

# **"To Bot or Not to Bot? Potenziale und Grenzen KI-gestützter Lernbegleiter in der Hochschullehre"**

## **1. Einleitung**

### **1.1 Der Chatbot in der Lehre - Hype oder Hoffnungsträger?**

Künstliche Intelligenz in der Lehre polarisiert. Auf der einen Seite steht die Euphorie: KI-Chat-bots sind aus dieser Perspektive Lernbegleiter, die rund um die Uhr verfügbar sind, individuell auf Fragen reagieren und Lehrende entlasten. Auf der anderen Seite fragen Skeptiker: Produzieren sie nur oberflächliche Antworten? Fördern sie Denkfaulheit? Ersetzen sie echte Interaktion durch algorithmische Scheinlösungen?

Der provokante Titel „To Bot or Not to Bot?“ trifft also den Kern einer Diskussion, die derzeit Hochschulen umtreibt. Large Language Models (LLMs) wie ChatGPT haben seit 2022 einen Innovationsschub ausgelöst, der Lehrende und Studierende gleichermaßen herausfordert. Während manche Hochschulen KI-Nutzung weitgehend verbieten, experimentieren andere mit Bot-gestützten Lernangeboten.

KI trifft dabei auf eine Hochschullehre im Spannungsfeld bzw. könnte verschiedene aktuelle Herausforderungen aufgreifen: Massifizierung, Digitalisierung, KI-Kompetenzen der Studierenden.

Die Corona-Pandemie hat Online-Lehre von der Notlösung zur Normalität gemacht. Besonders große Vorlesungen stehen damit aber vor strukturellen Problemen: Geringe Interaktion, anonyme Teilnehmermassen, Lehrende, die ins „schwarze Loch“ sprechen. Gleichzeitig steigen die Erwartungen: Studierende fordern flexible Lernwege, schnelle Rückmeldungen, niedrigschwellige Unterstützung - rund um die Uhr. KI-Chatbots versprechen hier eine Antwort. Sie können organisatorische Routinefragen übernehmen, Lernmaterial kontextualisieren oder als Sparring-Partner für Verständnisfragen dienen.

Was bislang fehlt ist empirische Evidenz aus dem komplexen realen Lehralltag: Welche Mehrwerte entstehen tatsächlich? Wo liegen die Grenzen? Und vor allem: Wie müssen solche Systeme didaktisch eingebettet werden? Der Beitrag greift genau diese Forschungslücke bzw. Fragen auf.

### **1.2 Forschungsfragen und Ziel dieses Beitrags**

Der Beitrag verfolgt die übergreifende Frage: In welcher Weise kann Künstliche Intelligenz genutzt werden, um personalisierte Lernpfade zu entwickeln und die Studierenden besser zu unterstützen? Dazu berichten wir Ergebnisse aus der wissenschaftlich begleiteten Implementation des KI-gestützten Lernbegleiters Nýt in einer Lehrveranstaltung an der FAU Erlangen-Nürnberg über ein Semester hinweg.

Unsere Teil-Forschungsfragen lauteten:

1. Welche Mehrwerte liefert ein KI-Chatbot als Lernbegleiter in großen Online-Vorlesungen?
2. Wo liegen seine Grenzen – technisch, didaktisch, interaktiv?
3. Welche Design-Prinzipien lassen sich für eine zukünftige erfolgreiche Bot-Integration in Hochschullehre ableiten?

Um Design-Prinzipien praxisorientiert fundiert ableiten zu können und eine Basis für eine nachhaltige Verankerung in der Hochschullehre zu schaffen, die von den Beteiligten als sinnvoll, machbar und bereichernd mitgetragen wird, wurden die Perspektiven aller Akteure aufgegriffen und trianguliert. Dabei wurden verschiedene Fragestellungen je Akteur aufgegriffen:

- Studierende: Wie wird der Chatbot tatsächlich genutzt? Welche Erwartungen werden erfüllt/enttäuscht?
- Lehrende: Welche Entlastung/Mehrbelastung entsteht? Verändert sich die Interaktivität?
- Entwicklerteam: Welche technischen und konzeptionellen Herausforderungen traten auf?

Damit verbunden verfolgt der Beitrag zwei **Ziele**. Einerseits geht es darum, **empirisch** Mehrwerte von KI-ChatBots zu zeigen, andererseits formulieren wir durch die Design-Prinzipien auch die **konzeptuelle Rahmung**, d.h. klare Bedingungen für eine zukünftige sinnvolle Integration in die praktische Hochschullehre.

## 2. Theoretischer Rahmen & Stand der Forschung

### 2.1 Conversational AI in der Bildung: Vom FAQ-Bot zum Lernbegleiter

Der Einsatz von Chatbots in der Bildung ist kein neues Phänomen. Bereits vor dem Durchbruch großer Sprachmodelle (LLMs) wie ChatGPT experimentierten Hochschulen mit regelbasierten Chatbots, die primär organisatorische Routinefragen beantworteten (Winkler & Söllner, 2018). Diese frühen Systeme funktionierten nach dem Prinzip des Pattern Matching: Sie erkannten Schlüsselwörter in Nutzeranfragen und lieferten vordefinierte Antworten. Ihre Grenzen waren offensichtlich – sobald Fragen komplexer wurden oder vom Skript abwichen, versagten sie. Mit dem Aufkommen von Large Language Models ab 2022 hat sich das Feld fundamental verändert. LLMs wie GPT-4, Claude oder Gemini können natürlichsprachliche Dialoge führen, Kontext über mehrere Gesprächsrunden hinweg behalten und auf offene Fragen reagieren, ohne auf vordefinierte Antwortmuster angewiesen zu sein (OpenAI, 2023). Diese technologische Transformation hat das Einsatzspektrum erweitert: Chatbots werden nicht mehr nur als FAQ-Maschinen eingesetzt, sondern als Lernbegleiter, die inhaltliche Verständnisfragen

beantworten, Reflexion anregen und individuell auf Lernende eingehen können. Einsatzgebiete sind das Sprachlernen (Fryer et al., 2020), Mathematik-Tutoring (Deng & Yu, 2023), Programmierunterricht (Prather et al., 2024) und organisatorische Unterstützung (NASPA, 2024; Winkler & Söllner, 2018).

Bisherige Studien zu LLM-basierten Chatbots in der Bildung zeigen dabei strukturelle Lücken. So untersuchen die meisten Studien Bot-Einsatz nur über wenige Wochen hinweg, oft in kontrollierten Laborumgebungen (Kuhail et al. 2023), nicht in realen Lehrveranstaltungen über ein Semester. Die überwiegende Mehrheit der Forschung erhebt ausschließlich die Sicht der Lernenden. Die Perspektiven von Lehrenden und Entwicklern bleiben unterbelichtet. Was also fehlt, ist insbesondere eine systematische Triangulation aus den drei Perspektiven über einen längeren Zeitraum hinweg - genau diese Lücke adressiert die vorliegende Studie.

## **2.2 Lerntheoretischer Rahmen: Warum Chatbots didaktisch Sinn machen können**

Der Einsatz von KI-Chatbots als Lernbegleiter lässt sich lerntheoretisch begründen. Drei zentrale Perspektiven legitimieren ihren didaktischen Mehrwert:

**Constructivism:** Aus konstruktivistischer Sicht ist Lernen kein passiver Wissenserwerb, sondern ein aktiver Prozess, bei dem Lernende neues Wissen mit bestehendem Vorwissen verknüpfen und eigene mentale Modelle aufbauen (von Glasersfeld, 1989 und 1995). Chatbots können diesen Prozess unterstützen, indem sie als Dialog-Partner fungieren. Sie beantworten nicht nur Fragen, sondern stellen Rückfragen, fordern zur Reflexion auf und regen zur Auseinandersetzung mit dem Stoff an. Die entscheidende Bedingung: Ein Bot muss so instruiert sein, dass er nicht direkt Lösungen liefert, sondern den Weg zur Lösung begleitend unterstützt.

