

SHRIMP Reports

Projektberichte und Begleitmaterialien aus dem
Lehrinnovationsprojekt SHRIMP

1: Projektdokumentation Laboruniversität (2016)

[PREPRINT VERSION]

Eine Übersicht aller im Rahmen des Projekts erstellten Publikationen (inkl. Artikel in Fachzeitschriften) ist online unter <http://www.shrimpp.de/publications> zu finden.

* * *

Das Lehrinnovationsprojekt SHRIMP (Social Hypertext Reader & Interactive Mapping Platform) wurde 2015 am Institut für Amerikanistik der Universität Leipzig ins Leben gerufen. Ziel des Projekts ist es, das didaktische Potenzial von *Social Hypertext* zu erforschen und nutzbar zu machen. Mehr Informationen zum Projekt finden Sie online unter www.shrimpp.de. Die Reihe *SHRIMP Reports* enthält die im Rahmen des Projekts erstellten Berichte und andere Begleitmaterialien.

Kontakt: Sebastian M. Herrmann, smherrmann@uni-leipzig.de bzw. smh@shrimpp.de.

Projektbericht zum SHRIMP-basierten Projekt „Machine Analytics & Interaction Patterns“

Regina Borovaya, Sarah Doberitz, Sebastian M. Herrmann, Franz Wendt

Abstract: Der vorliegende Bericht beschreibt Ziele, Arbeitsprozesse, Outcomes und Evaluationsergebnisse des Projekts „SHRIMP Machine Analytics & Interaction Patterns: Generalisierbare Learning Analytics in textbasierten Geisteswissenschaften.“ Das Projekt wurde im Rahmen des Verbundprojektes „Advanced Learning and Examination Spaces in Saxony (ALExS.sax)“ von Mai 2019 bis März 2021 mitfinanziert mit Steuermitteln auf Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes. Ziel war die Erforschung von Machine-Learning Verfahren für die Learning-Analytics-Auswertung von Lernmetadaten, sowie das Experimentieren mit neuen Interaktionsmustern zwischen Lernenden und Plattform, insb. durch Verwendung eines lernbegleitenden Chatbot.

Ziele und Aufbau der Förderphase

Das im Rahmen des Verbundprojektes Advanced Learning and Examination Spaces in Saxony (ALExS.sax) von Mai 2019 bis März 2021 durchgeführte Projekt nutzte die Lehr-Lernplattform SHRIMP als Testbed für zwei voneinander weitgehend unabhängige Projektziele: die testweise Implementierung und Überprüfung von Machine-Learning-Verfahren zur Auswertung von Learning Analytics und die Entwicklung von automatisierten, skalierenden tutoriellen Kommunikationsformen („Interaction Patterns“) über einen Chatbot. Es handelte sich dabei um das vierte SHRIMP-basierte Lehr-Lernprojekt im Rahmen von vier aufeinander folgenden und aufeinander aufbauenden Förderphasen seit 2015.¹

Beide Arbeitsstränge dieser Phase basieren somit auf den Ergebnissen aus früheren Förderphasen, in denen unter anderem die Lernplattform SHRIMP entwickelt worden war (1) und das grundsätzliche fachinterne Transferpotenzial der Technologie und der mit ihr verbundenen didaktischen Konzepte ausgelotet worden war (2). In der vergangenen, ebenfalls im Rahmen des Verbundprojektes ALExS.sax geförderten Projektphase (3) war eine auf Learning Analytics basierende automatische Gruppierung der Lernenden gemäß ihrer jeweiligen Lernpräferenzen entwickelt worden. Während die hier angewendeten statistischen Verfahren gute Ergebnisse lieferten, war das Ziel nun, Machine-Learning-Verfahren daraufhin zu prüfen, ob sie eine solche Klassifizierungsaufgabe in besser generalisierbarer Form durchführen können, was ein Portieren der Plattform und der zugehörigen Lern- und Interaktionsszenarien erleichtern würde.

Ebenfalls hatte sich in früheren Projektphasen gezeigt, dass eine auf konventionellen Benutzer:innenoberflächen basierende Interaktion der Plattform mit den Lernenden rasch an Grenzen stößt und von den Lernenden nicht als potenziell wertvolle, quasi-tutorielle Interaktion sondern lediglich als Benutzen einer Software verstanden wird. Versuche, diese Kommunikation per E-Mail zu führen, waren aber von den Studierenden als eher störend empfunden worden. Gleichzeitig hatten sich Studierende in der Evaluation früherer Projektphasen explizit gewünscht, dass die Plattform direkter mit ihnen kommunizieren soll. Entsprechend war hier das Ziel, die Nutzung von Interaktionsdaten über einen als digitalem Tutor fungierenden Chatbot neu aufzustellen. Das Interaktionsmuster des Chatbot, so die Annahme, würde für die Lernenden auf einer

¹ Vgl. <http://www.shrimpp.de/page/shrimp-milestones>

dezidiert anderen Ebene stattfinden als die reine Nutzung des Plattform-User-Interface und auf dieser Ebene auch stärker als tutorielles Angebot eingeordnet und angenommen werden.

Das Projekt verlief in vier großen Phasen: einer **Konzeptionsphase** (Mai-Sept. 2019) in der das Vorgehen in beiden Arbeitssträngen weiter geplant werden konnte; einer ersten **Durchführungsphase 1** (Okt. 2019-Feb. 2020), in der Muster ausprobiert und erste Erfahrungen gesammelt werden konnten; einer **Implementierungs- und Aufbauphase** (März-Sept. 2020), in der die einzelnen Komponenten umgesetzt werden konnten; und einer **Durchführungsphase 2** (Okt. 2020-Feb. 2021), in der die neu umgesetzten Techniken erprobt und bewertet werden konnten. Diese zweite Durchführungsphase beinhaltete entsprechend auch die Dokumentation des Projektes und seiner Ergebnisse.

Zusammenfassend lässt sich festhalten, dass zumindest mit den vorliegenden Datensatzgrößen Machine-Learning-Verfahren den in früheren Projektphasen erarbeiteten statistischen Verfahren unterlegen sind. Gleichzeitig können die Experimente zu einer Machine-Learning-basierten Erkennung von Lernneigungen die in der früheren Phase erarbeiteten Erkenntnisse noch einmal perspektivieren und, zumindest teilweise, in Frage stellen. Details hierzu finden sich im Folgenden, sowie in einem separat veröffentlichten Bericht zum Machine Learning.² Gleichzeitig lässt sich festhalten, dass die Nutzung des Chatbot als „Interaction Pattern“ enormes Potenzial im Kontext von E-Learning gezeigt hat. Der Chatbot wurde von den Studierenden sehr gut angenommen und hat sich, insbesondere in dem von der Corona-Pandemie gezeichneten Wintersemester 2020/21, außerordentlich gut bewährt. Details zur Implementierung des Chatbot finden sich in einem ebenfalls separat veröffentlichten Whitepaper.³

Auch wenn also detaillierte Berichte zu beiden Arbeitssträngen veröffentlicht wurden, soll mit diesem Projektbericht einerseits eine konzise Zusammenfassung dieser detaillierteren Berichte geleistet werden; andererseits soll eine Zusammenschau der Ergebnisse im Kontext des größeren Projektzusammenhangs geleistet werden. Insbesondere enthält dieser Projektbericht auch die abschließen gewonnenen Evaluationsergebnisse zum Einsatz des Chatbot.

