Ein wertesensibles Design für formatives Feedback mit Trusted Learning Analytics und KI[1]

Heike Karolyi,[2] Lars van Rijn,[3] Michael Hanses[4] und Claudia de Witt[5]

Der Einsatz von Künstlicher Intelligenz (KI) ermöglicht die Generierung von skalierbarem personalisiertem, hochinformativem Feedback (HIF) an Hochschulen. Die Verfügbarkeit von Large Language Models (LLMs) kann an Hochschulen durch lokale Installationen vertrauenswürdig umgesetzt werden. Beim Einsatz der LLMs im Zusammenhang mit Feedback in der Hochschullehre sollten jedoch eine Reihe von Rahmenbedingungen beachtet werden, um die von den Lehrenden definierten Inhalte und Kriterien qualitätsgesichert umzusetzen. Bei den im Forschungszentrum CATALPA an der FernUniversität in Hagen entwickelten KI-Anwendungen zu formativem Feedback wurden im Rahmen eines Value Sensitive Design (VSD)-Prozesses konzeptionelle, empirische und technologische Anforderungen integriert, die sich aus dem TLA-Ansatz von Hansen et a. (2020) ergeben. So wird ein vertrauenswürdiger Einsatz von Learning Analytics (LA) und KI bereits in der Anwendungsentwicklung sichergestellt.

A Value Sensitive Design for formative Feedback with Trusted Learning Analytics and AI

The use of artificial intelligence (AI) enables the generation of scalable, personalized, highly informative feedback (HIF) at universities. The availability of Large Language Models (LLMs) can be reliably implemented at universities through local installations. However, when using LLMs in connection with feedback, several conditions should be considered in order to implement the content and criteria defined by the teachers in a quality-assured manner.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID-ID: 0000–0002-8587-9530

3 ORCID-ID: 0000–0002-8381-2666

4 ORCID-ID: 0009–0004-3365-6273

5 ORCID-ID: 0000–0001-6478-8392

In the LA and AI applications for formative feedback developed at the CATALPA research center at the FernUniversität in Hagen, conceptual, empirical and technological requirements resulting from the TLA approach of Hansen et al. (2020) were integrated as part of a Value Sensitive Design (VSD) process. This ensures the trustworthy use of learning analytics (LA) and AI as early as the application development stage.

Einleitung

Um KI-Anwendungen ethisch, rechtlich sowie sozial verträglich zu gestalten und zu nutzen, wurde im Forschungszentrum CATALPA an der FernUniversität in Hagen der Ansatz zum wertesensiblen Design (eng. Value Sensitive Design (VSD), vgl. Spiekermann 2021; Friedman & Hendry 2019) für die Entwicklung von Feedbackanwendungen genutzt. Basierend auf einer Triangulation von (1) konzeptionellen, (2) empirischen und (3) technologischen Untersuchungen, unterstützt ein VSD ein umfassendes Verständnis von Anforderungen und Bedürfnissen in der Anwendungsentwicklung. Damit konnten systematisch Designentscheidungen werteorientiert getroffen werden, um eine automatisierte personalisierte Feedbackanwendung mit besonders lernförderlichem *hochinformativem Feedback* (HIF) (Wisniewski et al. 2020, Hattie 2024) durch Learning Analytics (LA) und KI umzusetzen. Ein solches Feedback zu skalieren, ist in Anbetracht der knappen Betreuungskontingente von Lehrenden und Dozierenden in großen Kursen an Hochschulen nicht ohne technische Unterstützung möglich, bietet aber einen klaren Mehrwert für Lernende. Rahmenbedingungen wie der Ansatz von Trusted Learning Analytics (TLA) (Hansen et al. 2020), Datenschutz, Urheberrecht und EU-KI-Verordnung (AI-Act) erfordern einen sensiblen Umgang mit KI und wurden in der wertesensiblen Entwicklung in zwei KI-basierten Anwendungen realisiert: eine Anwendung für das Selbstmonitoring und eine Anwendung zu korrektivem Feedback für automatisierte Rückmeldungen zu langen Freitextaufgaben.