**Self-Regulated Learning:** Selbstreguliertes Lernen beschreibt die Fähigkeit, den eigenen Lernprozess zu planen, zu überwachen und zu reflektieren (z.B. Zimmerman, 2002). Chatbots können diese metakognitiven Prozesse fördern, indem sie Studierende dazu anregen, ihren Lernstand zu reflektieren: Wo stehe ich? Was habe ich verstanden? Wo brauche ich Vertiefung? Die 24/7-Verfügbarkeit eines Bots ermöglicht es Lernenden, jederzeit Unterstützung abzurufen - unabhängig von Sprechstunden oder Verfügbarkeit von Mitstudierenden. Diese Flexibilität kann Selbststeuerung fördern, birgt aber auch das Risiko der Bequemlichkeit.

**Zone of Proximal Development:** Vygotskys Konzept der Zone of Proximal Development (ZPD) beschreibt die Differenz zwischen dem, was Lernende allein bewältigen können, und dem, was sie mit Unterstützung erreichen (Vygotsky, 1978). Ein gut konfigurierter Chatbot kann als Scaffolding-Instrument dienen: Er bietet Hilfestellungen, die genau im richtigen Maß dosiert sind - nicht zu wenig (Überforderung), nicht zu viel (keine Herausforderung). Die Vision personalisierter Bot-Instanzen, die sich an den individuellen Lernstand

anpassen, würde dieses Prinzip auf die Spitze treiben: Jeder Studierende erhält genau die Unterstützung, die er oder sie in diesem Moment braucht.

Diese lerntheoretischen Begründungen zeigen: Chatbots können didaktisch wertvoll sein – aber nur, wenn sie gezielt instruiert und in ein didaktisches Konzept eingebettet werden. Ohne diese Rahmung besteht das Risiko, dass sie zu reinen Lösungsautomaten verkommen.

### **2.3 Chancen & Risiken des Einsatzes von AI aus der Literatur**

Die bisherige Forschung zu Conversational AI in der Bildung zeichnet ein ambivalentes Bild. Auf der einen Seite stehen vielversprechende Potenziale, auf der anderen Seite strukturelle Risiken, die nicht ignoriert werden dürfen.

**Skalierbarkeit:** Chatbots können theoretisch unbegrenzt viele Studierende gleichzeitig betreuen – eine Dimension, die klassische Tutorate oder Sprechstunden nicht bieten können (Kuhail et al., 2023; Kasneci et al., 2023). Besonders in Massenveranstaltungen, wo Betreuungsrelationen ungünstig sind, können Bots Lücken schließen.

**Personalisierung:** Während traditionelle Vorlesungen auf die Masse zugeschnitten sind, können Bots individuell auf Fragen eingehen, Lernstand berücksichtigen und adaptive Hinweise geben (Holmes & Porayska-Pomsta, 2024)

**24/7-Verfügbarkeit:** Lernende sind nicht mehr an Sprechstunden gebunden. Wer nachts um 2 Uhr eine Verständnisfrage hat, kann sofort Unterstützung abrufen (Markel & Opfermann, 2021; NASPA, 2024). Niedrige Hemmschwelle: Studierende berichten, dass sie einem Bot eher "dumme Fragen" stellen als Dozierenden oder Peers – die Angst vor Blamage entfällt (Cooper et al., 2018)

**Oberflächlichkeit:** LLMs können plausibel klingende, aber falsche oder ungenaue Antworten geben (Halluzinationen) (Ji et al., 2023). Studierende, die keine kritische Reflexionskompetenz entwickeln, übernehmen diese Antworten unkritisch.

**Abhängigkeit und Bequemlichkeit:** Wenn ein Bot jederzeit sofort Lösungen liefert, besteht das Risiko der Instant Gratification – Studierende könnten verlernen, mit Unsicherheit umzugehen und selbst nachzudenken (Fan et al., 2025; Yang et al., 2024; Zhang et al., 2024).

**Datenschutz:** Chat-Verläufe enthalten sensible Informationen über Lernstände, Wissenslücken, persönliche Fragen. Wer hat Zugang zu diesen Daten? Wie werden sie geschützt? Diese Fragen sind bislang ungeklärt (Schiff, 2024).

**Equity:** Wenn Premium-Bots kostenpflichtig sind, entstehen neue Ungleichheiten zwischen Studierenden mit und ohne finanziellen Zugang. Hochschulen müssen sicherstellen, dass KI-Unterstützung allen zur Verfügung steht (Sullivan & Wilson, 2024).

**Verlust menschlicher Interaktion:** Wenn Bots die primäre Anlaufstelle für Fragen werden, könnte der direkte Austausch zwischen Studierenden und Lehrenden zurückgehen – mit

negativen Folgen für Mentoring, Netzwerkbildung und soziale Einbettung des Lernens (Al-Zahrani, 2025; Rhodes, 2024).

Diese Chancen und Risiken sind nicht deterministisch. Sie hängen davon ab, wie Chatbots eingesetzt werden. Hier setzt die vorgestellte Studie an, indem sie aus einer Bot-Erprobung und Evaluation abgeleitete Design-Prinzipien für den erfolgreichen Bot-Einsatz vorstellt. Mit dieser Ausrichtung positioniert sich der Beitrag als empirisch fundierte Bestandsaufnahme aus dem realen Lehralltag – mit dem Ziel, anderen Hochschulen praxisnahe Orientierung zu bieten.

### **3. Methodisches Vorgehen**

#### **3.1 Studiendesign**

Die Entwicklung, Implementation und Evaluation des Bots war eingebunden in das „methodologische Rahmenkonzept“ (Reinmann, 2022, S.2) der Design-Based-Research, die es sich zum Ziel setzt, durch eine kontinuierliche und mehrperspektivische Datensammlung Probleme der Implementation frühzeitig zu erkennen, die Nutzerinnen als aktive Beteiligte einzubinden (Reinmann, 2022, S.3), Design-Prinzipien über das Semester hinweg zu eruieren (Euler, 2017) sowie den Prototypen des Bots soweit als möglich den semesterbegleitenden Evaluationsergebnissen kontinuierlich anzupassen.

DBR verfolgt im erziehungswissenschaftlichen Kontext das Ziel, ausgehend von einem Praxisproblem (s. Kap. 2), einerseits forschungsbasiert, innovative Lösungen (didaktische Interventionen) zu entwickeln und zugleich didaktische Gestaltungsprinzipien zu liefern, die sich auf weitere (vergleichbare) Bereiche übertragen lassen (theoretischer Erkenntnisgewinn) (Euler 2014). Das Forschungsdesign eignete sich deshalb hervorragend für das skizzierte Projekt, da es darum ging, einen komplett neuen Chatbot für eine konkrete Lehrveranstaltung spezifiziert aufzusetzen, zu implementieren und über die Evaluation auch Erfahrungen zu sammeln, die für andere Lehrveranstaltungen wichtig sind, die ebenfalls KI einbinden möchten.

#### **3.2 Studienkontext/Lehrveranstaltung**

Die Studie fand im SS 2025 an der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) statt. Implementations- und Untersuchungskontext war der 5 ECTS Kurs "Nailing Your Thesis" (NYT) unter der Leitung von Prof. Dr. Dirk Riehle (Professur für Open Source Software). Der Kurs vermittelt die Fähigkeit zu wissenschaftlichem Arbeiten mit Fokus auf Forschung in den Ingenieurwissenschaften und die Anwendung dieses Wissens in wissenschaftlichen Bachelor- und Masterarbeiten. In Daten:

- 381 initiale (registrierte) Teilnehmende
- 249 Teilnehmende, die abgeschlossen haben
- 80 Teilnehmende durchschnittlich in den Online Vorlesungen

Der Kurs besteht aus wöchentlichen Vorlesungen (90min.) und Tutorien (90min.). Teilnehmende bereiten die Vorlesung mittels Lesen eines Artikels vor. Zur Fortschrittskontrolle werden wöchentliche Wissens-Quizze auf Basis der letzten Vorlesung eingesetzt. Geprüft wird mittels der Quizze sowie einer begleitenden wissenschaftlichen Methoden-Übung, welche einer Seminararbeit vergleichbar ist.

