

Human-Centered Data Science – Etablierung einer kritisch-reflexiven Praxis bei der Entwicklung von datengetriebener Software


Claudia Müller-Birn ¹ und Lars Sipos²

Abstract: Die Erfahrungen der letzten Jahre im Bereich Data Science zeigen immer deutlicher, dass es ein Umdenken bei der Ausbildung bedarf, da die sozialen Nuancen in Daten nicht erfasst oder ethische Kriterien bei der Entwicklung datengetriebener Software zu selten berücksichtigt werden. Diese Lücke soll mit dem Ansatz der Human-Centered Data Science geschlossen werden. Im Rahmen einer Lehrveranstaltung mit Informatikstudierenden haben wir daher diesen Ansatz umgesetzt. Die Ergebnisse einer im Anschluss durchgeführten Interviewstudie zeigen, dass Studierende ein gesteigertes Bewusstsein für die gesellschaftlichen Auswirkungen von Data Science haben. Nichtsdestotrotz hatten wir jedoch den Eindruck, dass sie nicht über eine „Toolbox“ verfügen, die es ihnen erlaubt, diese Einsicht nachhaltig in ihren Datenpraktiken zu verankern. Infolgedessen schlagen wir einen konzeptionellen Rahmen für die Förderung einer kritisch-reflexiven Praxis in der Data Science-Ausbildung vor, den wir in diesem Artikel näher vorstellen. Anhand einer beispielhaften Umsetzung erläutern wir jede der vier aufeinander aufbauenden Phasen, die in zukünftiger Forschung evaluiert werden.

Keywords: Human-Centered Data Science; Ethik

1 Die Motivation für Human-Centered Data Science

Das Forschungsgebiet Data Science hat sich in den letzten Jahren rasant entwickelt und dies stellt die akademische Ausbildung vor besondere (wenn auch nicht neue) Herausforderungen (siehe [Va17]). Die Erfahrungen der letzten Jahre zeigen immer deutlicher, dass der Fokus auf statistische und numerische Aspekte in der Ausbildung nicht ausreicht, um soziale Nuancen, affektive Beziehungen, ethische, werteorientierte Grundsätze oder die tatsächlichen Auswirkungen einer datengetriebenen Software auf die Gesellschaft zu erfassen [BA17, Ko20, Ar22]. Aus diesen Herausforderungen hat sich das Gebiet *Human-Centered Data Science* entwickelt, in welchem die komplexen Interaktionen zwischen Gesellschaft, Technologie und von Menschen erzeugten Daten erforscht werden [Ar22]. Im Mittelpunkt steht dabei, die herkömmlichen computergestützten Methoden zur Analyse großer Datensätze mit qualitativen Methoden zu verbinden, die mit ihrem Detailreichtum und Kontextwissen zu einem tieferen Verständnis von Daten und Gesellschaft beitragen

¹ Freie Universität Berlin, Institut für Informatik, Human-Centered Computing, clmb@fu-berlin.de, 
<https://orcid.org/0000-0002-5143-1770>

² Freie Universität Berlin, Institut für Informatik, Human-Centered Computing, lars.sipos@fu-berlin.de

können. Ein zentrales Anliegen in der Human-Centered Data Science ist es, den Studierenden eine kritisch-reflexive Datenpraxis zu vermitteln. Dieser Artikel soll einen Diskussionsbeitrag liefern, wie eine solche kritisch-reflexive Datenpraxis in der akademischen Ausbildung im Bereich Data Science verankert werden kann. Es wird dazu ein laufendes Lehrangebot skizziert und erste Erkenntnisse aus der Umsetzung vorgestellt.