Das Projekt IMPACT und Trusted Learning Analytics (TLA)

In dem Bund-Länder- geförderten Verbundprojekt IMPACT werden an fünf Hochschulen KI-basierte Lösungen zum Einsatz von Chatbots, zum formativen – bzw. summativen Feedback entwickelt und in die Implementierung gebracht. Das Verbundprojekt folgt dabei dem Ansatz von TLA nach Hansen et al.

(2020) mit einem umfassenden Verhaltenskodex, der Ziele und Prinzipien für den Einsatz von Daten in Bildungseinrichtungen festlegt. Die Hauptziele bestehen darin, Daten ethisch verantwortungsvoll, zielgerichtet und transparent einzusetzen, um sichere und vertrauenswürdige Systeme in die Anwendung zu bringen und sicherzustellen, dass alle Beteiligten sich mit dem Verhaltenskodex identifizieren.

Um diese Ziele zu erreichen, basiert der TLA-Ansatz (Hansen et al. 2020) auf sieben Prinzipien: Durch den Einsatz von LA sollen die *Bedingungen für Studium und Lehre verbessert (1)*, *Unterstützungsangebote für alle Studierende (2)* bereitgestellt, ein *transparenter Umgang mit Daten (3)* gepflegt, und ein *kritischer Umgang mit Daten (4)* gewährleistet werden. M*enschliche Kontrolle (5)* soll sicherstellen, dass Entscheidungen auf der Grundlage von Daten nur nach menschlichen Vorgaben und Kontrolle getroffen werden. Das Prinzip der *Führungsverantwortung (6)* stellt sicher, dass Daten ethisch und verantwortungsvoll genutzt werden, z. B. nach festgelegten Richtlinien und Standards. Die *Verpflichtung zu Weiterbildungsangeboten (7)* erfordert sowohl regelmäßige Teilnahme an als auch die Bereitstellung von Weiterbildungen.

Die Einhaltung dieser Prinzipien trägt dazu bei, dass vertrauenswürdige Anwendungen im Hochschulstudium nachhaltig im Einsatz bleiben. Bei der Gestaltung und Entwicklung von Feedback-Anwendungen müssen neben den Rahmenbedingungen zudem verschiedene Perspektiven in Abstimmung gebracht werden. Im Teilprojekt der FernUniversität in Hagen wurde für die Anwendungsentwicklung zum formativen Feedback nach dem Ansatz der wertesensiblen Gestaltung nach Friedman & Hendry (2019) gearbeitet.

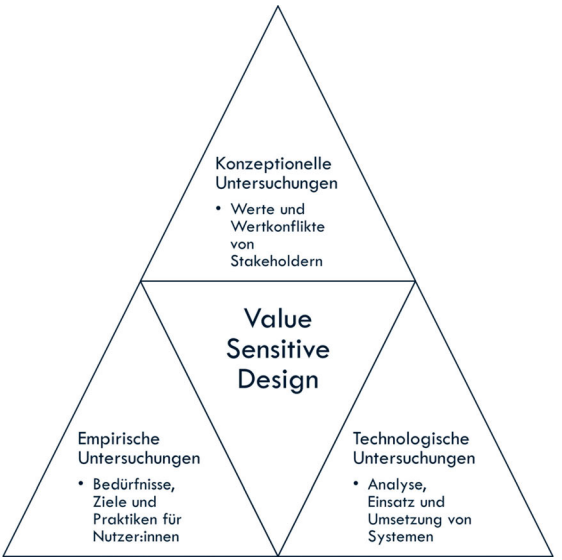
Wertesensible Gestaltung von Feedback-Anwendungen

Die wertesensible Gestaltung von Systemen ist ein zentrales Anliegen bei der Entwicklung von Technologien, um menschliche Werte und Bedürfnisse zu berücksichtigen (Spiekermann 2021; Friedman & Hendry 2019).

Die Triangulation aus konzeptionellen, empirischen und technologischen Untersuchungen, ermöglicht es, ein umfassendes Verständnis von Anforderungen, der aktuellen Praxis und Bedürfnissen zu TLA zu berücksichtigen und so systematische Designentscheidungen zu treffen. Der Ansatz des VSD fungiert als essenzieller Impuls auf dem Weg zu einer verantwortungsvolleren und nachhaltigeren Technologieentwicklung und bindet mehrere Disziplinen in den Designprozess ein.