### **3.3 Der Chatbot: Nýt**

Nýt (ausgesprochen "nite", angelehnt an "Nailing Your Thesis") ist ein LLM-basierter Chatbot, entwickelt vom Münchner Startup Unwritten in Kooperation mit der FAU. Technische Grundlage ist Claude (Anthropic) in Verbindung mit einer Retrieval-Augmented Generation (RAG)-Architektur. RAG ermöglicht es, externe Wissensdatenbanken - in diesem Fall vorlesungsspezifisches Material - kontextuell in die Generierung von Antworten einzubinden, ohne das gesamte Material im Prompt unterbringen zu müssen.

Materialgrundlage: Nýt wurde mit folgenden Inhalten „gefüttert“:

- Vorlesungsskripte und Folien
- Transkripte aufgezeichneter Vorlesungssitzungen
- Mitschriften zur Lehrveranstaltung
- Organisatorische Informationen (Termine, Deadlines, Prüfungsmodalitäten)
- Übungsaufgaben und Musterlösungen

Das Material wurde strukturiert und in semantische "Häppchen" aufbereitet, um das begrenzte Kontextfenster von LLMs optimal zu nutzen.

Die Persönlichkeit von Nýt orientierte sich an Sheldon Cooper aus der TV-Serie "The Big Bang Theory": präzise, sachlich, gelegentlich nerdig, aber niemals herablassend. Diese Tonalität sollte einerseits fachliche Kompetenz signalisieren, andererseits eine niedrighschwellige, nicht-einschüchternde Atmosphäre schaffen.

Nýt hatte drei primäre Einsatzfelder bzw. Funktionen: 1. Beantwortung organisatorischer und inhaltlicher Routinefragen der Studierenden. 2. Lernbegleiter, insbesondere Unterstützung bei der Vor- und Nachbereitung von Sitzungen, Reflexion des Lernprozesses, Prüfungsvorbereitung. 3. Sparring-Partner: Interaktion mit dem Dozenten während der Vorlesung (z.B. Diskussion kontroverser Thesen, Einwerfen von Zwischenfragen). Nýt war über eine eigene Weboberfläche zugänglich und konnte sowohl während als auch außerhalb der Vorlesungszeiten genutzt werden.

### **3.4 Datenerhebung**

Um die Mehrwerte und Grenzen KI-gestützter Lernbegleiter empirisch zu erfassen, wählten wir - wie oben bereits skizziert - einen Mixed-Methods-Ansatz - unter Einbindung der drei Perspektiven Studierende, Lehrende und Entwickler. Um die Triangulation zu realisieren, kombinierten wir quantitative und qualitative Erhebungsmethoden zu

verschiedenen Zeitpunkten, welche die verschiedenen Akteure sowie unterschiedliche Erhebungsziele einbezogen:

- Vorab-Fragebogen an Studierende (April 2025)

Über einen Fragebogen zu Beginn des Semesters wurden Charakteristika der Teilnehmenden (AI-Vorerfahrung, präferiertes Lernverhalten, Technikaffinität), Erwartungen an Nj̈t (z.B. Mehrwerte fürs Studium, kognitive Be-/Entlastung) sowie eine Selbsteinschätzung der AI-Kompetenz (z.B. Prompt-Formulierung, kritische Reflexion von Bot-Antworten) abgefragt. Zugleich fand eine Bewertung der Einführung von Nj̈t in der Kick-Off-Einheit der Lehrveranstaltung statt.

- Interims-Votings (Mai & Juni 2025)

In 2 Interims-Votings während der Lehrveranstaltung wurden über ein Live-Voting-Tool der FAU der Status Quo der Nutzung (Häufigkeit, Anwendungsfälle) sowie bislang aufgetretene Probleme (technische Schwierigkeiten, unklare Antworten, fehlende Features) erfasst. Da das Entwicklerteam neben der Evaluatorsin an der Online-Veranstaltung teilnahm, konnten Sie direkt auf die geäußerten Problemfälle eingehen und Handlungsmöglichkeiten der Lösung mit den Studierenden klären.

- Abschluss-Voting (Juli 2025)

In der abschließenden Einheit der Lehrveranstaltung erfolgte eine zusammenfassende Bewertung von Nj̈t über das bekannte Live-Voting-Tool. Zusätzlich hatten die Studierenden die Möglichkeit, zu gestellten Fragen auch über den Chat der Kommunikationsplattform Rückmeldung zu geben, dessen Verlauf ebenfalls in die Datenauswertung einfluss.

- Abschluss-Fragebogen (Juli - August 2025)

Über einen Fragebogen am Ende des Semesters erfolgte nicht nur eine summative Bewertung von Nj̈t (Mehrwerte, problematische Aspekte, Zufriedenheit), sondern auch ein Abgleich der Erwartungen und Erfahrungen der Studierenden. Der Fragebogen nutzte eine Kombination aus geschlossenen Items (Likert-Skalen, 1-5) und offenen Feldern, um Verbesserungsvorschläge frei zu erfassen.

- Interviews mit Dozenten und Entwickler (August 2025)

Zeitnah zum Abschluss der Erprobungsphase erfolgten semi-strukturierte Interviews (ca. 60 Minuten) mit dem Dozenten und dem hauptverantwortlichen Entwickler. Hierüber wurden im Sinne einer Retrospektive die ursprünglichen Erwartungen, besonders prägnante Erfahrungen sowie die Bewertung von Nj̈t mit Blick auf wahrgenommene Mehrwerte, Integration in die Lehre und Auswirkungen auf Interaktivität und Lernverhalten der Studierenden. Auch herausfordernde Aspekte wurden aus Perspektive der Akteure erfragt.

- Analyse Chat-Bot-Kommunikations-Verläufe

Alle Interaktionen zwischen Studierenden und Nýt wurden anonymisiert protokolliert. Die Analyse fokussierte auf: Häufigkeit und Timing der Nutzung Typen von Fragen (organisatorisch, inhaltlich-vertiefend, metakognitiv) Problemfälle (fehlerhafte Antworten, Missverständnisse). Eine detaillierte Analyse hierzu steht noch aus.

- Auswertungs-Workshops (September & Oktober 2025)

Nach Semesterende fanden zwei strukturierte Workshops mit dem Entwicklerteam und dem Dozenten statt, in dem technische, didaktische und organisatorische Herausforderungen systematisch reflektiert wurden. Durch die Rückkopplung der Evaluationsergebnisse erfolgte eine individuelle Überprüfung, Ergänzung bzw. Bestätigung der Interviews. Zugleich diente der Austausch aber auch einer gemeinsamen kommunikativen Validierung (Bortz & Döring, 2015) der abgeleiteten Design-Prinzipien.

### **3.5 Sample und Beteiligung**

Auf Dozenten- und Entwicklerseite erfolgte eine Einbindung aller zentralen Akteure, die an der Entwicklung bzw. Implementation des Bots in die Lehrveranstaltung beteiligt waren. Diese sind auch als Autoren an diesem Beitrag beteiligt.

Bei den Studierenden erfolgte der Versuch, die Grundgesamtheit aller Teilnehmenden einzubinden. Eine genauere Erhebung und Untersuchung des Sampling nach Studiengängen oder Fachsemestern erfolgte nicht. Dadurch sollte die Anonymität gegenüber den Studierenden unterstrichen und die Beteiligungsrate erhöht werden. Auch kamen im Laufe des Semesters Nachzügler verspätet dazu bzw. beendeten Studierende den Kurs vorzeitig.