Arbeitsbericht

Da die Prozesse und die sie leitenden Überlegungen für beide Arbeitsstränge in jeweils separaten Publikationen dargestellt wurden, erfolgt hier nur eine zusammenfassende Darstellung in Kurzform.

Machine Learning

Die Arbeiten zu Machine Learning waren strikt explorativ ausgelegt: Primäres Ziel war also nicht, eine schlüsselfertige Machine-Learning-Lösung zur Klassifikation von Lernenden zu entwickeln. Vielmehr war das primäre Ziel, die Nutzbarkeit von Machine-Learning-Verfahren für diesen Aufgabenbereich durch Erprobung und Einsatz auf Basis der bestehenden bzw. im Projektzeitraum gewonnenen Learning-Analytics-Daten zu erkunden. Dieser Aufgabenzuschnitt ergab sich einerseits aus der bestehenden Datenlage (mit ca. 80-90 Studierenden pro Jahr war die Zahl der nutzbaren Datensätze relativ gering), andererseits aus dem Forschungsstand zum Thema: Auch wenn Lear-

² Vgl. Wendt, „SHRIMP Machine Learning Report“

³ Vgl. Doberitz et al.

ning Analytics und Machine Learning keine neuen Themen mehr sind, gibt es bisher nur relativ wenige Arbeiten, die Learning-Analytics-Auswertungen mit Machine-Learning-Verfahren durchführen und insb. gibt es bisher kaum Beispiele zur erfolgreichen Nutzung von Learning Analytics in den Geisteswissenschaften. Diese explorative Ausrichtung wurde im Lauf der Förderphase durch Zwischenergebnisse insofern weiter bestätigt, als sich keiner der geprüften Algorithmen für den vorliegenden Use-Case als unmittelbar einsetzbar erwies.

In einem ersten Schritt wurde in diesem Arbeitsstrang entsprechend geprüft, ob sich die Annahme von charakteristischen Gruppen von Lernneigungen (Lernneigungsprofile) durch naive Clustering-Algorithmen bestätigen lassen. Hierbei waren zwei Überlegungen maßgebend: 1) Einerseits sollte so abgeschätzt werden, wie gut solche 'naiven' Verfahren, denen also keine Vorannahmen über die Beschaffenheit des Datenmaterials mitgegeben werden, mit der vorliegenden Klassifizierungsaufgabe zurecht kommen können. 2) Zweitens sollte die in der vorangegangenen Förderphase durch eine Faktorenanalyse erarbeitete Klassifizierung perspektiviert werden: einzelne ihr zugrundeliegende Annahmen, wie zum Beispiel Schwellwerte für gültige Zuordnungen, waren seinerzeit nach Augenmaß festgelegt worden, und die Anwendung von Clusteringverfahren versprach, die damals getroffenen Entscheidungen um eine zusätzliche Perspektive zu bereichern und somit ggf. auch abzusichern. Hierbei wurden zwei Clustering-Algorithmen erprobt: KMeans und OPTICS.

In einem zweiten Schritt wurde mit dem Hidden Markov Model ein beispielhaftes Machine-Learning-Verfahren angewendet, um zu prüfen, ob sich ein solches Modell dahingehend trainieren lässt, dass es die Lernneigungen der Lernenden adäquat einschätzt. Als Trainingsziel (Ground Truth) — und damit gleichzeitig als Validierungs-Datensatz — standen zwei unterschiedliche Datenmengen zur Wahl. Einerseits sollte probiert werden, das Modell darauf zu trainieren, die bestehende, statistische Zuordnung zu reproduzieren. Andererseits sollte versucht werden, das Modell darauf zu trainieren, die von den Lernenden zu Beginn der Nutzung in Form eines Tests vorgenommene Selbsteinschätzung nachzubilden.

Interaction Pattern (Chatbot)

In der Vergangenheit hatte sich gezeigt, dass die Lernenden die Plattform generell sehr gut annehmen und sie nutzen, um mit dem Lernstoff und — wenngleich in geringerem Umfang — miteinander zu interagieren. Gleichzeitig hatte sich aber auch gezeigt, dass es nur zu wenig Interaktionen kam, die nicht auf die konkrete Seminararbeit ausgerichtet waren. Insbesondere hatten tutorielle Interaktionsangebote seitens der Plattform nur wenig Aufmerksamkeit gefunden, die Ergebnisse der Learning Analytics konnten daher nur schwer an die Studierenden zurückgespiegelt werden, und zahlreiche Features waren von den Studierenden unverstanden und ungenutzt geblieben. So hatten die Evaluationen früherer Projektphasen gezeigt, dass die Studierenden zum Beispiel nur sehr unklare Vorstellungen davon hatten, warum sie für bestimmte Leistungen von der Plattform Auszeichnungen (Badges) erhielten. Entsprechend sollte in dieser Förderphase ein Chatbot als neues didaktisches 'Interaction Pattern' entwickelt und im praktischen Einsatz getestet werden.

Da sich auch der Einsatz von Chatbots in der Hochschullehre noch am Anfang befindet, verstand das Projekt-Team seine Arbeit hier explizit als eine 'Pathfinder Mission', die entsprechend umfangreich in einem eigenen Whitepaper dokumentiert wurde. Nach einer Übersicht über bestehende

Chatbot-Lösungen innerhalb und außerhalb des hochschuldidaktischen Bereichs wurde die Entscheidung gefällt ein User Interface zu implementieren, das sich im Großen und Ganzen an den den Studierenden vertrauten Messenger-Apps (Whatsapp, Telegram, Signal, etc.) orientiert.

Dabei stand schnell fest, dass kein Natural Language Processing zum Einsatz kommen sollte sondern stattdessen vorstrukturierte Dialogbäume mit Auswahlmöglichkeiten geschrieben werden sollten. Für diese Entscheidung gab es eine Reihe von Gründen: Erstens sollte der Fokus auf der inhaltlichen Gestaltung der didaktischen Interaktionen liegen, so dass der Verzicht auf ein Natural Language Interface hier rein programmiertechnische Komplexität reduzierte und es erlaubte, die zur Verfügung stehenden Ressourcen auf die Gestaltung der Dialoge und die Erstellung des eigentlichen User Interface zu konzentrieren. Zweitens sind die verfügbaren Natural Language Processing derzeit noch in einem zu wenig befriedigenden Stadium und sozusagen im 'uncanny Valley' verortet: zu wenig perfekt, als dass eine sich wirklich natürlich anfühlende Interaktion möglich wäre, aber perfekt genug, als dass man als Nutzer:in einem Natural Language Bot dessen Fehler beim Spracherkennen übel nimmt. Drittens entspricht das Modell, das solchen vollständig gescripteten Dialogen zugrunde liegt, der medialen Logik des Hypertext-Graphen, der das gesamte SHRIMP-Projekt bestimmt: einzelne, kurze Texte (Knoten), die durch Links (Kanten) miteinander verbunden sind.

