



# Abschlussbericht

## Responsible Academic Performance Prediction (RAPP):

### Ein sozialverträglicher Ansatz zur Entwicklung und Einführung studentischer Leistungsprognose an einer Deutschen Hochschule

**Förderkennzeichen:** 16DHB4020

**Laufzeit:** 01.03.2021 – 29.02.2024

- Prof. Dr. Stefan Conrad (Projektleitung)
- Prof. Dr. Frank Marcinkowski (Co-Leitung)
- Dr. Johannes Krause
- Prof. Dr. Ulrich Rosar
- Prof. Dr. Michael Leuschel
- Dr. Christopher Starke

**Heinrich-Heine-Universität Düsseldorf**

Universitätsstraße 1

40225 Düsseldorf

# Teil I: Kurzbericht

Das BMBF-geförderte Projekt RAPP wurde als interdisziplinäres Projekt zwischen Informatik und Sozialwissenschaften konzipiert und beschäftigte sich mit drei inhaltlichen Schwerpunkten: (i) der diskriminierungsfreien Vorhersage von studentischen Leistungen mittels Methoden des maschinellen Lernens zur Verbesserung der Beratung im Studium, (ii) der Analyse diskriminierungsbehafteter Daten als Variablen zur Bestimmung des Studienerfolgs und (iii) der Untersuchung der individuellen Wahrnehmung der Studierenden hinsichtlich des Einsatzes von KI-Methoden. Insgesamt wurde damit die Möglichkeit der Einführung eines RAPP-Systems in einer Hochschule untersucht.

Für die automatisierte Vorhersage von Studienleistungen und Studienerfolg wurden aufbauend auf bestehenden Arbeiten Methoden des maschinellen Lernens entwickelt, die Leistungsdaten der ersten Semester sowie weitere Merkmale der Studierenden (die durch Befragungen ermittelt wurden) verwenden. Damit kann z. B. die Gefahr eines Studienabbruchs oft nach einem oder zwei Semestern vorhergesagt werden um gefährdeten Studierenden eine passende Beratung und geeignete Unterstützungsmaßnahmen anzubieten. Dazu wurden auch Erklärungsmodelle untersucht, die die Ergebnisse der maschinellen Vorhersage so erläutern, dass diese in der Beratung direkt genutzt werden können. Da hier personenbezogene Daten verarbeitet wurden, wurde in intensiver Absprache mit der Datenschutzbeauftragten der Heinrich-Heine-Universität ein Datenschutzkonzept entwickelt, das im Kern auf einer Pseudonymisierung der Daten beruht. Für die Nutzung und Verarbeitung der Daten wurden die erforderlichen Einträge im Verfahrensverzeichnis der Universität erstellt.

Zum Trainieren der Vorhersage-Modelle wurde eine Anwendung prototypisch implementiert. Auch für die Anwendung der Vorhersage-Methoden sowie ein auf Entscheidungsbäumen basierendes Erklärungsmodell wurden prototypisch in einer Anwendung implementiert. Diese Anwendungen wurden in öffentlich zugänglichen Repositories veröffentlicht.

Ferner wurde untersucht, inwieweit diskriminierungsbehaftete Variablen zur Bestimmung des Studienerfolgs genutzt werden können. Die Erklärungskraft soziodemographischer Variablen, der sozialen Integration an der Universität sowie schulischer Leistungen für den Studienerfolg wurden untersucht. Vor allem die Abbruchsintention, der Besuch eines Gymnasiums sowie die Note der Hochschulzugangsberechtigung dienen als Prädiktor für den Erfolg der Studierenden.

Auch die individuellen Wahrnehmungen der Studierenden gegenüber der KI-basierten Leistungsprognose waren Gegenstand der Untersuchungen. In einer Reihe von standardisierten und teil-standardisierten Befragungen sowie faktoriellen Surveys wurden neben den Fairness- und Schadenswahrnehmungen ebenso Anforderungen der Studierenden an das KI-System untersucht. Dabei zeigte sich unter anderem, dass Studierende dem Einsatz der KI-basierten Leistungsprognose tendenziell skeptisch gegenüberstehen und eine Vielzahl potenzieller Schäden sehen, die von *Academic Performance Prediction* (APP) ausgehen können. Dazu zählen insbesondere die Gefahr der Demotivation, ein mangelndes menschliches Einfühlungsvermögen bzw. die fehlende Berücksichtigung besonderer Umstände sowie die Entstehung von Missverhältnissen und fehlerhafte Vorhersagen. Studierende legen zudem hohen Wert auf ihre persönliche

Autonomie und die Erklärung ihrer persönlichen Leistungsprognose. Ergebnisse einer Experimentalstudie verdeutlichen darüber hinaus, dass nicht nur die Erklärung an sich, sondern auch das eigene Verständnis der Erklärungen wichtig für eine Fairnesseinschätzung von APP-Systemen ist.

Auf Basis der im Rahmen des RAPP-Projekts gewonnenen Erkenntnisse wurden empirisch fundierte Schlussfolgerungen für einen sozialverantwortlichen Einsatz von KI-Systemen an Hochschulen gezogen: So müssen 1) die Bedenken der relevanten Stakeholdergruppen – hier in besonderem Maße der Studierenden – in den Mittelpunkt einer sozialverantwortlichen KI-Implementierung gestellt werden. Verschiedene ethische Ansichten müssen ernst genommen werden. 2) Hochschulentscheider und politische Entscheidungsträger sind dafür verantwortlich, ein ausreichendes Maß an Transparenz zu gewährleisten, sodass der Einsatz von KI in der Bildung für die Betroffenen zum einen durchschaubar wird, sie zum anderen jedoch nicht überfordert werden. So erfordert dies auch aus Forschungsperspektive eine Herangehensweise, die den Teilnehmenden das Anwendungsfeld so wenig abstrakt wie möglich vermittelt, damit das Verständnis für die tatsächliche Anwendung höchstmöglich ist, ohne dabei mit zu vielen Informationen für Verwirrung zu Sorgen. In Anbetracht der wachsenden Bedeutung erklärbarer KI, können in diesem Zusammenhang White-Box-KI-Verfahren, wie beispielsweise Entscheidungsbäume, die Nachvollziehbarkeit und das Verständnis der Prozesse verbessern. 3) Es bedarf unabhängiger (Kontroll- &) Beratungsinstanzen, um gerade im Falle schlechter Leistungsprognosen die Bereitschaft zur Inanspruchnahme von Unterstützung zu erhöhen und somit zu einer verbesserten Studienleistung beizutragen. 4) Neben der Planung, Implementierung und Erklärung des Verfahrens ist eine fortlaufende Evaluation unter Einbezug der universitären Stakeholdergruppen notwendig, um das KI-System an neu gewonnene Erkenntnisse und Anforderungen anpassen zu können.

# Teil II: Eingehende Darstellung

## Einleitung

Das Forschungsprojekt „Responsible Academic Performance Prediction (RAPP)“ wurde mit dem Ziel durchgeführt, verantwortungsbewusste Methoden zur Vorhersage von akademischer Leistung zu entwickeln und zu erproben. Das Projekt begann am 01.03.2021 und wurde am 29.02.2024 erfolgreich abgeschlossen. In diesem Bericht werden die wichtigsten Ergebnisse, Erkenntnisse und Erfahrungen des Projekts zusammengefasst.

Das Projekt RAPP beschäftigte sich mit drei inhaltlichen Schwerpunkten: (i) der diskriminierungsfreien Vorhersage von studentischen Leistungen mittels Methoden des maschinellen Lernens zur Verbesserung der Beratung im Studium (Arbeitspaket AP I), (ii) der Analyse diskriminierungsbehafteter Daten als Variablen zur Bestimmung des Studienerfolgs (AP II) und (iii) der Untersuchung der individuellen Wahrnehmung der Studierenden hinsichtlich des Einsatzes von KI-Methoden (AP III). Entsprechend diesen inhaltlichen Schwerpunkten ist dieser Bericht gegliedert. Die Ergebnisse sind in diverse wissenschaftliche Publikationen eingeflossen bzw. werden noch publiziert. Die Publikationen sowie die eingereichten oder in Vorbereitung befindlichen Beiträge sind jeweils am Ende der Abschnitte zu den APs.

Alle durchgeführten Arbeiten basieren auf der Arbeitsplanung im Antrag und wurden entsprechend durchgeführt. Sie haben jeweils für die Projektziele sinnvolle und nützliche Ergebnisse hervorgebracht. Alle Arbeitsschritte waren zur Erreichung der Projektziele notwendig und wurden angemessen (inhaltlich sowie hinsichtlich des Mitteleinsatzes) durchgeführt.

## AP I

Das Arbeitspaket AP I behandelte schwerpunktmäßig die diskriminierungsfreie Vorhersage von studentischen Leistungen.

## Datenschutzkonzept

Als Basis der Datennutzung und Datenverarbeitung im Projekt wurde zusammen mit der Datenschutzbeauftragten der HHU sowie den Datenschutzkoordinator\*innen der beteiligten Fakultäten sowie der zentralen Verwaltung ein Datenschutzkonzept entworfen und konkretisiert (T1.1). Hierzu wurde eine detaillierte Klärung von Fragen des Datenschutzes hinsichtlich der durchzuführenden Befragungen sowie der Nutzung von personenbezogenen Befragungs- und Leistungsdaten im Projekt vorgenommen. Die Ergebnisse sind unmittelbar in die konkrete Gestaltung der Durchführung der Befragung eingeflossen und leiteten die weitere Verarbeitung der Daten im Projekt. Während die Daten der Studierenden, wenn diese in der Befragung die entsprechende Einwilligung gegeben haben, mit den über sie an der HHU gespeicherten Leistungsdaten zusammengeführt werden dürfen (unter Nutzung der Matrikelnummer), wurde für die weitere Nutzung der zusammengeführten Daten im Projekt mit der Datenschutzbeauftragten der HHU ein Pseudonymisierungskonzept entwickelt. Darüber hinaus wurden die erforderlichen Einträge ins Verzeichensverzeichnis der HHU

erstellt.

## Entwicklung von Vorhersagemodellen

**Datengrundlage.** Für das Projekt standen pseudonymisierte Leistungsdaten von Studierenden der Fächer *Informatik (Bachelor)* und *Sozialwissenschaften – Medien, Politik, Gesellschaft (Bachelor)* zur Verfügung. Die Daten umfassten alle Studierenden, die an den drei Befragungswellen (siehe AP II) teilgenommen hatten. Zusätzlich waren anonymisierte Leistungsdaten von Informatik-Studierenden, die innerhalb des Zeitraums 2010-2021 mindestens eine Hochschulprüfung absolviert hatten, verfügbar, die für das Training der Vorhersagemodell basierend auf Leistungen verwendet werden konnten, aber nicht mit den Befragungsdaten verknüpft werden konnten. Pro Studierenden standen Informationen zu jeder im genannten Zeitraum absolvierten Prüfung (Modul, Note, ECTS, Anzahl Versuche, Datum, Bestanden/Durchgefallen) sowie die Information, ob das Studium abgeschlossen wurde, zur Verfügung. Darüber hinaus waren schützenswerte Attribute wie das Geschlecht (männlich, weiblich oder divers), das Alter bei der Einschreibung (Geburtsjahr, Immatrikulationsdatum) und die Staatsbürgerschaft (deutsch oder ausländisch) ebenfalls zugreifbar.

Die Daten wurden statistisch ausgewertet (T1.1), wodurch Differenzen in der prozentualen Verteilung der Absolventen zwischen den Gruppen mit deutscher und ausländischer Staatsbürgerschaft ermittelt wurden, wie Tabelle 1 zu entnehmen ist. Ebenso wurde ein Unterschied in der Qualität des Abschlusses (Gesamtnote) festgestellt. Unterschiede zwischen den Geschlechtern von deutschen Studierenden erwiesen sich als weniger signifikant. Den größten Unterschied gab es zwischen männlichen, deutschen Studierenden und weiblichen, ausländischen Studierenden.

Tabelle 1: Statistiken über Informatikstudierende des Datensatzes im Hinblick auf verschiedene, demografische Gruppen. Es werden Geschlecht und deutsche Staatsbürgerschaft der durchschnittlichen Abschlussnote, der Bachelor-Abschlussquote und der Master-Zulassungsquote von Bachelor-Absolventen gegenübergestellt.

Geschlecht	Deutsch	Note (Durchschnitt)	BA-Abschluss (%)	MA-Zulassung (%)
weiblich	nein	2,78	12,35	38,89
	ja	2,28	17,21	62,65
männlich	nein	2,90	14,66	25,93
	ja	2,29	20,64	62,41

**Anforderungsprofil.** Es wurde ein Anforderungsprofil an das Vorhersagesystem erstellt (T1.1). Die Modelle müssen mit den verschiedenen Dateiformaten (kontinuierliche sowie kategoriale Werte) als Eingabeformat umgehen können. Das gemessene Vorhersageverhalten trainierter Modelle darf hierbei keine signifikanten Unterschiede für geschützte Gruppen (Geschlecht, Staatsbürgerschaft) aufweisen. Neben binären Vorhersagemodellen (Studienabbruch, Studienerfolg, Erreichen der Masterzulassung), sollen

auch Regressionsmodelle für die Vorhersage kontinuierlicher Werte (Studiendauer, Durchschnittsnote) trainiert werden. Solche kontinuierlichen Vorhersagewerte wurden durch die Umfrageergebnisse von AP III als von der Studierendenschaft präferiert herausgearbeitet (T1.4).

**Vorhersagebasis und Vorhersagewerte.** Da anfänglich unklar war, welche Art von Vorhersagen für das Projekt am geeignetsten wäre, wurden verschiedene Sätze von unabhängigen Variablen sowie Zielvariablen definiert. Vorhersagemodelle wurden für jede entstehende Kombination trainiert, um einen qualitativen Überblick über alle Vorhersage-Tasks zu erhalten. Folgende Sätze an unabhängigen Variablen wurden genutzt:

- Klausuren (Info/SoWi): Leistungsdaten über relevante Klausuren des entsprechenden Fachs,
- Leistungen: Anzahl bestandener Prüfungen insgesamt oder pro Semester (dieser Punkt ist modul-unspezifisch),
- Versuche: Daten über angetretene Prüfungen insgesamt oder pro Semester (modul-unspezifisch),
- ECTS: Erhaltene ECTS-Punkte insgesamt oder pro Semester.

Als mögliche Zielvariablen wurden genutzt:

- Studienabbruch,
- Abschlussnote,
- Erreichte ECTS-Punkte nach dem 4. Semester,
- Benötigte Semesteranzahl,
- Masterzulassung.

**Verwendete Algorithmen zur Vorhersage.** Zur Erstellung von Vorhersagemodellen wurden traditionelle Machine-Learning-Methoden eingesetzt. Da als Vorhersagegröße sowohl kategoriale als auch kontinuierliche Werte in Frage kommen, unterscheiden wir zwischen den Algorithmen für die Klassifikation der kategorischen und für die Regression der kontinuierlichen Werte. Für die Klassifikation kamen zum Einsatz: Entscheidungsbäume mit Cost-Complexity-Pruning (Breiman et al., 1984), Logistische Regression (Cox, 1958), Random Forests (Breiman, 2001), Support-Vector-Maschinen (Cortes & Vapnik, 1995), Gaussian Naive Bayes (Rish et al., 2001), und Multi Layer Perceptron (Haykin, 1994). Im Falle der Regression wurden genutzt: Lineare Regression (Hutcheson, 2011), Linear Ridge Regression (Hoerl & Kennard, 1970), Bayesian Ridge Regression (MacKay, 1992), Entscheidungsbäume für Regression (Breiman, 2001), Multi Layer Perceptron (Haykin, 1994) für Regression, und Elastic Net Regression (Kim et al., 2007).

Für das Training wurden jeweils 80% der Studierendendaten verwendet. Die resultierenden Modelle wurden auf den verbleibenden 20% (dem Testset) mit verschiedenen Metriken evaluiert (vgl. Zheng, 2015). Im Rahmen der Klassifikation wurden primär Metriken über die Konfusionsmatrix der Modelle erstellt: Accuracy und Balanced Accuracy, Precision, Recall und F1-Score sowie die ROC–Area under Curve. Für die Regression wurden genutzt: Mean absolute error, mean squared error und root mean squared error.

**Fairnessanalyse der Modelle.** Im Rahmen des Projekts wurde die Fairness-Metrik Equalised Odds als geeignet und empfehlenswert herausgearbeitet. Dieses Resultat ergab sich nach Sichtung der relevanten Literatur zur algorithmischen Fairness, welche sich damit beschäftigt, in welchen Fällen welche Fairness-Metriken relevant sind.

Hierfür wurde analysiert, welcher sogenannten Weltansicht (engl.: Worldview) die Academic Performance Prediction unterliegt. Zwei Extrema hierbei sind "We're all equal" (WAE) und "What you see is what you get" (WYSIWYG). WAE entspricht einer Weltsicht, in der jegliche Unterschiede zwischen demographischen Gruppen durch diskriminierende Verzerrungen in der Datenerhebung oder -zuordnung entstehen. Dies könnte dadurch entstehen, dass diskriminierte Gruppen durch die genutzten, unabhängigen Variablen nicht korrekt abgedeckt werden oder diese Variablen nur Annäherungen von latenten Werten sind, die eigentlich genutzt werden müssten. Im Fall von historischen Daten können auch vormalige, diskriminierende Zuweisungen der Vorhersagegröße zu diesen Verzerrungen führen. In WYSIWYG wird konträr davon ausgegangen, dass die gesammelten Daten keinen Verzerrungen unterliegen und die richtigen unabhängigen Variablen gesammelt wurden.

Während des Projekts haben wir eine Zuordnung von APP zur Ansicht WYSIWYG herausgearbeitet. Dies lässt sich dadurch begründen, dass vorherige akademische Leistung zu den vorhersagestärksten Merkmalen für APP zählen (Alyahyan & Düstegör, 2020), diese jedoch auch im direkten Zusammenhang mit finaler akademischer Performance stehen. Während bspw. die Notenvergabe innerhalb einer Vorlesung durchaus diskriminierend sein kann und eine WAE-Weltsicht eingenommen werden sollte, steht APP im Kontext des Projekts einen Schritt dahinter. Prüfungsleistungen und -bewertungen wurden bereits abgelegt. Die Feststellung der Eignung zum Master (als Beispiel) ist ab diesem Punkt ein Prozess, der bürokratischen Regeln folgt und Unterschiede über die Gruppen ergeben sich aus Unterschieden in ihren Leistungen. Als Konsequenz bietet sich WYSIWYG als Weltanschauung für das Vorhaben in RAPP an. Diskriminierende Prozesse in der Notenvergabe stehen hierbei außerhalb der Reichweite von möglichen Interventionen eines APP-Systems.

Eine Literaturrecherche zeigt, dass sich für WYSIWYG Equalised Odds anbietet und das Anwenden anderer Fairness-Normen eventuelle Diskriminierung sogar weiter propagieren könnte (Yeom & Tschantz, 2021). Die Erkenntnis, dass WYSIWYG für RAPP anwendbar ist und Fairness durch die Metrik von Equalised Odds quantifiziert werden kann, wurde auf dem "Workshop Fairness, Accountability, and Transparency in Educational Data" (FATED'22) der "Educational Data Mining Conference" (EDM'22) nach Peer-Review veröffentlicht und vorgestellt (Dunkelau & Duong, 2022).

**Fairness-Methoden.** Um dem Anforderungsprofil gerecht zu werden, wurden verschiedene Fairness-Methoden entwickelt. Generell können unbalancierte Daten zu einer Benachteiligung von demografischen Gruppen führen. Die entwickelten Methoden (Duong & Conrad, 2023; Duong & Conrad, 2024) zielen darauf ab, durch die Erzeugung von synthetischen Daten eine bessere Balance zwischen den Gruppen zu erreichen ohne dabei die Vorhersagegenauigkeit von echten Daten zu verschlechtern. In den veröffentlichten Publikationen wurde die Wirksamkeit der Methoden auf mehreren realen und öffentlichen Datensätzen nachgewiesen. Die initiale Methode (Duong & Conrad, 2023) behandelt nur binäre Attribute, während die neuere Methode (Duong & Conrad, 2024) auch mit nicht-binären Attributen umgehen kann. Dies ermöglicht es, die Methode auf den erhaltenen Datensatz der Studierenden anzuwenden, welche ebenfalls nicht-binäre Attribute (bspw. Geschlecht) beinhaltet. Die initiale Methode wurde auf der "Big Data Analytics and Knowledge Discovery (DaWaK'23)"-Konferenz nach Peer-Review veröffentlicht und vorgestellt (Duong & Conrad,

2023), während die neuere Methode auf der “*The Australasian Data Science and Machine Learning Conference (AusDM’23)*”-Konferenz nach Peer-Review präsentiert wurde (Duong & Conrad, 2024). Die Methoden wurden in Python implementiert, sind dokumentiert und unter der MIT-Lizenz frei verfügbar (<https://github.com/mkduong-ai/fairdo>). Da die Fairness-Methoden für alle Tabular-Daten geeignet sind, können die Methoden auch auf allen Datensätzen (Leistungsdaten, Umfragedaten) im Anforderungsprofil angewendet werden. Die Entwicklung der Methoden erfüllt auch die Teilaufgabe T2.4.

**RAPP-Tool.** Für das Training der Modelle im Kontext der Hochschuldaten und Fairness-Analyse ist ein spezialisiertes Tool entwickelt worden, das RAPP-Tool (Duong et al., 2023). Dieses Tool zielt darauf ab, durch eine benutzerfreundliche Oberfläche die einfache Erstellung verschiedener Datensätze aus einer einzigen Datenbank mit gewünschten Merkmalen und Labels zu ermöglichen. Die Anwendung ermöglicht das Trainieren, Speichern und Evaluieren von Machine-Learning-Algorithmen auf den erzeugten Datensätzen. Die geforderten Merkmale und Zielvariablen können durch SQL-Abfragen der Datenbank abgeleitet werden. Darüber hinaus unterstützt das Tool die Vorhersagegenauigkeit sowie die Fairness der Modelle zu bewerten und zu vergleichen. So werden sogenannte Pareto-Fronten visuell dargestellt, welche dem Nutzer den Trade-off zwischen Fairness und Vorhersagegenauigkeit veranschaulichen und bei der Auswahl eines finalen Modells unterstützen.

Das Tool wurde 2023 als wissenschaftlicher Fachartikel auf der “*Database Systems for Business, Technology and Web (BTW’23)*”-Konferenz nach Peer-Review veröffentlicht und vorgestellt (Duong et al., 2023). Der Quelltext wurde unter der MIT-Lizenz publiziert (<https://github.com/hhu-rapp/rapp-tool>) und ist frei nutzbar (T1.6).

Abbildung 1: Startbildschirm vom RAPP Tool. Machine Learning-Klassifikatoren können auf Leistungsdaten von Studierenden trainiert werden, um Abbrecher zu erkennen.