Konzeptionelle Untersuchungen helfen Werte und Wertkonflikte von Stakeholder:innen in der Anwendungsentwicklung zu identifizieren und zu verstehen. Grundlage dafür bildeten die Befragungen von Studierenden und Mitarbeitenden an der FernUniversität in Hagen, die nach dem SHEI-LA-Framework (Tsai et al., 2018) ermittelt wurden. Als zweite Grundlage diente der KI-Leitfaden und die Handlungsempfehlungen zu generativer KI (Biederbeck et al. 2023; Biederbeck et al. 2024), die Grundsätze für den Umgang mit KI formulieren und die Hochschule als Akteur in Wissenschaft und Gesellschaft betrachtet. Ein verantwortungsvoller Umgang mit KI, trotz Technologieoffenheit, gilt dabei als essenziell.

Abb. 1: Aspekte des Value Sensitive Design (eigene Darstellung)



Empirische Untersuchungen hingegen konzentrieren sich auf Bedürfnisse, Ziele und Praktiken für Nutzer:innen. Damit fließen die Erkenntnisse der empirischen Bildungsforschung u.a. zu Feedback, zu Feedback Literacy und zu Online-Engagement mit ein. Die empirische Bildungsforschung zeigt, dass insbesondere HIF, ein entscheidender Faktor für den Lernerfolg ist, da es kognitive Leistung, Motivation, Verhalten und Emotionen positiv beeinflusst

(Hattie 2009; Wisniewski et al. 2020). HIF hat mit einer Effektstärke von *d* = 0,99 den größten Einfluss auf den Lernerfolg und zeichnet sich dadurch aus, dass es nicht nur korrektive Rückmeldungen, mit Informationen zur Selbstregulation, Aufmerksamkeit und Selbstreflexion verknüpft. Grundvoraussetzung dafür, dass Studierende Feedbackinformationen für sich nutzbar machen können, sind aber auch hinreichende Kenntnisse von Feedback Literacy (Woit et al. 2023).

Die gängige Praxis wird aber auch durch gesetzliche Rahmenbedingungen, wie die Datenschutzgrundverordnung (DSGVO) und AI-Act, mit der Erwartung zu Transparenz, der Bereitstellung von Informationen zu und der Aufklärung über den Umgang mit Daten sowie deren Verarbeitung bestimmt. Diese Informationen wurden in unterschiedliche Komponenten des Userinterfaces integriert. Für die Erlangung informierter Einwilligungen wurde ein sogenannte Privacy by Default Systematik umgesetzt und im Zuge dessen eine direkte Verknüpfung zu Datenbanken angelegt. Dies hat direkte Konsequenzen für die Datennutzung und Visualisierung im System. Die enge Abstimmung mit Datenschutzbeauftragten diente der Sicherstellung, dass alle erforderlichen Schritte eingehalten wurden. Die Gewährleistung menschlicher Aufsicht und Transparenz bei der Nutzung von KI-Systemen ist zentral für die Compliance zum AI-Act. Ebenso ist eine systemseitige Kontrolle unerwünschter oder fehlerhafter Ausgaben bei der Nutzung von kennzeichnungspflichtiger generativer KI (genKI) erforderlich.

Technologische Untersuchungen, die die Anwendungsentwicklung durch Vorabanalysen und Möglichkeiten des technischen Einsatzes sowie deren Umsetzung in Systemen begleiten, sind zentral für die Bereitstellung von robusten und vertrauenswürdigen KI-Anwendungen. In der technischen Umsetzung bestand die Herausforderung darin, Lösungen zu finden, um die in den konzeptionellen und empirischen Untersuchungen identifizierten Aspekte als zentrale Bedarfe sowie didaktische Elemente oder prozedurale Strukturen zu berücksichtigen. Informationen zu Feedback Literacy sowie Transparenzinformationen zu genutzten Kennzahlen und Daten werden in beiden Anwendungen beispielsweise kontextbezogen in Infoboxen bereitgestellt.