Die Vorab-Befragung der Studierenden zu Semesterbeginn skizzierte wichtige Charakteristika der Teilnehmenden, was ihre Vorerfahrungen und bisherige Nutzung von KI betrifft: Nur 15% gaben an, KI-Tools regelmäßig im Alltag zu nutzen. Im Studium kommt AI vor allem für Prüfungsvorbereitung, schnellen Überblick und Literaturrecherche zum Einsatz. Auch die Selbsteinschätzung der AI-Kompetenz fiel moderat aus: 59% der Befragten trauten sich zu, geeignete Prompts für AI-Bots zu formulieren. Jedoch konnten nur 37% einschätzen, welcher Bot für welchen Use-Case geeignet ist. Bei der kritischen Reflexion von Bot-Antworten sahen sich 56% kompetent, ebenso bei der Verifikation von Ergebnissen (56%). Programmieren konnten nur 22% der Studierenden selbst.

Trotz der aktuell geringen Nutzung bestand ein sehr hohes Interesse, AI zukünftig im Studium einzusetzen. Beim Lernverhalten präferierte die Mehrheit weiterhin die aktive Teilnahme am Kurs und schätzte die Interaktion mit anderen Studierenden sowie regelmäßiges Lernen über das Semester hinweg statt kurzfristigem Lernen vor den Prüfungen.

Von den 380 registrierten Teilnehmenden beteiligten sich durchschnittlich 45-75 Personen an den Befragungen (12-20%). Die geringe Beteiligungsquote variierte je nach Erhebungszeitpunkt und entwickelte sich absteigend (Vorab-Fragebogen: ~75 Personen,

Interims-Votings: ~50-60 Personen, Abschluss-Fragebogen: ~45 Personen). Zudem wurden viele Fragen übersprungen, was die Auswertung einzelner Items zusätzlich erschwerte. Eine systematische Nachverfolgung der Beteiligung anhand von Benutzerkennungen fand nicht statt, sodass keine Aussagen über die Kontinuität der Teilnahme möglich sind.

### **3.5 Analysemethode**

Bei den quantitativen Daten (Fragebögen) wurden über Deskriptive Statistik lediglich Häufigkeiten, Zustimmungsraten zur Charakterisierung der Stichprobe und Bewertung des Bots ermittelt. Aufgrund der geringen Fallzahl und hohen Anzahl fehlender Werte verzichteten wir auf inferenzstatistische Verfahren.

Die Chat-Verläufe wurden mittels Pattern-Extraktion analysiert: Identifikation typischer Fragetypen, Problemfälle und unerwarteter Nutzungsformen.

Die Auswertung der Interviews erfolgte mithilfe der Qualitativen Inhaltsanalyse (Mayring, 2015). Angewendet wurde die formale Strukturierung nach thematischen, induktiv entwickelten Kriterien (Gläser-Zikuda, 2013, S. 143-144), die thematisch zu Clustern verdichtet (Braun & Clarke, 2006) und schließlich als Gestaltungsprinzipien abstrahiert wurden. Die Kodierung des gesamten Datenmaterials wurde dazu mit der Software MAXQDA vorgenommen.

## **4. Ergebnisse**

In diesem Kapitel werden zentrale Ergebnisse aus den drei einbezogenen Perspektiven jeweils getrennt voneinander zusammengefasst vorgestellt. Der Fokus liegt dabei auf Erkenntnissen, die mit Implikationen für zukünftige Implementationen von KI-Chat-Bots in der Hochschullehre verbunden sind.

### **4.1 Perspektive Studierende**

Die Erwartungen der Studierenden an Nýt waren zu Beginn des Projektes hoch: 84% der Befragten erwarteten Unterstützung bei Fragen zum Kurs, 80% bei der Nachbereitung, 78% bei der Vorbereitung. 71% erhofften sich Hilfe bei der Reflexion des eigenen Lernprozesses, 69% sahen Nýt als Raum, um Fehler zu machen und daraus zu lernen. 65% gaben an, Nýt würde ihnen das Selbstvertrauen geben, Fragen zu stellen, die sie sonst nicht stellen würden. Interessanterweise erwarteten nur 20%, dass Nýt die Kursteilnahme ersetzen könnte. Die Studierenden sahen Nýt also von Anfang an als Ergänzung, nicht als Ersatz für die Veranstaltung.

Bei der erwarteten kognitiven Wirkung der Nutzung des Bots hielten sich anspruchsvoll (44%) und entlastend (49%) nahezu die Waage. Im Vergleich zu anderen Bots erwarteten

die Studierenden von Nýt: kursspezifischere und korrektere Antworten, schnellere Responses, Vorschläge für weiterführende Ressourcen und eine menschlichere Konversation.

Das tatsächliche Nutzungsverhalten blieb hinter den Erwartungen zurück. Nýt wurde primär pragmatisch für organisatorische Routinefragen eingesetzt: Termine, Deadlines, Materialzugang. Für tiefere Lernbegleitung oder reflexive Auseinandersetzung mit Inhalten wurde Nýt seltener genutzt. Diese geringe Intensität sowie Oberflächlichkeit der Nutzung war eine zentrale Enttäuschung sowohl für Lehrende als auch Entwickler.

Zieht man die verschiedenen Erhebungen hinzu, erscheinen die Gründe hierfür vielfältig gewesen zu sein: Nýt wurde nicht von Anfang an als integraler Bestandteil der Veranstaltung eingeführt, sondern eher als Add-on wahrgenommen. Studierende hatten "keine Zeit für ein zusätzliches Tool" und sahen "keinen Bedarf neben Vorlesung und Kursmaterialien". Dennoch meldeten die Studierenden verschiedene **positive Aspekte** zurück:

- **24/7-Verfügbarkeit:** Nýt war jederzeit ansprechbar, unabhängig von Sprechstunden.
- **Keine Hemmschwelle:** Studierende schätzten, dass sie dem Bot "dumme Fragen" stellen konnten, ohne sich zu blamieren.
- **Kursspezifischer Kontext:** Nýt benötigte weniger Kontext zur Beantwortung von Fragen, da er auf Kursmaterial trainiert war – zumindest eine "grobe Ahnung" vom Kurs hatte er.
- **Arbeitserleichterung:** Material musste nicht selbst eingepflegt werden, was die Nutzung vereinfachte.
- **Höhere Sicherheit:** Studierende fühlten sich sicherer, dass die Antworten korrekt waren (im Vergleich zu generischen Bots).

Aber auch verschiedene **problematische Aspekte** zeigten sich in den Befragungen:

- **Fehlende Benutzerfreundlichkeit:** Technische Probleme, langsame Antworten, fehlende Upload-Möglichkeiten.
- **Keine wahrgenommene "Uniqueness":** Im Vergleich zu ChatGPT & Co. fehlte ein klarer Mehrwert.
- **Organisatorische Lücken:** Nýt hatte kein integriertes Datum-Bewusstsein, d.h. keine Informationen über Zeitplanung und Termine
- **Antwortqualität:** Manche Antworten waren fehlerhaft, unvollständig, zu lang oder zu unspezifisch. Die Balance zwischen Präzision und Ausführlichkeit stimmte nicht immer.
- **Mismatch mit Prüfungsleistungen:** Nýt passte nicht zu den Quizze. Bei Klausurvorbereitung hätte er mehr Nutzen gebracht.
- **Tonalitäts-Spannung:** Während einige Studierende die Sheldon-Cooper-Persönlichkeit als auflockernd empfanden, wünschten sich andere extrem nüchterne, präzise Antworten ohne "Persönlichkeit". Wieder andere hätten gerne mehr Storytelling gehabt.