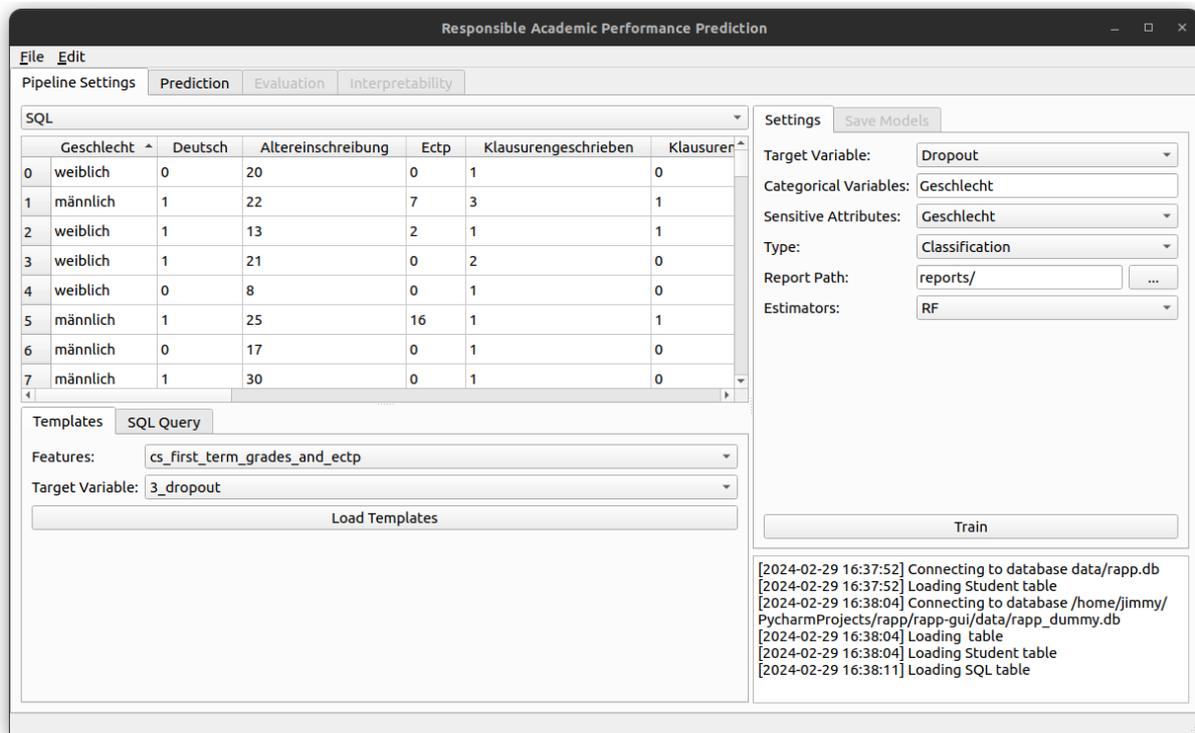
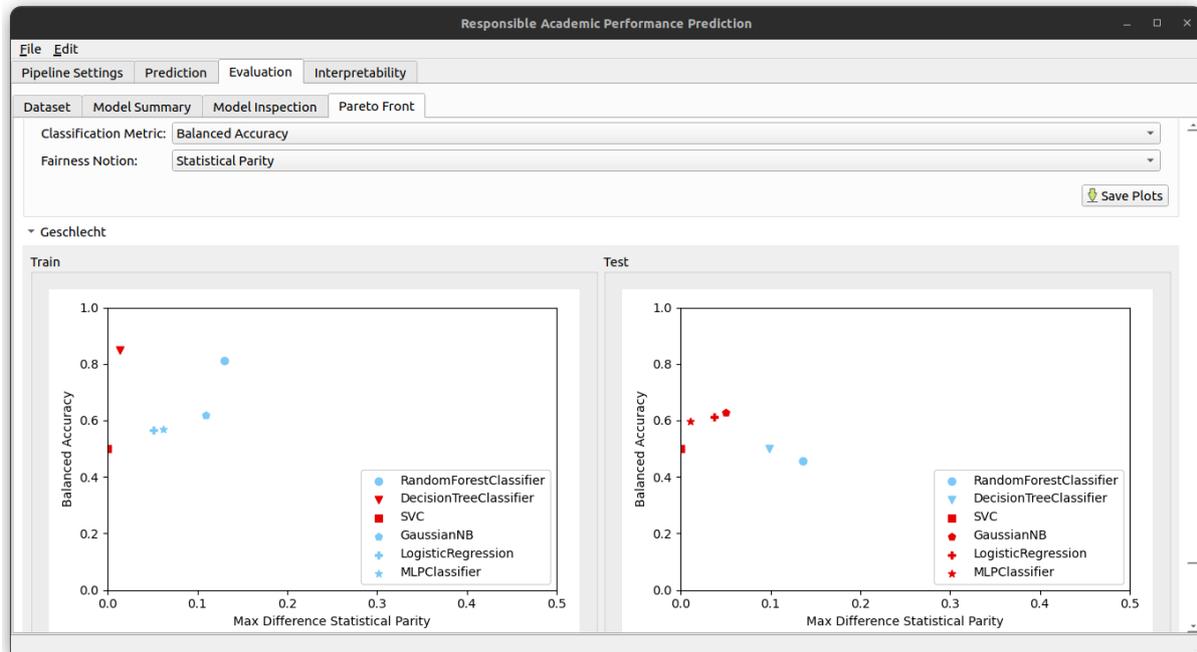


Abbildung 2: In diesem Beispiel werden Klassifikatoren im RAPP Tool miteinander verglichen (Genauigkeitsmaß in y-Achse, Fairness Metrik in x-Achse). Je höher die Genauigkeit und je niedriger die gemessene Diskriminierung des Klassifikators, desto besser ist dieser Klassifikator. Alle Klassifikatoren wurden im Tool selbst trainiert und evaluiert. Die Pareto-Front wird hier rot dargestellt.



## Ergebnisse

Die Ergebnisse sind in den Tabellen 2 und 3 aufgeführt. Es zeigte sich, dass bei Klassifikationsaufgaben, wie der Vorhersage des Studienabbruchs, Entscheidungsbäume entweder die besten Modelle waren, oder mindestens mit den besten Modellen kompetitiv waren. Entscheidungsbäume sind darüber hinaus die einzigen Modelle, die eine überschaubare Erklärung für die Vorhersage liefern können. Anders als bei anderen Modellen, wie Neuronale Netze, ist es möglich, die Entscheidungsregeln zu visualisieren. Black-Box-Modelle wie Neuronale Netze sind nicht in der Lage, eine Erklärung für die Vorhersage zu liefern und benötigen approximative Verfahren zur Erklärungsfindung, was zu einer ungenauen Erklärung führt und irreführend sein kann. Hierdurch könnten ethische Probleme verschleiert werden. Entscheidungsbäume sind gleichzeitig auch die einzigen Modelle, die die Fairness-Metrik Equalised Odds erfüllen. Entscheidungsbäume haben über verschiedene Datensätze hinweg die besten oder zweitbesten Ergebnisse erzielt und waren stets an der Pareto-Front der Fairness-Metrik und der balancierten Accuracy. Bei der Regression, wie bspw. der Vorhersage der Abschlussnote, waren die Ergebnisse der linearen Regression am besten. Die Ergebnisse der anderen Modelle waren ähnlich, jedoch waren die Ergebnisse der linearen Regression stets am stabilsten. Dies deutet darauf hin, dass zwischen den Leistungsdaten in den ersten Semestern und der Abschlussnote offensichtlich ein linearer

Zusammenhang besteht.

Tabelle 2: Vorhersageergebnisse für das Fach Informatik. Es werden die Ergebnisse der Klassifikation als Balanced Accuracy dargestellt und als Mean Square Error für die Regression.

Die vorhergesagten Werte sind Studienabbruch (Abbruch), Abschluss in Regelstudienzeit (RSZ), Erreichen von mindestens 100 ECTP nach dem 4. Semester (4. Sem. ECTP), Erreichen der Master-Zulassung (MA-Zulassung) und die erreichte Abschlussnote im Bachelor (Note). Alle unabhängigen Variablen (UVs) beziehen sich auf das Ende des ersten Semesters.

UVs	Abbruch	RSZ	4. Sem. ECTP	MA-Zulassung	Note
ECTS	0,625	0,670	1,000	0,628	0,323
Leistungen	0,678	0,612	0,999	0,667	0,288
ECTS + Leistungen	0,683	0,589	1,000	0,674	0,296
Klausuren	0,739	0,627	0,793	0,620	0,265

Tabelle 3: Vorhersageergebnisse für das Fach Sozialwissenschaften. Es werden die Ergebnisse der Klassifikation als Balanced Accuracy dargestellt und als Mean Square Error für die Regression.

Die vorhergesagten Werte sind Studienabbruch (Abbruch), Abschluss in Regelstudienzeit (RSZ), Erreichen von mindestens 100 ECTP nach dem 4. Semester (4. Sem. ECTP), Erreichen der Master-Zulassung (MA-Zulassung) und die erreichte Abschlussnote im Bachelor (Note). Bis auf die Klausuren, beziehen sich alle unabhängigen Variablen (UVs) auf das Ende des ersten Semesters. Da konkrete Klausurnoten in dem Studiengang erst nach dem zweiten Semester zur Verfügung standen, steht die Zeile "Klausuren" im Bezug auf das Ende des zweiten Semesters.

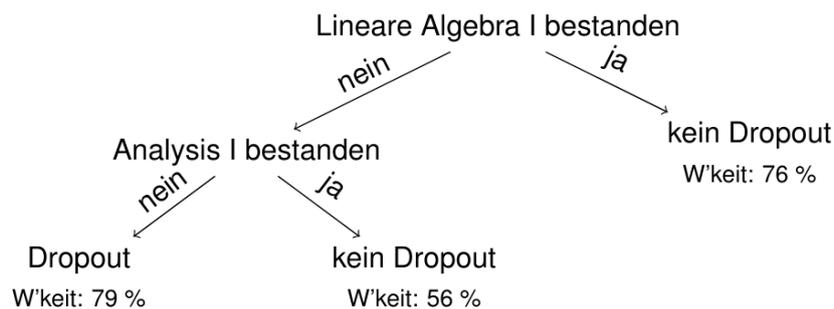
UVs	Abbruch	RSZ	4. Sem. ECTP	MA-Zulassung	Note
ECTS	0,632	0,519	1,000	0,624	0,258
Leistungen	0,640	0,559	1,000	0,729	0,150
ECTS + Leistungen	0,657	0,541	1,000	0,735	0,153
Klausuren	0,811	0,630	0,670	0,850	0,071

Bei Betrachtung der Ergebnistabellen zeigen sich Unterschiede in der Vorhersagegüte zwischen den Fächern. So ist die Vorhersage der Masterzulassung für die Sozialwissenschaftler in der Mehrheit der Experimente höher als für die Informatiker. Dies korreliert natürlich mit geringeren Mean-Squared-Error-Werten in der Vorhersage der Bachelor-Abschlussnote. Die Vorhersage der erzielten ECTP nach dem vierten Semester ist problematisch zu bewerten. Zwar scheinen die Vorhersagewerte sehr gut zu sein mit 100 % Genauigkeit, jedoch ist die Klasse an Studierenden mit 100 ECTP nach dem vierten Semester verschwindend gering; die Modelle sind somit nicht aussagekräftig.

## Erklärung von Vorhersagen

Die Entwicklung eines transparenten Vorhersagemodells ergab sich natürlich durch die Nutzung von Entscheidungsbäumen, welche bereits eine Whitebox-KI sind (T1.3). Eine Regelinduktion findet durch das Ablaufen der Verzweigungen im Entscheidungsbaum statt. Durch die Darstellung, welche Verzweigungen während der Vorhersage gewählt wurden, lässt sich eine akkurate Erklärung für die Modellvorhersage erstellen, welche zusätzlich mit minimalen Gegenbeispielen ergänzt werden kann, die zu einer anderen Vorhersage geführt hätten (T1.5). Abbildung 3 zeigt einen solchen Entscheidungsbaum.

Abbildung 3: Beispielhafter Entscheidungsbaum für die Abbruchvorhersage (Dropout) mit Konfidenzangabe, wie wahrscheinlich die Zuordnung korrekt ist.



## Anwendung für Endnutzer

Im November 2022 wurde ein Workshop mit universitären Stakeholdern (Prorektor für Studium und Lehre, Zentrale Studienberatung, Studiendekanate zweier Fakultäten, Studierendenvertreter\*innen, ...) durchgeführt, um die Anforderungen von RAPP zu ermitteln und um auf die Bedürfnisse der Endnutzer einzugehen.

Für Endnutzer in der Hochschuladministration ist eine einfache und benutzerfreundliche Oberfläche zu bevorzugen, die gezielt auf nicht-technische Benutzer im Universitätsumfeld ausgerichtet ist. Dies wurde in Form einer Webanwendung verwirklicht. Die Anwendung bietet eine Login-Funktion mit unterschiedlichen Nutzertypen, um verschiedenen Anwendern selektiv Zugriff auf die Ihnen zugeteilten Studierenden und Studienfächer zu gewähren.

Die Hauptfunktionen des Tools umfassen eine übersichtliche Dashboard-Ansicht mit relevanten Informationen wie Studierendendaten, durchschnittliche ECTS und demografischen Grafiken. Historische Daten werden in Form von Leistungshistorien präsentiert, sowohl auf individueller Ebene mit dem ECTS-Verlauf mit Vergleich der durchschnittlichen Leistung von Studierenden desselben Fachs als auch auf Gruppenebene im Studienfachvergleich. Eine inkludierte Risikoanalyse umfasst demografische Aspekte wie Geschlecht und Nationalität und hebt nennenswerte Unterschiede innerhalb dieser Gruppen hervor.

Darüber hinaus ermöglichen Prognosefunktionen der Anwendung Studierende anhand eines Ampelsystems (rot, grün) und Risiko-Scores zu bewerten. Hierbei werden Zielvariablen wie

die Klausurleistung berücksichtigt. Durch die Identifikation von Risikostudierenden können gezielte Interventionsangebote gemacht werden. Die unterliegenden Vorhersagemodelle, die in der Webanwendung zum Einsatz kommen, wurden zuvor durch das RAPP Tool erzeugt und analysiert.

Die Nutzung des Tools zielt auf präventive Maßnahmen im Hochschulumfeld ab, indem es umfassende Einblicke in die Studierendenpopulation und deren Leistung bietet. Die Webapp wurde auf <https://github.com/hhu-rapp/rapp-webapp> publiziert und steht nun frei zugänglich unter der MIT Lizenz zur Verfügung.

Abbildung 4: Homebildschirm von der Webanwendung für Endnutzer. Überblick über die soziodemographische Verteilung des Datensatzes.

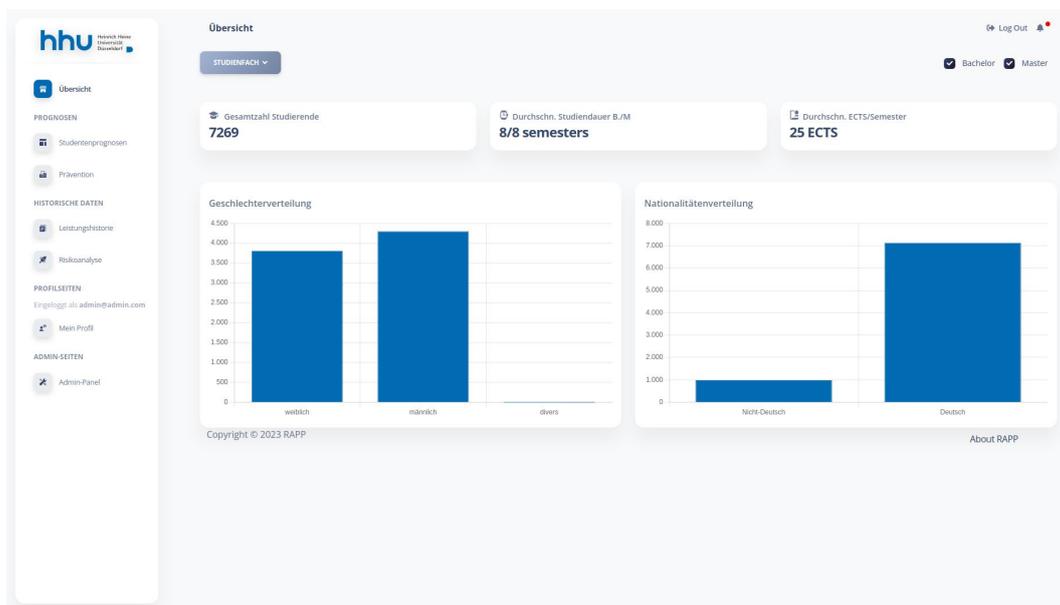


Abbildung 5: Mit den trainierten Klassifikatoren aus dem RAPP Tool werden in dieser Webanwendung Vorhersagen für Studierende gemacht. Studierende werden vorab automatisch als gefährdet/nicht-gefährdet von der KI gekennzeichnet. Der Endnutzer hat dann die Möglichkeit, sich alle Studierenden genauer anzuschauen. Der Endnutzer markiert dann die aus seiner/ihrer Sicht gefährdeten Studierenden manuell. Bei den markierten Studierenden kann dann eine Präventionsmaßnahme eingeleitet werden. Die Webanwendung unterstützt Entscheidungen mit KI-Vorhersagen, erlaubt jedoch volle Kontrolle für den Endnutzer.



Duong, M.K. and Stefan Conrad (2024). (Un)certainly of (Un)fairness: Preference-Based Selection of Certainly Fair Decision-Makers. In ECAI 2024. IOS Press. (begutachtet, zur Publikation angenommen)

## AP II

### Erarbeitung der ethischen Anforderungen an ein APP-System im Hochschulkontext

Das Projekt begann in AP II mit der Identifikation erster ethischer Anforderungen an ein APP-System im Hochschulkontext. Auf Grundlage eines umfangreichen Literaturstudiums konnte eine erste Sammlung (potenziell) diskriminierender Variablen angelegt werden, die den Studienerfolg beeinflussen. Zu diesen zählen in erster Linie Geschlecht, Ethnizität und religiöse Orientierung (vgl. Rzepka et al., 2023), ergänzt durch weitere ethisch fragwürdige Merkmale wie politische Ansichten, Intelligenz, Drogenabhängigkeit, Krankheit oder Behinderung (vgl. Mühlhoff 2020). Außerdem wurden sogenannte Proxy-Variablen identifiziert, deren Berücksichtigung in einem APP-System Diskriminierung indirekt perpetuieren (vgl. Fazelpour & Danks, 2021).

Im Anschluss an die theoretische Ableitung möglicher relevanter Variablen wurden diese mit theoretischen Konzepten verknüpft, die verschiedene vorläufige Ideen zu Anforderungen an ein APP-System im Hochschulkontext relevant werden ließen. Diese speisen sich aus vier theoretischen Perspektiven. Die interaktionistische Perspektive verweist auf die Bedeutung von Interaktionen und Kontakten an der Hochschule für den Studienerfolg der Studierenden. Die Perspektive des kulturellen Kapitals betont hingegen die Bedeutung der Höhe des kulturellen Kapitals und die Ähnlichkeit von persönlichem und institutionellem Habitus für den Studienerfolg. Psychologische Perspektiven rücken hingegen Persönlichkeitsmerkmale in den Mittelpunkt, während die Perspektive der rationalen Wahl die wahrgenommenen materiellen und immateriellen Kosten und Erträge des Studiums als Variablen in den Mittelpunkt rücken (vgl. Sarceletti & Müller, 2011).

Der Erkenntnis folgend, dass diskriminierende Unterscheidungen handlungsfeldtypisch sind (vgl. Heinemann & Mecheril, 2023) und damit sehr gezielt untersucht werden müssen, konnten mit Hilfe leitfadengestützter Interviews die Anforderungen an ein App-System weiter differenziert, Lücken in der bisherigen Sammlung relevanter Anforderungen geschlossen und Ideen priorisiert werden. Schließlich wurden die sich daraus ergebenden neuen Aspekte und Konkretisierungen zentraler Konstrukte in die Umsetzung des APP-Systems zur Vorhersage von Studienleistungen und Abbruchgefährdungen integriert.

### Feldzugang und Feldexploration: Interviews mit Menschen mit Expertise

Lag der Fokus der bisherigen Forschung in erster Linie auf der Verbesserung der Vorhersagegenauigkeit von APP-Systemen, wird die Frage nach diskriminierenden Aspekten der Vorhersage von Studienleistungen kaum berücksichtigt (vgl. Schiefner-Rohs et al., 2023; Schmohl et al., 2023). Um Wissenslücken zu schließen und das Literaturstudium zu ergänzen, wurden zwischen Mai 2021 und Juni 2022 insgesamt 35 qualitative Leitfadeninterviews durchgeführt, um in Kooperation mit relevanten Stakeholdern konkrete Anforderungen an ein APP-System zu erarbeiten. Die Hauptfeldphase lag zwischen Mai und September 2021.

Der Leitfaden adressierte die ethischen Anforderungen an ein APP-System im Hochschulkontext und wurde auf Grundlage der Sichtung des Forschungsstandes erstellt.

Zugleich wurde der Leitfaden offen genug gestaltet, um Raum für ergänzende Einschätzungen der Menschen mit Expertise zu weiteren relevanten Aspekten zu bieten. Damit konnte eine systematische Herangehensweise, effiziente Zeitnutzung der Menschen mit Expertise, die Vermeidung von Informationslücken, eine Vergleichbarkeit der Ergebnisse, aber auch individuelle Schwerpunktsetzung und Anpassung an verschiedene Zielgruppen gewährleistet werden (vgl. Przyborski & Wohlrab-Sahr, 2014).

Den Interviews ging die Identifikation relevanter Stakeholder voraus, zu denen neben VertreterInnen der zentralen Universitätsverwaltung, der Studiendekanate, (Vertrauens-)Dozierende und Studierende der Studiengänge Informatik und Sozialwissenschaften der Heinrich-Heine Universität (HHU) Düsseldorf gezählt wurden. Darüber hinaus wurde das Sample durch VertreterInnen der Wissenschaftscommunity mit Expertise zu den Forschungsthemen Diskriminierung, Studienverläufe und ethische KI ergänzt. Die Interviewten gelten uns im Projekt als ExpertInnen, da sie Studienverläufe begleiten bzw. diese selbst durchlaufen oder Studienverläufe erforschen. Der nicht unproblematische Begriff der ExpertIn wird an dieser Stelle also herangezogen, um Personen zu bezeichnen, die in Relation zur Fragestellung über einen privilegierten Zugang zu Informationen verfügen (Meuser & Nagel, 1991).

Das Sampling erfolgte gezielt nach dem Prinzip der maximalen Kontrastierung der Interviewten hinsichtlich ihrer fachlichen Zuordnung sowie ihrer Position im Hochschulsystem. Die Fächerauswahl orientierte sich an den forschungsleitenden Überlegungen des Gesamtprojekts, das an zwei sehr unterschiedlichen Studiengängen die Effektivität des Systems erprobt. Diese wurden als die beiden BA-Studiengänge Informatik und Sozialwissenschaften identifiziert. Im Gegensatz zum Bachelorstudiengang Informatik unterliegt der Studiengang Sozialwissenschaften an der HHU einer Numerus Clausus (NC)-Beschränkung, wodurch der Anteil der aktiv an diesem Studium partizipierenden Studierenden vergleichsweise hoch ist. Zudem weist der Studiengang Sozialwissenschaften im Unterschied zur Informatik eine hohe Abschluss- und eine niedrige Abbruchquote auf. Im Verlauf der Datenerhebung wurde das Interviewmaterial durch teilnehmende Beobachtungen während des Stakeholder-Workshops des Projekts am 04.11.2022 an der HHU Düsseldorf ergänzt.

## Untersuchungsergebnisse des qualitativen Teilprojekts

Die Auswertung des Interviewmaterials erweiterte den Blick auf ethische Anforderungen an APP-Systeme. Neben der Reflexion der Vorhersage von Studienleistungen durch die Interviewten, tritt der thematische Fokus der Verbesserung von Studienleistungen bzw. Dropoutprävention und damit den präventiven Einsatz individueller Unterstützungsmaßnahmen. Die in der Fachdiskussion zentralen kritischen Momente der Datengenerierung und Datenverarbeitung eines KI-Systems zur Vorhersage von Studienleistungen (vgl. Fazelpour & Danks, 2021; Muñoz et al., 2016; Calders & Žilobaitė, 2013; Dwork et al., 2012; Daniel, 2015) bleiben Hintergrundthemen einer Fokussierung auf die Konsequenzen einer Leistungsvorhersage. Ethische Anforderungen an ein APP-System bedeuten für die Interviewten vor allem den verantwortungsvollen Umgang mit deren Ergebnissen. Die Interviewanalysen konstatieren drei kritische Momente für die Interviewten, die die Kontingenz von Leistungsprognosen betonen und als Übersetzungsleistungen der Prognosen von und für Lernende, Lehrende, Systemeigner und Institution verstanden werden können. Diese Übersetzungsleistungen resultieren für die Interviewten primär aus der Differenz von Leistungsprognose und Problemanalyse.

