständige Bearbeitung der beiden Beweise die vollen Punkte für das Übungsblatt vergeben, unabhängig von der mathematischen Korrektheit. Mit dieser Regelung sollte vermieden werden, dass die Studierenden einfach die korrekten Beweise im Internet oder der Literatur nachschauen und unreflektiert übernehmen, sondern sich stattdessen selbstständig mit diesen beschäftigen. Neben den beiden Beweisen sollten die Studierenden zudem ihre Chatprotokolle und einen Fragebogen einreichen. In diesem wurden die Vorerfahrungen mit ChatGPT, die Vorstellungen zu dessen Funktionsweise, die Zufriedenheit mit dessen Antworten, die Einschätzungen zur Korrektheit der eingereichten Beweise und die abschließende Reflexion der Studierenden über den Einsatz von ChatGPT für das Beweisen in dieser Vorlesung erhoben.

Ergebnisse

Von den Studierenden haben $n=129$ an der Fallstudie teilgenommen und jeweils Beweise und Fragebögen eingereicht. Während die Fragebögen von allen Studierenden einzeln beantwortet wurden, wurden die Beweise in Teilen in Gruppen abgegeben, sodass die Gesamtzahl der Beweise geringer ausfällt. Langfristig ist eine systematische qualitative Auswertung aller Fälle im Vergleich vorgesehen, um fallübergreifende Aussagen treffen zu können. Im Rahmen dieses Beitrages werden ausgewählte Fälle im Detail betrachtet, um erste Ergebnisse der Studie sichtbar zu machen.

Darstellung Fall 1

Als ersten Fall betrachten wir den Studierenden A. Dieser gibt bei den Vorerfahrungen an, bereits mehrfach mit ChatGPT gearbeitet und dieses primär zur Zusammenfassung von Texten verwendet zu haben. Im Fall des ersten Beweises zum Innenwinkelsatz im Dreieck, fragt er ChatGPT zuerst nach möglichen Axiomen und Sätzen, welche für den Beweis relevant sein könnten und lässt sich diese nachfolgend mit mehreren Nachfragen erläutern. Hierbei benennt ChatGPT alle relevanten Sätze und Axiome für den Beweis der Innenwinkelsumme durch die Konstruktion einer geeigneten Parallele, was einen der möglichen Beweise darstellt. A hat nun jeweils eigene Ideen zur Nutzung der Sätze für den Beweis und befragt ChatGPT jeweils zu diesen. ChatGPT liefert hier auch darüberhinausgehende Hinweise zur Struktur des Beweises, sodass A diesen vollständig und richtig führt. ChatGPT wird hier in der Funktion eines Tutors verwendet, welcher Hilfestellungen gibt und die Korrektheit der eigenen Ideen des Studierenden beurteilt. Beim zweiten Beweis zu den Basiswinkeln im gleichschenkeligen Dreieck wurde der zu beweisende Satz in folgender Form gestellt:

Ein Dreieck ABC ist gleichschenkelig mit $AC=BC$ genau dann, wenn $\alpha = \beta$ gilt.

Für den Beweis bedeutet dies, dass beide Implikationen gezeigt werden müssen. ChatGPT interpretiert den Satz auf Nachfrage des Studierenden so, dass sowohl die Winkelgleichheit als auch die Seitengleichheit gegeben sind und hieraus zu folgern sei, dass das Dreieck gleichschenkelig ist. Auf diese Antwort hin fragt A nach, ob nicht die Gleichheit der Seiten die Gleichheit der Winkel voraussetzen würde. ChatGPT korrigiert darauf seine Antwort und behauptet, dass es tatsächlich umgekehrt wäre und die Gleichheit der Winkel die Gleichheit der Seiten voraussetze. ChatGPT bezieht sich hier auf eine übliche Definition, bei welcher nur die Implikation, dass aus der Gleichheit der Seiten die Gleichheit der Winkel folgt als ein Satz gegeben ist und ignoriert, dass der Studierende eigentlich beide Implikationen zeigen müsste. Der Studierende fragt daraufhin nach, ob sein Ansatz, das Dreieck mithilfe von einer Mittelsenkrechten zu teilen und über die Teildreiecke zu argumentieren geeignet ist. ChatGPT bejaht dies und empfiehlt mehrere Schritte eines Beweises, die in sich nicht logisch sind. Der Studierende fragt nicht weiter nach und übernimmt die Schritte des Beweises weitestgehend, versucht hierbei aber die logischen Fehler anzupassen. Hierbei schafft er es nicht, diese vollständig zu beheben und gibt letztendlich einen falschen Beweis ab. Es zeigt sich, dass der Studierende zwar produktiv mit ChatGPT umgeht, die Qualität des letztendlichen Beweises durch seine Fähigkeit, die Korrektheit der Lösungen einzuschätzen, jedoch limitiert ist. Dies wird auch daran deutlich, dass der Studierende in der Reflexionsfrage einschätzt, dass beide seiner Beweise vermutlich richtig seien.