## 4.2 Perspektive Lehrender

Die Vorlesung "Nailing Your Thesis" stand vor dem Projekt vor den vorne skizzierten typischen Herausforderungen großer Online-Veranstaltungen. Der Dozent beschrieb das Gefühl, "ins schwarze Loch zu sprechen" – kaum Interaktion, anonyme Masse, fehlendes Feedback. Die Frage war: Kann ein Bot hier Abhilfe schaffen?

Seine Erwartungen waren eher vage und es war ihm unklar, ob Nýt tatsächlich Mehrwert bringt oder nur zusätzlichen Aufwand bedeutet. Erhofft wurde, dass Nýt die Veranstaltung "irgendwie auflockert" und das Gesprächstempo verändert. Nýt sollte zudem als Ansprechpartner für das Selbststudium dienen, wenn Studierende nicht an Live-Sessions teilnehmen.

Der Dozent äußerte zudem eine grundsätzliche Skepsis gegenüber Bots mit Blick auf verschiedene didaktische Risiken: Studierende könnten den Bot zum Schummeln bei den Hausaufgaben nutzen, statt tatsächlich zu lernen. Der Bot liefert sofort Antworten – das fördert eine Kultur der schnellen Befriedigung statt Ausdauer und Konzentration. Verlust von Stille und Zuhören: Wenn der Bot immer verfügbar ist, könnten Studierende verlernen, mit Unsicherheit umzugehen und selbst nachzudenken.

Die Erprobung war für den Dozenten mit mehreren **positiven Erfahrungen** verbunden:

- **Entlastung bei Routinefragen:** Nýt übernahm organisatorische Anfragen, die sonst per E-Mail oder in Sprechstunden geklärt werden müssten.
- **Sparring-Partner in der Vorlesung:** Die Interaktion zwischen Dozenten und Bot während der Live-Sessions lockerte die Atmosphäre auf und bot neue Gesprächsimpulse. Zudem regte der Einsatz den Dozenten an, Inhalte aus neuer Perspektive zu betrachten – ein unerwarteter Lerneffekt für den Lehrenden selbst.

Demgegenüber stehen verschiedene **negative Erfahrungen/Bewertungen:**

- **Interaktivität nicht gestiegen:** Entgegen der Hoffnung führte Nýt aus seiner Sicht nicht zu mehr studentischer Beteiligung. Im Gegenteil: Manche Studierende "lehnten sich zurück" und wurden passiver.
- **Integration unklar:** Nýt wurde nicht systematisch genug eingebunden und blieb ein Fremdkörper. Aufgrund fehlenden Feedbacks fehlten Einblicke in die tatsächliche Nutzung durch die Studierenden.

## 4.3 Perspektive Entwickler

Für das Entwicklerteam war der Einsatz in der Großgruppe ein Experiment mit unsicherem Ausgang:

- **Funktionalität und Nutzen:** Würde Nýt in einer realen Lehrveranstaltung über ein ganzes Semester hinweg funktionieren? Wie nutzen Dozent und Studierende ihn tatsächlich? Gibt es Features, die Nýt im Vergleich zu ChatGPT besonders wertvoll machen?

- **Technische Herausforderungen:** LLMs haben ein begrenztes Kontextfenster, d.h. Vorlesungsinhalte, Skripte, Transkripte oder Übungen mussten strukturiert und in semantische "Häppchen" aufbereitet werden, damit Nýt relevante Informationen gezielt abrufen kann. Diese Vorbereitung war zeitaufwändig.
- **Zwei Zielgruppen:** Nýt musste sowohl für Studierende als auch für den Dozenten funktionieren und deren unterschiedliche Ansprüche aufgreifen. Während Studierende eher pragmatisch-nüchterne Antworten wollten, suchte der Dozent einen Sparring-Partner, der neue Perspektiven einbringt.
- **Tonalität:** Die Wahl von Sheldon Cooper erwies sich als tragfähiger Kompromiss, stieß aber auf heterogene Resonanz.
- **Material-Abhängigkeit:** Nýt funktioniert am besten mit umfangreichem, sprachlich formuliertem Material. Nicht-sprachliche Formate wie Grafiken oder Tabellen sind für aktuelle Sprachmodelle schwer zu verarbeiten. Gerade aber beim Thema des Kurses (wissenschaftliches Arbeiten) sind diskontinuierliche Texte zentral.

Das Entwicklerteam war überrascht von der Qualität mancher Dialoge – wenn Nýt richtig genutzt wurde, entfaltete er als Companion echtes Potenzial. Durch das Sparring zwischen Professor und Bot in der Lehrveranstaltung entstand eine Dynamik, die Studierende ermutigte, selbst kreative oder naive Fragen zu stellen. Nýt konnte "dumme Antworten" geben, was die Hemmschwelle senkte.

Als negativ zeigte sich die mangelnde Vorbereitung des Dozenten auf die Arbeitsweise des Bots, was die Integration erschwerte. Damit verbunden fand nur eine oberflächliche Nutzung statt, was lediglich zu einer eher pragmatischen Zufriedenheit führte, d.h. die Intensität der Nutzung hätte aus Sicht der Entwickler höher sein können, wenn die Integration klarer gewesen wäre. Nýt wurde ferner zu spät in die Veranstaltung integriert und nur als externes Add-on wahrgenommen – ein strukturelles Problem.

#### **4.4 Abgeleitete Design-Prinzipien**

Aus der Triangulation der drei Perspektiven lassen sich konkrete Design-Prinzipien ableiten, die in vergleichbaren Settings den Unterschied zwischen erfolgreichem und gescheitertem Bot-Einsatz ausmachen können. Diese Prinzipien strukturieren wir entlang des Einsatzzyklus: Vor, während und nach dem Semester.

##### **Vor dem Einsatz:**

**Klare Erwartungs-Absprachen:** Dozierende und Entwicklerteam müssen vorab definieren, was mit dem Bot erreicht werden soll. Welche Rolle soll er spielen? Welche

Aufgaben übernehmen? Welche didaktischen Ziele unterstützen? Ohne diese Zielvereinbarung bleibt der Bot-Einsatz diffus und wird schnell zum frustrierenden Add-on.

**Didaktisches Design anpassen:** Der Bot darf nicht einfach "dazugestellt" werden. Die gesamte Veranstaltung muss auf seine Integration ausgerichtet werden: Welche Aufgaben werden über den Bot gestellt? Wo wird er in der Vorlesung eingebunden? Welche Reflexionsanlässe entstehen durch seine Nutzung? Diese Planung erfordert Zeit. Eine „Last-Minute-Integration“ führt zu oberflächlicher Nutzung.

**Last-Minute-Mentalität aushebeln:** Studierende neigen dazu, Aufgaben kurz vor der Deadline zu erledigen. Ein Bot, der nur als optionales Zusatzangebot wahrgenommen wird, fällt diesem Verhalten zum Opfer. Lösung: Den Bot in kontinuierliche Lernaktivitäten einbinden und Prüfungsformate entwickeln, die seine Nutzung voraussetzen (z.B. prozessorientierte Prüfungen statt reiner Reproduktion).

**"Weg zur Lösung"-Problematik:** LLMs sind darauf trainiert, sofort Lösungen zu liefern ("pleasing the user"). Lehre zielt aber auf den Weg zur Lösung - Reflexion, Auseinandersetzung, Fehler machen dürfen. Diese Spannung muss explizit thematisiert und didaktisch gerahmt werden. Storytelling kann hier helfen: Warum ist der Prozess wichtiger als das Ergebnis? Wie nutze ich den Bot als Sparring-Partner statt als Lösungsautomat?

**Material vorab bereitstellen:** Die Qualität der Bot-Antworten hängt maßgeblich von der Materialgrundlage ab. Vorlesungsskripte, Transkripte, Übungsaufgaben und - zentral - organisatorische Informationen (Termine, Deadlines, Prüfungsmodalitäten) müssen strukturiert und in Textform aufbereitet werden (auch diskontinuierliche Formate). Diese Vorbereitung ist zeitaufwändig, aber unverzichtbar.