Eine erste Übersetzung erfolgt durch die Überführung der zu Grunde liegenden Daten in eine konkrete Leistungsprognose (1). Eine zweite Übersetzung ergibt sich aus der Zielgruppe der entsprechenden Informationen, da eine Leistungsprognose durch die beteiligten Akteure sehr unterschiedlich verstanden werden kann (2). Schließlich lässt sich eine dritte Übersetzung

identifizieren, indem eine Leistungsprognose in konkrete Empfehlungen und Interventionen überführt und gleichzeitig mit gebotener Vorsicht kommuniziert werden muss (3).

- 1) Schon die Überführung von Leistungsdaten in eine Leistungsprognose scheint voraussetzungsvoll. Eine grundsätzliche Orientierung bewegt sich zwischen zwei verschiedenen Ansätzen der Bildungsdatenanalyse, die jeweils unterschiedliche Ziele verfolgen (vgl. Keller, Lünich & Marcinkowski, 2022, S.128ff.). Während sich die Dropout-Prediction darauf konzentriert, potenzielle Studienabbrecher zu identifizieren, um die Abschlussquoten zu erhöhen, zielt die Academic Performance Prediction darauf ab, die akademische Leistung von Studierenden vorherzusagen und zu verbessern. Beide Ansätze sind wichtige Instrumente im Bereich des Bildungsmanagements, unterscheiden sich aber hinsichtlich der Reichweite und Art der naheliegenden Interventionen. Mit der Zieldefinition geht einher, ob bestimmte Studierende beobachtet und gefördert werden, um den Abschluss zu sichern, oder alle Studierenden beobachtet werden, um deren Leistungen zu verbessern.

Darüber hinaus stellt sich die Frage nach der konkreten Ausgestaltung der Leistungsprognose, also in welchen Ausprägungen sie sichtbar wird und welche Kategorien von Studierenden sie erzeugt. Neben dem Ziel scheint also auch die Art der Information relevant. Denkbar ist hier eine dualistische Einteilung der Studierenden in eine gefährdete und nicht gefährdete Gruppe. In Konkurrenz dazu treten graduelle Ampel- oder Notensysteme, die Zwischenkategorien schaffen und die Komplexität der Reichweite und Art der Interventionen erhöhen. In beiden Fällen handelt es sich dennoch um leicht verständliche, klare, komplexitätsreduzierende Bewertungen, die individuelle Werte nach bestimmten Kriterien ordnen, um als sinnvolle Informationen bearbeitet werden zu können (vgl. Zweig & Krafft, 2018).

- 2) Zweitens betonen die Interviewten die Effekte einer Leistungsvorhersage im Sinne einer self-fulfilling Prophecy. Leistungsvorhersagen durch ein APP-System werden also einmal als (Fremd-)Positionierung der Studierenden verstanden, da die Vorhersagen von Studienverläufen auf die Leistungen der Studierenden zurückwirken können. Auf der anderen Seite werden sie für Lehrende und Studienberatung zu Wegweisern in der Auseinandersetzung und ermöglichen den zielorientierten Einsatz begrenzter Ressourcen. Die Kommunikation der Ergebnisse durch wen und an wen wird hier zum kritischen Moment eines APP-Systems.

Wird die Leistungsprognose selbst direkt an die Studierenden übermittelt, werden zwei Szenarien von den Interviewten skizziert. Eine positive Leistungsvorhersage kann dann Motivation und Selbstvertrauen stärken, was zu einer tatsächlich verbesserten Leistung führen kann. Umgekehrt kann eine negative Leistungsvorhersage zu Selbstzweifeln und geringerem Engagement führen, was die Wahrscheinlichkeit erhöht, dass die Vorhersage tatsächlich eintritt. Nicht nur Studierende selbst sind demnach mit Vorhersagen konfrontiert, auch Lehrende, Studienberatung und Hochschule handeln in Abhängigkeit der für sie sichtbaren Informationen und Vorhersagen über die Studierenden. Eine Anforderung an ein APP-System ist demzufolge eine bewusste Steuerung des Zugangs zu dessen Vorhersagen (vgl. Olteanu et al., 2019).

- 3) Drittens zeigte sich, dass die Leistungsvorhersagen eines APP-Systems für die Interviewten auslegungsbedürftig sind und in konkrete Empfehlungen und Maßnahmen übersetzt werden müssen. Hier betonen die Interviewten die Bedeutung professioneller Programme und Strategien der Unterstützung von Studierenden, die schon vor dem Eintritt ins Studium ansetzen, das Onboarding, den Studienverlauf und den Abschluss des Studiums begleiten. Es stellen sich also Fragen danach, wer, wann, welche Interventionen veranlasst. Hier sind nicht zuletzt Bedenken der

Interviewten hinsichtlich der Ressourcenknappheit der Hochschulen zu erwähnen, was die Notwendigkeit einer integrierten Strategie, die Interventionen mit bedenkt, bei der Entwicklung und Implementierung eines APP-Systems unterstreicht. Eine Leistungsvorhersage zieht also nicht automatisch eine Intervention nach sich, sondern muss aktiv mit Maßnahmen verknüpft werden. Eine Anforderung an ein APP-System ist demzufolge die Verknüpfung von Leistungsvorhersage und Interventionen bzw. die Integration der Leistungsvorhersage in das etablierte System der Studienberatung (vgl. Ekowo & Palmer, 2017, Chapter 5; Fazelpour & Danks, 2021).

APP-Systeme wurden in den qualitativen Interviews als ein Instrument charakterisiert, das keine eigenständige Problemanalyse generiert: Wie auch die Auswertung der *teilnehmenden Beobachtung* während des Stakeholder Workshops bestätigt, wird eine notwendig menschliche Interaktion eingefordert (vgl. Starke & Lünich, 2020). Die Übersetzung von Leistungsvorhersagen in konkrete Empfehlungen und Interventionen wird als Aufgabe der Hochschule insb. der (Studien-)Beratung angesehen. Deren Beratungsleistung werde durch die Vorhersagen eines APP-Systems allerdings nicht nur unterstützt, indem Fälle identifiziert würden, die Empfehlungen erhalten und die man im Anschluss beraten könne. Die Beratenden müssten festlegen, welche Interventionen für wen bei welchen Werten ausgelöst werden und damit die Prognosen des APP-Systems interpretieren. Nur der direkte Austausch zwischen Studierenden und Beratenden über die Leistungsprognose des APP-Systems könne die Studierenden unterstützen, vor Fehlinterpretationen und Demotivation schützen.

Als grundlegendes Problem wurde die mangelnde Nachfrage von Beratungsangeboten thematisiert, die auch ohne APP-Unterstützung die eigentlich abbruchgefährdeten Studierenden nicht erreichen (vgl. Schneider et al., 2021). Trotz zahlreicher Strategien (klare Kommunikation der Vorteile, persönliche Ansprache, Workshops und Informationsveranstaltungen, Kooperation mit Dozierenden, Social Media Auftritte und der Zusicherung von Vertraulichkeit) gestaltete sich der Zugang zur Beratung schwierig. Die Überführung einer Leistungsprognose in eine Intervention wird insb. dann als Herausforderung angesehen, wenn die Vorhersage den Ausgangspunkt der Beratung bilde und weniger das Problembewusstsein der Studierenden. Dieses Problembewusstsein müsse nur unter bestimmten Umständen mit der Prognose des APP-Systems übereinstimmen und dann in der Beratung entwickelt werden.

Zusammenfassend kann festgehalten werden, dass mit der Prognose von Studienleistungen durch ein APP-System von den Stakeholdern in erster Linie Anforderungen an die sich anschließende Bearbeitung der Prognose verbunden werden. Die Leistungsanalyse und daraus resultierende Leistungsprognose von Studienleistungen durch ein APP-System produziert für die Stakeholder keine Problemanalyse. Die Ergebnisse müssen demnach übersetzt werden, was eine zentrale Anforderung des APP-Systems an die beteiligten Akteure darstellt. Diese betonen die Bedeutung der Art der Kommunikation der Prognosen und plädieren für ein persönliches Beratungsgespräch, in dem ein Problembewusstsein unter Umständen erst erarbeitet werden muss. Den Zugang zur Beratung zu organisieren und die Beratungsleistung entsprechend zu gestalten, wurde von den StakeholderInnen als zentrales Problem benannt.

Zu vermeiden ist eine abschreckende Wirkung durch die (automatisierte) Kontaktaufnahme. Insbesondere diese wurde immer wieder problematisierend thematisiert. Hier wurde auf das notwendige „Fingerspitzengefühl“ verwiesen, weswegen eine generalisierte und automatisierte Kontaktaufnahme durch das APP-System als kritisch betrachtet wurde. Für die StakeholderInnen innerhalb des Workshops war klar, dass es spätestens bei der Kontaktaufnahme einer menschlichen Instanz bedarf, damit die Notwendigkeit und die Zielsetzung des angedachten Beratungsgesprächs mit der richtigen Empathie kommuniziert werden kann.

Neben den Ergebnissen der teilnehmenden Beobachtung gerinnt in den qualitativen Interviews eine weitere Erkenntnis: Studienerfolg ist ein mehrdimensionales Konstrukt, welches immer aus mehreren Perspektiven betrachtet werden kann und auch muss. Unsere Forschungsergebnisse fügen sich damit in den von der aktuellen Studienerfolgsforschung favorisierten Blick auf die Komplexität ihres Gegenstands (Bornkessel, 2018; Heinze, 2018; Weber et al., 2018). Die befragten ExpertInnen teilten zwar die Bedenken der Personen aus dem Stakeholder Workshop hinsichtlich der Problemstellung: Wie kann man wen unter welchen Bedingungen zielführend kontaktieren und wie verpflichtend sollen solche Gespräche gemacht werden. Die ExpertInnen betonten allerdings vielfach, und das auch unabhängig von der institutionellen SprecherInnenposition, dass das Konzept des Studienerfolgs perspektivabhängig ist, man also mindestens zwischen einer institutionen- und einer studierendenzentrierten Perspektive differenzieren müsse (vgl. Ülpenich et al., 2023):

Aus der Sicht der Institution sind hohe Abschlusszahlen in Regelstudienzeit als ein Erfolg zu betrachten. Der institutionenzentrierte Erfolgsbegriff favorisiert eine bildungspolitische Perspektive, die als institutionell verankert beschrieben wird. Hier wird der Studienabschluss als Indikator zentral gesetzt, wodurch vor allem der Studienabbruch als Problem zu betrachten ist. Der Fokus auf den Abschluss wird hingegen durch eine weitere Perspektive auf den Studienerfolg kontrastiert. Der studierendenzentrierte Erfolgsbegriff macht sich die Perspektive der Bildungssubjekte zu eigen und betont individuelle Entscheidungen und persönliche Ziele als relevante Aspekte von Lebenswegen; durchaus auch in Abgrenzung zum Abschluss des Studiums mit einem Bildungszertifikat. Studienabbrüche werden hier vielmehr als Entwicklungsanlass, Orientierungsphase oder Findungsprozess gelesen.

Die offene Zielperspektive des Studiums verweist implizit auf eine zentrale Kompetenz der Studierenden als Indikator eines erfolgreichen Studiums: die Kompetenz, Ziele für sich zu definieren, diese zu verfolgen und nicht lediglich die universitär definierten Kompetenzziele zu erreichen. Es lässt sich vermuten, dass diese Kompetenzanforderungen je nach Herkunft und Lebenssituation unterschiedlich erfüllt werden (Dahm et al., 2018, S. 157). Sie scheinen besonders gut zu Studierenden zu passen, deren Lebensmittelpunkt das Studium bildet und die sich Suchbewegungen und Neuorientierungen im Studium leisten können. Dennoch gerinnt hier eine Ausdeutung, die sich aus der institutionellen Wahrnehmung schwerlich messen lässt. Die studierendenzentrierte Perspektive favorisiert eine flexible Studierpraxis, die für die Studierenden Selbstverwirklichung bedeutet und auf Seiten der Studierenden Selbststeuerungsdispositionen voraussetzt.

Diese Perspektive auf den Studienerfolg erschwerte die Zielsicherheit der automatisierten Identifikation von Studierenden, die eine Unterstützungsleistung benötigen. Studienerfolg aus einer studierendenzentrierten Perspektive wurde somit zu einem beweglichen Ziel und ggf. gar zu einer retrospektiven Einschätzung. Was die Kontrastierung der Erfolgsorientierungen aber vor allem verdeutlichte, war die zentrale Anforderung an ein APP-System, Studierende mit abweichenden Orientierungen nicht zu diskriminieren. Eine Operationalisierung verschiedener Erfolgsperspektiven wurde zur zentralen Herausforderung, der wir mit der Erweiterung des Fragebogens im Zuge der quantitativen Erhebung begegnet sind.

Darüber hinaus verdeutlicht die herausgearbeitete Perspektivenvielfalt die Bedeutung der Verknüpfung einer Leistungsvorhersage mit einer persönlichen Beratungsleistung, in der die jeweilige Orientierung reflektiert und Fehlinterpretationen der Prognose begegnet werden kann. Letztlich verweist der Befund der gegensätzlichen Erfolgsorientierungen auf die Bedeutung adaptiver APP-Systeme, die sich an die Nutzenden und ihre Bedürfnisse anpassen lassen (zu adaptiven Lernumgebungen vgl. Wild & Schiefele, 1994).

Zusammenfassend lassen sich die qualitativen Ergebnisse wie folgt beschreiben:

- 1) Es gibt eine Grundskepsis der ExpertInnen hinsichtlich der Anwendungsgebiete eines

APP-Systems im universitären Feld, die sich vor allem aus einem Mangel an Problemanalyse durch das APP-System speisen.

- 2) Die automatisierte Identifikation von Studierenden mit einem APP-System wird kritisch betrachtet, da vor allem die Gefahr einer Self-fulfilling prophecy bei abbruchgefährdeten Studierenden erkannt wird.
- 3) Eine automatisierte Kontaktaufnahme sollte nicht durch ein APP-System geschehen. Die persönliche Beratung und Einordnung wird favorisiert, um den zuvor genannten kritischen Momenten zu begegnen.
- 4) Im Feld existieren gegensätzliche Erfolgsorientierungen, denen ein APP-System und dessen Implementierung Rechnung tragen sollte.
- 5) Kann das APP-System nicht für verschiedene Erfolgsorientierungen adaptiert werden, so muss die zugrundeliegende Orientierung transparent gemacht werden. Dies verbessert die Akzeptanz und die Interpretation der Ergebnisse durch die beteiligten Akteure (zu Explainable AI vgl. Keller et al., 2022).

Neben diesen inhaltlichen Ergebnissen haben die qualitativen Forschungsergebnisse auch einen Einfluss auf die quantitativen Erhebungen gehabt. Die Zufriedenheit mit dem Studium und die Wahrscheinlichkeit eines anschließenden Masterstudiums wurden in der folgenden Welle als Item ergänzt sowie mehrere Variablen zur persönlichen Lebenssituation. Ebenso wurden Fragen zu Zielen des Studiums und weitere mögliche Beratungsangebote ergänzt. Der quantitative Teil von Arbeitspaket 2 modelliert Erfolgsvorhersagen auf Basis von Studierendendaten. Gleichzeitig wird die Frage zu beantworten sein, welche diskriminierungsfreien Variablen gute Prädiktoren für Studienerfolg sind und wie diese möglichst datensparsam, also ohne die Not kontinuierlicher und umfangreicher Erhebungen, verwendet werden können.

In AP II stand die Frage der Verwendbarkeit diskriminierungsbehafteter Variablen zur Bestimmung des Studienerfolgs im Mittelpunkt. In vorheriger Forschung wurden bereits eine Reihe von Variablen auf ihren Zusammenhang zum Studienerfolg untersucht, allerdings unterlagen diese immer bestimmten Restriktionen (Grunschel & Dresel, 2021). In den meisten Fällen wurden lediglich Querschnittsdaten erhoben, womit sich keinerlei kausaler Zusammenhang belegen lässt (Bornkessel, 2018; Pannier et al., 2020). Eine inhaltliche Verknüpfung der erforschten Items ist also nicht gesichert. Bei dem zusätzlichen Aufwand, der für die Messung des Abbruchrisikos entsteht, sollte vorher sichergestellt werden, dass sich dieser tatsächlich lohnt und die richtigen Studierenden, i.S.v. leistungsschwach, erreicht werden. Eine detaillierte Auseinandersetzung mit den einzelnen Einflussfaktoren ist hier also notwendig.

Für die Vorhersage des Studienerfolgs ist eine Messung während des Studiums essenziell, da nur so eine Unterstützung durch die Universität angeboten werden kann, um erfolgreich zum Studienabschluss zu gelangen. Gleichzeitig sollte eine Vorhersage möglichst auf Daten beruhen, die die Universität bereits erfasst hat, damit kein Mehraufwand, i.S.v. Ressourceneinsatz für die Prognose, betrieben werden muss (Schneider et al., 2019). Um zu einer Orientierung über die Daten zu gelangen, die die Universität bereits erfasst hat, bietet sich der Blick in das Hochschulstatistikgesetz (HStatG) an. Nach HStatG§3 werden eine Reihe von Merkmalen über Studierende erfasst, unter anderem: Geschlecht, Geburtsjahr, Staatsangehörigkeit, Ort und Art der Hochschulzugangsberechtigung, berufspraktische Tätigkeit vor Aufnahme des Studiums, Prüfungserfolg und Gesamtnote abgelegter Prüfungen[1]. Ob sich anhand dieser Merkmale eine Vorhersage über den Studienerfolg bestimmen lässt, soll im Folgenden untersucht werden.

Für eine Vorhersage des Studienerfolgs dürfen nur solche Faktoren verwendet werden, die kausal mit einem erfolgreichen Studium in Verbindung stehen und bestehende Ungleichheiten nicht zusätzlich verstärken. Deshalb sollten viele der benannten Einflussfaktoren auf den

Studienerfolg, wie bspw. das Geschlecht oder akademischer familiärer Hintergrund, nicht verwendet werden, da diese eine Dimension sozialer Ungleichheit darstellen (Pannier et al., 2020). Dabei kann zwischen drei Ebenen der Diskriminierung durch eine KI unterschieden werden: (1) kann bereits in den Basisdaten Diskriminierung stattfinden, wenn Personengruppen mit bestimmten Merkmalen, zum Beispiel das Erreichen der Hochschulzugangsberechtigung (HZB) an einer Gesamtschule, unterrepräsentiert sind. (2) kann ein Algorithmus Diskriminierung reproduzieren, wenn er Muster findet, die durch die soziale Ungleichheit bedingt sind, und diese seiner Logik zugrunde legt, zum Beispiel das geringere Vorkommen von Frauen an der mathematisch-naturwissenschaftlichen Fakultät. (3) kann der Algorithmus auch soziale Ungleichheit produzieren, indem die Mustererkennung Zusammenhänge findet, die nicht aus sozialer Ungleichheit entstehen und daher erst mithilfe der KI entdeckt und genutzt werden. Zum Beispiel kann eine Variable, die als erklärend angesehen wird, bereits durch Diskriminierung beeinflusst sein. Diese als Erklärung zu verwenden, produziert im Folgenden Ergebnisse, die durch ebendiese Diskriminierung bedingt sind. Hier kann beispielsweise an den Ort der HZB gedacht werden; ein Gymnasium wird vor allem von Personen besucht, deren Eltern bereits einen höheren Bildungsabschluss besitzen. Dabei muss es sich nicht zwingend um Leistungsunterschiede handeln (zu weiteren Risiken der Diskriminierung im Kontext von Machine Learning vgl. Suresh & Guttag, 2021).

Ein vollständiges Erklärungsmodell für die Leistungen im Studium ist zu unterscheiden von einem System, das potenzielle StudienabbrecherInnen identifiziert. Bei ersterem werden alle Einflussfaktoren ermittelt und ihre Effekte nachgewiesen, während bei letzterem der Fokus auf für das Studium relevante Indikatoren liegt, die eine Intervention ermöglichen. Mithilfe soziodemographischer Variablen lässt sich erklären, warum manche Studierende schlechtere Leistungen erzielen. Hier kann weniger Vorwissen bestehen oder die Personen haben weniger zeitliche Ressourcen für die Universität. Diese Erklärungen sollten jedoch nicht einem System, das zur Vorhersage von Studienleistungen dienen soll, zugrunde gelegt werden. Hier bestünde die Gefahr, dass jede Person, die aus einem nicht-akademischen Haushalt kommt, als potenzielle Studienabbrecherin identifiziert wird. Jedoch ist nicht die soziale Herkunft der Grund für einen Studienabbruch, sondern die fehlende Passung oder Motivation. Die soziale Herkunft ist also nicht kausal für den Abbruch, stellt aber einen mittelbaren Einfluss dar. Zwischen diesen beiden Phänomenen kann ein Algorithmus jedoch nicht unterscheiden, weshalb die Auswirkungen im Ergebnis identisch wären. Ein System, das gefährdete Studierende ermitteln soll, sollte als Grundlage jene Faktoren verwenden, die einen unmittelbaren Einfluss auf den Studienerfolg haben. Diese Faktoren sollten im Vordergrund eines Systems zur Vorhersage von Studienleistungen stehen.

## Methodik des quantitativen Teilprojekts

Um die theoretischen Annahmen empirisch zu prüfen, ist das quantitative Teilprojekt von Arbeitspaket 2 als Querschnitts- und Panelstudie mit drei Wellen angelegt. Die Studierenden wurden Anfang 2022, Ende 2022 und Ende 2023 zur Befragung eingeladen. Es wurden Studierende des ersten, dritten, fünften und siebten Fachsemesters (wieder-)befragt. Gleichzeitig war die Befragung für jedes Semester offen. Es ist somit innerhalb des Projektes möglich, Studienverläufe präzise nachzuvollziehen. Um eine exakte Messung zu ermöglichen, wurden jeweils die vorliegenden Prüfungsdaten der Studierenden beim Prüfungsamt angefragt. Diese wurden dem Projekt nach datenschutzrechtlicher Prüfung und unter Nutzung eines klaren Datenverarbeitungsschemas zur Verfügung gestellt<sup>[2]</sup>. Somit liegen Daten zum Titel der Veranstaltung, Semester der Veranstaltung, Art der Veranstaltung, Anzahl gesammelter ECTS, Note bei Prüfung für jeden Studierenden über einen Zeitraum von drei Jahren vor. Ergänzend liegen ebenfalls Daten zu Prüfungsrücktritten sowie nicht bestandenen Prüfungen vor. Durch die Ausführlichkeit dieser Daten kann zumindest für die Leistungen der Studierenden ein reliables und valides Ergebnis produziert werden.

Gleichzeitig sollen alle relevanten Daten zum Studienverhalten sowie Soziodemographie

durch eine Befragung erhoben werden. Die Ausgestaltung des Fragebogens teilt sich dabei in zwei Teile: Items, die zeitlich stabil sind und demnach nur einmal erhoben wurden, und Items, die sich im Laufe der Zeit verändern (können), die in jeder Befragungswelle erneut erhoben wurden. Sollte es also zum Beispiel bei der Abbruchsintention der Studierenden zu einer Veränderung kommen, kann nachverfolgt werden, ob sich dadurch die Leistungen im Studium verändern beziehungsweise ob diese vielleicht schon vorher schlechter waren und dadurch die Abbruchsintention erst anstieg. Kausale Rückschlüsse sind also möglich.