Der Chatbot wurde entsprechend im Herbst 2019 konzeptionell geplant. Im Sommer 2020 wurde das User Interface implementiert und getestet. Ebenfalls im Sommer wurde der Großteil der Dialoge geplant und geschrieben. Hierbei wurden Dialoge mit sehr unterschiedlichen primären Zielsetzungen erstellt; dadurch, dass zu unterschiedlichen Zeitpunkten im Semesterablauf unterschiedliche Zielsetzungen dominant wurden, ergab sich hierbei eine Art Semester-Dramaturgie: Zu Beginn des Semesters (und damit auch: zu Beginn der Nutzung der Plattform) standen Usability-Fragen im Vordergrund. Viele Dialoge konzentrierten sich darauf, Funktionen der Plattform zu erläutern oder den didaktischen Zweck einzelner Features, wie zum Beispiel der Leistungsauszeichnungen durch 'Badges', zu besprechen. Mit voranschreitendem Semester rückten Fragen von Motivation, von Lernstrategien, aber auch von Compliance ins Zentrum. Ziel der automatisierten e-tutoriellen Interaktion war hierbei, durch die im Chat vorgeschriebenen Frage-Antwort-Abläufe Selbstreflexion bei den Lernenden anzuregen. Für die auch angesichts des Corona-Semesters eingeführten „Mood-Dialoge“, die die Lernenden nach ihrem mentalen Wohlbefinden befragten, wurde ein neues Feature eingeführt: ein Freitext-Feld, das allerdings natürlich nur manuell und erst im Nachgang zum Dialog ausgewertet werden konnte. Zum Ende des Semesters hin rückte dann, unter anderem mit einem Quiz-Dialog, die Prüfungsvorbereitung ins Zentrum.

Der Chatbot wurde im Wintersemester 2020/21 erstmalig in der Lehre eingesetzt. Im Rahmen dieser abschließenden Durchführungsphase wurde die Interaktion der Studierenden mit dem Chatbot auf unterschiedliche Weise evaluiert: eine quantitative Auswertung der Metadaten zu Interaktionen zeigte, wie dieses neue Feature angenommen wurde. Eine Fragesektion innerhalb der regulären Seminarevaluation gab Aufschluss darüber, wie die Studierenden die Interaktion mit dem Chatbot empfunden haben. Schließlich wurde mit einer (leider sehr kleinen) Focus Group von zwei Studierenden ein Feedbackgespräch geführt, in dem noch etwas weiter in die Tiefe gegangen werden konnte.

Outcomes

Wie schon in der vorangegangenen Förderphase „SHRIMP Analytics“ hat sich das Förderformat dahingehend bewährt, als es erlaubt hat, eine interdisziplinäre Arbeitsgruppe fluktuierender Größe (vier bis sechs Personen) zusammenzustellen, in der Studierende unterschiedlicher Qualifikationsstufen, Promovierende und Mitarbeiter gemeinsam über einen längeren Zeitraum hinweg am Projekt arbeiten konnten. Zur Organisation der gemeinsamen Arbeit hat sich die Kombination aus Slack (Projektkommunikation), Kanboard (Task Management), Nextcloud (Datenspeicherung) und git (Quellcode-Verwaltung) bewährt.⁴ Die Organisation der Abläufe orientierte sich in Grundzügen an den Prinzipien des agile development.

Über den Großteil der Laufzeit war die Arbeit in Teams (‘Departments’) organisiert, die die Bereiche Coding & Analysis, Mediendidaktik und Outreach abdeckten. Die Aufgabe, die Dialoge zu gestalten und zu schreiben (scripting) wurde dabei als Querschnittsaufgabe von den zwei Departments Mediendidaktik und Outreach gemeinsam bearbeitet. Auch wenn sich die ausgesprochen hierarchiearme, selbstorganisierte Struktur bewährt hat, ist für mögliche künftige Projektphasen zu prüfen, ob eine dezidierte Funktion als ‘Scrum Master’ geschaffen werden sollte.

Networking & Outreach

Neben der Arbeit in den zwei Projektsträngen bildete Outreach, also die Vernetzung mit anderen hochschuldidaktischen Akteurinnen und Akteuren, die Sichtbarmachung der eigenen Arbeit und das Schaffen der Voraussetzungen für Verstetigung und Transfer, einen wichtigen Teil der Projektarbeit in der Förderphase.

Eine besonders herauszuhebende Vernetzungskomponente stellt die in dieser Förderphase begonnen Kooperation mit der Professur für Allgemeine Didaktik und Schulpädagogik des Sekundarbereichs am Institut für Bildungswissenschaften der Universität Leipzig dar. Dort hat ein Projekt-Team im Rahmen einer Förderung durch die Leipziger Labor-Universität damit begonnen, eine separate Installation von SHRIMP für die Lehre zu verwenden (vgl. [Interview mit Dr. Christian Herfter](#)). Durch den gleichzeitigen Einsatz in zwei fachlich unterschiedlichen Lehrkontexten ergeben sich erhebliche Synergien und neue Potenziale zur Erforschung der didaktischen Potenziale von Social Hypertext.

Als weitere Outcomes im Bereich Vernetzung und Outreach in der aktuellen Förderphase sind insbesondere zu nennen:

- Einrichtung eines **Twitter Accounts** unter dem handle [@projektshrimp](#).
- Präsentation der SHRIMP-Plattform bei der **Jahrestagung der Deutschen Gesellschaft für Hochschuldidaktik** im März 2019
- Präsentation bei der „**European Summer University ESU** Digitale Geisteswissenschaften“ in Leipzig im Juli 2019 im Rahmen einer [Poster Session](#).
- Präsentation beim „**Tag der Lehre**“ der Universität Leipzig im November 2019
- **Digital Humanities Day** der Universität Leipzig im Dezember 2019

⁴ Vgl. <https://slack.com>, <https://kanboard.org>, <https://speicherwolke.uni-leipzig.de>, <https://git-scm.com>.