Gestaltung von KI-gestütztem formativem Feedback

In dem entwickelten IMPACT-Feedbackzentrum wird HIF über zwei Anwendungen an Studierende ausgegeben: Es handelt sich zum einen um ein **M**onitoring **In**formation **D**ashboard (MIND) mit interaktiven Visualisierungen zum Self-Monitoring, das in Hanses et al. (2024) ausführlich beschrieben wurde. Zum anderen stellt das Feedbackzentrum eine KI-gestützte Web-Anwendung zu »**Co**rrective **F**ormative **Fee**dback« (COFFEE) zur Verfügung, das hier im Weiteren beschrieben wird. COFFEE nutzt eine Kombination aus vorgegebenen Regeln und Kriterien sowie der Einsatz lokaler Large Language Models (LLM). Das Resultat ist ein Feedback zu den von Studierenden verfassten Freitextantworten, das eine Rückmeldung dazu gibt, wie die Lösung in Bezug auf vorab definierte Kriterien verbessert werden kann. Es lässt sich damit nahtlos in ein verbreitetes hochschuldidaktisches Lehrformat einsetzen. Bei einem LLM-generierten »korrektiven« Feedback ohne direkte menschliche Supervision ist der Prozess, über den das Feedback umgesetzt wird, entscheidend.

*Die Grundidee des formativen Feedbacks in COFFEE liegt darin, Studierende **nicht** durch die KI bewerten zu lassen, sondern ein LLM zu nutzen, um Studierenden zu verdeutlichen, was eine gute Lösung ausmacht.*

Studierenden wird mit COFFEE ein schnelles Überprüfen von Lösungsvorschlägen für spezifizierte Freitextaufgaben ermöglicht.

Regelbasiertes und KI-gestütztes korrektives Feedback

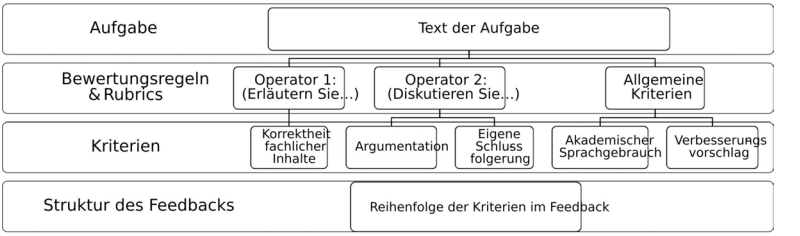
COFFEE kombiniert vordefinierte Regeln und Kriterien mit lokalen Large Language Models (LLM).

Der Prozess, durch den Lehrende ein korrektives Feedback in COFFEE anlegen (vgl. Abb. 2), besteht aus mehreren Schritten:

1. Zunächst wird eine Aufgabe im Feedbackzentrum angelegt, bei der Informationen wie Titel, Beschreibung und Anforderungen eingetragen werden.
2. Im nächsten Schritt werden Bewertungsregeln und Rubrics für eine Aufgabe definiert. Lehrende können verschiedene Kriterien wie z.B. Inhalt, Struktur, Stil oder Grammatik im System mit einem Prompt festlegen.

3. Die Kriterien ordnen die Lehrenden dann einer Aufgabe zu, sodass für jede Aufgabe nur die von ihnen erwünschten Kriterien verwendet werden.
4. Nachdem die Aufgabe und die Bewertungskriterien sowie deren Reihenfolge definiert sind, kann die Aufgabe veröffentlicht werden.

Abb. 2: Systematisierung für ein regelbasiertes formatives Feedback



Bewertungskriterien werden über einen formulierten Prompt an ein ausgewähltes LLM übermittelt. In dem Prompt können auch nicht erwünschte Ausgaben definiert werden. Die Kenntlichmachung der Nutzung von genKI und Informationen zu den verwendeten Prompts werden transparent sowie kontextbezogen automatisiert aktualisiert und auf der Nutzeroberfläche bereitgestellt. Das korrektive Feedback, das durch diesen Prozess generiert wird, ist skalierbar und kann für eine Vielzahl von Aufgaben und Studierenden angepasst werden. Es kombiniert eine von Lehrenden vorgegebene regelbasierte Bewertung mit der Einbindung von LLMs, um Feedback bereitzustellen, das die Erwartung an eine gute Aufgabenlösungen aufzeigt. Voraussetzung für die Umsetzung des regelbasierten Ansatzes ist, dass es zu den Aufgaben klare Bewertungsregeln z.B. durch Operatoren oder Rubrics (English et al., 2022; Nordrum, Evans & Gustafsson, 2013) gibt, die auf die Ebene eines Kriteriums heruntergebrochen werden können (vgl. Abbildung 2).