Die Akquirierung von Befragten stellte sich als Herausforderung dar. Besonders im Hochschulkontext werden viele Befragungen durchgeführt, weshalb Studierende eine gewisse Teilnahmemüdigkeit per se aufweisen. Gleichzeitig war die zu befragende Grundgesamtheit recht klein, weshalb wir innerhalb dieser eine hohe Rückmeldung benötigten. Erreicht wurden die Studierenden über Vorlesungen in ihren respektiven Fächern, die für alle verpflichtend sind. So konnte sichergestellt werden, dass zumindest theoretisch alle Studierenden erreicht werden konnten. Gleichzeitig muss erwähnt werden, dass Vorlesungen nicht von jedem besucht werden. Um diesem Aspekt zu begegnen, wurden die Studierenden ebenso über das Forum der entsprechenden Lehrveranstaltung eine Nachricht über die Befragung geschaltet. Ebenso wurde eine Nachricht im Studierendenportal platziert. Bei den Wiederbefragungen wurden bereits befragte Personen per E-Mail auf die Erhebung hingewiesen. Da die Rücklaufquote in der ersten Welle trotz aller dieser Maßnahmen niedrig blieb, wurde eine Incentivierung ergänzt. Alle Teilnehmenden hatten die Möglichkeit, 500€ zu gewinnen, diese wurden fünfmal unter den Befragten verlost. Diese Maßnahme zeigte hinsichtlich der Rückläufe eine deutliche Erhöhung, weswegen an der Incentivierung für die Wellen zwei und drei festgehalten wurde. Da es notwendig ist, möglichst vollständige Datensätze zu haben, wurden die Incentives nur unter den Studierenden verlost, die (a) die Umfrage komplett abgeschlossen haben und (b) plausible Antworten zeigten. Letzteres wurde im Prozess der Datenbereinigung mit gängigen Mitteln der Sozialwissenschaften geprüft (s.u.). Diese Information wurde transparent an die Studierenden kommuniziert. Die Höhe der Incentives ergibt sich aus der Länge des Fragebogens in Kombination mit der Notwendigkeit einer hohen Ausschöpfungsquote der kleinen Grundgesamtheit. Eine weitere Vorgehensweise, die eine hohe Ausschöpfung gewährleisten sollte, war die persönliche Bewerbung der Umfrage in den Lehrveranstaltungen. Insgesamt erhielten wir durch dieses Bündel an Maßnahmen Stichprobengrößen, mit denen multivariate Analysen durchgeführt werden konnten. Zwar sind auch unsere Befragten im Notendurchschnitt und hinsichtlich des erreichten European Credit Transfer System (ECTS)[3] etwas besser als die Grundgesamtheit, diese Unterschiede sind aber nicht signifikant. Mit einer solchen Befragung erreicht man klassischerweise vor allem erfolgreiche Studierende (Pannier et al., 2020). Der Unterschied ist jedoch nicht so groß, dass die Repräsentativität gefährdet ist. Weitere Analysen auf Basis der Daten der Grundgesamtheit zeigen ebenso nur marginale Abweichungen, die statistisch zu vernachlässigen sind. Hierbei wurden das Alter, Geschlecht und die Nationalität verglichen.

Die Fragebögen jeder Welle waren jeweils circa zwei Monate im Feld und wurden in diesem Zeitraum beworben. Die vorherigen Teilnehmenden wurden mehrmals per Mail auf die Erhebung hingewiesen. Nach Abschluss der Feldphase wurden die Matrikelnummern der Befragten mit ihren Leistungsdaten vom Prüfungsamt verknüpft. Dabei wurden die Daten durch die Projektleitung anonymisiert. Der endgültige Datensatz enthielt somit keine Daten mehr, die einen individuellen Rückschluss ermöglichten.

Die Daten wurden im Anschluss bereinigt. Alle Befragten, die nicht Sozialwissenschaften oder Informatik studieren, wurden entfernt. Die Befragungsdaten wurden auf auffällige Antwortmuster untersucht und diese Fälle ebenso entfernt. Wenn mehr als 10 Fragen nicht beantwortet wurden oder die Bearbeitungszeit sehr kurz/lang war, wurden diese Fälle entfernt.

Für den Fragebogen wurden in der Bildungsforschung bekannte Items verwendet. Nach der

Literaturrecherche konnten die wichtigsten Variablen für den Studienerfolg ermittelt werden und wurden im Folgenden in Fragebogenitems übersetzt. Hier konnte in den meisten Fällen auf bestehende Skalen zurückgegriffen werden: Soziale Integration (Dahm & Lauterbach, 2016), Akademische Passung (Weber et al., 2018), Allgemeine Selbstwirksamkeit (Beierlein et al., 2014), Efficacy (Kovaleva et al., 2014), die BIG 5 (Rammstedt et al., 2014), Selbstreflexivität (Knispel et al., 2021). Für all diese Konstrukte wurden Indizes gebildet, nachdem die Reliabilität mithilfe von Cronbachs Alpha geprüft wurde. Ergänzend wurde die Studienfachabbruchsintention erfasst (Blüthmann et al., 2011), um diese als moderierenden Faktor in die statistischen Modelle integrieren zu können; wer sowieso geneigt ist sein Studium abzubrechen, wird sich aller Voraussicht nach weniger engagiert zeigen, Leistungen zu erbringen. Der Fragebogen umfasst ebenso Variablen, die in der Bildungsforschung als klassische Einflussfaktoren benannt werden, wie die Note der Hochschulzugangsberechtigung, den Besuch eines Gymnasiums und die Bildung der Eltern. Für das Kulturkapital wurde mit einer offenen Angabe gefragt „Wie viele Bücher gab es ungefähr bei Ihnen zu Hause, als Sie 16 Jahre alt waren?“ (Sieben & Lechner, 2019). Für das Leben in einem Einelternhaushalt wurde gefragt: „Als Sie 15 Jahre alt waren, haben Sie damals mit Ihren beiden Eltern gemeinsam in einem Haushalt gelebt?“.

Als intervenierende Variablen der Universität wurden unterstützende Angebote wie Mentoringprogramme und das Orientierungstutorium angesehen, weshalb die Studierenden nach einer Nutzung dieser Angebote gefragt wurden. Weitere Interventionsmöglichkeiten wurden in der dritten Befragungswelle auf Basis der teilnehmenden Beobachtung im Stakeholder-Workshop sowie der qualitativen Interviews ergänzt, um weitere erfolgreiche Konzepte zu ermitteln. Das Angebot an der Universität ist hier sehr breit aufgestellt, weshalb eine dezidierte Auseinandersetzung mit den verschiedenen Unterstützungsangeboten sinnvoll ist, um mögliche Lücken zu schließen.

Durch die Verknüpfung von qualitativen und quantitativen Methoden konnten Ergänzungen für den Fragebogen aus den Interviews gezogen werden. Bereits in der zweiten Welle wurde eine Itematterie zum Thema Studieneinstiegsselbstwirksamkeit, da früh deutlich wurde, dass besonders der Studienbeginn wichtig für den weiteren Studienverlauf ist (Petri, 2020). In der dritten Welle wurden zum einen Items zu eigenen Pflgetätigkeiten und Fahrtzeit zur Universität ergänzt. Zum anderen wurde angemerkt, dass die Abschlussnote des Studiums besonders für die Personen wichtig ist, die ein Masterstudium anstreben. Hier wurde dementsprechend eine Frage ergänzt, inwiefern sich die Befragten ein Masterstudium vorstellen können. Ebenso wurden Fragen zur Zufriedenheit und zur individuellen Zielorientierung im Studium ergänzt.

Im Rahmen der institutionenorientierten Operationalisierung der Variablen wurde der Studienerfolg sowohl über die gesammelten ECTS (im Studienverlauf) sowie der durchschnittlichen Note messbar gemacht. Dabei können die ECTS eine generalisierte Idee über den Studienverlauf geben, während die Note spezifischer den Erfolg darstellt. Dabei ist zu beachten, dass Noten je nach Fach stark variieren (Müller-Benedict & Tsarouha, 2011). Ebenso mag für manche Studierende der Notendurchschnitt nicht in gleichem Ausmaß relevant sein, wie das Abschließen des Studiums. Die Berechnung der gesammelten ECTS ist ebenso wichtig, um einzuschätzen, ob das Studium tatsächlich in Vollzeit absolviert wird. Denn ein Studienabbruch wird durch das Studieren eines Vollzeitstudiums in Teilzeit erhöht (Dahm et al., 2018). Dabei sind dahinter liegende Effekte wie eingeschränkte Möglichkeiten für das Studium noch nicht näher untersucht. Mithilfe der gemachten Befragungsdaten können (kausale) Zusammenhänge zwischen Studierverhalten, soziodemographischen Merkmalen, Lebenssituation und Studienleistungen untersucht werden.

## Untersuchungsergebnisse des quantitativen Teilprojekts

Die Studierenden mussten im Verlauf der Umfrage ein eindeutiges Pseudonym vergeben,

damit sie (a) im Zuge weiterer Wellen identifiziert und (b) kontaktiert werden können. Die initiale Einladung für die erste Welle erfolgte zu Beginn des Jahres 2022 und die Feldphase endete im März 2022. Die zweite Welle wurde von November 2022 bis Januar 2023 erhoben, die dritte Welle lief von Dezember 2023 bis Februar 2024. Die Ergebnisse werden separiert für die Wellen vorgestellt.

In der ersten Welle haben insgesamt 367 Personen den Fragebogen begonnen und 250 mit der Zustimmung zur Datenabstimmung mit der Universitätsverwaltung beendet. Nach der Datenbereinigung verbleiben 157 Fälle für die finalen Analysen. Ebenso kam es vor, dass Studierende nicht im ersten oder dritten Semester waren, diese wurden im Datensatz belassen und lediglich für entsprechende Analysen gefiltert. Diese Daten wurden im Folgenden mit den Prüfungsdaten der Studierenden verknüpft.

Tabelle 4: Variablenbeschreibung Welle 1

Variable	Kodierung	Min	Max	M	SD
Gesundheitszustand	1 – sehr schlecht 5 – sehr gut	1	5	3,81	0,96
soziales Kapital	N Anzahl Bücher im Elternhaushalt	0	350	305,3	529,4
soziales Netzwerk	N Unterstützung	0	7	1,02	1,54
Bildung Eltern (Mittelwert)	100 – ohne Schulabschluss 244 – Abschluss Sekundarbereich I 344 – Abschluss Sekundarbereich II 645 – Bachelor bzw. gleichwertiges Programm 844 – Promotion	241	844	421,1	190,1
Ort HZB	1 – in Düsseldorf 2 – 50 km um Düsseldorf 3 – NRW >50 km von Düsseldorf 4 – Deutschland aber außerhalb NRW 5 – außerhalb von Deutschland	1	5	2,4	1,05
Note HZB	Angabe Durchschnittsnote	1	3,8	2,25	0,59
Studienabbruchintention	1 – niedrige Abbruchintention 10 – hohe Abbruchintention	1	10	3	2,26
Alter	Angabe Alter	18	66	23,26	5,58

Im Datensatz befinden sich 66 Studierende der Sozialwissenschaften und 91 aus der Informatik. 89 Personen identifizieren sich als männlich, 67 als weiblich und eine Person als divers. Für die Zwecke der Analyse wird diese Person als missing bei der Dummy Variable kodiert. Die Form der Hochschulzugangsberechtigung kann über die Universität ausgegeben werden. Nach diesen Angaben haben 70,1 % der Studierenden ihre Hochschulzugangsberechtigung an einem Gymnasium absolviert, 5,8 % an einem Abendgymnasium, 0,6 % ist beruflich qualifiziert, 4,5 % an einem Fachgymnasium, 13,6 % Gesamtschule, 3,8 % haben eine ausländische HZB und 1,3 % werden als sonstige Studienberechtigte ausgegeben. Für die Analyse wird eine Dummy Variable erstellt, die zwischen einem HZB an einem Gymnasium und allen anderen unterscheidet. Ergänzend wurden die Studierenden gefragt, ob sie vorher eine Ausbildung absolviert haben, das war für 11,5 % der Fall. 22,9 % der Befragten gaben an, dass sie in einem Einelternhaushalt aufgewachsen sind.

Um mögliche Einflussfaktoren auf die Leistungsmerkmale im Studium zu bestimmen, wurden lineare Regressionen berechnet. Hierbei wurden die in der Bildungsforschung klassischen Variablen als unabhängige Variablen ins Modell integriert. Als abhängige Variablen wurden die durchschnittlichen ECTS pro Semester und die durchschnittliche Note der Prüfungen verwendet. Der Einfluss der oben beschriebenen Variablen wurde untersucht. Aus Platzgründen werden in den hier dargestellten Modellen nicht alle analysierten Variablen abgetragen. Für die nicht dargestellten Variablen finden wir keinen signifikanten Effekt für die ECTS oder die durchschnittliche Note.

Tabelle 5: Lineare Regression auf die durchschnittlichen ECTS pro Semester

AV: durchschnittliche ECTS pro Semester	B	Std. Fehler	B	Std. Fehler	B	Std. Fehler
Konstante	19,60 *	5,545	27,82 ***	7,211	16,15	10,808
	4 *		2		0	
Big 5 Extraversion	1,049	0,989	0,684	0,942	1,417	0,966
Big 5 Neurozentrismus	0,001	0,833	0,073	0,909	0,833	0,968
Big 5 Offenheit für Erfahrungen	-0,761	0,944	-0,720	0,919	-0,965	0,927
Big 5 Gewissenhaftigkeit	1,813	1,172	0,254	1,187	-0,092	1,252
Big 5 Verträglichkeit	-2,132 *	0,985	-1,295	0,947	-1,409	0,970
Soziale Integration			-0,500	1,192	-0,033	1,230
Akademische Passung			1,629	1,609	2,022	1,631
Allgemeine Selbstwirksamkeit			-0,966	1,461	-1,407	1,513
Efficacy			1,138	1,484	1,883	1,556
Abbruchsintention			-1,415 **	0,464	-1,625 *	0,480
soziales Netzwerk			-0,291	0,629	0,206	0,687
Selbstreflexivität			-0,245	2,431	-0,124	2,451

HZB Gymnasium (Ref. alle anderen)		3,520 +	1,824	0,962	1,997
HZB Note		-0,407 **	0,145	-0,304 *	0,152
Staatsangehörigkeit Deutsch (Ref. alle anderen)				5,961	4,038
Geschlecht (Ref. männlich)				-0,714	1,826
Alter				-0,421 *	0,207
Bildung Eltern (ISCED, z-standardisiert)				0,884	0,975
Sprache deutsch (Ref. alle anderen)				-0,806	2,743
Broken Home (Ref. ja)				2,625	2,045
Kulturkapital				0,000	0,002
Gesundheitszustand				1,862 +	0,996
vorherige Ausbildung				2,673	2,870
n	144	144	144	144	144
Adjusted R <sup>2</sup>	0,022	0,143	0,177	0,177	0,177
Änderung in F	1,634	3,172***	1,584	1,584	1,584

Anmerkung: +  $p < 0,1$ ; \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; OLS Regression; p-Werte berechnet auf Basis des T-Tests

In Tabelle 5 werden die Ergebnisse der Regression für die durchschnittlichen ECTS pro Semester abgetragen. Dafür wurden drei Modelle geschätzt, in welche 144 Befragte Eingang fanden. Zunächst wurden die Big 5 aufgenommen, um die Persönlichkeitsmerkmale der Studierenden zu beachten. Hier kann jedoch nur ein signifikanter Effekt für die *Verträglichkeit* festgestellt werden. Dementsprechend liegt das korrigierte R<sup>2</sup> bei 0,022, das Modell erklärt also lediglich 2,2% der Streuung der abhängigen Variablen „durchschnittliche ECTS pro Semester“.

Durch die Ergänzung der Variablen zur Studiensituation und vorherigen akademischen Leistungen im zweiten Modellschritt steigt die Varianzaufklärung signifikant um 12,1% auf in Summe 14,3%. Die Abbruchsintention ist signifikant und hat einen negativen Effekt auf die ECTS, die durchschnittlich pro Semester erreicht werden. Wer also darüber nachdenkt, das Studium abzubrechen, erreicht signifikant weniger ECTS pro Semester. Ebenso hat der Ort der Hochschulzugangsberechtigung einen signifikanten Einfluss auf die ECTS pro Semester. Hat eine Person ihre Hochschulzugangsberechtigung an einem Gymnasium erhalten, erreicht sie durchschnittlich 3,5 ECTS (pro Studiensemester) mehr als eine Person, die an einer anderen Bildungseinrichtung die Berechtigung zum Studium erhalten hat. Geht man von einer Regelstudienzeit von sechs Semestern aus, bedeutet dies einen ECTS-Vorsprung von 21 Punkten für Personen, die von einem Gymnasium kommen. Anders formuliert: Personen, die die Hochschulzugangsberechtigung auf einem anderen Weg erhalten haben, brauchen im Mittel fast ein Semester länger für den Abschluss. Ergänzend ist auch die Note der Hochschulzugangsberechtigung signifikant für die ECTS, die im Semester durchschnittlich erreicht werden. Je besser die Abiturnote ist, desto mehr ECTS werden im Semester erreicht. Eine hypothetische Person, die ihr Abitur mit 1,0 abgeschlossen hat, erreicht in der Regelstudienzeit durchschnittlich 48 ECTS mehr als eine Person, die ihr Abitur mit 3,0 abgeschlossen hat, weshalb letztere zwei Semester mehr brauchen wird[4].

Im dritten Modell werden soziodemographische Variablen ergänzt, um hier diskriminierende Variablen zu identifizieren. Die gemeinsame Erklärkraft aller Variablen beträgt 17,7% der Streuung der abhängigen Variablen. Neben der Abbruchsintention und der Note der Hochschulzugangsberechtigung ist hier außerdem das Alter signifikant. Bei höherem Alter fällt die Anzahl der durchschnittlichen Anzahl an ECTS, die erreicht werden, niedriger aus. Personen, die darüber nachdenken, das Studium abzubrechen, erreichen pro Semester weniger ECTS als Personen, die keinen Abbruch erwägen. Ebenso ist der Gesundheitszustand signifikant und hat einen positiven Zusammenhang mit den erreichten Leistungspunkten. Wird der Gesundheitszustand subjektiv besser eingeschätzt, werden pro Semester durchschnittlich mehr ECTS erbracht.

Tabelle 6: Lineare Regression auf die durchschnittliche Note

AV: durchschnittliche Note	B	Std. Fehler	B	Std. Fehler	B	Std. Fehler
Konstante	2,525***	0,440	0,923 +	0,499	1,620 *	0,778
Big 5 Extraversion	0,068	0,081	0,085	0,066	0,069	0,070
Big 5 Neurozentrismus	0,079	0,067	0,015	0,063	0,022	0,070
Big 5 Offenheit für Erfahrungen	-0,045	0,077	-0,001	0,065	0,002	0,067
Big 5 Gewissenhaftigkeit	-0,179 +	0,094	-0,023	0,082	-0,003	0,088
Big 5 Verträglichkeit	0,108	0,080	0,052	0,066	0,036	0,069
Soziale Integration			0,027	0,084	0,038	0,088
Akademische Passung			-0,080	0,112	-0,128	0,116
Allgemeine Selbstwirksamkeit			0,159	0,101	0,119	0,108
Efficacy			-0,116	0,103	-0,052	0,111
Abbruchsintention			0,077 *	0,035	0,094 *	0,037
soziales Netzwerk			0,032	0,044	0,013	0,048
Selbstreflexivität			0,224	0,170	0,266	0,174
HZB Gymnasium (Ref. alle anderen)			-0,355 **	0,129	-0,330 *	0,145
HZB Note			0,063 **	0,010	0,056 **	0,011
			*		*	
Staatsangehörigkeit Deutsch (Ref. alle anderen)					-0,650 *	0,298
Geschlecht (Ref. männlich)					-0,044	0,132
Alter					-0,012	0,015
Bildung Eltern (ISCED, z-standardisiert)					-0,052	0,071
Sprache deutsch (Ref. alle anderen)					0,090	0,193
Broken Home (Ref. ja)					-0,061	0,144
Kulturkapital					0,000	0,000
Gesundheitszustand					0,078	0,071

vorherige Ausbildung			-0,063	0,202
n	139	139	139	
Adjusted R <sup>2</sup>	0,009	0,349	0,349	
Änderung in F	1,255	8,705***	0,997	

Anmerkung: +  $p < 0,1$ ; \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; OLS Regression; p-Werte berechnet auf Basis des T-Tests

In Tabelle 6 wurde eine Regression für die durchschnittliche Note, die im Studium erreicht wird, berechnet. Bei dieser Regression unterscheidet sich die betrachtete Personenzahl, da einzelne Studierende noch keine benotete Leistung erbracht haben und somit nicht Teil dieser Analyse sind.

Es wurden erneut drei Modelle geschätzt. Zunächst wurden die Persönlichkeitsmerkmale der Big 5 eingeführt. Hier wird ein korrigiertes R<sup>2</sup> von 0,009 (= 0,9%) ausgegeben. *Gewissenhaftigkeit* hat hier einen signifikant negativen Effekt auf die Durchschnittsnote. Das bedeutet, dass bei einem höheren Wert für *Gewissenhaftigkeit* die Note besser wird.[5]

Im zweiten Modell wurden die Variablen zur Studiensituation und den vergangenen akademischen Leistungen ergänzt. Mithilfe dieser Variablen können 34,9% der Varianz der durchschnittlichen Note erklärt werden. Hier sind sowohl die Abbruchsintention, der Ort der Hochschulzugangsberechtigung als auch die Note der Hochschulzugangsberechtigung signifikant. Bei einer höheren Abbruchsintention verschlechtert sich die durchschnittliche Note der Studierenden. Wenn jemand die Hochschulzugangsberechtigung nicht an einem Gymnasium erreicht hat, verschlechtert sich die durchschnittliche Note dadurch um durchschnittlich 0,355 Notenpunkte im Vergleich zu einer Person, die das Gymnasium besucht hat. Studierende, die einen alternativen Bildungsweg gewählt haben, sind also ungefähr eine Drittelnote schlechter als die Studierenden, die ein Gymnasium besucht haben. Für jede Verbesserung der Abiturnote um 0,1 verbessert sich die durchschnittliche Note im Studium um 0,063. Zwischen einem 1,0 Abitur und einem 2,0 Abitur ergeben sich im Studium also ein Unterschied von 0,6 Notenpunkten. Dieser Unterschied kann nicht nur, aber insbesondere im Hinblick auf die Zulassung zu einem Master relevant werden.

Im dritten Modell werden soziodemographische Variablen hinzugefügt und erbringen keinen wesentlichen Erklärungszuwachs. Die einzige neue Variable, die hier einen signifikanten Effekt hat, ist die Staatsangehörigkeit. Besitzt eine Person keine deutsche Staatsangehörigkeit, verschlechtert sich die durchschnittliche Note im Studium um 0,650 im Vergleich zu einer Person mit deutscher Staatsangehörigkeit. Dabei handelt es sich um einen Unterschied von mehr als einer halben Note, der allein durch die Staatsangehörigkeit erklärt werden kann.[6]

Für unseren Datensatz der Welle 1 finden sich unter den untersuchten Einflussvariablen nur vereinzelt signifikante Effekte. Für die durchschnittlichen Leistungspunkte, die pro Semester erreicht werden, können wir konstatieren, dass die Note der Hochschulzugangsberechtigung sowie die Abbruchsintention die stärksten Einflüsse haben.