- Erstellen der **OpenAccess-Reihe** *SHRIMP Reports*, innerhalb derer sukzessive alle Projektberichte, die innerhalb des Projekts erstellt werden, publiziert werden sollen. Die Reihe umfasst bereits die Projektberichte früherer Förderphasen und, unter dem Titel „Ansätze für eine datengestützte Analyse von SHRIMP“, eine erste Vorstudie zum Thema Learning Analytics. Die Reihe ist im Open-Access-Repository der Universität Leipzig langzeitarchiviert unter: <https://ul.qucosa.de/id/qucosa:72272>.
- Die **Zeitschriftenveröffentlichung** „Non-Linear Reading and Fragmented Texts: Leipzig’s Social Hypertext Reader SHRIMP and the ‘Introduction to American Studies’“ im *American Studies Journal*.
- Präsentation beim digitalen „**Tag der Lehre**“ der Universität Leipzig im Juni 2020
- Präsentation der Plattform durch die SHRIMP-Arbeitsgruppe der Leipziger Bildungswissenschaften bei der **Jahrestagung der Deutschen Gesellschaft für Hochschuldidaktik** im Mai 2020 in Berlin auf einer [Postersession](#).
- Die **Zeitschriftenveröffentlichung** „SHRIMP (Social Hypertext and Interactive Mapping Platform): Ein exemplarisches digitales Lehr-Lern-Setting in den textbasierten Geisteswissenschaften“, die für *Blickpunkt Hochschuldidaktik* angenommen wurde.
- Präsentation des Projektes im Rahmen des 2020 (digital stattfindenden) **Workshop on E-Learning**.
- Das **Video** <https://www.youtube.com/watch?v=PqKze42B8fE>, das SHRIMP der weiteren Öffentlichkeit vorstellt.
- Zwei Beiträge beim **Digital Humanities Day** im November 2020 in Leipzig: Sarah Doberitz zu „Digitales Tutoring per Chatbot“ und Christian Herfter zu „Lehren und Lernen mit SHRIMP: Social Hypertext im einführenden bildungswissenschaftlichen Modul“

Machine Learning

Im Rahmen dieses explorativ angelegten Arbeitsstrangs konnten wichtige Ergebnisse erzielt werden, auch wenn diese auf den ersten Blick ernüchternd scheinen: Die getesteten naiven Clustering-Verfahren (Optics, KMeans) konnten keine befriedigenden Cluster innerhalb der vorliegenden Lerndaten identifizieren. Auch hierzu sind Details im entsprechenden Bericht veröffentlicht. Zusammenfassend lässt sich sagen, dass die Algorithmen dazu neigten, die Lernenden entlang der Achse „intensive Nutzung vs. geringe Nutzung“ aufzuteilen. Innerhalb des Clusters „intensive Nutzung“ ließen sich vereinzelt Andeutungen von Clustern unterschiedlicher Nutzungspräferenzen identifizieren. Diese waren aber (auch angesichts der zur Verfügung stehenden Gesamtnutzer:innenmenge) zu klein und zu divers, um als gültige Cluster anerkannt zu werden.

Ebenso konnte das geprüfte Machine-Learning-Verfahren (Hidden Markov Model) nicht so implementiert werden, dass es die als Ground Truth dienenden Zuordnungen (das Ergebnis der statistischen SHRIMP-Klassifikation oder das Ergebnis des Lernneigungstests zu Beginn des Semesters) mit befriedigender Genauigkeit hätte erlernen und reproduzieren können. Je nach Parametrisierung ergaben sich unterschiedliche gute Lernerfolge. In jedem Fall aber blieb ein Großteil der Datensätze unklassifiziert bzw., gemessen an der Ground Truth, ‘falsch’ klassifiziert.

Sowohl beim Clustering als auch beim Trainieren eines Machine-Learning-Modells hätte es noch weitere Spielräume zur Parametrisierung gegeben, die die Ergebnisse möglicherweise in voraussichtlich engen Grenzen verbessert hätten. Aus Sicht des explorativen Projektzuschnitts war aber das Experimentieren mit unterschiedlichen Herangehensweisen und die Reflexion der daraus ableitbaren Schlüsse das übergeordnete Ziel. Dies gilt umso mehr, als rückblickend fraglich ist, ob die zur Analyse zur Verfügung stehenden Daten in Qualität und Volumen ausreicht, um Machine-Learning-Verfahren erfolgreich zum Einsatz zu bringen: insgesamt standen Daten aus drei Jahren mit je ca. 80-90 Studierenden zur Verfügung. Diese Datensätze waren allerdings in sich noch einmal heterogen, da die Plattform in dieser Zeit stetig weiterentwickelt worden war, was wiederum dazu geführt hatte, dass die die zur Verfügung stehenden Interaktionen und die dabei anfallenden Nutzungsdaten variabel waren.

Interessanter als die Frage nach der Qualität der zugrundeliegenden Daten ist aber die zusätzliche Perspektivierung der zugrundeliegenden Klassifikation der Lernenden, die der Einsatz von Machine-Learning-Verfahren leistet: Die naiven Clustering-Verfahren haben die Unterscheidung zwischen intensiver und zurückhaltender Nutzung als ein besonders salientes Unterscheidungsmerkmal identifiziert. Es liegt nahe, diese Unterscheidung in enthusiastische und vermeidende Nutzer:innen noch stärker ins Zentrum zu rücken und auf ihre Korrelation mit anderen Faktoren und auf ihre didaktische Nutzbarkeit hin abzuklopfen. Wichtiger noch ist, dass bei der Auswertung der Machine-Learning-Ergebnisse weitere, bisher nicht erfasste potenziell charakteristische Parameter in den Blick rückten (insb. gibt es Indizien, dass der zeitliche Verlauf von Interaktionen, ob also Features zum Beispiel kontinuierlich oder schubweise genutzt werden, dabei helfen kann, Lernneigungen und Lernneigungsprofile zu charakterisieren).

Die Schwierigkeiten innerhalb der vorliegenden Datensätze stabile Cluster zu identifizieren, wirft aber noch grundsätzlicher die Frage auf, ob die Lernneigungen der Studierenden als stabile Charakteristika oder als sich dynamisch entwickelnde Reaktionen auf mediale Affordanzen und Lern-Interaktions-Angebote zu konzeptionalisieren sind. Eine solche dynamischere Konzeptualisierung stünde wiederum in einem Spannungsverhältnis zur Eigenwahrnehmung der Studierenden (und zum didaktischen Potenzial einer Selbstreflexion), wie sie sich in den Evaluationsergebnissen dieser und der vergangenen Projektphase zeigt. Dort erklärten die Studierenden in dieser und der vorangegangenen Projektphase, dass die Klassifizierung beim Verstehen und Reflektieren des eigenen Lernverhaltens hilfreich gewesen sei und auch dazu beigetragen habe, sich als Teil einer Gruppe innerhalb des sozialen Gefüges des Seminars zu fühlen.

Interaction Pattern (Chatbot)

Wichtigstes und greifbarstes Ergebnis dieses Arbeitsstrangs ist selbstverständlich das in dieser Förderphase entwickelte, implementierte und im praktischen Einsatz bewährte Chat-Interface, durch das der Chatbot im Wintersemester 2020-21 erstmalig mit den Lernenden kommuniziert hat.