### Während des Einsatzes:

**Bot motivierend einführen; Integration leben:** Studierende müssen verstehen, warum der Bot wertvoll ist. Welche Mehrwerte bietet er gegenüber ChatGPT & Co.? Welche Möglichkeiten eröffnet er? Diese Einführung sollte nicht technisch, sondern anwendungsorientiert erfolgen: Konkrete Use-Cases zeigen, Möglichkeitsraum aufspannen, zum Ausprobieren anregen.

**Regelmäßige Nutzung in diversen Rollen:** Der Bot sollte nicht nur als FAQ-Maschine fungieren, sondern in unterschiedlichen Settings zum Einsatz kommen: Als Sparring Partner in der Vorlesung (Dozent stellt Bot Fragen), als Reflexionstool für Studierende (Bot stellt Rückfragen zu Aufgaben), als Tutor für Nachbereitung. Diese Rollenvielfalt zeigt das Potenzial und reduziert die Wahrnehmung als bloßes Hilfstool.

**Fragen und Antworten systematisch aufgreifen:** Studierende sollen ihre Bot-Interaktionen in der Vorlesung teilen. Welche Fragen wurden gestellt? Welche Antworten hat der Bot gegeben? War die Antwort hilfreich? Diese gemeinsame Reflexion schafft

Lernmomente, korrigiert Fehlinformationen und zeigt, wie der Bot didaktisch sinnvoll genutzt werden kann.

**Regelmäßige Problemerkennung:** Durch Interims-Votings oder kurze Pulse-Checks sollten Probleme zeitnah identifiziert werden: Technische Schwierigkeiten? Unklare Antworten? Fehlende Features? Diese Rückmeldungen ermöglichen Anpassungen während des Semesters, statt erst am Ende festzustellen, dass der Bot nicht funktioniert hat.

**Verpflichtende Rückmeldung zur Nutzung:** Studierende sollten möglichst regelmäßig Feedback zur Bot-Nutzung geben. Diese Rückmeldungen sind essentiell für die Weiterentwicklung und sollten nicht optional, sondern verpflichtend eingebaut werden (z.B. als Teil der Studienleistung).

### **Nach dem Einsatz:**

**Evaluation von Benutzerfreundlichkeit, Lerneffekten und Herausforderungen:** Am Semesterende sollte systematisch erhoben werden: War der Bot benutzerfreundlich? Hat er zu Lerneffekten beigetragen? Welche Herausforderungen traten auf? Diese Evaluation dient nicht nur der Rechenschaft, sondern der Verbesserung für den nächsten Durchlauf.

**Auswertung der Kommunikation:** Das Entwicklerteam sollte zentrale Daten aus den Chat-Verläufen aufbereiten und dem Dozenten zur Verfügung stellen: Welche Fragen wurden häufig gestellt? Wo gab der Bot fehlerhafte Antworten? Welche Themen wurden intensiv diskutiert? Diese Analyse liefert didaktische Erkenntnisse: Wo gab es Verständnisprobleme? Welche Inhalte müssen im nächsten Durchlauf klarer aufbereitet werden?

**Reflexion der Lehrenden:** Der Bot-Einsatz ist kein einmaliges Experiment, sondern ein iterativer Prozess. Lehrende müssen bereit sein, aus den Erkenntnissen des ersten Durchlaufs Konsequenzen zu ziehen: Material überarbeiten, didaktisches Design anpassen, Prüfungsformate weiterentwickeln. Diese Reflexion ist anspruchsvoll, aber entscheidend für nachhaltigen Erfolg.

Diese Design-Prinzipien sind nicht als starre Checkliste zu verstehen, sondern als Orientierungsrahmen für Bot-Einsatz in der Hochschullehre. Sie fassen die zentralen Learnings aus der Nüt-Evaluation zusammen und bieten Ansatzpunkte für andere Lehrende, die KI-Chatbots in ihre Veranstaltungen integrieren möchten.

## **5. Diskussion**

Die Evaluation von Nýt liefert differenzierte Einblicke in Mehrwerte und Grenzen KI gestützter Lernbegleiter in der Hochschullehre. Dieses Kapitel interpretiert die Ergebnisse zusammenfassend und zeigt Limitationen auf.

## **5.1 Interpretation der Ergebnisse**

### **Mehrwerte von KI in der Hochschullehre sind real, aber kontextabhängig**

Die Evaluation zeigt, KI-Chatbots können sowohl für Studierende als auch Dozierende verschiedene Mehrwerte liefern. Doch diese Mehrwerte entstehen nicht automatisch. Sie sind kontextabhängig: Ein Bot, der als optionales Add-on wahrgenommen wird, bleibt ungenutzt. Ein Bot, der technisch fehlerhaft ist oder keine organisatorischen Informationen enthält, frustriert statt zu unterstützen. Ein Bot, der didaktisch nicht eingebettet ist, fördert Passivität statt Aktivierung.

Oder anders ausgedrückt: Der Bot ist kein Plug-and-Play-Tool, sondern ein didaktisches Instrument, das Planung, Integration und kontinuierliche Reflexion erfordert. Integration ist dabei der kritische Erfolgsfaktor, wie alle drei eingebundenen Akteursgruppen im Projekt übereinstimmend betonten. Ohne diese systematische Einbettung verkommt der Bot zum "nice-to-have", das in der Prüfungsstress-Phase der Studierenden schnell untergeht.

### **Personalisierungsoptionen sind von Anfang an mitzudenken**

Ein überraschendes Ergebnis der Evaluation war die Heterogenität der Erwartungen auf Seiten der Studierenden an die Tonalität. Darin zeigt sich kein Defizit, sondern ein Designproblem: Unterschiedliche Lerntypen haben unterschiedliche Präferenzen. Die Lösung kann nicht darin bestehen, einen Mittelweg zu finden (der niemanden zufriedenstellt), sondern in Customization. Dies kann beispielsweise durch einen Tonalitäts-Slider gelingen, mit dem Nutzer vorab einstellen können, ob sie sachlich-präzise (0%) oder kreativ-emotional (100%) begleitet werden möchten. Diese Erkenntnis hat Implikationen über Nýt hinaus: Jeder Bot-Einsatz in der Lehre sollte von vornherein Personalisierungsoptionen vorsehen, statt eine Einheitslösung zu erzwingen.

### **Vermeidung von Passivität ist anspruchsvoll aber machbar**

Ein zentrales Spannungsfeld unserer Erprobung zeigt sich in der Bewertung der Interaktivität. Während das Entwicklerteam begeistert von der Tiefe mancher Dialoge berichtet, äußert der Dozent, Studierende seien durch den Bot eher passiver geworden. Dieser Widerspruch lässt sich auflösen: Beide Beobachtungen sind korrekt, beziehen sich aber auf unterschiedliche Nutzungsmuster. Wo der Bot didaktisch gerahmt und als Sparring-Partner genutzt wird, entsteht tatsächlich Aktivierung. Wo er jedoch als Lösungsautomat missbraucht wird, fördert er Bequemlichkeit. Diese Spannung muss explizit thematisiert werden. Dozierende müssen daher nicht nur den Bot instruieren, sondern auch die Nutzungshaltung der Studierenden prägen.

### **Widersprüchliche Erwartungshaltungen müssen zusammengebracht werden**

Das Auffälligste in unseren Daten ist die unterschiedliche Bewertung des Gesamtimpacts: Der Dozent sieht "momentan eher einen negativen Impact" – Studierende werden

passiver, Interaktivität sinkt. Das Entwicklerteam hingegen berichtet von Begeisterung über die "Tiefe der Interaktionen" und betont, dass der Bot die Vorlesung "grundlegend verändert" habe. Wie lässt sich diese Diskrepanz erklären?