Ergänzend wurde eine *backward Regression* berechnet, wobei schrittweise alle Variablen entfernt werden, die keinen signifikanten Regressionskoeffizienten aufweisen (Spilke & Mielenz, 2006). Bei der Verwendung von vielen unabhängigen Variablen ist eine *backward Regression* besonders hilfreich, um explorativ den Einfluss unterschiedlicher Variablen zu

testen. Dies ist hier sinnvoll, da es so viele potenzielle Einflussfaktoren gibt, wobei aber nur wenige signifikante Effekte auftreten. Für die durchschnittlichen ECTS pro Semester ergeben sich dabei signifikante Ergebnisse für akademische Passung, Abbruchsintention, HZB Note, der Besuch eines Orientierungstutoriums, Alter, Gesundheitszustand und der vorherige Abschluss einer Ausbildung. Für die durchschnittliche Note finden wir signifikante Einflüsse für die allgemeine Selbstwirksamkeit, Selbstreflexivität, HZB Gymnasium, HZB Note und den Gesundheitszustand.

Die Note der Hochschulzugangsberechtigung muss logischerweise als Indikator für die akademischen Leistungen der Studierenden betrachtet werden (Heublein et al., 2017). Insofern ist ein Effekt hier nicht nur erklärbar, sondern auch erwünscht. Allerdings muss hierbei beachtet werden, dass schon die Abiturnote stark durch die soziale Herkunft beeinflusst wird (Hovestadt & Eggers, 2007), wie auch die *backward Regression* nahelegt. Somit kann davon ausgegangen werden, dass sowohl beim Studienzugang als auch im Studienverlauf Kinder aus akademischen Haushalten bessere Chancen haben. Dadurch kann hier indirekte Diskriminierung stattfinden, indem Personen aus nicht-akademischen Haushalten gar nicht die passende Note erreichen, um bestimmte Fächer zu studieren. Ebenso kann die Selbsteinschätzung der Studierenden durch schlechtere Noten leiden. Da dieses Problem jedoch vor der Aufnahme des Studiums entsteht, liegt es außerhalb der Einflussphäre von Universitäten. Mögliche Ansätze für gerechtere Bildungschancen müssen früher ansetzen, wo sie auch Erfolge verzeichnen können (Erdmann et al., 2022).

Die Abbruchsintention hat ebenso einen Effekt auf die erreichten Leistungspunkte. Hierbei ist die Richtung des Effekts unklar. So kann es auf der einen Seite durch fehlende Leistungen und nicht bestandene Klausuren zu einer höheren Abbruchsintention kommen. Ebenso kann eine Abbruchintention auf der anderen Seite durch weitere Faktoren wie fehlende Motivation bedingt sein und führt dann im Folgenden zu weniger Leistungen, da der Abschluss des Studiums sowieso in Frage gestellt wird. In beiden Fällen lässt sich dieser Effekt erklären und wird durch die Theorie gestützt (Theune, 2021). Es ist davon auszugehen, dass die anderen Variablen zur Studiensituation ebenso einen Effekt haben, jedoch von dem Effekt der Abbruchsintention mediiert werden. Zwischen diesen Variablen besteht eine hohe Korrelation, jedoch ist die Abbruchsintention die Variable mit der höchsten Polarisierung. Im Zuge weiterer Wellen wird der Versuch unternommen, diesen aktuell nur als Korrelation beschreibbaren Zusammenhang in Panelmodellen kausal zu quantifizieren.

Für die Bestimmung des Studienerfolgs kann festgestellt werden, dass die gesammelten ECTS deutlich weniger durch diskriminierende Variablen bestimmt werden. Zwar finden sich auch hier Effekte der sozialen Herkunft, diese lassen sich jedoch nicht von schulischen Leistungen trennen. Eine bessere Variable ist nicht einfach zu finden, also sollte bei einer Implementierung eines Vorhersagemodells hierauf zurückgegriffen werden.

In Welle 2 unserer Erhebung haben wir zum einen die zuvor befragten Personen erneut befragt und zum anderen wieder über Vorlesungen der ersten und dritten Semester rekrutiert. Insgesamt ergab sich daraus eine Stichprobe von 273.

Tabelle 7: Variablenbeschreibung Welle 2

Variable	Kodierung	Min	Max	M	SD
Gesundheitszustand	1 – sehr schlecht 5 – sehr gut	1	5	3,82	0,94

soziales Kapital	N Anzahl Bücher im Elternhaushalt	0	2000	306,72	760,68
soziales Netzwerk	N Unterstützung	0	7	1,07	1,65
Bildung Eltern (Mittelwert)	100 – ohne Schulabschluss 244 – Abschluss Sekundarbereich I 344 – Abschluss Sekundarbereich II 645 – Bachelor bzw. gleichwertiges Programm 844 – Promotion	241	844	461,89	203,34
Ort HZB	1 – in Düsseldorf 2 – 50 km um Düsseldorf 3 – NRW >50 km von Düsseldorf 4 – Deutschland aber außerhalb NRW 5 – außerhalb von Deutschland	1	5	2,47	1,09
Note HZB	Angabe Durchschnittsnote	1	3,6	2,25	0,67
Studienabbruchintention	1 – niedrige Abbruchintention 10 – hohe Abbruchintention	1	10	3,35	2,36
Alter	Angabe Alter	17	66	21,77	4,89

Im Datensatz befinden sich 120 Studierende der Sozialwissenschaften und 153 aus der Informatik. 141 Personen identifizierten sich als männlich, 117 als weiblich und zwei Personen als divers. Für die Zwecke der Analyse werden diese beiden Personen als missing bei der Dummy Variable kodiert. 73,7 % der Studierenden haben ihre Hochschulzugangsberechtigung an einem Gymnasium absolviert, 4,6 % an einem Abendgymnasium, 3,9 % an einem Fachgymnasium, 10,0 % Gesamtschule, 2,3 % haben eine ausländische allgemeine HZB, 3,9 % eine ausländische fachspezifische HZB, 0,8 % werden als sonstige Studienberechtigte ausgegeben, 0,4 % haben eine HZB aus einer deutschen Schule im Ausland und 0,4 % haben ein Studienkolleg besucht. Für die Analyse wird eine Dummy Variable erstellt, die zwischen einem HZB an einem Gymnasium und allen anderen unterscheidet. Ergänzend wurden die Studierenden gefragt, ob sie vorher eine Ausbildung absolviert haben, das war für 5,6 % der Fall. 23,5 % der Befragten gaben an, dass sie in einem Einelternhaushalt aufgewachsen sind.

Tabelle 8: Lineare Regression auf die durchschnittlichen ECTS pro Semester

AV: durchschnittliche ECTS pro Semester	B	Std. Fehler	B	Std. Fehler	B	Std. Fehler
Konstante	7,609	5,365	19,832 *	8,507	6,852	12,190
Big 5 Extraversion	0,593	0,929	-0,138	0,951	0,439	0,966

Big 5 Neurozentrismus	0,396	0,755	1,071	0,874	0,991	0,910
Big 5 Offenheit für Erfahrungen	- *	0,877	-1,949 *	0,867	-2,640 **	0,899
	1,915					
Big 5 Gewissenhaftigkeit	2,728 *	0,927	1,671	1,039	1,145	1,066
	*					
Big 5 Verträglichkeit	0,933	1,015	0,525	1,022	0,634	1,046
Soziale Integration			-0,769	1,156	-0,431	1,238
Akademische Passung			-0,541	1,559	-0,110	1,663
Allgemeine Selbstwirksamkeit			-1,276	1,346	-1,657	1,389
Efficacy			-0,043	1,448	-0,393	1,483
Abbruchsintention			-0,789 +	0,419	-0,837 +	0,428
soziales Netzwerk			-0,370	0,566	-0,274	0,577
Selbstreflexivität			0,757	2,082	1,754	2,131
HZB Gymnasium (Ref. alle anderen)			3,212 +	1,810	3,151	2,017
HZB Note			-0,169	0,139	-0,111 *	0,142
Staatsangehörigkeit Deutsch (Ref. alle anderen)					4,476	3,291
Geschlecht (Ref. männlich)					-0,909	1,754
Alter					0,411 *	0,200
Bildung Eltern (ISCED, z-standardisiert)					1,750 *	0,840
Sprache deutsch (Ref. alle anderen)					-3,314	2,402
Broken Home (Ref. ja)					3,928 *	1,951
Kulturkapital					-0,001	0,001
Gesundheitszustand					-0,053	1,048
vorherige Ausbildung					0,350	3,688
n	169		169		169	
Adjusted R <sup>2</sup>	0,059		0,114		0,139	
Änderung in F	3,088*		2,134*		1,496	

Anmerkung: +  $p < 0,1$ ; \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; OLS Regression; p-Werte berechnet auf Basis des T-Tests

In Tabelle 8 sind die Ergebnisse der Regression für die durchschnittlichen ECTS pro Semester abgetragen. Dafür wurden erneut drei Modelle geschätzt, das Vorgehen ist analog zur Welle 1. Im Vergleich zur ersten Welle kann ein signifikanter Effekt für Big 5 *Offenheit* und *Gewissenhaftigkeit* festgestellt werden, Big 5 *Verträglichkeit* ist nicht signifikant. Wie im ersten Modell sind Abbruchsintention, HZB Gymnasium, HZB Note und Alter signifikant. Der Gesundheitszustand ist in diesem Modell nicht signifikant, dafür finden wir Effekte für die Bildung der Eltern und das Aufwachsen in einem Einelternhaushalt. Die Bildung der Eltern hat einen positiven Effekt auf die ECTS, haben die Eltern also einen höheren Bildungsabschluss, sammelt das Kind durchschnittlich mehr ECTS pro Semester. Eine Person, die in einem

Einelternerhaushalt aufgewachsen ist, sammelt circa 4 ECTS weniger pro Semester als eine Person, die mit beiden Eltern aufgewachsen ist. Alle anderen Variablen sind nicht signifikant.

Tabelle 9: Lineare Regression auf die durchschnittliche Note

AV: durchschnittliche Note	B	Std. Fehler	B	Std. Fehler	B	Std. Fehler
Konstante	3,541 ** *	0,534	1,206	0,729	2,188 *	1,041
Big 5 Extraversion	-0,042	0,096	0,068	0,083	0,044	0,085
Big 5 Neurozentrismus	0,027	0,076	0,023	0,075	0,021	0,078
Big 5 Offenheit für Erfahrungen	0,068	0,089	0,086	0,075	0,090	0,079
Big 5 Gewissenhaftigkeit	-0,273 **	0,094	-0,159 +	0,089	-0,115	0,091
Big 5 Verträglichkeit	-0,040	0,104	0,026	0,091	0,043	0,094
Soziale Integration			0,181 +	0,104	0,173	0,111
Akademische Passung			0,072	0,136	0,068	0,144
Allgemeine Selbstwirksamkeit			0,118	0,116	0,180	0,121
Efficacy			-0,154	0,127	-0,204	0,129
Abbruchsintention			-0,006	0,036	-0,008	0,037
soziales Netzwerk			0,017	0,051	0,022	0,051
Selbstreflexivität			0,182	0,181	0,141	0,186
HZB Gymnasium (Ref. alle anderen)			-0,606 ** *	0,157	-0,668 ** *	0,177
HZB Note			0,072 ** *	0,012	0,071 ** *	0,012
Staatsangehörigkeit Deutsch (Ref. alle anderen)					0,382	0,283
Geschlecht (Ref. männlich)					0,015	0,155
Alter					-0,027	0,017
Bildung Eltern (ISCED, z-standardisiert)					-0,005	0,074
Sprache deutsch (Ref. alle anderen)					-0,278	0,204
Broken Home (Ref. ja)					-0,369 *	0,172
Kulturkapital					0,000	0,000
Gesundheitszustand					-0,062	0,093
vorherige Ausbildung					-0,050	0,315
n	160		160		160	
Adjusted R <sup>2</sup>	0,038		0,338		0,356	

Änderung in F	2,247+	8,753***	1,446
---------------	--------	----------	-------

Anmerkung: +  $p < 0,1$ ; \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; OLS Regression; p-Werte berechnet auf Basis des T-Tests

In Tabelle 9 wurde eine Regression für die durchschnittliche Note, die im Studium erreicht wird, berechnet. Es wurden erneut drei Modelle geschätzt, um einen Vergleich mit den Ergebnissen der ersten Welle zu ermöglichen. Wir finden erneut einen signifikanten Effekt für Big 5 *Gewissenhaftigkeit*, HZB Gymnasium sowie HZB Note. In der ersten Welle waren ergänzend die Abbruchsintention und die Staatsangehörigkeit signifikant, diese Effekte finden wir in der zweiten Welle nicht. Stattdessen finden wir signifikante Einflüsse für die soziale Integration und das Aufwachsen in einem Einelternhaushalt. Sind Studierende besser sozial integriert, verbessert sich ihre durchschnittliche Note. Wächst eine Person mit einem Elternteil im Haushalt auf, verschlechtert sich die durchschnittliche Note im Studium um 0,369 im Vergleich zu einer Person mit beiden Elternteilen. Dabei handelt es sich um einen Unterschied von mehr als einer Drittelnote, der allein durch das Aufwachsen in einem Einelternhaushalt erklärt werden kann.

Der Effekt für soziale Integration verschwindet im dritten Modell, hier ist denkbar, dass dieser Effekt über die Soziodemographie moderiert wird. Wenn das Elternhaus bereits für ein Studium vorbereitet, ist eine soziale Integration einfacher und beeinflusst dadurch weniger den Studienerfolg. Kommt die betreffende Person aber aus einem akademikerfernen Elternhaus, spielt die soziale Integration eine größere Rolle, da mehr Unterstützung für den Studienbeginn benötigt wird.

Zu dem separaten Effekt für den Einelternhaushalt gibt es bisher wenig Literatur (Brake, 2014; Peuckert, 2019; Schutter & Schweda-Möller, 2017). Eine Überlegung ist, dass die gleichzeitige Belastung durch ein bildungsfernes Elternhaus, schlechtere ökonomische Lage sowie ein kleineres soziales Netzwerk neues Risikopotenzial birgt. Diese Kinder sind seltener Teil von Vereinen oder gehen anderen teuren Hobbies nach, weisen ein geringes Selbstwertgefühl auf und neigen eher zu gesundheitlichen Beeinträchtigungen (Brake, 2014). Diese Effekte können auch den Studienerfolg beeinträchtigen.

Ergänzend wurde auch hier eine *backward Regression* berechnet. Für die durchschnittlichen ECTS pro Semester finden wir signifikante Ergebnisse für Big 5 Offenheit, allgemeine Selbstwirksamkeit, Abbruchsintention, Bildung der Eltern und das Aufwachsen in einem Einelternhaushalt. Für die durchschnittliche Note ergeben sich signifikante Einflüsse für die allgemeine Selbstwirksamkeit, HZB Gymnasium, HZB Note, Besuch O-Tutorium, Alter, Aufwachsen in einem Einelternhaushalt und dem kulturellen Kapital.

Für unseren Datensatz finden sich unter den untersuchten Einflussvariablen nur vereinzelt signifikante Effekte. Für die durchschnittlichen Leistungspunkte, die pro Semester erreicht werden, können wir konstatieren, dass die Note der Hochschulzugangsberechtigung sowie die Abbruchsintention die stärkste Erklärungskraft haben. Daneben finden wir in beiden Wellen einen Effekt für das Erhalten der Hochschulzugangsberechtigung an einem Gymnasium.

Für die Bestimmung des Studienerfolgs kann festgestellt werden, dass die gesammelten

ECTS deutlich weniger durch diskriminierende Variablen bestimmt werden. Zwar finden sich auch hier Effekte der sozialen Herkunft, diese lassen sich jedoch nicht von schulischen Leistungen trennen. Eine bessere Variable ist nicht einfach zu finden, also sollte bei einer Implementierung eines Vorhersagemodells hierauf zurückgegriffen werden.

Wir finden für keine der geschützten Merkmale einen direkten und signifikanten Einfluss auf den Studienerfolg. Für die Staatsangehörigkeit finden wir lediglich in einem Modell einen signifikanten Effekt, die Fallzahl der Personen ohne deutsche Staatsangehörigkeit ist jedoch sehr gering. Somit ist eine Verwendung dieser im Rahmen einer KI zur Erkennung des Abbruchrisikos weder zielführend noch notwendig. Indirekt können sie eine Rolle spielen, wenn Faktoren wie der Ort der Hochschulzugangsberechtigung sowie die Note der Hochschulzugangsberechtigung betrachtet werden. Es ist bekannt, dass nicht nur Leistung diese beiden Punkte beeinflusst, sondern soziale Ungleichheit hier ebenso eine Rolle spielt. Kinder, deren Eltern kein Gymnasium besucht haben, werden deutlich seltener ein Gymnasium besuchen als jene, deren Eltern selbst Erfahrung damit haben (Hovestadt & Eggers, 2007). Ebenso werden Noten von einer Reihe weiterer Faktoren beeinflusst, die nicht immer mit dem meritokratischen System legitimiert werden können. Bei der Verwendung dieser Indikatoren muss dies immer mitberücksichtigt werden. Selbst für die Wahl des Nebenjobs im Laufe des Studiums finden sich Unterschiede nach sozialer Herkunft, nämlich ob es sich um einen studienorientierten Job handelt oder welche Art von Qualifizierung benötigt wird (Staneva, 2017).

Zum Ausgleich unterschiedlicher und sozial bedingter Startbedingungen wäre auch ein verstärkter Fokus auf die Eingangsphase im Studium wichtig, um für alle Studierenden den gleichen Startpunkt zu ermöglichen. Vorkurse und Orientierungstutorien setzen hier an der richtigen Stelle an und vermindern den Effekt sozialer Ungleichheit. So haben sie einen direkten Einfluss auf den Studienerfolg. Allein das Besuchen eines Orientierungstutoriums erhöht also die Wahrscheinlichkeit eines erfolgreichen Studiums (Bülow-Schramm, 2018; Schubarth et al., 2018). Die bestehenden Angebote in dieser Richtung sollten so ausgeweitet werden, dass allen Studierenden eine Teilnahme ermöglicht wird.

Ergänzend können weitere Angebote in dieser Richtung hinzugefügt werden. So sollte jeder Studiengang eine Veranstaltung zu Techniken wissenschaftlichen Arbeitens anbieten, um besonders die Sorge um die schriftlichen Ausarbeitungen zu nehmen.

In der Studiengestaltung sollten Studierende, die vom klassischen Studierendentypus abweichen, stärker bedacht werden. Dabei könnte beispielsweise über Veranstaltungen nachgedacht werden, die spezifisch für Studierende ausgelegt sind, deren Erwerb ihrer Hochschulzugangsberechtigung länger zurückliegt. Dies wird als besonderes Risiko für die Bewältigung der Anforderungen durch das Studium gesehen (Majer, 2018). Die steigende Anzahl von Teilzeitstudierenden in Vollzeitstudienfächern muss ebenso beachtet werden. Ein richtiger Umgang muss hier gefunden werden, ein ausreichendes Angebot an Teilzeitstudiengängen ist aktuell nicht gegeben. Dabei ist ein Teilzeitstudium an sich nicht nachteilig, sondern lediglich, wenn es in einem formellen Vollzeitstudiengang umgesetzt wird (Dahm et al., 2018).

Ein erfolgreicher Ansatz der Universität ist das Orientierungstutorium, welches im Rahmen der möglichen Minderung sozialer Ungleichheit an der Universität angeboten wird. Hiermit soll das Onboarding an der Universität erleichtert und das Wissen um organisatorische Angelegenheiten angeglichen werden. Insbesondere Studierende, die kein Gymnasium

besucht haben, erfahren weniger Vorbereitung für die Prozesse an der Universität (Heublein et al. 2010). Sowohl die akademische als auch die soziale Integration sind wichtige Determinanten für ein erfolgreiches Studium (Tinto, 1975). Das Orientierungstutorium scheint hier eine geeignete Maßnahme zu sein, um mögliche Wissenslücken zu schließen und die universitäre Lehr- und Lernkultur kennenzulernen.

Wie kann die Lebenswelt der Studierenden erfolgreich mit der hochschulischen Praxis verbunden werden? Die soziale Integration und die akademische Passung der Studierenden werden als zentrale Faktoren für die Zufriedenheit mit dem Studium betrachtet. Es gilt, die Studierenden bei ihrem Weg an die Hochschule zu unterstützen. Ebenso müssen Wissenslücken in Bezug auf den akademischen Bereich entdeckt und geschlossen werden. Die Studieneingangsphase ist hier am wichtigsten, um Angebote zu machen, die in diesen Bereichen unterstützen.

Die Erhebung derart vielfältiger Items stellt sich bereits für einzelne Studiengänge als aufwändig dar. Selbst mit einem fertigen Fragebogen ist die Bearbeitung durch die Studierenden langwierig und wird nicht von jeder Person erfolgen. Zusätzlich ist die Auswertung der Daten nicht zu vernachlässigen, wer diese Aufgabe übernehmen würde, ist fraglich. Hier eine KI zu verwenden, würde den Arbeitsaufwand deutlich vermindern und so eine realistische Option für Universitäten darstellen. So könnten die Daten analysiert und die Ergebnisse der Studienberatung zugespielt werden. Hierbei muss auf Transparenz und Verständlichkeit des Outputs geachtet werden (vgl. Arbeitspaket 3).

Mit der vorliegenden Arbeit kann festgestellt werden, dass bereits wenige einzelne Indikatoren für die Vorhersage der Studienleistung der meisten Studierenden gut funktionieren. Neben den Leistungsindikatoren, die durch das Prüfungsamt bezogen werden können, sind nur die Abbruchsintention, die Teilnahme an Orientierungstutorien sowie vorherige schulische Leistungen aussagekräftig. Der Besuch eines Gymnasiums ist besonders vorteilhaft für das Studium. Für die bisherigen schulischen Leistungen bestehen Korrelationen zu geschützten Merkmalen, weshalb hier soziale Ungleichheiten an der Universität sichtbar werden. Statt diese als Grundlage für Erfolgsvorhersagen zu benutzen, erscheint es als sinnvoll, diese Differenzen zu begleichen. In unserer Analyse finden wir dafür mehrere Möglichkeiten, wie beispielsweise den Besuch eines Orientierungstutoriums. Durch die Angleichung der Voraussetzungen für alle Studienanfänger können in Zukunft eventuell andere Indikatoren für den Studienerfolg entdeckt werden, die unter aktuellen Bedingungen verschleiert werden.

Im Antrag wurden für AP II folgende Forschungsfragen formuliert:

- 1) Welche Input-Daten sind verfügbar bzw. müssen erhoben werden, die eine gültige Vorhersage erlauben, ohne datenschutzrechtliche Bestimmungen und/oder die Privatsphäre der Betroffenen zu verletzen?
- 2) Welche Merkmale müssen geschützt werden, um Diskriminierung zu vermeiden?

Für die erste Forschungsfrage kann festgehalten werden, dass die Universität Daten zum Alter, Geschlecht, Nationalität, Hochschulzugangsberechtigung, bisherigen Leistungen vorliegen hat. Aus datenschutzrechtlicher Perspektive muss für die Verwendung dieser Daten in einem Vorhersagesystem das Einverständnis der Betroffenen eingeholt werden und diese über den Zweck des Systems informiert werden. Darüber hinaus handelt es sich bei einigen dieser Daten um geschützte Merkmale, die aus ethischen Gründen nicht zu verwenden sind. Gleichzeitig finden wir in unseren Ergebnissen, dass diese Merkmale weder für eine

Vorhersage benötigt werden noch die Qualität dieser verbessern. Relevanter sind der Besuch eines Gymnasiums, die Bildung der Eltern, die Note der HZB sowie das Leben in einem Einelternhaushalt. Sowohl der Ort der HZB als auch die Note liegen dem Prüfungsamt vor und können mit Einverständnis der Studierenden verwendet werden.

Die zweite Forschungsfrage betrifft den Kern unseres Arbeitsprojekts, die Vermeidung der Diskriminierung bei der Vorhersage. Hier soll zunächst erwähnt werden, dass die bisherige Note im Studium stärker durch Diskriminierungseffekte der sozialen Herkunft behaftet ist. In erster Linie sollten also die bisherigen ECTS verwendet werden. Hier lassen sich keine direkten Effekte von geschützten Merkmalen finden. Gleichzeitig lässt sich nicht ausschließen, dass bereits der Besuch eines Gymnasiums nicht gleich wahrscheinlich für jede Person ist, weshalb hier indirekt Ungleichheiten Einfluss nehmen können. Diese lassen sich zum Teil durch eine Intervention, wie beispielsweise ein Orientierungstutorium, ausgleichen.