Neben der reinen software-technischen Umsetzung konnte hier eine Reihe für vergleichbare Projekte wegweisender konzeptioneller Entscheidungen getroffen werden, die jeweils als 'Lessons Learned' in dem vom Projekt veröffentlichten Whitepaper zu finden sind. Zu nennen ist hier zum Beispiel, wie sehr die Akzeptanz des Chatbot als Kommunikationspartner / der Eindruck einer tatsächlichen Chat-Experience von vergleichsweise kleinen Entscheidungen im UI/UX-Design abhing. So konnte beispielsweise eine inhaltsabhängige Verzögerung der Antworten des Bot und die

Einführung eines 'typing-Indicators' den Realismus der Simulation eines Gesprächspartners deutlich erhöhen. Ein weiteres wichtiges Ergebnis der konzeptionellen Arbeit war die Einführung von 'Channels' in denen die Dialoge ablaufen. Dadurch, dass die Dialoge jeweils in sich geschlossen modelliert und als einzelne, voneinander unabhängige Channels angeboten werden konnten, konnte sichergestellt werden, dass die Nutzer:innen zu mehreren Themen gleichzeitig mit dem Chatbot im Dialog stehen konnten. Schließlich zeigte sich bei der Umsetzung der eigentlichen Dialoge auch, wie wichtig die Arbeit mit 'Personas', also mit ausformulierten Charakteren für die Lernenden und für die Figur des digitalen Tutors Shrimpy war (vgl. im Whitepaper, Abschnitt „3.2 Virtuelle Charaktere: Die Persona des Chatbot und der User:innen“, S. 5-7).

Beim Experimentieren mit der Persona des Chatbot und dem Genre des Chattens zeigte sich ein weiteres, in der Projektplanung noch unvorhergesehenes Einsatzgebiet des Chatbots: die Einbeziehung des mentalen Wohlbefindens der Lernenden. Wie sich zeigte, ist ein Chatbot gut geeignet, nicht einfach nur Lernhandlungen zu thematisieren oder zur Reflexion über das eigene Lese- und Lernverhalten einzuladen und anzuleiten. Innerhalb der Genrekonventionen des Chattens liegt es nahe, diese Aspekte mit Fragen nach dem Wohlbefinden zu verknüpfen und so unter anderem den Blick der Lernenden für den Einfluss des eigenen Wohlbefindens auf das Lesen und Lernen zu schärfen. Diese Ebene der Interaktion erhielt durch die während der zweiten Hälfte der Förderphase andauernde Covid-19-Pandemie, in der viele Studierende sich erheblichen psychischen Belastungen ausgesetzt sahen, zusätzliches Gewicht.

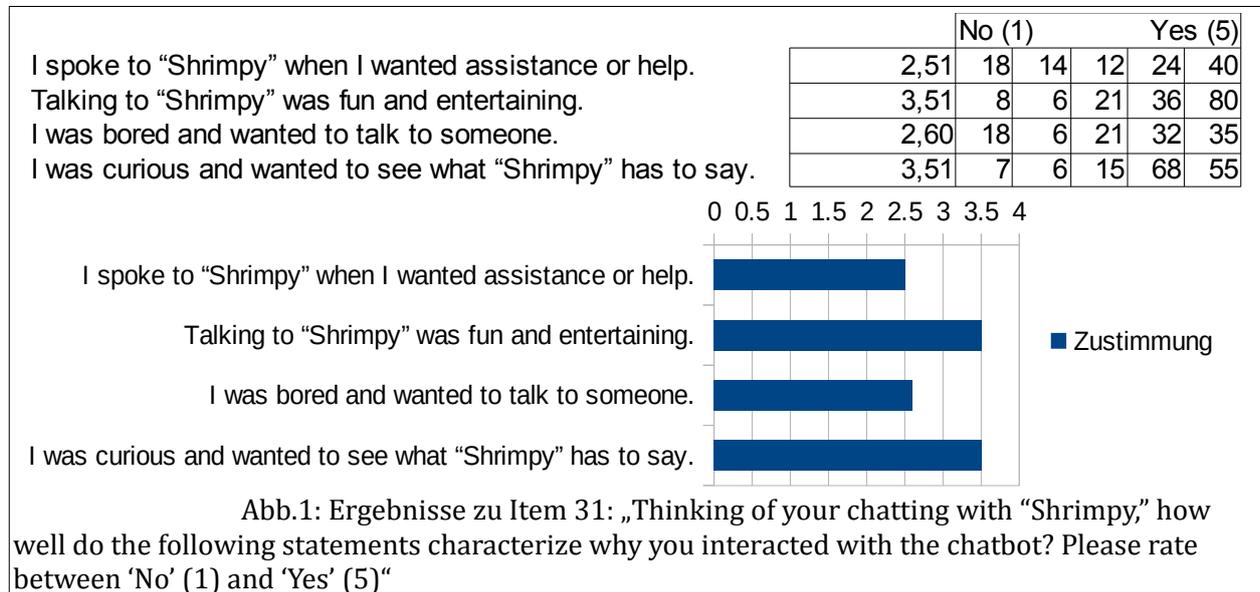
Evaluationen

Am Ende der zweiten Durchführungsphase wurde SHRIMP umfangreich evaluiert. Hierzu wurde erstens eine längere Sektion in die reguläre Seminarevaluation (per Fragebogen) eingefügt, in der die allgemeinen SHRIMP-Features, die Learning Analytics und die in diesem Semester neue Chatbot-Funktion thematisiert wurden. Zusätzlich wurde mit einer Focus Group aus zwei freiwilligen Studierenden ein 90-minütiges qualitatives Interview durchgeführt. Schließlich gab auch eine statistische Auswertung der Interaktionen mit dem Chatbot Auskunft über Nutzungsmuster dieses neuen Features.