Prompt Engineering und Qualitätssicherung

Eine der größeren Herausforderungen besteht in der Sicherung der Qualität, die durch die Nachteile der generativen KI, z.B. fehlerhafte Ausgaben, entstehen. In Ermangelung gesicherter Erkenntnisse zum Prompt-Engineering für formative Feedbacks durch LLMs haben sich im Projekt nach intensiver und

sorgfältiger Exploration folgende Aspekte bei der Formulierung von Prompts bewährt. Ein Prompt sollte:

- inhaltlich strukturiert, in kurzen, vollständigen Sätzen formuliert sein und wenige Formatierungen verwenden.
- eine reduzierte Komplexität für das LLM aufweisen (nur ein Kriterium enthalten).
- Kontextinformationen zur Aufgabe bereitstellen (z.B. Lehrtext).
- Antwortstil und Verbotsregeln beschreiben (z.B. sachlich, akademisch, keine Punktzahlen).
- die Struktur des Feedbacks beschreiben (z.B. erst positive Aspekte, dann Fehler, dann (?) Inhalt guter Lösung beschreiben).
- konsistente Begrifflichkeiten verwenden (z.B. nicht wahlweise »Aufgabe« und »Lösungsversuch«).

Um sicherzustellen, dass das korrektive Feedback sowohl fair, transparent und wirksam ist, als auch die menschliche Aufsicht gewährleistet, müssen Bewertungskriterien und Prompts transparent und nachvollziehbar sein. Der Ansatz respektiert die Autonomie der Studierenden und unterstützt die persönliche Entwicklung in einem geschützten Rahmen, da alle Feedbackbereiche so gestaltet sind, dass Lehrende keinen Einblick in die individuellen Ergebnisse erhalten.

Die EU-Vorgaben wurden mit dem AI-Act Compliance Checker überprüft und nachgehalten. Transparenz zur Anwendung, Unterdrückung ungewollter Inhalte sowie die Kennzeichnung von genKI wurden durch die Implementierung von geeigneten Mechanismen und kontextbezogenen Systemhinweisen sichergestellt, in denen auch über den Einsatz von LLMs und die (didaktische) Funktion informiert wird. Die bisherigen Rückmeldungen der Studierenden belegen, dass COFFEE einen deutlichen Mehrwert im Lernprozess bietet und eine hohe Akzeptanz hat.

Fazit & Ausblick

Der Beitrag zeigt, wie ein VSD-Ansatz dazu beiträgt, die Potenziale von KI für eine skalierbare Bereitstellung von formativem Feedback zu nutzen und gleichzeitig Aspekte für einen vertrauenswürdigen Einsatz von LA und KI bereits in der Anwendungsentwicklung zu berücksichtigen. Die Feedback-

anwendung COFFEE stellt Studierenden KI-geschriebenes HIF innerhalb weniger Sekunden bereit und ermöglichten es ihnen, gezielt ihre Fähigkeiten und Leistungen selbstverantwortlich zu verbessern. Orientiert an den Grundsätzen von Lehre und Didaktik wird die menschliche Kontrolle gewahrt, um sicherzustellen, dass die bereitgestellten Informationen sachgerecht und für die Studierenden relevant sind, aber auch ein Deskilling (Reinmann 2023) vermieden wird. Autonomie, Transparenz und Interpretierbarkeit werden durch kontextnahe und -relevante Informationen in der Anwendung gewährleistet.