Soziale Ungleichheiten finden wir besonders in den Effekten eines Gymnasiumsbesuches (spezifische Vorbereitung), Bildung der Eltern (spezifisches Kulturkapital) und dem Einelternhaushalt (Überschneidung Risikolagen). Der Zusammenhang mit den schulischen Leistungen mag an einer grundlegenden Passung im Bildungssystem liegen. Studierende, die bereits in der Schule mit Vorgaben zurechtgekommen sind, finden auch an der Universität einen erfolgreichen Umgang mit den Herausforderungen.

Die Beantwortung der Forschungsfragen dient als Vorarbeit für das AP I. So sollte zunächst auf Grundlage vorhandener Daten geprüft werden, inwiefern diese tatsächlich im Zusammenhang mit dem Studienerfolg stehen, bevor dann ein Modell mit Hilfe von KI erstellt wurde. Der Fokus lag besonders auf geschützten Merkmalen, die soziale Ungleichheit im Rahmen der Verwendung von KI bestärken könnten.

Zum aktuellen Zeitpunkt werden Daten der Welle 3 noch analysiert. Besonders durch die Ergänzung dieser Daten ergibt sich die Möglichkeit, unsere bisherigen Analysen zu vertiefen. Mit den vollständigen Daten können Strukturgleichungsmodelle berechnet werden, um den zeitlichen Ablauf der Effekte zu prüfen.

Die **Ergebnisse** des Projekts (hier AP II) wurden sowohl über wissenschaftliche Publikationen als auch über wissenschaftliche Vorträge zugänglich gemacht:

Ülpenich, Bettina; Junghänel, Luisa; Krause, Johannes; Rosar, Ulrich (2023): Studienerfolg und -misserfolg aus institutionen- und studierendenzentrierter Perspektive: Eine qualitative Untersuchung. In: Beiträge zur Hochschulforschung. Jg. 45. Heft 2, S.54-74.

Ülpenich, Bettina; Junghänel, Luisa (2022): Studienabbrüche prognostizieren. Zur Rechtfertigung von Leistungsvorhersagen im Studium, Vortrag auf dem 41. Kongress der Deutschen Gesellschaft für Soziologie, Sektion Bildung und Erziehung, Bielefeld.

Ülpenich, Bettina; Junghänel, Luisa; Krause, Johannes; Rosar, Ulrich (2022): Studienabbrüche prognostizieren. Zur Rechtfertigung von Leistungsvorhersagen im Studium, Vortrag auf der Herbsttagung der Düsseldorfer Soziologie auf Schloss Mickeln, Düsseldorf.

Junghänel, Luisa; Ülpenich, Bettina; Krause, Johannes; Rosar, Ulrich (eingereicht): Diskriminierungsfreie Vorhersage des Studienerfolgs - Eine empirische Untersuchung möglicher Leistungsindikatoren. In: Zeitschrift für Erziehungswissenschaft.

[1] [https://www.gesetze-im-internet.de/hstatg\\_1990/index.html](https://www.gesetze-im-internet.de/hstatg_1990/index.html), zuletzt abgerufen am 26.02.24.

[2] Hierbei wurde gewährleistet, dass nur die Projektleitung eine Zuordnung von Studierenden zu Prüfungsleistungen vornehmen konnte. Ergänzend wurde die Zustimmung der befragten Studierenden eingeholt, damit Befragungsdaten (s.u.) mit den Leistungsdaten kombiniert werden konnten.

[3] Die Begriffe ECTS und CP werden hier synonym verwendet.

[4] Dabei ist zu beachten, dass das Fach Sozialwissenschaften durch einen Numerus Clausus zulassungsbeschränkt ist, weshalb hier die Spannweite der Abiturnoten deutlich kleiner ist.

[5] Die schwache Erklärungskraft der BIG 5 Variablen in beiden Modellen (vgl. Tabellen 5 und 6) ist insofern ein relevanter Befund, da viele der ExpertInnen in den qualitativen Vorstudien Persönlichkeitsmerkmale von Studierenden in den Vordergrund rückten, wenn es um Problemlösungskompetenzen geht.

[6] Hierbei muss beachtet werden, dass im Datensatz lediglich 9 Personen keine deutsche Staatsbürgerschaft haben. Dieses Ergebnis ist also mit Vorsicht zu interpretieren.

## AP III

### Problemaufriss und Vorarbeiten

Ein sozialverträglicher APP-Einsatz erfordert den Einbezug aller Stakeholdergruppen. Dazu zählen neben Entwickler:innen und Verantwortlichen an Hochschulen ebenso und in besonderem Maße die Studierenden als direkte Betroffene des KI-Systems. Vor dem Hintergrund der Diskriminierungsgefahren und anderen ethischen Bedenken (Slade & Prinsloo, 2013), die von Algorithmen ausgehen und die beispielsweise bereits in Großbritannien zur leidvollen Erfahrung zahlreicher Schülerinnen und Schüler geworden sind (Smith, 2020), galt es die Wahrnehmungen und Anforderungen der Studierenden empirisch zu untersuchen und im Rahmen eines iterativen Prozesses für die Ausgestaltung des soziotechnischen Systems zu berücksichtigen.

Zu diesem Zweck konnten wir auf einer Reihe von Vorarbeiten aufbauen. So zeigte bereits eine nicht-repräsentative Untersuchung an der Universität Düsseldorf aus dem Jahr 2019, dass 78,8% der befragten Studierenden die Nutzung von KI-basierten Studienabbruchsprognosen ablehnen (Kieslich et al., 2019). Dabei variiert jedoch die generelle Akzeptanz von KI-Anwendungen an Hochschulen in Abhängigkeit des Einsatzbereiches in hohem Maße. Eine hohe Ablehnungsquote seitens der Studierenden kann die Hoffnung auf die Senkung von Studienabbruchsquoten schnell zunichte machen. So ist zum einen ein APP-System auf die Verarbeitung persönlicher Daten angewiesen, und es sind zum anderen negative Folgen für die Motivation der Studierenden denkbar. Eine weitere Vorstudie weist zudem darauf hin, dass der Einsatz von KI-Systemen Konsequenzen für das Protestverhalten gegenüber der Hochschule sowie für ihre Reputation mit sich bringen können (Marcinkowski et al., 2020). Die individuellen Fairnesswahrnehmungen spielen dabei neben der faktischen Fairness der KI-Systeme eine besondere Rolle für eine erfolgreiche Implementierung (Marcinkowski & Starke, 2019). An dieser Stelle wird daher deutlich, dass

sich die Gesamtbeurteilung eines sozialverträgliches APP-System aus ethischer Perspektive sowohl aus der Berücksichtigung manifester Fairnessmaßnahmen, als auch aus subjektiven Eindrücken der Betroffenen speist.

In einem systematischen Literaturreview, welches im Laufe der Projektzeit durch die beteiligten Forschenden im internationalen Fachjournal „Big Data & Society“ veröffentlicht wurde, wurden das theoretische und empirische Forschungswissen über individuelle Fairnesswahrnehmungen gegenüber KI-Systemen synthetisiert (Starke et al., 2022). Dabei zeigt sich zum einen, dass individuelle Wahrnehmungen mitunter von algorithmischen Fairnessdefinitionen abweichen (Saha et al., 2020; Saxena et al., 2020; Srivastava et al., 2019) und zum anderen, dass die Wahrnehmung davon, ob Menschen oder Algorithmen fairere Entscheidungen treffen, maßgeblich vom Anwendungsfall abhängt (Araujo et al., 2020; Lee, 2018). Insgesamt mangelt es jedoch an empirischen Erkenntnissen über Fairnesswahrnehmungen im Bildungskontext, fokussieren sich viele Studien doch beispielsweise auf KI-Systeme im Strafverfolgungskontext oder im Personalwesen. Darüber hinaus lassen sich einerseits in der vergleichenden Betrachtung keine einheitlichen Erkenntnisse des Einflusses soziodemographischer Merkmale auf die Fairnesswahrnehmungen ableiten und andererseits fehlt es an empirischer Forschung der Folgen wahrgenommener algorithmischer (Un-)Fairness (Starke et al., 2022). Vor diesem Hintergrund wird das Bestreben des Projekts, die Wahrnehmungen der Studierenden gegenüber einem APP-System sowie der Untersuchung der daraus resultierenden Folgen samt sich ableitenden Handlungsempfehlungen für Hochschulen und Entscheider:innen als besonders notwendig und wertvoll erachtet.

## Konzeptuelle Erarbeitung der Herausforderungen eines sozialverträglichen APP-Systems unter Berücksichtigung der Auswirkungen auf die Fairnesswahrnehmung von Studierenden

Um die relevanten Dimensionen der individuellen Wahrnehmungen und Einschätzungen der Betroffenen zu identifizieren, welche wiederum später systematisch in das Design des soziotechnischen Systems eingespeist werden sollen, wurden in einer konzeptuellen Arbeit zunächst die ethischen Herausforderungen eines sozialverträglichen APP-Systems herausgearbeitet (T3.2a). In Anbetracht der Diskriminierungsgefahr von Algorithmen lag ein besonderer Fokus auf den Fairnesswahrnehmungen verschiedener Stakeholder gegenüber einem Academic Performance Prediction (APP) System, mit besonderem Fokus auf Studierenden als direkte Betroffene der KI-Anwendung. Zu diesem Zweck wurde ein idealtypisches APP-System skizziert, das zum einen verschiedene Designfaktoren auf der Input-, Throughput- und Output-Ebene des soziotechnischen algorithmischen Systems benennt. Zum anderen wurde auf jeder Ebene aufgezeigt, inwiefern die (studentische) Fairnesswahrnehmung durch die Justierung der einzelnen Parameter beeinflusst werden kann. Konzeptuelle Basis dieser Überlegungen bildeten die vier Fairnessdimensionen der *Organizational Justice*-Literatur (Greenberg, 1990), die zwischen einer distributiven, prozeduralen, interaktionalen und informationellen Fairness unterscheiden. Konkret wurden auf Input-Ebene die Problemspezifizierung, die Entscheidung für die Nutzung eines White-Box- oder Black-Box-Algorithmus, die Wahl der Inputdaten sowie die Freiwilligkeit der Datenabgabe als Einflussfaktoren identifiziert. Auf der Ebene des Throughputs beeinflussen insbesondere die Sicherheit des APP-Systems, aber auch die Wahl einer konkreten Fairnessnorm die Wahrnehmung der Studierenden. Die Output-Ebene wird schließlich durch die Interaktion zwischen dem soziotechnischen System und dem Menschen, Fragen der Transparenz und der Erklärbarkeit des APP-Systems sowie der Kategorisierung des Ergebnisses bestimmt. Darüber hinaus beeinflussen ebenso die zugeschriebene Verantwortlichkeit im Falle von Fehlentscheidungen sowie die aus der Entscheidung des APP-Systems resultierende Verteilung von Interventionen die Fairnesswahrnehmungen der

Studierenden.

Die Darstellung eines idealtypischen APP-Systems zeigt die Fallstricke in der Entwicklung und dem Einsatz eines soziotechnischen Systems im Bildungskontext auf. Neben den Risiken, werden auch Lösungsvorschläge unterbreitet, die einen fairen, sozialverträglichen Einsatz des APP-Systems ermöglichen sollen, wenngleich zu berücksichtigen bleibt, dass nicht alle Parameter gleichermaßen und zur gleichen Zeit maximiert werden können. So ist beispielsweise mit Trade-offs zwischen der Genauigkeit der Vorhersage (Accuracy) und der Erklärbarkeit (Explainability) des Systems zu rechnen. Neben den Erkenntnissen für die spezielle Anwendung des APP-Systems kann der hier entwickelte Rahmen, welcher im Sammelband „Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions“ veröffentlicht wurde, auch als Vorlage für die Bewertung weiterer KI-Anwendungen im Hochschulkontext herangezogen werden (Keller et al., 2022).

Diese konzeptuelle Vorarbeit lieferte die Grundlage der im weiteren Projektverlauf durchgeführten Studien, in denen empirische Evidenz für die individuellen Fairnesswahrnehmungen der Studierenden gewonnen, sowie deren Auswirkungen auf Verhaltensintentionen getestet wurden (T3.3a). Zur Vorbereitung der empirischen Untersuchungen und der Messung der Fairnesswahrnehmungen wurde ein Itempool der vier Fairnessdimensionen (distributiv, prozedural, informationell, interaktional) auf Basis der konzeptuellen Arbeit sowie verschiedener unserer Vorarbeiten zusammengestellt (T3.2a; M3.1). Die Skalen der einzelnen Dimensionen wurden in den nachfolgenden Studien fortlaufend validiert (T3.2b; M3.2).

## Bewertung unterschiedlicher Verteilungsnormen: Wahrnehmung der Vergabe von Unterstützungsmaßnahmen auf Basis von APP

In Anbetracht der distributiven Fairnesswahrnehmungen stellen sich dabei zunächst Fragen der Verteilungsgerechtigkeit. Zwar liegt das formale Ziel eines APP-Einsatzes zunächst nicht in der Verteilung von Ressourcen, sondern in einer statistischen Prognose potenzieller Studienabbrecher:innen, jedoch bietet ein solches System für Hochschulen darüber hinaus auch die Möglichkeit, gerade leistungsschwächeren Studierenden individuelle Unterstützungsangebote auf Basis ihrer individuellen Leistungsprognose anbieten zu können, wodurch der Studienerfolg im besten Fall erhöht werden kann (Mishra et al., 2014). Gleichzeitig steht dieser Chance jedoch das Risiko gegenüber, dass sich an der Verteilung der Unterstützungsangebote distributive (Un-) Fairnesswahrnehmung entzünden können. Dies ist in jedem Fall unter Annahme der begrenzten Ressourcen moderner Hochschulen zu befürchten, wenn Leistungsprognosen dazu führen, dass nicht alle Studierenden die für sie gewünschte Unterstützung erhalten, sondern beispielsweise nur die, die ein KI-System für sie als angemessen erachtet (Keller et al., 2022).

Vor diesem Hintergrund wurden in einer empirischen Studie die Präferenzen der Studierenden für verschiedene Verteilungsnormen in den Blick genommen. Mit der Unterscheidung zwischen *Equality*, *Equity* und *Need* (Deutsch, 1975) wurden unterschiedliche Distributionsnormen in den Blick genommen. Während *Equality* eine Gleichverteilung von Unterstützungsmaßnahmen für alle Studierende beschreibt, spricht sich *Equity* für eine Verteilung nach Leistungsprinzipien aus. Demnach soll jeder Studierende Unterstützungsmaßnahmen entsprechend des selbst eingebrachten Inputs erhalten – sodass leistungsstärkere Studierende bevorzugten Zugang zu weiteren Maßnahmen erhalten. *Need* hingegen beschreibt eine Verteilung entlang der Bedürftigkeit, also einer vorrangigen Berücksichtigung von leistungsschwächeren Studierenden. Auf Basis einer Onlinebefragung mit deutschen Studierenden (n = 1.378), welche mit Unterstützung des Befragungsinstituts Innofact durchgeführt wurde, wurden direkte und indirekte Effekte der Präferenz von Verteilungsnormen getestet. In den Blick genommen wurde, inwiefern

Verteilungsnormpräferenzen sich zum einen auf die Befürwortung eines APP-Systems auswirken (direkter Effekt) sowie inwiefern dieser Effekt durch die Befürchtung von Ungleichheiten, ausgelöst durch den APP-Einsatz, moderiert wird (indirekter Effekt). Die Ergebnisse zeigen, dass je stärker die *Need*-Norm bevorzugt wird, desto eher wird der Einsatz eines APP-Systems abgelehnt. Hinsichtlich der Befürwortung der *Equality*- und *Equity*-Norm zeigt sich hingegen ein gegensätzliches Bild: Je eher diese befürwortet werden, desto eher wird auch der APP-Einsatz befürwortet. Eine mögliche Erklärung dieser Befunde ist, dass Studierende, die eine Verteilung nach dem Bedürftigkeitsprinzip befürworten, Wert auf ein gewisses Maß an Fingerspitzengefühl und menschlichem Einfühlungsvermögen legen, während eine Verteilung nach *Equality* oder *Equity* nach klaren, konsistenten Regeln erfolgen soll. Letzteres beschreibt dabei eher Eigenschaften, die KI-Systemen zugeschrieben werden und in der Einschätzung von Studierenden durch ein APP-System besser erreicht werden können, als durch menschliche Entscheidungen.

Die Studie erweitert die bislang vereinzelt vorliegenden Forschungserkenntnisse, die die Bewertung unterschiedlicher Fairnessnormen (häufig auch mathematischer Natur) gegenüber spezifischen KI-Anwendungen in den Blick nehmen. Welche individuell stabilen Fairnesspräferenzen jedoch die Basis für diese Bewertungen bilden, blieb bislang in der Literatur unberücksichtigt. Die hier gewonnen empirischen Erkenntnisse deuten darauf hin, dass in Kontexten, in denen eine Verteilung nach dem *Need*-Prinzip präferiert wird (sprich: es soll auf die Bedürftigkeit geachtet werden), eine KI-Anwendung möglicherweise mit erheblichen Vorbehalten hinsichtlich der wahrgenommenen Fairness des Systems rechnen muss. Die Erkenntnisse dieser Studie wurden im internationalen Workshop „(Un)fairness of Artificial Intelligence“, welcher von der Forschungsgruppe „Human(e) AI“ organisiert wurde, im Oktober 2022 in Amsterdam vorgestellt (Lünich et al., 2022).

Darüber hinaus wurde eine vertiefende explorative Analyse im Rahmen des faktoriellen Surveys durchgeführt, die (1) neben der zuvor genannten Frage der distributiven Fairnesswahrnehmung unterschiedlicher Verteilungsnormen bei der Bewertung eines APP-Systems (2) zusätzlich Unterschiede in der Fairnessbewertung zwischen KI-basierten Leistungsprognosesystemen und menschlichen Leistungsvorhersagen sowie (3) Unterschiede in der Fairnesswahrnehmung in Abhängigkeit des Studienfaches in den Blick nimmt. Es zeigt sich, dass deutsche Studierende eine Verteilung von Unterstützungsmaßnahmen auf Basis von APP entlang der *Equality*-Norm präferieren. Gleichzeitig bestehen keine Unterschiede in der distributiven Fairnesswahrnehmung in Abhängigkeit davon, ob die Leistungsprognose auf algorithmischer Basis oder menschlicher Entscheidung beruht. Auch die Wahl des Studienfachs – unterschieden wurde hier zwischen Studierenden aus MINT- und aus geistes- und gesellschaftswissenschaftlichen Fächern (im Weiteren SHAPE genannt) – nimmt keinen signifikanten Einfluss auf die Fairnessbewertung des APP-Systems. Ausgenommen der Erkenntnis, dass Studierende der MINT-Fächer eine Verteilung nach *Need* etwas fairer bewerten, als SHAPE-Studierende.

Auf Basis dieser Erkenntnisse kann daher die Frage gestellt werden, ob eine Einführung von KI-basierten APP-Systemen überhaupt sinnvoll ist. Denn folgt man der Präferenz von Studierenden für eine Gleichverteilung von Unterstützungsmaßnahmen, so bedarf es keiner vorherigen Leistungsprognose. Insofern wäre ein zweckmäßiger Einsatz eines APP-Systems mit dem Ziel einer gerechten Verteilung von Fördermaßnahmen in Frage zu stellen. Die Erkenntnisse der Untersuchung sowie die Diskussion dieser Ergebnisse wurden in einem Paper in der internationalen Fachzeitschrift „Technology, Knowledge, and Learning“ publiziert (Lünich et al., 2024).

## Individuelle Schadenswahrnehmung der Studierenden

Um vertiefende Hinweise über die Gründe mangelnder Fairnesswahrnehmungen zu gewinnen

und mögliche Bedenken der Studierenden zu identifizieren, wurden neben den individuellen Fairnesswahrnehmungen der Studierenden in der zuvor genannten Studie auch ihre persönlichen Einschätzungen gegenüber potenziellen Schäden, die durch den Einsatz eines APP-Systems ausgelöst werden können, untersucht. Die Studierenden wurden gebeten, die folgende Frage mit einer offenen Antwortmöglichkeit zu beantworten: „Wenn Sie einmal an mögliche Schäden der KI-basierten Leistungsvorhersage denken: Welche ganz konkreten Schäden oder negativen Konsequenzen fallen Ihnen im Zusammenhang mit der Leistungsvorhersage ein?“

Mithilfe der Software MAXQDA wurde nach Datenbereinigung eine induktive Kategorienbildung mit  $n = 1.165$  von ursprünglich  $n = 1.467$  Antworten durchgeführt. Daraus resultieren 77 inhaltliche Kategorien auf neun Abstraktionsebenen. Die von Studierenden wahrgenommenen Schäden lassen sich in eine technische und eine soziale Dimension unterteilen. Erstere befasst sich insbesondere mit der Sorge eines fehlerhaften KI-Systems und unterscheidet zwischen Input-, Throughput- und Output-Phase der Dropout Detection. Am häufigsten befürchteten Studierende dabei, dass die Inputdaten fehlerhaft oder zu selektiv in das System eingespeist würden, sodass nicht alle aus ihrer Sicht notwendigen Parameter berücksichtigt würden (beispielsweise, weil diese nicht messbar sind oder nur eine Momentaufnahme darstellen würden). Daraus resultiert gewissermaßen zugleich der Schaden auf Output-Ebene, dass das APP-System eine falsche Vorhersage trifft. Vorbehalte gegenüber der technischen Sicherheit des Systems, zum Beispiel durch Hackerangriffe oder allgemeine Datenschutzproblematiken, werden hingegen selten genannt. Auf sozialer Ebene werden Bedenken individueller sowie gesellschaftlicher Folgen für Studierende betont. Auf individueller Ebene wird ein hohes Ausmaß psychischer Schäden durch die KI-basierte Leistungsvorhersage befürchtet. So wird insbesondere die Gefahr gesehen, dass der Einsatz von APP-Systemen negative Konsequenzen für die Motivation der Studierenden haben sowie psychischen Druck und negative Emotionen auslösen könnte. Auf gesellschaftlicher Ebene besteht ein besonderes Problembewusstsein zudem für das Risiko der Entstehung von Missverhältnissen, beispielsweise in Form von Stigmatisierung oder Ungleichbehandlung sowie dem Verlust von Möglichkeiten, beispielsweise Masterzulassungen oder Stipendien zu erhalten. Des Weiteren werden normative und ethische Fragen von den Studierenden aufgeworfen, die zum Beispiel die grundsätzliche Quantifizierung von Leistung oder die Vereinheitlichung von Studierenden kritisieren.