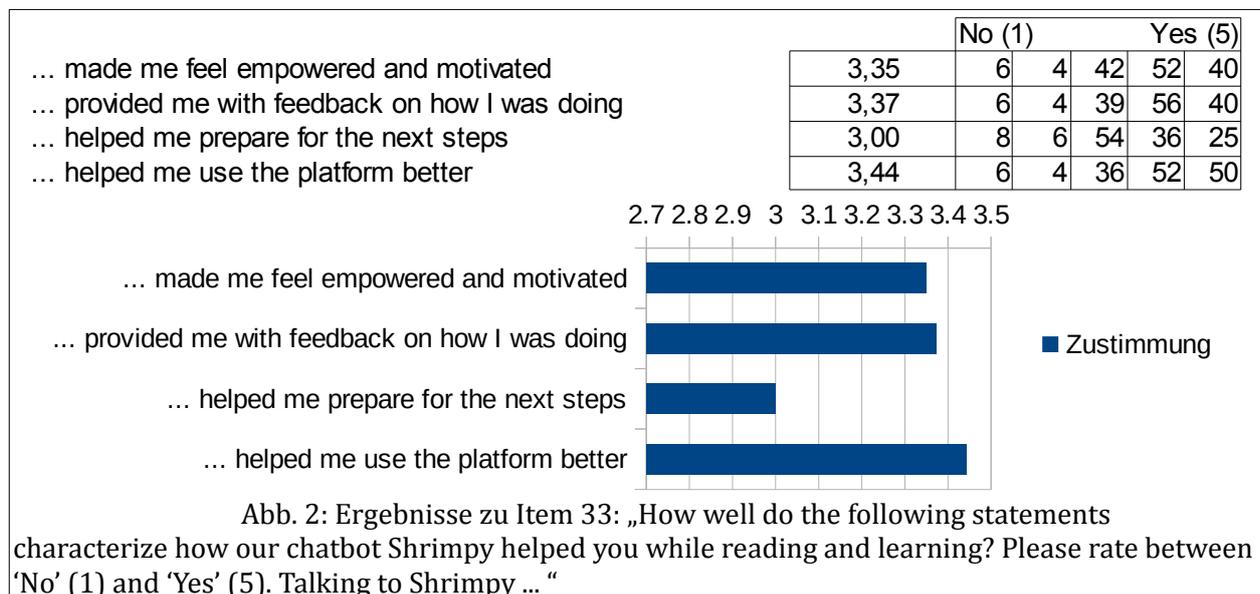
Die **statistische Auswertung** der Nutzungsdaten ergab zum Beispiel, dass die Nutzer:innen die Interaktion mit dem Chatbot mehrheitlich zeitlich konzentriert vornehmen. Es ergeben sich also Nutzungsphasen, in denen die Nutzung der Plattform unterbrochen wird, um stattdessen zu chatten. Außerdem zeigte sich, dass nur ein Teil der Dialoge bis zum Ende (also der Verabschiedung) durchgespielt werden. Selbstverständlich reagieren die Nutzer:innen ohnehin nicht auf alle Gesprächsangebote, so dass ein gewisser Anteil der Dialoge komplett unbeantwortet bleibt. Je Dialogbaum haben aber nie mehr als ca. 15% der Nutzer:innen einen Dialog bis einer Verabschiedung durchlaufen, selbst wenn sie das Gespräch eine Weile am Laufen gehalten haben. Beide Ergebnisse sprechen dafür, was sich auch in den anderen Evaluationen widerspiegelt, dass die Nutzung als 'Pausentätigkeit' innerhalb des Lesens genutzt wurde.

Innerhalb der **Seminar-Evaluation per Fragebogen** zeigte sich ebenfalls, dass die Studierenden mit dem Chatbot vor allem aus Neugier und aus einem Unterhaltungsbedürfnis heraus interagiert haben. Nur vergleichsweise selten war die Suche nach Hilfe oder Unterstützung ein entscheidendes Motiv bei der Interaktion (Item 31, vgl. Abb. 1). Exemplarische Freitextantworten zu weiteren

Gründen, mit dem Chatbot zu sprechen, unterstreichen dies („It's fun! Shrimpy is very cute!“; „General trolling and lols“), sprechen aber auch eine weitere Motivation an, auf die unten noch einmal (u.A. im Punkt „Ausblick“) eingegangen werden wird: „He asked how I feel“ (Item 32)



In der Tat waren eine gesteigerte Motivation und eine positive Grundhaltung für die Lernenden zentrale Aspekte, die sie an der Interaktion mit dem Chatbot schätzten. Auf die Frage, wie der Chatbot ihnen geholfen habe, werden neben dem offensichtlichen Zweck „helped me use the platform“ in dichter Folge die Punkte „provided me with feedback on how I was doing“ und „made me feel empowered and motivated“ genannt (Item 33. vgl. Abb. 2). Das spiegelt sich auch in den Freitextantworten zur Frage „How would you describe your experience with our chatbot Shrimpy?“ wider. Exemplarische Antworten sind hier: „It was fun and Shrimpy is cute. It makes the whole website more friendly. Just good vibes“; „I really liked the interaction with Shrimpy and I would not change anything, I was always motivated when Shrimpy asked about my mood or how I am doing. The Memes were also hilarious“ (Item 34).



Auch wenn die Seminar-Evaluation zeigte, dass noch immer ein zu kleiner Teil der Studierenden die Learning Analytics Features in vollem Umfang wahrgenommen hat, ergab die Evaluation hier insgesamt positive Ergebnisse. Auf die Freitext-Frage „We identified four groups of users on SHRIMP [...] How did it feel to be characterized as one of them? Would you propose another name for the way you interact with SHRIMP? If so, please tell us which and briefly describe your use of SHRIMP“ waren exemplarische Antworten: „I liked the groups and can see why I got identified as a ‘member’ of a certain group“; „I like the idea. It helped me better understand my reading habits“; „I think it was helpful to be categorized in a way that I felt motivated to explore the assets of other reading strategies“; „It was helpful and empowering since it has a positive connotation! It is a nice feature“; etc. (Item 30).

Neben diesen Aussagen zur Nutzung des Chatbot und den Learning Analytics bestätigte die restliche Seminarevaluation das der Plattform zugrundeliegende didaktische Konzept. Auf die Freitext-Frage, ob die Plattform die Lesegewohnheiten verändert hätte, antworteten die Studierenden beispielsweise: „It changed my reading habits in a positive way. I am now more able to read and comment on something at the same time and already make up my mind about what the specific sentence might mean and how different topics interact with each other“; „it helped to explore more content on a specific topic and also helped to see connections with different other readings“; „a bit because of the bookmarks and the green progress markers – those are really motivating“ (Item 26). Ergebnisse, die von der ebenfalls offenen Frage „Do you think SHRIMP helped you improve certain skills? If yes, which ones?“ bestätigt wurden: „I really learned to make annotations“; „it made me more of an explorer when reading“; „Reading faster, annotating, and thinking more critically while reading“; etc. (Item 27).

Viele der Ergebnisse der Seminar-Evaluation, insbesondere zum Thema der Chatbot-Nutzung, wurden schließlich auch durch das **qualitative Interview** mit zwei Studierenden bestätigt und zusätzlich perspektiviert. Beide gaben an, dass ihre Nutzung der Chatbot-Funktion vor allem spielerisch war und das Ziel hatte „erst einmal so [eine] kleine Pause, kurze Ablenkung“ zu leisten bevor „man wieder zurück zum [...] Lesen gehen“ kann. Anders gesagt: der mediale Bruch zwischen Lesen und Chatten, und die Möglichkeit, das Lesen durch das Chatten zu unterbrechen, erlaubten es, sich vom Lesen zu erholen, „erstmal ganz kurz [den] Kopf frei zu bekommen“, und „trotzdem irgendwie bei der Sache [zu] bleiben“ (Zeile 301-5; auch 88-100, 258-70). In den Augen der Studierenden erlaubte die Chatfunktion also kleine Ausweichbewegungen, die im Lernprozess notwendig waren. Während ohne den Chatbot die Studierenden womöglich auf andere Ablenkungsangebote ausgewichen wären (YouTube, Social Media), konnten sie durch die Chatbot-Funktion „irgendwie bei der Sache bleiben“.