1. Unterschiedliche Beobachtungsebenen: Der Dozent bewertet das aggregierte Verhalten der gesamten Kohorte (geringe Nutzung, wenig sichtbare Aktivierung). Das Entwicklerteam fokussiert auf einzelne intensive Dialoge (die tatsächlich stattfanden, aber von einer Minderheit geführt wurden).

2. Unterschiedliche Maßstäbe: Der Dozent misst am Ideal aktiver, selbstständiger Lernender. Das Entwicklerteam misst am technischen Potenzial des Bots. Beide haben recht - aber die Diskrepanz zeigt, dass technisches Potenzial und didaktische Wirkung nicht identisch sind.

3. Erfahrungshorizont: Der Dozent ist seit Jahrzehnten in der Lehre tätig und hat einen klaren Maßstab für gute Interaktion. Das Entwicklerteam kommt aus der Technologie-Perspektive und ist von der Qualität der Bot-Antworten beeindruckt. Diese unterschiedlichen Horizonte führen zu unterschiedlichen Bewertungen. Die Lehre daraus: Erfolgreicher Bot-Einsatz erfordert gemeinsame Reflexion zwischen Lehrenden, Studierenden und Entwicklern. Nur wenn diese Perspektiven systematisch aufeinander bezogen werden, können Missverständnisse geklärt und Erwartungen justiert werden.

## **5.2 Limitationen der Studie**

Vier zentrale Einschränkungen sind bei der Interpretation der Ergebnisse zu berücksichtigen:

### 1. Geringe und schwankende Stichprobe

Von über 380 registrierten Studierenden beteiligten sich durchschnittlich 45-75 Personen an den Befragungen (12-20%). Die Beteiligungsquote war hoch zu Semesterbeginn und sank kontinuierlich. Zudem wurden viele Fragen übersprungen, was die Auswertung einzelner Items erschwerte. Es ist unklar, ob die Teilnehmenden repräsentativ für die Gesamtkohorte sind. Vermutlich sind besonders motivierte oder besonders kritische Studierende überrepräsentiert. Die Ergebnisse sind daher als explorative Momentaufnahmen zu verstehen, nicht als repräsentative Vollerhebung.

### 2. Generalisierbarkeit eingeschränkt

Die Studie untersucht einen spezifischen Fall: eine große Online-Vorlesung im Bereich Informatik. Die Übertragbarkeit auf andere Settings ist eingeschränkt. Fraglich ist beispielsweise, funktioniert ein Bot anders, wenn Studierende physisch anwesend sind? Braucht man in interaktiven Kleingruppen überhaupt einen Bot? Gelten die Erkenntnisse auch für andere Fachkulturen? Diese Fragen können nur durch Folgestudien in diversen Kontexten beantwortet werden.

### 3. Technische Beschränkungen

Die RAG-Architektur mit begrenztem Kontextfenster erfordert, dass Material vorab strukturiert und in "Häppchen" aufbereitet wird. Nichtsprachliche Formate (Grafiken, Tabellen) sind für aktuelle LLMs schwer verarbeitbar. Diese technische Limitation beeinflusst die Antwortqualität in bestimmten Themenbereichen. Zukünftige Modelle mit größeren Kontextfenstern und multimodalen Fähigkeiten könnten diese Einschränkungen überwinden – aber zum Zeitpunkt der Evaluation war dies eine reale Grenze.

#### 4. Fehlende Langzeitdaten

Die Evaluation endete mit dem Semester. Aussagen über langfristige Lerneffekte waren damit nicht möglich. Spannende Fragen wären beispielsweise gewesen: Hat der Bot tatsächlich zu besseren Abschlussarbeiten geführt? Haben Studierende, die den Bot intensiv nutzten, andere Kompetenzen entwickelt als jene, die ihn ignorierten? Haben sich die Erwartungen und Nutzungsmuster im zweiten oder dritten Durchlauf verändert? Diese Fragen bleiben offen und erfordern längsschnittliche Forschungsdesigns.

Trotz dieser Limitationen liefert die Studie empirische Evidenz aus dem realen Lehralltag und bietet zugleich Impulse für die Zukunft, welche im abschließenden Kapitel skizziert werden.

## 6. Fazit & Ausblick

### 6.1 Zusammenfassung: To Bot or Not to Bot? - Ja, aber mit didaktischem Konzept

Die provokante Eingangsfrage "To Bot or Not to Bot?" lässt sich auf Basis der Nüt Evaluation klar beantworten: Ja zum Bot – aber nur mit einem klaren didaktischen Konzept. Die Evaluation zeigt, KI-gestützte Lernbegleiter können in drei Dimensionen Mehrwerte liefern: Organisatorische Entlastung, 24/7-Lernbegleitung und Interaktivitätssteigerung. Wo der Bot als Sparring-Partner didaktisch eingebunden wird, können damit neue Dynamiken in der Lehre entstehen.

Doch diese Mehrwerte sind kein Selbstläufer. Ohne Integration ab Tag 1, ohne didaktische Rahmung und ohne kontinuierliche Reflexion bleibt der Bot ein ungenutztes Add-on. Die zentrale Formel lautet: "Miteinander oder gar nicht" – der Bot muss von Anfang an Teil der Veranstaltung sein, nicht nachträglich hinzugefügt werden.

### 6.2 Implikationen für Lehrende

Lehrende müssen für eine sinnvolle Integration den Bot als didaktisches Instrument begreifen und nutzen:

**Bot in didaktisches Konzept integrieren:** Konkret bedeutet dies, Lehrende brauchen klare Lernziele, Meilensteine und Aha-Effekte, an denen sich die Arbeit mit dem Bot ausrichtet. Dozierende müssen hierfür in der Lage sein, Lernziele in systematische Instruktionen für den Bot zu übersetzen.

**Prüfungsformate anpassen:** Wenn Studierende jederzeit Zugang zu einem Bot haben, der Lösungen liefert, sind klassische Reproduktionsprüfungen obsolet. Stattdessen müssen prozessorientierte Prüfungen (Open-Book mit KI-Zugang, Reflexion statt Reproduktion) entwickelt werden.

**Reflexionsbereitschaft:** Ein Bot-Einsatz ist kein einmaliges Experiment, sondern ein iterativer Prozess. Lehrende müssen bereit sein, aus Learnings des ersten Durchlaufs Konsequenzen zu ziehen.

### 6.3 Implikationen für Hochschulen

Auch für die Hochschulinstitutionen ergeben sich Anforderungen, wenn KI systematisch in die Lehre integriert und Mehrwerte realisiert werden sollen:

**Infrastruktur schaffen:** Hochschulen müssen Strukturen für Finanzierung, Betreuung und Integration von KI-Bots schaffen. Das erfordert Investitionen in LMS-Integration, technischen Support und didaktische Begleitung.

**Kooperationen eingehen:** Hochschulen sollten frühzeitig überlegen, mit welchen Anbietern sie kooperieren wollen - ob kommerzielle Anbieter, Open-Source-Lösungen oder eigene Entwicklungen. Die Kommoditisierung (Bots als Standardfunktion in Moodle, StudOn etc.) wird den Markteintritt erleichtern.

**Weiterbildung für Lehrende:** Entscheidend für eine motivierte Beteiligung einer möglichst hohen Anzahl von Lehrenden wird dabei aus den Projekterfahrungen heraus das eigene Erleben sein, um Potenzial und transformative Kraft von Bots möglichst praxisnah nachzuvollziehen. Hochschulen sollten deshalb Möglichkeiten schaffen, Bots als Hands-on-Experience in der eigenen Lehrveranstaltung ausprobieren zu können.

**Anleitung für Studierende:** Um einen Bot als Sparring-Partner nutzen zu können, und nicht als Lösungsautomat zu missbrauchen, müssen Studierende lernen, den Bot als Reflexionstool einzusetzen, das den Lernprozess unterstützt - nicht als Abkürzung, die Nachdenken ersetzt. Diese Nutzungshaltung muss explizit vermittelt und eingefordert werden. Hochschulen müssen dazu AI Literacy als Querschnittskompetenz verankern.