Darstellung Fall 2

Der zweite Fall behandelt die Studierende B. Diese gab bei den Fragen an, ChatGPT in der Vergangenheit für das Zusammenfassen von Texten, das Umformulieren von Sätzen und die Konvertierung von Stichpunkten zu Fließtext verwendet zu haben. In beiden Fällen hat die Studierende die Beweise von ChatGPT direkt übernommen (Screenshots der Chats) und nur im ersten Fall mit einer Skizze versehen. Für den Innenwinkelsatz wurde innerhalb der ersten Schritte des ChatGPT-Beweises eine Konstruktion erklärt, welche von der Studierenden als Skizze zur Veranschaulichung ebenfalls durchgeführt wurde. Auf dieser Skizze aufbauend argumentiert ChatGPT mit einem Zirkelschluss und weiteren unlogischen Schritten, welche keinen sinnvollen Beweis darstellen. Dies bleibt von B unkommentiert. Im Fall des zweiten Satzes bleibt eine Skizze völlig aus, hier reicht die Studierende nur die Screenshots des Chats ein. Sichtbar ist auch die Rückfrage nach dem Beweis, ob dies als ein Beweis ausreiche. Nachdem ChatGPT dies bejaht, folgt

keine weitere Nachfrage. Während ChatGPT hier die Behauptung des Beweises zumindest auf die Frage nach dem Basiswinkelsatz korrekt wiedergibt, also die Notwendigkeit des Nachweises der Äquivalenz erkennt, ist der darauffolgende Beweis nur wenig sinnvoll. So entstehen ohne weitere Konstruktion plötzlich mehrere weitere Dreiecke, die sogar gleichseitig sind und Implikationen für die Winkel im gleichschenkligen Dreieck mit sich bringen. Auch dieser offensichtliche logische Fehlschluss bleibt von der Studierenden unkommentiert. Die Studierende geht ihren Antworten im Fragebogen nach auch davon aus, dass beide Beweise grundsätzlich richtig sind und nur etwas formaler notiert werden müssten. Die logischen Fehler bleiben unbeachtet.

Fazit

Im Vergleich der beiden Studierenden zeigt sich, dass A einen produktiven Umgang mit ChatGPT als eine Art Tutor pflegt, während B die Lösungen unreflektiert und ohne tiefere Auseinandersetzung mit diesen übernimmt. In beiden Fällen muss festgestellt werden, dass den Studierenden die Kompetenz fehlt, falsche Lösungen von ChatGPT korrekt einzuordnen, sodass diese teilweise falsch übernommen werden. Die zum Teil logisch nicht korrekten Schlüsse von ChatGPT stellen somit eine Herausforderung dar (Schorcht et al., 2023), können aber unter Umständen positiv gewendet werden.

Aus einer ersten Kodierung der Fragebögen zeigt sich allgemein, dass ein signifikanter Anteil der Studierenden davon ausgeht, ChatGPT sei eine Suchmaschine, welche die Antworten auf Basis von Suchanfragen im Internet generiert. Aus diesem Grund haben die Studierenden ein hohes Vertrauen in die Korrektheit der generierten Beweise, was vermutlich auch mit dem Problem der beiden oben dargestellten Fälle in Zusammenhang steht. Grundsätzlich kann es aus unserer Sicht für das Beweisverständnis produktiv sein, Impulse von ChatGPT zum Beweisen zu nutzen oder logische Fehler von generierten Beweisen zu identifizieren. Hierzu ist aber ein Grundverständnis zu Beweisen und der Funktionsweise generativer künstlicher Intelligenz notwendig. Eine tiefere Analyse und weitere Erhebungen sollen die Potenziale und Herausforderungen in diesem Zusammenhang genauer aufzeigen.

Literatur

- Bauer, T., Müller-Hill, E., Neuhaus-Eckhardt, S. & Rach, S. (2022). Beweisverständnis im Mathematikstudium unterstützen: Vergleich unterschiedlicher Varianten der Strategie „Illustrieren am Beispiel“. *Journal für Mathematik-Didaktik*, 43, 311–346.
- Schorcht, S., Baumanns, L., Buchholtz, N., Huget, J., Peters, F. & Pohl, M. (2023). Ask to Get Smart: Mathematische Ausgaben generativer KI-Sprachmodelle verbessern durch gezieltes Prompt Engineering. *Mitteilungen der GDM*, (115), 12–23.
- Winter, H. (2016). *Entdeckendes Lernen im Mathematikunterricht. Einblicke in die Ideengeschichte und ihre Bedeutung für die Pädagogik*. Springer.