Die wertesensibel gestalteten Feedbackanwendungen des IMPACT-Feedbackzentrums bieten eine wertvolle Ergänzung zum Lehren und Lernen an Hochschulen. Dennoch ergeben sich Herausforderungen, da für die lernförderliche Nutzung Kenntnisse von Feedback Literacy bei Lernenden Voraussetzung sind. Lehrende, die die Feedbackanwendung in ihr Fach übertragen möchten, können strukturierte Bewertungsregeln für ihre Aufgaben in einen Prompt überführen, müssen dann aber über aufwendiges Testen sicherstellen, dass ihre Vorgaben auch tatsächlich zuverlässig umgesetzt werden. Die Qualitätssicherung stellt damit eine zentrale Notwendigkeit beim Einsatz generativer KI im Hochschulstudium, die in Zukunft noch verbessert und systematisiert werden muss.

Danksagung

Die Autorinnen und Autoren bedanken sich für die Förderung des Verbundprojekts IMPACT im Rahmen der Bund-Länder-Förderinitiative Künstliche Intelligenz in der Hochschulbildung. Das Teilprojekt der FernUniversität in Hagen mit dem Förderkennzeichen 16DHBKI043 wurde durch das Bundesministerium für Bildung und Forschung (BMBF) und das Land Nordrhein-Westfalen für den Zeitraum Dezember 2021 bis November 2025 unterstützt.

Literatur

Biederbeck, A., Bils, A., Gisbert, A., Hinte, Hofmann, U., Karolyi, H., Kempka, A., Opel, S., Kempka, A., Remic, O., Schröder, A., Sorichter, N., Sperl, A., Terbeck, M., De Witt, C., 2024. Handlungsempfehlungen für den didaktischen Einsatz von generativer KI in der Hochschullehre.

Biederbeck, A., Bils, A., Gisbert, A., Karolyi, H., Kempka, A., Opel, S., Sperl, A., De Witt, C., 2023. KI-Leitfaden der FernUniversität in Hagen Grundsätze und Orientierungshilfen für die Nutzung von Künstlicher Intelligenz in Lehre und Studium.

English, N., Robertson, P., Gillis, S., Graham, L., 2022. Rubrics and formative assessment in K-12 education: A scoping review of literature. International Journal of Educational Research 113, 101964. https://doi.org/10.1016/j.ijer.2022.101964

Friedman, B., Hendry, D., 2019. Value sensitive design: shaping technology with moral imagination. The MIT Press, Cambridge, Massachusetts.

Hansen, J., Rensing, C., Herrmann, O., Drachsler, H., 2020. Verhaltenskodex für Trusted Learning Analytics. Version 1.0. Entwurf für die hessischen Hochschulen. https://doi.org/10.25657/02:18903

Hanses, M., Van Rijn, L., Karolyi, H., De Witt, C., 2024. Guiding Students Towards Successful Assessments Using Learning Analytics From Behavioral Data to Formative Feedback, in: Sahin, M., Ifenthaler, D. (Eds.), Assessment Analytics in Education, Advances in Analytics for Learning and Teaching. Springer International Publishing, Cham, pp. 61–83. https://doi.org/10.1007/978-3-031-56365-2_4

Hattie, J., 2024. Visible learning 2.0. Schneider Verlag Hohengehren GmbH, Baltmannsweiler.

Nordrum, L., Evans, K., Gustafsson, M., 2013. Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. Assessment & Evaluation in Higher Education 38, 919–940. https://doi.org/10.1080/02602938.2012.758229

Reinmann, G., 2023. Deskilling durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik. Diskussionspapier Nr. 25, Hochschulforum Digitalisierung, Berlin.

Spiekermann, S., 2021. Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme. Informatik Spektrum 44, 247–256. https://doi.org/10.1007/s00287-021-01378-4

Woitt, S., Weidlich, J., Jivet, I., Orhan Göksün, D., Drachsler, H., Kalz, M., 2023. Students' feedback literacy in higher education: an initial scale validation study. Teaching in Higher Education 1–20. https://doi.org/10.1080/13562517.2023.2263838

Wisniewski, B., Zierer, K., Hattie, J., 2020. The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. Front. Psychol. 10, 3087. https://doi.org/10.3389/fpsyg.2019.03087