Die explorative Analyse, die im Rahmen einer Bachelorarbeit durchgeführt wurde und anschließend im Sinne der wissenschaftlichen Nachwuchsförderung auf der 68. Jahrestagung der Deutschen Gesellschaft für Publizistik und Kommunikationswissenschaft (DGPK) vorgestellt wurde (Maxhuni et al., 2023), belegt eine Vielzahl von potenziellen Schäden, die Studierende bei der Einführung von APP-Systemen befürchten. Darüber hinaus legt die Analyse auch divergierende Schadenswahrnehmungen zwischen Studierenden und Forschenden offen. Neben den in zahlreichen ethischen Richtlinien diskutierten Herausforderungen der KI-Entwicklung – insbesondere der Transparenz, Fairness und Verantwortlichkeit (Hagendorff, 2020; Jobin et al., 2019) – kritisieren Studierende verstärkt die operative Logik der KI aus einer sozialen und ethisch-normativen Perspektive sowie widerwillig in Kauf zu nehmende Kosten für das, durch die KI zu erreichende Ziel, wie die Quantifizierung von Leistung. So bleiben zum Beispiel Fragen der individuellen Folgen – wie Demotivation oder entstehender Druck – des Einsatzes prädiktiver Analytik im wissenschaftlichen Diskurs häufig unbenannt. Fragen von Effizienz und Optimierung kreieren so eine Ethik kontinuierlicher Verbesserung, die wenig Platz für (ggf. divergierende) Ansichten von Betroffenen lässt. Bestehende hegemoniale Verhältnisse zeigen sich nicht nur in der Appropriation der Definitionsmacht, sondern auch im Umgang mit der kommunikativen Vermittlung von Informationen über algorithmische Vorhersagesysteme. Gerade im Lichte der heterogenen Wissensbestände der Studierenden kommen Fragen politisch-partizipatorischer und informationeller Fairness auf. Diskursive Konfliktbewältigung setzt somit voraus, dass Ängste und Bedenken der Betroffenen ernst genommen und zur Diskussion gebracht werden.

Die Ergebnisse der qualitativen Inhaltsanalyse sowie die kritische Diskussion des hegemonialen Ungleichgewichts zwischen Bedenken der Studierenden und Zielen der Entwickler:innen und Entscheider:innen von KI-Systemen werden derzeit in einer wissenschaftlichen Publikation zusammengefasst und für die Einreichung in einem Fachjournal vorbereitet. Außerdem wurden die Erkenntnisse der explorativen Analyse genutzt, um auf dieser Basis Items für eine standardisierte Erhebung der studentischen Schadenswahrnehmung in Ergänzung zur Fairnessmessung abzuleiten (T3.2). Aus diesem Schritt ergaben sich 38 Items einer Schadensmessung, die in einer Befragung, die im März 2023 mit Unterstützung des Befragungsinstituts Talk durchgeführt wurde, standardisiert erhoben wurden.  $n = 845$  Studierenden wurden gebeten zum einen die Wahrscheinlichkeit, dass der genannte Schaden eintritt, sowie zum anderen das Ausmaß des Schadens, wenn dieser eintritt, auf einer fünf-stufigen Skala (sehr unwahrscheinlich – sehr wahrscheinlich bzw. sehr geringer Schaden – sehr großer Schaden) zu bewerten. Eine erste Analyse deutet darauf hin, dass die Studierenden insgesamt ein eher hohes Schadensausmaß sowie eine hohe Eintrittswahrscheinlichkeit der Schäden sehen – der Mittelwert der Antworten befindet sich fast ausschließlich über dem Skalenmittelpunkt. Lediglich die Wahrscheinlichkeit, dass Studierende diskriminiert werden oder aufgrund persönlicher Merkmale benachteiligt werden, wird etwas geringer eingeschätzt ( $M < 3$ ). Die höchste Schadenswahrscheinlichkeit in Kombination mit einem hohen empfundenen Schadensausmaß sehen Studierende darin, dass vorübergehende Probleme sowie nicht alle Stärken und Schwächen berücksichtigt würden, die Leistungsprognose zu Selbstzweifeln führe und sie unter gesteigerten Leistungsdruck geraten. Darüber hinaus wurden Unterschiede in der Schadenswahrnehmung zwischen verschiedenen Studienfächern (MINT vs. SHAPE) sowie zwischen männlichen und weiblichen Studierenden analysiert. Dabei zeigt eine erste explorative Analyse, dass SHAPE-Studierende tendenziell höhere Befürchtungen in Hinblick auf Leistungsdruck und Selbstzweifel empfinden und ebenso einen höheren Schaden durch mangelndes menschliches Einfühlungsvermögen einer KI sehen. In Hinblick auf Unterschiede zwischen Geschlechtern zeigt sich deutlich, dass weibliche Studierende im Aggregat den erhobenen Schäden sowohl eine höhere Eintrittswahrscheinlichkeit als auch ein höheres Ausmaß zuschreiben. Die Ergebnisse der Studie wurden bislang nicht publiziert und lediglich zu projektinternen Informationszwecken verwendet. Eine Publikation ist jedoch beabsichtigt.

## Vergleichende Risiko- & Schadenswahrnehmung von Studierenden & Restbevölkerung

Nachdem die qualitative Analyse der studentischen Schadenswahrnehmung bereits Abweichungen beziehungsweise eine unterschiedliche Fokussierung im Vergleich zu wissenschaftlich und politisch diskutierten ethischen Richtlinien aufweist, erschien eine vergleichende Betrachtung der Wahrnehmungen sinnvoll. Vor diesem Hintergrund wurde eine Onlinebefragung mit dem Befragungsinstitut forsa durchgeführt, in der neben studentischen Wahrnehmungen ( $n = 751$ ) auch die Wahrnehmungen einer repräsentativen Bevölkerungsstichprobe ( $n = 1.008$ ) in den Blick genommen wurden. Vor dem Hintergrund des Einflusses der wahlberechtigten Bevölkerung auf Policy-Entscheidungen der KI-Entwicklung und Einführung im Kontrast zur direkten Betroffenheit der Studierenden von KI-Anwendungen im Hochschulkontext, wurde die Risikowahrnehmungen beider Gruppen auf unterschiedlichen Ebenen verglichen.

Auf Ebene der generellen Befürwortung des KI-Einsatzes in verschiedenen Anwendungsfeldern, zeigt sich, dass Studierende im Vergleich zum Rest der Bevölkerung den Einsatz von KI in vielen Bereichen signifikant eher befürworten (Banken, Gesundheitswesen, industrielle Produktion, Verkehr, öffentlichen Verwaltung und im persönlichen Leben). Die Befürwortung des KI-Einsatzes in sicherheitskritischen Bereichen (Polizei, Militär, Geheim- & Sicherheitsdienste) sowie für politische Entscheidungen wird jedoch von beiden Gruppen eher abgelehnt. In Schulen und Universitäten bewerten sowohl die Bevölkerung als auch die

Studierenden einen KI-Einsatz als ambivalent – es herrscht also eine gewisse Uneinigkeit über die Befürwortung des KI-Einsatzes im Bildungswesen. Auf Ebene von APP als konkrete Anwendung in diesem Bereich, wurden die Befragten zudem gebeten das generelle Risiko, das von APP ausgeht, zu bewerten (minimales Risiko, begrenztes Risiko, hohes Risiko, inakzeptables Risiko). Während die EU APP-Systeme im Rahmen des AI Acts als Hochrisiko-System einstuft (European Commission, 2021), zeigt sich, dass Studierende APP-Systemen im Durchschnitt eher ein begrenztes bis hohes Risiko zuschreiben, während die Gesamtbevölkerung das Risiko insgesamt etwas geringer einschätzt – wenngleich keine signifikanten Unterschiede zutage treten. Auf der nächstgelagerten Ebene werden nach allgemeinen Risiken, gesellschaftliche und individuelle Schäden sowie die Bewertung potenzieller Diskriminierung durch APP untersucht. An dieser Stelle zeigt sich, dass Studierende sowohl für die Gesellschaft als auch für sie persönlich signifikant größere Schäden befürchten als der Rest der Bevölkerung, wobei in beiden Gruppen die gesellschaftlichen Schäden höher bewertet werden als individuelle Schäden, die von APP ausgehen. Ebenso erwarten Studierende eine höhere Diskriminierungsgefahr durch APP. Die dennoch in den Ergebnissen vorzufindende ambivalente Haltung der Studierenden deutet auf ein gewisses Maß an Unsicherheit bezüglich der aus der APP-Implementierung resultierenden Konsequenzen hin.

## Erklärungsfaktoren der studentischen Schadenswahrnehmung

Schließlich wurden in einem weiteren Schritt mögliche Erklärungsfaktoren der studentischen Schadenswahrnehmung im Rahmen eines Strukturregressionsmodells untersucht. Nach theoretischer Begründung und auf Grundlage empirischer Evidenz wurden Bedenken in Bezug auf die Privatsphäre (*Privacy Concerns*), Vertrauen in KI, persönliche Diskriminierungserfahrung sowie Institutionenvertrauen neben den soziodemographischen Merkmalen Geschlecht und Alter sowie der Unterscheidung zwischen MINT- und SHAPE-Studienfächern in das Modell mit aufgenommen. Die Analyse belegt, dass erlebte Diskriminierung die Wahrnehmung individueller Schäden positiv beeinflusst. Je eher Studierende demnach bereits persönlich von Diskriminierung betroffen waren, desto eher befürchten sie auch persönlichen Schaden durch ein APP-System zu erleiden. Ebenso zeigt sich, dass institutionelles Vertrauen einen negativen Einfluss sowohl auf die Wahrnehmung gesellschaftlicher, als auch individueller Schäden nimmt. Je geringer also das Vertrauen in die eigene Hochschule, desto eher werden Schäden durch das KI-System befürchtet. Die übrigen getesteten Faktoren nahmen keinen signifikanten Einfluss auf die Schadenswahrnehmung der Studierenden.

## Verhaltensintentionen der Studierenden als Reaktion auf die APP-Einführung

Neben der Berücksichtigung der Bedenken und Fairnesswahrnehmungen der Studierenden im Entwicklungsprozess der KI-Anwendung ist insbesondere für Hochschulen die Abschätzung möglicher Folgeaktionen von großer Bedeutung (T3.3a). Um diese Verhaltensintentionen bewerten zu können, wurden die Studierenden gebeten sich vorzustellen, dass die Hochschule, an welcher sie eingeschrieben sind, ein APP-System einführt. Anschließend bewerteten die Studierenden die Reputation einer APP-einführenden Hochschule sowie ihre persönliche Protestbereitschaft oder gar die Bereitschaft die Hochschule zu verlassen. Es zeigt sich, dass die Befragten die Reputation einer Hochschule, die ein APP einsetzt, gering einschätzen, wenngleich die Protestbereitschaft und die Meidungshaltung ambivalent ausfallen. Setzt man diese Erkenntnisse in Beziehung zu einer unserer früheren Studien zeigt sich, dass Verhaltensintentionen der Studierenden besonders durch (distributive und prozedurale) Fairnesswahrnehmungen beeinflusst werden (Marcinkowski et al., 2020).

## Anforderungen der Studierenden an APP

Nachdem die Fairness- und Schadenswahrnehmungen der Studierenden umfassend untersucht wurden, stellt sich die Frage nach konkreten Anforderungen, die Studierende an die APP-Entwicklung und Einführung stellen, um die Gestaltung eines sozialverträglichen Systems zu gewährleisten. Abgefragt wurden 16 verschiedene Anforderungen, die aus einem im Rahmen des Projekts durchgeführten Stakeholder-Workshop an der Universität Düsseldorf abgeleitet wurden und die die Umsetzung der KI-Einführung hinsichtlich Fragen der Selbstbestimmtheit der Datenfreigabe und Transparenz auf der einen Seite sowie des Umgangs mit auf Basis von APP resultierenden Fördermaßnahmen auf der anderen Seite aufgreifen. Hinsichtlich der Erwartungshaltung der Studierenden an die Implementierung von KI-Systemen an Hochschulen zeigt sich, dass Studierende sowohl hohen Wert auf ihre Selbstbestimmtheit legen, indem sie selbst entscheiden möchten, ob und welche ihrer persönlichen Daten zur KI-basierten Leistungsprognose verwendet werden. Die generelle Bereitschaft, ihre persönlichen Daten preiszugeben, fällt sehr gering aus. Die größte Zustimmung findet der Wunsch, dass Studierende ihre individuelle Leistungsprognose erklärt bekommen. Außerdem fordern sie, dass ihre persönliche Vorhersage jeder Zeit für sie – und nur für sie persönlich – einsehbar ist. In Hinblick auf die Inanspruchnahme von Fördermaßnahmen, die auf Basis von APP angeboten werden, fordern Studierende ebenfalls ein hohes Maß an Selbstbestimmtheit. So lehnen sie es ab, dass APP grundsätzliche Konsequenzen für ihren Studienverlauf hat und sprechen sich auch deutlich gegen verpflichtende Fördermaßnahmen bei einer schlechten Leistungsprognose aus. Auch eine anschließende Beratung durch Dozierende des eigenen Studiengangs wird eher abgelehnt. Eine Beratung durch unabhängige Studienberater:innen der Universität wird hingegen als weniger kritisch erachtet.

In der nachfolgenden vergleichenden Betrachtung der studentischen Anforderungen zwischen Studierenden mit hohen gegenüber Studierenden mit niedrigen Schadenswahrnehmungen, stechen insbesondere vier signifikante Unterschiede heraus. Studierende mit niedriger Schadenswahrnehmung sind signifikant eher dazu bereit ihre Daten freiwillig zur Verfügung zu stellen und sind außerdem verstärkt der Meinung, dass individuelle Fördermaßnahmen auf Basis der KI-basierten Leistungsvorhersage angeboten werden sollten. Daran anschließend zeigen sie eine höhere Bereitschaft, freiwillige Beratungsangebote beim Wissen über ihre persönliche Prognose auch in Anspruch zu nehmen. Studierende mit hoher Schadenswahrnehmung befürchten hingegen signifikant häufiger, dass sie schlechter von ihren Dozierenden bewertet werden könnten, wenn diese Einsicht in ihre Prognose nehmen könnten.

Die Ergebnisse der vergleichenden Untersuchung der Bevölkerungsmeinung und der Studierendenmeinung, der Auswirkungen der APP-Einführung auf Verhaltensintentionen der Studierenden sowie der Einflussfaktoren auf die Schadenswahrnehmung und die Anforderungen der Studierenden an die Einführung eines APP-Systems wurden in Teilen auf der 2. Learning AID-Konferenz mit wissenschaftlichen Kolleg:innen, Hochschuldidaktiker:innen, Lehrenden und Qualitätsverantwortlichen diskutiert (Lünich et al., 2023). Zudem wurden die Erkenntnisse in einer wissenschaftlichen Publikation aufbereitet. Das daraus resultierende Paper befindet sich derzeit im peer-review Prozess eines internationalen Fachjournals (Lünich et al., im Review).

## Erklärung der KI-basierten Leistungsprognose mithilfe von Entscheidungsbäumen

Basierend auf der empirischen Evidenz, dass Studierende hohen Wert auf die Erklärung ihrer persönlichen Leistungsprognose legen, wurde in einem [präregistrierten](#) Experimentaldesign

die Wirkung von unterschiedlichen Ausprägungen von Entscheidungsbäumen, die zur Leistungsprognose eingesetzt werden, auf die individuelle Fairnesswahrnehmung der Studierenden untersucht (T3.3a). Entscheidungsbäume bilden dabei den häufigsten Ansatz zur Erstellung von erklärbarer KI (XAI) im Kontext von Leistungsprognosen (Issah et al., 2023) und bilden die Basis der in AP I entwickelten Modelle. Im Rahmen der Experimentalstudie wurden daher zwei verschiedene Entscheidungsbäume, die jeweils in ihrer Tiefe – das heißt in ihrer Komplexität beziehungsweise Einfachheit – variieren, als Stimuli verwendet. Während die Abbildung der Entscheidungsbäume den realen Modellen des Projekts entsprachen, wengleich die Darstellung an einigen Stellen vereinfacht wurde, wurde als zweiter variierter Faktor die Vorhersagegenauigkeit (Accuracy) manipuliert. Neben einer hohen und einer niedrigen Einfachheit variierten die Entscheidungsbäume somit zusätzlich zwischen einer Vorhersagegenauigkeit von 65 % und 95 %. Mit Unterstützung des Befragungsinstitut Talk wurde eine faktorielle Befragungsstudie mit  $n = 1.047$  deutschen Studenten durchgeführt, um zu untersuchen, wie die Eigenschaften der Entscheidungsbäume (Einfachheit und Genauigkeit) die wahrgenommene distributive und informationelle Fairness beeinflussen, mediiert durch *Causability* (das heißt die selbst eingeschätzte Verständlichkeit der Ursache-Wirkungs-Zusammenhänge des Vorhersagemodells). Zusätzlich wurde die moderierende Rolle des institutionellen Vertrauens in diesen Beziehungen in den Blick genommen.

Die Ergebnisse deuten darauf hin, dass die Einfachheit des Entscheidungsbaums sowohl die informationelle als auch die distributive Fairnesswahrnehmung positiv beeinflusst, was durch die *Causability* mediiert wird. Im Gegensatz dazu beeinflusst die Vorhersagegenauigkeit diese Wahrnehmungen weder direkt noch indirekt. Wengleich die zuvor erwarteten Effekte nur gering oder nicht vorhanden sind, zeigen die Ergebnisse dennoch, dass der mittlere positive Effekt der *Causability* auf die Bewertung der distributiven Fairnesswahrnehmung vom institutionellen Vertrauen abhängt.

Letztlich deuten die komplexen Ergebnisse darauf hin, dass nicht nur die Erklärung der Leistungsprognose, sondern ebenso das Verständnis dieser Erklärung von großer Bedeutung für die wahrgenommene Fairness von APP sind. Es zeigt sich auch, dass ein hohes Vertrauen der Studierenden in ihre Hochschule, dieses persönliche Verständnis in Hinblick auf die informationelle Fairnesswahrnehmung nicht ersetzen kann, wengleich es sich als wichtig zur Schaffung eines transparenten und unterstützenden Umfelds erweist. Der mangelnde Einfluss der Vorhersagegenauigkeit auf die Fairnesswahrnehmung weist hingegen darauf hin, dass sich Studierende des potenziellen Diskriminierungsrisikos, welches von ungenau arbeitenden KI-Systemen ausgeht, nicht ausreichend bewusst sind. Dies erweist sich insbesondere vor dem Hintergrund der im Rahmen der Untersuchung wahrgenommener Schäden nachgewiesenen Sorge vor fehlerhaften Entscheidungen als erklärungs-würdig. Darüber hinaus wird deutlich, dass die an vielen Stellen geforderte Transparenz von KI-Systemen zwar notwendig erscheint, ein zu hohes Maß an Informationen die Betroffenen jedoch überfordern kann, indem es zu Lasten der Verständlichkeit geht. Dabei hängt die Angemessenheit der Erklärung auch vom Kontext ab. So haben beispielsweise Studierende einen anderen Wissens- und Erfahrungshintergrund als andere Teile der Bevölkerung, was wiederum zu unterschiedlichen Reaktionen auf XAI führen kann.

Die Ergebnisse der Studie sowie die diskutierten Implikationen für Fairness und Transparenz bei der Implementierung von APP-Systemen wurden in einem Paper aufbereitet und bei der ACM Conference on Fairness, Accountability, and Transparency (FAccT) vom 03. – 06. Juni 2024 in Rio de Janeiro, Brasilien, einem internationalen Fachpublikum vorgestellt. Das Paper wurde im Anschluss an die Konferenz veröffentlicht (Lünich & Keller, 2024). Des Weiteren wurde eine weitere Studie mit Unterstützung des Befragungsinstituts Talk durchgeführt, in der die Erkenntnisse der ersten Experimentalstudie um kontraintuitive Erklärungen erweitert werden. So erweisen sich in denen im Projekt entwickelten Modellen beispielsweise Faktoren als signifikante Entscheidungspunkte der Leistungsprognose, die aus Studierendensicht intuitiv nicht plausibel erscheinen. Zum Beispiel kann eine höhere Anzahl an

Prüfungsversuchen in einem Modul dennoch zu einer geringeren Abbruchwahrscheinlichkeit führen, als weniger Versuche. Die Untersuchung, die Einblick in die Wirkung entsprechender Erklärungen gibt, befindet sich derzeit in der Auswertung und soll zeitnah publiziert werden.

## Weiteres (Vorgehen)

Die hier genannten und weitere bisher unausgewertete Datenbestände geben Gelegenheit für vertiefte Analysen zu vielen der hier aufgeworfenen Forschungsfragen, sei es in Form von Fachaufsätzen oder Qualifikationsarbeiten. So befasst sich beispielsweise das laufende Dissertationsprojekt der Projektmitarbeiterin Birte Keller mit der Analyse von politischen Rechtfertigung des KI-Einsatzes im Bildungswesen. Insofern ist mit fortgesetzter Erkenntnisproduktion des Projektes durch den ausführenden Lehrstuhl (Kommunikations- und Medienwissenschaft I der HHU) auch über sein offizielles Ende hinaus zu rechnen.

## Implikationen der konzeptuellen Arbeit und empirischen Erkenntnisse

Unter Berücksichtigung der aus dem dreijährigen Projekt resultierenden konzeptuellen Arbeiten sowie empirischen Erkenntnisse lassen sich eine Reihe von Handlungsempfehlungen für eine faire und sozialverträgliche Entwicklung und Implementierung von APP-Systemen an deutschen Hochschulen ableiten (T3.4; M3.3). Dabei adressieren die Forschungsarbeiten und Erkenntnisse einige der von Tsai und Gasevic (2017) aufgezeigten Herausforderungen der Implementierung von Learning Analytics and Hochschulen, die die Autoren auf Basis eines Reviews verschiedener Policy-Paper für die Einführung von KI an Hochschulen erarbeitet haben. So stellen sie heraus, dass Learning Analytics an Hochschulen nur selten strategisch geplant ist und die Implementierung nur unzureichend durch ein evaluierendes Monitoring begleitet und kontrolliert wird. Daher wird ein stärkerer Einsatz und die darauffolgende Evaluation von pädagogischen Ansätzen gefordert, die geeignet sein müssen, identifizierte Lernbarrieren tatsächlich zu überwinden. In diesem Sinne kommt dem Einsatz von Fördermaßnahmen auf Basis von APP sowie der Untersuchung ihrer Akzeptanz eine besondere Bedeutung zu. Schließlich stellen Tsai und Gasevic (2017) auch fest, dass es bis dato nur sehr wenige empirische Studien gibt, die den tatsächlichen (und implizit immer *positiv* gedachten) Einfluss von auf Learning Analytics fußenden Interventionen validieren.

Vor diesem Hintergrund können die im Zuge des Projekts durchgeführten Studien eine Reihe von bedeutsamen Erkenntnissen liefern, die die sozialverträgliche APP-Entwicklung und Einführung im Speziellen, gleichzeitig aber auch die KI-Implementierung im Bildungssektor im Allgemeinen informieren. Zusammenfassend zeigt sich dafür zunächst:

- Studierende stehen APP eher kritisch gegenüber.
- Studierende sehen eine Vielzahl potenzieller Schäden, die von APP ausgehen können. Dazu zählen insbesondere die Gefahr der Demotivation, ein mangelndes menschliches Einfühlungsvermögen bzw. die fehlende Berücksichtigung besonderer Umstände sowie die Entstehung von Missverhältnissen und fehlerhafte Vorhersagen.
- Verteilungsnormpräferenzen haben einen Einfluss auf die Bewertung von APP, wobei Studierende eine Gleichverteilung von Förderangeboten befürworten.
- Die Leistungsvorhersage soll kein Ausschlusskriterium sein, darf jedoch als Grundlage für eine gezielte Ansprache bedürftiger Studierender genutzt werden.
- Studierende legen hohen Wert auf ihre persönliche Autonomie und die Erklärung ihrer persönlichen Leistungsprognose.
- Ergebnisse einer Experimentalstudie verdeutlichen, dass nicht nur die Erklärung an sich, sondern auch das eigene Verständnis der Erklärungen wichtig für eine Fairnesseinschätzung von APP-Systemen ist.