Ein zweiter Punkt, der im Interview auffällt, ist dass die beiden Studierenden explizit einen Zusammenhang zwischen der Nutzung von Learning Analytics und Chatbot herstellen. Da der Chatbot bisher nicht zur Rückspiegelung der Learning-Analytics-Ergebnisse genutzt werden konnte, ist der Überschneidungspunkt hier die „Mood Survey“ die zwei Mal im Semester per Dialog durchgeführt wurde. Hier fragte der digitale Tutor Shrimpy die Studierenden, wie es gerade um ihre emotionale Situation bestellt sei. Die Ergebnisse wurden manuell ausgewertet und in einem daraufhin neu geschriebenen Dialog an die Studierenden zurückgespiegelt. Unter Bezugnahme auf diese Mood Surveys nannten die Studierenden als zentrales Motiv für ihre Interaktion mit dem

Chatbot, ein soziales Interesse: mehr darüber zu erfahren, wie es den anderen gerade geht bzw. das soziale Gefüge der Seminargruppe und die eigene Rolle darin besser zu verstehen (368-401).

Bemerkenswerterweise wurde an mehreren Stellen im Interview von den Studierenden thematisiert, dass sie dabei ihr Gegenüber, den Chatbot Shrimpy, als eine 'Person' wahrgenommen haben, und dass die Charakterzüge dieser Person ein wichtiger, wenn nicht sogar der zentrale Faktor für die Interaktion war (226-250). Mehrere Passagen machen das deutlich, zum Beispiel antworten die Studierenden auf die Frage „zuerst, würden wir gerne wissen, wie würdet ihr den Chatbot 'Shrimpy' beschreiben?“ wie folgt:

- 220 B: ~((lachen)) #00:15:30-1#
- 222 A: ~ ((schmunzeln)) #00:15:30-1#
- 226 B: Also es ist ja so ne Art ähm Wesen so ähnlich jetzt äh (...) ((stöhnt)) dieses Duolingo Bird, also so in diese Kategorie hab ich das mir selber so eingeordnet. Äh. Ja aber äh, schon irgendwie nett Ähm, manchmal n bisschen so äh (...) ähm (o...noxious) sag ich mal, also so aufpoppend. Äh (..) ja ähm, aber also wenn man's jetzt zum Beispiel mit diesem Duolingo Bird vergleicht, äh- also da hat man ja diese ganzen Witze gemacht, dass es so passiv-aggressiv wirkt. Äh also das finde ich jetzt bei Shrimpy nicht so. Ja, das wirkt eigentlich ganz nett. Ja. #00:16:38-0#
- 233 A: Ja, das find ich auch und ich fand's auch nicht irgendwie krass penetrant sag ich mal, sondern ich fand's voll in Ordnung. Weil auch wenn dann irgendwas neues im Chat kam, kam dann nicht so ne Riesenmeldung irgendwie, dass da was Neues ist. Sondern man hat's halt einfach gesehen, wenn man hochgeschaut hat. Äh und ich geb B voll recht, im Vergleich zu anderen ähnlichen Bots, sag ich mal, wie zum Beispiel dem Duolingo Bird ((schmunzeln)), fand ich's auch sehr entspannend. Man hat sich auch irgendwie nicht-, also keine Ahnung. Man hat sich auch irgendwie nicht schlecht gefühlt, wenn man mal doch was verpasst hat und irgendne Meldung kam. Also ja, es war einfach alles sehr positiv aus-, ausgelegt. Was ich sehr gefreut hat, dass fand ich sehr angenehm. Ähm ja. #00:17:27-8#
- 242 Frage: Gibt's es noch besondere Charaktereigenschaften oder sowas ((lachen)), die ihr noch nennen wollt oder wie würdet ihr den so charaktermäßig beschreiben kurz? #00:17:36-7#
- 245 A: Sehr unterstützend, finde ich ((lachen)). Das fand ich immer sehr niedlich, ähm, wenn dann () sowas kleines Unterstützendes kam oder auch wenn, wenn man- bei den Umfragen war das glaube ich-, wenn man dann doch irgendwie nicht alles gelesen hat oder so (dann) "Ah ja leider nicht so viel wie geplant, aber das ist völlig in Ordnung" und so. Also keine Ahnung, dass fand ich sehr unterstützend und das hat einen auch nicht so irgendwie schuldig fühlen lassen oder so. Ähm ja. #00:18:04-8#
- 252 B: Hmm ja, unterstützend. Ich hatte als Wort so fürsorglich im Kopf. Äh (...) Ja ((lachen)) #00:18:18-1#

Wie erhofft nennen die Studierenden dabei die visuellen Aspekte des Interface und die 'Rundheit' des Charakters des Chatbot als einen wichtigen Faktor, der die Interaktion 'natürlich' gemacht hat, auch ohne dass ein natursprachliches Interface zur Verfügung gestanden hätte:

- 272 Frage: Okay. Und, was hat, hat euch noch so gefallen-, also was hat euch besonders gefallen beim Chatten mit Shrimpy? Was würdet ihr sagen? #00:20:01-2#
- 275 B: Äh, also ich habs ja auch schon mal angesprochen. Dieses äh, das wenn man jetzt ne Antwort schickt sozusagen aus diesen Antworten, die man als Optionen hat, das dann äh dieses äh Prozess, äh dass die Antwort von Shrimp vorbereitet wird, das es so sichtbar ist und jetzt nicht sofort kommt so. Also das hat so äh organischer ((schmunzeln)) gemacht vielleicht. Ähm ja genau (...) und die Memes wie schon gesagt. Äh. Ja. ((schmunzeln)) #00:20:50-4#
- 281 A: Ich fand auch tatsächlich, dass es sich manchmal auch nicht so wirklich angefühlt hat, als würde man unbedingt mit nem Chatbot schreiben. Weil halt, wie B schon gesagt hat, dass dieser Prozess, sag ich mal, sehr organisch gestaltet ist. Äh dadurch das man halt immer n paar Antwortmöglichkeiten irgendwie hat, kann halt auch, also ka-, könn hal-. Wenn's richtig vorbereitet ist, kann das halt sehr schön sein, dass das einfach nicht so wie n Bot scheint, der einfach Antworten raushaut, sondern es fühlt sich wirklich so n bisschen an, als würde man mit jemand so schreiben. Ähm also ich fands einfach sehr angenehm wie organisch das Ganze war und sich nicht unbedingt wien Chatbot angefühlt hat teilweise. Das fand ich sehr angenehm. #00:21:32-2#

290 B: Und auch dieses ähm- Also oft war's dann so am Ende von einem bestimmten Gespräch, dass sowas kam wie "Ja ich bin jetzt wieder im Ozean schwimmen" oder sowas. Das war eigentlich auch immer ganz süß ((schmunzeln)). Ja. #00:21:50-4#

Auffällig ist hierbei, dass auch in der Interaktion mit dem Chatbot die soziale Komponente und, letztlich, die Beziehungsebene mit der — natürlich trotzdem zu jedem Zeitpunkt als künstlich verstandenen — Figur Shrimpy im Vordergrund steht. Auf einer zweiten Ebene zeigt sich hier noch einmal die Interaktion mit dem Chatbot als eine Form des Spiels, in der die Studierenden bereit sind, die Künstlichkeit der Interaktion zu ignorieren, um sich in diesen spielerischen Austausch begeben zu können.