### 6.4 Mögliche Zukunftsszenarien

Die Evaluation von Nýt wirft nicht zuletzt die Frage auf: Wohin entwickelt sich KI-gestützte Hochschullehre? Im Reflexions-Workshop wurden drei zeitliche Horizonte diskutiert. Die Szenarien sind keine fixen Prognosen, sondern Möglichkeitsräume, die zeigen, wie sich Lehre verändern könnte - wenn die entsprechenden Weichenstellungen erfolgen.

#### **Kurzfristig (2-3 Jahre):**

Bereits In den nächsten Jahren werden Dashboards entstehen, die Dozierenden einen Metablick auf den Kursverlauf geben. Diese Learning Analytics ermöglichen es, Lehre

datengetrieben zu verbessern, ohne dass Dozierende selbst durch hunderte Chat-Verläufe scrollen müssen.

### **Mittelfristig (3-5 Jahre):**

Mittelfristig werden Chatbots zu Standardfunktionen in Lernmanagementsystemen werden. Die Frage wird nicht mehr sein, ob ein Bot eingesetzt wird, sondern wie gut er konfiguriert ist. Eine mögliche Vision dazu: Jeder Studierende erhält eine personalisierte Bot-Instanz, die über das Semester hinweg mit ihm lernt. Der Lehrende fungiert als "Modell-Spender" - seine didaktischen Prinzipien, sein Fachwissen, seine Tonalität prägen den Bot. Aber der Bot passt sich an den individuellen Lernstand, das Lerntempo und die Präferenzen des Studierenden an. Adaptives Lernen wird damit Realität. Die Verbindung von Chat-Daten und Learning Analytics ermöglicht zudem eine datengetriebene Verbesserung der Lehre: Wo gibt es systematische Verständnisprobleme? Welche Studierenden brauchen zusätzliche Unterstützung? Diese Insights fließen zurück in die Kursgestaltung und schaffen einen kontinuierlichen Verbesserungskreislauf.

### **Langfristig (5-10 Jahre):**

Langfristig verschiebt sich die Rolle des Dozierenden vom Content-Deliverer zum Modell-Kurator. Statt Vorlesungen zu halten, kuratieren Dozierende das didaktische Modell, das dem Bot zugrunde liegt: Welche Lernziele sind zentral? Welche Aha-Effekte sollen entstehen? Welche kritischen Fragen müssen gestellt werden?

Wenn Bots zur Normalität werden, verändert sich Hochschullehre fundamental. Der Fokus verschiebt sich von Reproduktion zu Reflexion, von Wissensvermittlung zu kritischem Denken, von Frontalunterricht zu individueller Begleitung. Vorlesungen mit statischen Skripten werden abgelöst durch dynamische Lern-Experiences, bei denen Bots als permanenter Companion fungieren. Studierende lernen nicht mehr von der Vorlesung, sondern mit dem Bot, der die Vorlesung kontextualisiert, vertieft und auf individuelle Fragen eingeht.

Diese mögliche Transformation von Hochschullehre wirft Fragen auf, die über Didaktik hinausgehen:

- Datenschutz: Wer hat Zugang zu den Chat-Verläufen? Wie werden Lernprofile geschützt?
- Equity: Wer hat Zugang zu Premium-Bots? Entstehen neue Ungleichheiten zwischen Hochschulen mit und ohne KI-Infrastruktur?
- Abhängigkeit: Verlernen Studierende selbstständiges Denken, wenn sie sich auf Bots verlassen?
- Rolle der Lehrenden: Werden Dozierende zu Kuratoren degradiert oder zu didaktischen Strategen aufgewertet?

Diese Fragen müssen frühzeitig diskutiert werden, um die Transformation der Hochschullehre sozial gerecht und pädagogisch verantwortungsvoll zu gestalten.

## Literaturverzeichnis

Al-Zahrani, A. M. (2025). Exploring the impact of artificial intelligence chatbots on human connection and emotional support among higher education students. *SAGE Open*, 15(1). <https://doi.org/10.1177/21582440251340615>

Bortz, J., & Döring, N. (2015). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Springer-Lehrbuch. Springer

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, 5(23). <https://doi.org/10.1186/s40594-018-0123-6>

Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), 2940. <https://doi.org/10.3390/su15042940>

Euler, D. (2014). Design-Research – a paradigm under development. In D. Euler & P. F. E. Sloane (Hrsg.), *Design-Based Research* (S. 15–44). Franz Steiner Verlag.

Euler, D. (2017). Design principles as bridge between scientific knowledge production and practice design. *EDeR - Educational Design Research*, 1(1), 1–15. <https://doi.org/10.15460/eder.1.1.1024>

Fan, Y., Wang, X., & Zhu, M. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2). <https://doi.org/10.1111/bjet.13544>

Fryer, L. K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24(2), 8–22. <https://doi.org/10.64152/10125/44719>

Gläser-Zikuda, M. (2013). Qualitative Inhaltsanalyse in der Bildungsforschung – Beispiele aus diversen Studien. In K. Aguado, L. Heine, & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (S. 136–159). Lang.

Holmes, W., & Porayska-Pomsta, K. (2024). The impact of artificial intelligence on personalized learning in higher education: A systematic review. *Contemporary Educational Technology*, 4(2), ep573. <https://doi.org/10.30935/cedtech/15089>

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55 (12), 1-38. <https://doi.org/10.1145/3571730>

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2023). Estimating the carbon footprint of BLOOM, a 176B parameter language model. *Journal of Machine Learning Research*, 24(253), 1-15

Markel, J. M., & Opfermann, M. (2021). Chatbots as a tool to scale mentoring processes: Individually supporting self-study in higher education. *Frontiers in Artificial Intelligence*, 4, 668220. <https://doi.org/10.3389/frai.2021.668220>

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz.

NASPA. (2024). *The impact of generative AI educational chatbots on the academic support experiences of students in U.S. research universities*. <https://www.naspa.org/blog/the-impact-of-generative-ai-educational-chatbot>

OpenAI. (2023). GPT-4 Technical Report. arXiv:2303.08774. <https://arxiv.org/abs/2303.08774>

Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Viking Press.

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. *Computers and Education Open*, 6, 100196. <https://doi.org/10.1016/j.caeo.2024.100196>

Reinmann, G. (2022). *Design-Based Research am Beispiel hochschuldidaktischer Forschung*. <https://doi.org/10.25656/01:25651>

- Rhodes, J. E. (2024). To bot or not to bot? How AI companions are reshaping human services and connection. *The Chronicle of Evidence-Based Mentoring*. <https://www.evidencebasedmentoring.org/to-bot-or-not-to-bot/>
- Schiff, D. (2024). OpenAI hit with €15 million fine by Italy's data protection authority over ChatGPT privacy violations. *TechCrunch*. <https://techcrunch.com/2024/12/20/openai-hit-with-15-million-fine-by-italys-data-protection-authority-over-chatgpt-privacy-violations/>
- Sullivan, M., & Wilson, S. (2024). *AI and the next digital divide in education*. Brookings Institution. <https://www.brookings.edu/articles/ai-and-the-next-digital-divide-in-education/>
- von Glasersfeld, E. (1989). Constructivism in education. In T. Husen & N. Postlethwaite (Eds.), *International encyclopedia of education* (Supplement Vol. 1, pp. 162- 163). Pergamon Press.
- von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. Falmer Press. <https://doi.org/10.4324/9780203454220>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings*, 2018(1), 15903. <https://doi.org/10.5465/AMBPP.2018.15903abstract>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, 21, 67. <https://doi.org/10.1186/s41239-024-00467-0>
- Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 16. <https://doi.org/10.1186/s40561-024-00316-7>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)