Auf Basis der Erkenntnisse, dass Studierende erstens APP-Systemen kritisch gegenüberstehen und zweitens eine darauf basierende Verteilung von Fördermaßnahmen nach der *Equality*-Norm befürworten, nach welcher keine Unterschiede zwischen Studierenden gemacht werden sollen und auch keine Konsequenzen einer individuellen Leistungsprognose gewünscht werden, bleibt die Frage der Notwendigkeit und Sinnhaftigkeit einer APP-Einführung aus Sicht der Studierenden im Raum. Aus dieser Perspektive dient der APP-Einsatz ausschließlich Informationszwecken, wobei das Ziel der Erhöhung des Studienerfolgs vor dem Hintergrund der durch Studierende neben anderen Faktoren befürchteten Sorge der Demotivation fraglich erscheint. In Anbetracht der Abhängigkeit des Systems von der freiwilligen Preisgabe studentischer Daten - um unter anderem der Reputation der Hochschule nicht zu schaden und Proteste zu vermeiden - für eine erfolgreiche Implementierung müssen die Bedenken der Studierenden, ihre Fairness- und Risikowahrnehmungen daher ernst genommen und in den Mittelpunkt der Entwicklungs- und Implementierungsprozesse gerückt werden. Eine sozialverträgliche APP-Entwicklung erfordert daher:

- 1) Bedenken im Mittelpunkt: Die Implementierung von KI erfordert eine verantwortungsvolle Berücksichtigung der (studentischen) Bedenken, wobei variierende ethische Ansichten von verschiedenen Studierenden ernst genommen und beachtet werden müssen.
- 2) Hochschulentscheider und politische Entscheidungsträger sind dafür verantwortlich, ein ausreichendes Maß an Transparenz zu gewährleisten, sodass der Einsatz von KI in der Bildung für die Betroffenen zum einen durchschaubar wird, sie zum anderen jedoch nicht überfordert werden. So erfordert dies auch aus Forschungsperspektive eine Herangehensweise, die den Teilnehmenden das Anwendungsfeld so wenig abstrakt wie möglich vermittelt, damit das Verständnis für die tatsächliche Anwendung höchstmöglich ist, ohne dabei mit zu vielen Informationen für Verwirrung zu Sorgen. In Anbetracht der wachsenden Bedeutung erklärbarer KI, können in diesem Zusammenhang White-Box-KI-Verfahren, wie beispielsweise Entscheidungsbäume, die Nachvollziehbarkeit und das Verständnis der Prozesse verbessern.
- 3) Unabhängige (Kontroll- &) Beratungsinstanzen: Unabhängige Beratungsangebote, gerade im Falle schlechter Leistungsprognosen, können die Bereitschaft zur Inanspruchnahme von Unterstützung erhöhen und somit zu einer verbesserten Studienleistung beitragen.
- 4) Fortlaufende Evaluation: Der Einsatz von APP-Systemen muss nicht nur geplant, implementiert und durch verständliche Erklärungen des Verfahrens flankiert werden, sondern sollte auch fortlaufend unter Einbezug der universitären Stakeholdergruppen evaluiert und an die gewonnenen Erkenntnisse angepasst werden.

Abschließend lässt sich feststellen, dass die Entwicklung einer sozialverträglichen akademischen Leistungsprognose eine komplexe Herausforderung darstellt, die eine umfassende Berücksichtigung ethischer, technischer und sozialer Aspekte erfordert. Die hier vorgeschlagenen Maßnahmen – von der zentralen Berücksichtigung studentischer Bedenken über die Gewährleistung von Transparenz und Verständlichkeit bis hin zur Einrichtung unabhängiger Beratungsinstanzen und der Notwendigkeit einer kontinuierlichen Evaluation – bilden einen vielversprechenden Ansatzpunkt, um die Akzeptanz und Wirksamkeit von APP-Systemen in der Hochschulbildung zu verbessern. Die Implementierung dieser Empfehlungen erfordert ein kooperatives Engagement aller Beteiligten, von Hochschulverantwortlichen über politische Entscheidungsträger:innen bis hin zu den Studierenden selbst. Nur durch einen solchen integrativen Ansatz kann sichergestellt werden, dass die Einführung von KI in der Bildung nicht nur technologische Fortschritte bringt, sondern auch die Bildungschancen aller Studierenden fördert und zu einer gerechteren und inklusiveren Hochschulbildung beiträgt. Die vorliegende Analyse und die daraus resultierenden Empfehlungen bieten somit einen wichtigen Beitrag zur Diskussion um eine ethisch verantwortungsvolle Integration von Technologie in den akademischen Sektor und legen den Grundstein für weitere Forschung in

diesem entscheidenden Bereich.

Die **Ergebnisse** des Projekts (hier AP III) wurden sowohl über wissenschaftliche Publikationen als auch über wissenschaftliche Vorträge zugänglich gemacht:

Keller, B., Lünich, M. & Marcinkowski, F. (2022). How Is Socially Responsible Academic Performance Prediction Possible? Insights From a Concept of Perceived AI Fairness. In F. Almaraz-Menéndez, A. Maz-Machado, C. López-Esteban & C. Almaraz-López (Hrsg.), *Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions* (S. 126–155). IGI Global. <https://doi.org/10.4018/978-1-7998-9247-2.ch006>

Lünich, M., Keller, B. & Marcinkowski, F. (2022, 27.–28. Oktober). *The effects of students' distributive justice norm preferences on the evaluation of Artificial Intelligence (AI) in higher education*. Vortrag auf dem Workshop "(Un)fairness of Artificial Intelligence" der Research Priority Area Human(e) AI, Amsterdam, Niederlande.

Lünich, M., Keller, B. & Marcinkowski, F. (2024). Fairness of Academic Performance Prediction for the Distribution of Support Measures for Students: Differences in Perceived Fairness of Distributive Justice Norms. *Technology, Knowledge and Learning*, 29(2), 1079–1107. [doi.org/10.1007/s10758-023-09698-y](https://doi.org/10.1007/s10758-023-09698-y)

Lünich, M., Keller, B. & Marcinkowski, F. (2023, 28. August). *Die studentische Wahrnehmung von Learning Analytics und ihre Konsequenzen für Einstellungen, Präferenzen und Verhaltensintentionen am Beispiel von Academic Performance Prediction - Ergebnisse einer Repräsentativbefragung und Implikationen für die Einführung von KI an der Hochschule*. Vortrag auf Learning AID 2023 "Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung". KI:edu:nrw. Learning AID, Bochum, Deutschland.

Lünich, M. & Keller, B. (2024). Explainable Artificial Intelligence for Academic Performance Prediction. An Experimental Study on the Impact of Accuracy and Simplicity of Decision Trees on Causability and Fairness Perceptions. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT '24)*, 3–6 Juni, 2024, Rio de Janeiro, Brasilien. <https://doi.org/10.1145/3630106.3658953>

Lünich, M., Keller, B. & Marcinkowski, F. (im Review). *Diverging Perceptions of Artificial Intelligence in Higher Education: A Comparison of Student and Public Assessments on Risks and Damages of Academic Performance Prediction*.

Lünich, M., Maxhuni, A., Keller, B. & Marcinkowski, F. (zur Einreichung vorbereitet). Focusing on Student Harm Perceptions in Academic Performance Prediction: A Participatory Technology Assessment Framework to Prevent Reductionist and Hegemonic AI Implementation in Higher Education [Arbeitstitel].

Maxhuni, A., Lünich, M., Keller, B. & Marcinkowski, F. (2023, 18.–20. Mai). *Hegemoniale Technologieimplementierung an der Hochschule - Eine qualitative Analyse der Schadenswahrnehmung betroffener Studierender bei der Einführung von Dropout Detection*. Vortrag auf der 68. DGPK-Jahrestagung in Bremen. Deutsche Gesellschaft für Publizistik und Kommunikationswissenschaft, Bremen, Deutschland.

Starke, C., Baleis, J., Keller, B. & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2), 1-16. <https://doi.org/10.1177/20539517221115189>

## Weitere Veröffentlichungen und Präsentationen

Neben den fachlichen Publikationen zu den Arbeiten der drei APs hat das Projektteam auch an verschiedenen Vernetzungstreffen teilgenommen und sich mit anderen geförderten Projekten aus der Förderlinie Digitale Hochschulbildung sowie mit andersweitig geförderten Projekten ausgetauscht. Hier wurden auch mehrere, in der Regel kürzere Vorträge zum Stand dieses Projektes gehalten. Hervorzuheben sind zwei besondere Präsentationen, in denen – nach offiziellem Abschluss des Projekts – das Gesamtprojekt und seine Ergebnisse dargestellt wurden. Hierbei handelte es sich um folgende Vorträge:

- Stefan Conrad: Responsible Academic Performance Prediction (RAPP). Konferenz „Digitalisierung weiterdenken – künstliche Intelligenz in Lehrorganisation und Hochschulverwaltung“, HRK, Projekt Modus (BMBF), 10./11.04.2024, Karlsruhe
- Stefan Conrad, Johannes Krause, Marco Lünich: Responsible Academic Performance Prediction (RAPP). HeiCAD-Lecture, Heinrich-Heine-Universität Düsseldorf, 15.05.2024, Düsseldorf

Letztgenannter Vortrag diente der Präsentation der Projektergebnisse gegenüber dem Rektorat und den das Projekt begleitenden Studiendekanaten und Verwaltungsstellen der Heinrich-Heine-Universität und der Diskussion hinsichtlich der Möglichkeiten zur Einführung eines RAPP-Systems an der Hochschule.

## Literaturverzeichnis

Alyahyan, E. and Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1).

Araujo, T., Helberger, N., Kruikemeier, S. & Vreese, C. H. de (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35(3), 611–623. <https://doi.org/10.1007/s00146-019-00931-w>

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). Allgemeine Selbstwirksamkeit Kurzsкала (ASKU). In: *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*, 10.

Blüthmann, I., Thiel, F., & Wolfgramm, C. (2011). *Abbruchtendenzen in den Bachelorstudiengängen. Individuelle Schwierigkeiten oder mangelhafte Studienbedingungen?*, In: *Die Hochschule: Journal für Wissenschaft und Bildung* 20(1), 110-126.

Bornkessel, P., & für Hochschul, D. Z. (2018). Erfolg im Studium: Konzeptionen, Befunde und Desiderate, wbv Media GmbH & Co. KG, Bielefeld.

Brake, A. (2014). Der Wandel familiären Zusammenlebens und seine Bedeutung für die (schulischen) Bildungsbiographien der Kinder. In: Rohlf, C., Harring, M., Palentien, C. (Hrsg.), *Kompetenz-Bildung: Soziale, emotionale und kommunikative Kompetenzen von Kindern und Jugendlichen*, 113-151. Springer VS, Wiesbaden.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC, Monterey, CA.

Bülöw-Schramm, M. (2018). Faktoren für Studienerfolg – eine kritische Analyse von Bachelorstudiengängen. In: Hericks, N. (Hrsg.), *Hochschulen im Spannungsfeld der Bologna-Reform*, 311-334. Springer VS, Wiesbaden.

Calders, T., & Žliobaitė, I. (2013). Why unbiased computational processes can lead to discriminative decision procedures. In: Custers, B., Calderys, T., Schermer, B., Zarsky, T. (Hrsg.), *Discrimination and Privacy in the Information Society: Data mining and profiling in large databases*, 43-57. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-30487-3>

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.

Dahm, G., & Lauterbach, O. (2016). Measuring students' social and academic integration—Assessment of the operationalization in the National Educational Panel Study. In: Blossfeld, HP., von Maurice, J., Bayer, M., Skopek, J. (Hrsg.), *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, 313-329. Wiesbaden: Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-11994-2\\_18](https://doi.org/10.1007/978-3-658-11994-2_18)

Dahm, G., Becker, K., & Bornkessel, P. (2018). Determinanten des Studienerfolgs nichttraditioneller studierendenqualifizierter Studierender ohne Abitur. In: Bornkessel, P. (Hrsg.), *Erfolg im Studium*, 108-174. Bielefeld: wbv Media GmbH & Co. KG. <https://doi.org/10.3278/6004654w>

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, 46(5), 904-920.

Deutsch, M. (1975). Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? *Journal of Social Issues*, 31(3), 137–149. <https://doi.org/10.1111/j.1540-4560.1975.tb01000.x>

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012, January). Fairness through awareness. In: *Proceedings of the 3rd innovations in theoretical computer science conference*, 214-226.

Ekowo, M., & Palmer, I. (2017). Predictive analytics in higher education. *Five Guiding Practices for Ethical Use*. New America.

Erdmann, M., Pietrzyk, I., Helbig, M., Jacob, M., & Stuth, S. (2022). Können intensive Beratungsprogramme soziale Ungleichheit beim Übergang in die Hochschule reduzieren? Ergebnisse eines Feldexperiments. *Schweizerische Zeitschrift für Soziologie*, 48(1), 137-162.

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. <https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585->

[01aa75ed71a1.0001.02/DOC\\_1&format=PDF](https://doi.org/10.1177/014920639001600208)

Greenberg, J. (1990). Organizational Justice: Yesterday, Today, and Tomorrow. *Journal of Management*, 16(2), 399–432. <https://doi.org/10.1177/014920639001600208>

Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), e12760.

Grunschel, C., & Dresel, M. (2021). Studienerfolg und Studienabbruch: Gruppenspezifische Untersuchungen ihrer Bedingungen–Einführung in das Themenheft. *ZeHf–Zeitschrift für empirische Hochschulforschung*, 4(1), 5-6.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

Heinemann, A. M., & Mecheril, P. (2023). Erziehungswissenschaftliche Diskriminierungsforschung. In: Scherr, A., Reinhardt, A.C., El-Mafaalani, A. (Hrsg.), *Handbuch Diskriminierung*, 115-130. Wiesbaden: Springer Fachmedien Wiesbaden.

Heinze, D. (2018). *Die Bedeutung der Volition für den Studienerfolg*. Springer Fachmedien Wiesbaden.

Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J., & Woisch, A. (2017). Zwischen Studiererwartungen und Studienwirklichkeit. *Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen*, 1. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hannover.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen*. Hannover: HIS, Forum Hochschule F2/2010.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hovestadt, G., & Eggers, N. (2007). Soziale Ungleichheit in der allgemein bildenden Schule. *Ein Überblick über den Stand der empirischen Forschung unter Berücksichtigung berufsbildender Wege zur Hochschulreife und der Übergänge zur Hochschule*. Im Auftrag der Hans Böckler Stiftung, Rheine.

Hutcheson, G. D. (2011). Ordinary least-squares regression. *L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research*, pages 224–228.

Issah, I., Appiah, O., Appiahene, P. & Inusah, F. (2023). A Systematic Review of the Literature on Machine Learning Application of Determining the Attributes Influencing Academic Performance. *Decision Analytics Journal*, 7, 100204. <https://doi.org/10.1016/j.dajour.2023.100204>

Jobin, A., Lenca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

Keller, B., Lünich, M., & Marcinkowski, F. (2022). How Is Socially Responsible Academic Performance Prediction Possible?: Insights From a Concept of Perceived AI Fairness. In *Strategy, Policy, Practice, and Governance for AI in Higher Education Institutions*, 126-155.

IGI Global.

Kieslich, K., Lünich, M., Marcinkowski, F. & Starke, C. (2019). *Hochschule der Zukunft: Einstellungen von Studierenden gegenüber Künstlicher Intelligenz an der Hochschule*. Précis. Düsseldorfer Institut für Internet und Demokratie. [https://diid.hhu.de/wp-content/uploads/2019/10/DIID-Precis\\_Kieslich-et-al\\_Fin.pdf](https://diid.hhu.de/wp-content/uploads/2019/10/DIID-Precis_Kieslich-et-al_Fin.pdf)

Kim, S.-J., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. (2007). An interior-point method for large-scale  $\ell_1$ -regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617.

Knispel, J., Wittneben, L., Slavchova, V., Arling, V. (2021). Skala zur Messung der beruflichen Selbstwirksamkeitserwartung (BSW-5-Rev). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2014). Internale-Externale-Kontrollüberzeugung-4 (IE-4). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668. <https://doi.org/10.1177/2053951718756684>

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation*, 4(3):415–447.

Majer, S. (2018). Studienerfolg von traditionellen und beruflich qualifizierten Studierenden. *Erfolg im Studium*, 175.

Marcinkowski, F., Kieslich, K., Starke, C. & Lünich, M. (2020). Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In: M. Hildebrandt, C. Castillo, E. Celis, S. Ruggieri, L. Taylor & G. Zanfir-Fortuna (Hrsg.), *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT\* '20)* (S. 122–130). ACM. <https://doi.org/10.1145/3351095.3372867>

Marcinkowski, F. & Starke, C. (2019). Wann ist Künstliche Intelligenz (un-)fair? Ein sozialwissenschaftliches Konzept von KI-Fairness. In: J. Hofmann, N. Kersting, C. Ritzi & W. J. Schünemann (Hrsg.), *Politik in der digitalen Gesellschaft: Zentrale Problemfelder Und Forschungsperspektiven* (Bd. 1, S. 269–288). transcript Verlag. <https://doi.org/10.14361/9783839448649-014>

Mishra, T., Kumar, D. & Gupta, S. (2014). Mining Students' Data for Prediction Performance. In *Fourth International Conference on Advanced Computing & Communication Technologies (ACCT 2014)* (S. 255–262). IEEE. <https://doi.org/10.1109/ACCT.2014.105>

Meuser, M. & Nagel, U. (1991). ExpertInneninterviews – vielfach erprobt, wenig bedacht.: Ein Beitrag zur qualitativen Methodendiskussion. In: D. Garz & K. Kraimer (Hrsg.), *Qualitativ-empirische Sozialforschung*, 441–471. VS Verlag für Sozialwissenschaften.

Mühlhoff, R. (2020). Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus. *Deutsche Zeitschrift für Philosophie*, 68(6), 867-890.

Müller-Benedict, V., & Tsarouha, E. (2011). Können Examensnoten verglichen werden? Eine Analyse von Einflüssen des sozialen Kontextes auf Hochschulprüfungen/are grades in exams comparable to each other? The impact of social context on grading in higher education.

*Zeitschrift für Soziologie*, 40(5), 388-409.

Muñoz, C., Smith, M., & Patil, D. J. (2016). Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. Executive Office of the President of USA.

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in big data*, 2, 13.

Pannier, S., Rendtel, U., & Gerks, H. (2020). Die Prognose von Studienerfolg und Studienabbruch auf Basis von Umfrage- und administrativen Prüfungsdaten: Das Projekt „Studienverläufe“ am FB Wirtschaftswissenschaft der FU Berlin. *ASTA Wirtschafts- und Sozialstatistisches Archiv*, 14, 225-266.

Petri, P. S. (2020). Skala zur Erfassung der Studieneinstiegsselbstwirksamkeit (SESW-Skala). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. <https://doi.org/10.6102/zis274>

Peuckert, R. (2019). *Familienformen im sozialen Wandel*. Springer VS Wiesbaden.

Przyborski, A. & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung: Ein Arbeitsbuch*. München: Oldenbourg Wissenschaftsverlag.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). Big Five Inventory (BFI-10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*.

Rish, I. et al. (2001). An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46.

Rzepka, N., Simbeck, K., Pinkwart, N. (2023). Learning Analytics und Diskriminierung. In: Mandy Schiefner-Rohs/ Sandra Hofhues/ Andreas Breiter (Eds.), *Datafizierung (in) der Bildung*, 211-228. Bielefeld: transcript Verlag.

Saha, D., Schumann, C., McElfresh, D. C., Dickerson, J. P., Mazurek, M. L. & Tschantz, M. C. (2020). Human Comprehension of Fairness in Machine Learning. In A. Markham (Hrsg.), *ACM Digital Library, Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (S. 152). Association for Computing Machinery. <https://doi.org/10.1145/3375627.3375819>

Sarcletti, A., & Müller, S. (2011). Zum Stand der Studienabbruchforschung. Theoretische Perspektiven, zentrale Ergebnisse und methodische Anforderungen an künftige Studien. *Zeitschrift für Bildungsforschung*, 1(3), 235-248.

Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C. & Liu, Y. (2020). How do fairness definitions fare? Testing public attitudes towards three algorithmic definitions of fairness in loan allocations. *Artificial Intelligence*, 283, 1–15. <https://doi.org/10.1016/j.artint.2020.103238>

Schiefner-Rohs, M., Hofhues, S., & Breiter, A. (2023). Datafizierung (in) der Bildung: Kritische Perspektiven auf digitale Vermessung in pädagogischen Kontexten.

Schmohl, T., Watanabe, A., & Schelling, K. (2023). *Künstliche Intelligenz in der Hochschulbildung: Chancen und Grenzen des KI-gestützten Lernens und Lehrens* (p. 286). transcript Verlag.

Schneider, K., Berens, J., & Burghoff, J. (2019). Drohende Studienabbrüche durch Frühwarnsysteme erkennen: Welche Informationen sind relevant? *Zeitschrift für Erziehungswissenschaft*, 22(5), 1121-1146.

Schneider, K., Berens, J., & Görtz, S. (2021). Maschinelle Früherkennung abbruchgefährdeter Studierender und Wirksamkeit niedrigschwelliger Interventionen. *Studienerfolg und Studienabbruch*, 369-392.

Schubarth, W., Wagner, L., Mauermeister, S., Berndt, S., Erdmann, M., Schmidt, U., ... & Pohlenz, P. (2018). Verbundprojekt StuFo: der Studieneingang als formative Phase für den Studienerfolg: Analysen zur Wirksamkeit von Interventionen: erste Befunde und Empfehlungen. Universitätsbibliothek Johannes Gutenberg-Universität Mainz.

Schutter, S., & Schweda-Möller, A. (2017). Wo Risiken zusammentreffen: Bildungsbenachteiligung in Einelternfamilien. *Bildung und Ungleichheit in Deutschland*, 139-154.

Sieben, S., & Lechner, C. M. (2019). Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences*, 1(1), 1-6.

Slade, S. & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>

Smith, H. (2020). Algorithmic bias: should students pay the price? *AI & SOCIETY*, 35(4), 1077–1078. <https://doi.org/10.1007/s00146-020-01054-3>

Spilke, J., & Mielenz, N. (2006, February). Vergleich von Kriterien und Verfahren zur Modellwahl bei der multiplen linearen Regression. In *Statistik und Datenanalyse mit SAS. Proceedings der 10. Konferenz der SAS Anwender in Forschung und Entwicklung*, Shaker-Verlag, Aachen, 215-223.

Srivastava, M., Heidari, H. & Krause, A. (2019). Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In: A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi & G. Karypis (Hrsg.), *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)* (S. 2459–2468). ACM. <https://doi.org/10.1145/3292500.3330664>

Staneva, M. (2017). Studentische Erwerbstätigkeit: Elternhaus spielt bei Wahl des Studentenjobs eine Rolle. *DIW Wochenbericht*, 84(41), 914-924.

Starke, C., & Lünich, M. (2020). Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, 2, 1-17.

Suresh, H., & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In: *Equity and access in algorithms, mechanisms, and optimization*. 1-9.

Theune, K. (2021). Determinanten und Modelle zur Prognose von Studienabbrüchen. *Studienerfolg und Studienabbruch*, 19-41.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1), 89-125.

Tsai, Y.-S. & Gasevic, D. (2017). Learning analytics in higher education --- challenges and policies. In A. F. Wise (Hrsg.), *ACM Digital Library, Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (S. 233–242). ACM. <https://doi.org/10.1145/3027385.3027400>

Weber, A., Daniel, A., Becker, K., & Bornkessel, P. (2018). Proximale Prädiktoren objektiver wie subjektiver Studienerfolgsindikatoren. *Erfolg im Studium. Konzeptionen, Befunde und Desiderate*, 59-107.

Wild, K.-P. & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 185-200.

Yeom, S. and Tschantz, M. C. (2021). Avoiding disparity amplification under different worldviews. In *Conference on Fairness, Accountability, and Transparency, FAccT '21*, pages 273–283, New York, NY, USA. ACM.

Zheng, A. (2015). *Evaluating machine learning models: a beginner's guide to key concepts and pitfalls*. O'Reilly Media.

Zweig, K. A., & Krafft, T. D. (2018). Fairness und Qualität algorithmischer Entscheidungen. 57518, 204-227. Mohabbat-Kar, Resa, Basanta E. Thapa und Peter Parycek, (Hrsg.) (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft. 1. Auflage. Berlin: Kompetenzzentrum Öffentliche IT.