Ausblick und künftige Forschungsfragen

Aus den Erfahrungen dieser Förderphase und aus den Evaluationen ergeben sich eine Reihe von Forschungsfragen und möglichen Perspektiven für eine Weiterentwicklung der Plattform und der genutzten Interaktionsmuster.

Das Interesse der Studierenden an der sozialen Dimension der Learning Analytics, das sich schon in der vergangenen Förderphase abzeichnete, und das nun noch einmal stark unterstrichen wurde, lohnt, weiter erforscht zu werden: offensichtlich ist für die Studierenden ein zentraler, motivierender Reiz der Learning-Analytics-Komponente etwas über sich selbst und insbesondere über sich selbst im Kontext des Sozialverbunds des Seminars zu erfahren. In Zukunft sollten daher Wege gefunden werden, kontinuierlicher und in kleineren Dialogen einzelne, laufende Ergebnisse aus der Learning-Analytics-Komponente an die Studierenden zurückzuspiegeln.

Insgesamt hat sich gezeigt, dass die medialen Affordanzen des Chattens (und insbesondere der automatisierten tutoriellen Betreuung innerhalb dieses Mediums) weiterer Erforschung bedürfen. Schon jetzt zeichnet sich ab, dass Unvollständigkeit und Wiederholung von uns in dieser Projektphase unterschätzte Aspekte dieses Mediums ist. Während die Studierenden beim Lesen der Seminarlektüre, trotz aller Ermutigung zum kursorischen Lesen ein starkes Bedürfnis nach Linearität, Abgeschlossenheit und Vollständigkeit an den Tag legen, scheinen diese Bedürfnisse beim Chatten außer Kraft gesetzt. Entsprechend eignet sich das Medium möglicherweise besonders für potenziell repetitive, dafür aber kürzere 'Gespräche'.

In dieser Projektphase waren vergleichsweise komplexe Dialogbäume konstruiert worden, um die thematische Variabilität eines natürlichen Gesprächs möglichst gut abzubilden. Nach derzeitigem Stand sollte in Zukunft dagegen eher mit kürzeren, repetitiveren Dialogbäumen experimentiert werden. Dies gilt insbesondere für das Feedback zu den Learning Analytics, das Anbieten von Lerntips, und für Compliance-Maßnahmen (Mahnen zu regelmäßigerer / intensiverer Arbeit). Die Tatsache, dass das Machine Learning vor allem gut darin war, Geringnutzer:innen zu identifizieren, ließe sich hier möglicherweise nutzen, um motivationell abhängende Studierende frühzeitig zu identifizieren und mit kurzen Dialogen vor einem weiteren Abrutschen in die Passivität zu bewahren.

Das wohl unerwartetste Ergebnis dieser Förderphase ist der Bedarf der Studierenden nach emotionaler Unterstützung durch den automatischen Chatbot. Bereits im Zusammenhang mit der ersten „Mood Survey“ hatten Studierende vereinzelt das Feedback gegeben, dass die Frage nach

dem Wohlbefinden ihnen viel bedeutet habe (Eine Freitext-Antwort in der Survey lautete: „I just wanted to say thank you for asking how I was, it’s just a simple question but it means a lot to me.“). Die Seminar-Evaluation und das Interview haben diesen Aspekt noch einmal stark unterstrichen.⁵

Weder das Bedürfnis nach spielerischer Ablenkung, noch das Bedürfnis nach zusätzlicher Motivation durch den digitalen Tutor, noch das Bedürfnis nach sozialer Einbettung und emotionaler Unterstützung sind dabei als didaktische Nebenschauplätze zu bewerten. Vielmehr zeigen die Ergebnisse dieser Förderphase, dass diese vermeintlich ‘weichen’ Faktoren von den Lernenden selbst als wichtige Elemente des eigenen Lernprozesses wahrgenommen werden. Bei der Konzeption des Chatbot zu Beginn der Förderphase war dessen emotionale Dimension, seine ‘Persona’, noch vor allem als Ersatz für das fehlende natursprachliche Interface gedacht. Zum Ende der Förderphase zeigt sich aber das enorme didaktische Potenzial, das in sozialen Dimension dieses Interaktionsmusters liegt. Ob dieses Potenzial zuverlässig und dauerhaft genutzt werden kann, oder ob das Interesse der Studierenden an dieser spezifischen Ebene auch mit den emotionalen Herausforderungen des zweiten Corona-Semesters und dem hohen Grad an sozialer Isolation durch die damit einhergehenden ‘Lock-Downs’ zusammen hängt, können nur künftige Arbeiten zu diesem Thema klären.

5 In der Tat ist aus den Frühzeiten der Forschung zu Chatbots überliefert, dass Nutzer:innen die Gespräche mit dem Chatbot ELIZA als so intim einstufen, dass sie ein gewisses Maß an Schutz der Privatsphäre für diese Gespräche einforderten (vgl. Weizenbaum 6-7).

Referenzen

- Doberitz, Sarah, et al. "Implementierung eines Chatbots in eine Online-Lernplattform im Hochschulbereich am Praxisbeispiel von SHRIMP." *SHRIMP Reports*, (zur Veröffentlichung angenommen).
- Doberitz, Sarah, and Sebastian M. Herrmann. "SHRIMP (Social Hypertext and Interactive Mapping Platform): Ein exemplarisches digitales Lehr-Lern-Setting in den textbasierten Geisteswissenschaften." *Ursprünge hinterfragen - Vielfalt ergründen - Praxis einordnen: ReGeneration Hochschullehre*, edited by Angelika Thielsch et al., vol. 138, 2021, pp. 131-43.
- Herrmann, Sebastian M. "Non-Linear Reading and Fragmented Texts: Leipzig's Social Hypertext Reader SHRIMP and the 'Introduction to American Studies.'" *American Studies Journal*, vol. 69, 2020.
- Weizenbaum, Joseph. *Computer Power and Human Reason : From Judgment to Calculation*. San Francisco : W. H. Freeman, 1976. *Internet Archive*, <http://archive.org/details/computerpowerhum0000weiz>. Accessed 27 Mar. 2021.
- Wendt, Franz. "Ansätze für eine datengestützte Analyse von SHRIMP." *SHRIMP Reports*, no. 3, 2017, <https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-725045>.
- . "SHRIMP Machine Learning Report." *SHRIMP Reports*, (zur Veröffentlichung angenommen).