Human-Centered Data Science ist ein interdisziplinäres Forschungsgebiet, das sich auf Erkenntnisse und Methoden aus den Bereichen der Mensch-Computer-Interaktion, den Sozialwissenschaften, der Statistik und des maschinellen Lernens stützt [Ar22]. Der Human-Centered Data Science liegt dabei ein menschenzentrierter Ansatz bei der Technologiegestaltung - das Human-Centered Design - zugrunde, um die Praktiken der Data Science zu verbessern. Dieser Ansatz der menschenzentrierten Gestaltung basiert auf einer Reihe von Leitsätzen, die Kling & Star bereits vor über 20 Jahren aufgestellt haben [KS98]. Danach sollte datengetriebene Software menschliche Fähigkeiten sinnvoll ergänzen, aber diese nicht ersetzen oder automatisieren. Soziale Konstrukte (z.B. Fairness) sollten nicht in mathematische Konzepte übersetzt werden, da mathematische Operationalisierungen die Vielfältigkeit der sozialen Realität nicht umfassend zu beschreiben vermögen. Somit sollte bei der Entwicklung datengetriebener Software nicht nur die Optimierung der statistischen Modelle im Mittelpunkt stehen, sondern auch der Kontext berücksichtigt werden, in dem die Software letztlich eingebettet ist. Dies erfordert es, dass auch Fragen der ökologischen Nachhaltigkeit berücksichtigt werden. Kling & Star verweisen bereits darauf, dass Entwickler:innen Bescheidenheit („modesty“) in Bezug auf die Fähigkeiten von datengetriebener Software entwickeln sollten, denn Technologie allein kann keine sozialen Probleme, wie beispielsweise die der Gerechtigkeit, lösen. Eine wesentliche Grundlage der menschenzentrierten Gestaltung bildet daher ein Verständnis über die in die datengetriebene Software inkludierten Werte, die der zukünftigen direkt oder indirekt durch den Softwareeinsatz betroffenen Personengruppen, aber auch die der Datenwissenschaftler:innen selbst. Durch eine solche menschenzentrierte Perspektive auf Data Science können eine Vielzahl von Methoden aus dem Bereich der Mensch-Computer Interaktion in der Data Science verankert werden, wie beispielsweise die der partizipativen Gestaltung. Eine grundlegende Voraussetzung in der Vermittlung dieser Methoden ist es aber, die zumeist positivistisch geprägte Haltung von Studierenden um eine kritisch-reflexive Datenpraxis zu bereichern.

2 Erste Umsetzung und Interviewstudie

Ausgehend von diesen Überlegungen haben wir im Wintersemester 2020/21 erstmalig die Veranstaltung Human-Centered Data Science für Studierende der Informatik und der Data Science angeboten, um sie mit dieser Perspektive vertraut zu machen. Aktuell werden bereits Veranstaltungen angeboten, die die Human-Centered Design-Perspektive einbeziehen (z. B. in den Bereichen Mensch-Computer-Interaktion, Datenvisualisierung) sowie Veranstaltungen im Bereich des Maschinellen Lernens. Wir haben jedoch festgestellt, dass

ein Kurs, der die gesellschaftlichen Auswirkungen von bestehenden Datenpraktiken hervorhebt, in der derzeitigen Ausbildung bisher noch fehlt.

Die Veranstaltung hatte eine typische Struktur und bestand aus einer Vorlesung (2 SWS), in der die Konzepte, Techniken und Ansätze theoretisch eingeführt wurden, und einer Übung (2 SWS), in der ausgewählte Themen praktisch angewandt wurden. Die Themen wurden primär aus den bestehenden Diskursen der ACM Conference on Human Factors in Computing Systems (CHI) und der ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) Community abgeleitet.³ Eine disziplinär heterogene Gruppe von Studierenden (vier Frauen und zehn Männer) nahm an der Veranstaltung teil: neben Informatik und Data Science gab es Studierende der Wirtschaftsinformatik, Ergonomie und Politikwissenschaft.

2.1 Studienaufbau

Nach der ersten Durchführung der Veranstaltung waren wir an den Eindrücken der Studierenden interessiert und führten daher eine Interviewstudie durch. Wichtig war dabei, Studierenden die Möglichkeit zu bieten, ihre persönlichen Erfahrungen direkt mit uns zu teilen und die Effektivität unseres Lehransatzes, bei der Vermittlung einer Human-Centered Data Science zu bewerten. Wir luden alle Studierenden ein, an einem Interview teilzunehmen, wovon sechs Studierende der Einladung folgten. Die Interviews folgten einem semi-strukturiertem Format, um genügend Raum für Diskussionen über unerwartete Aspekte zu schaffen. In den Interviews fragten wir beispielsweise Studierende, ob sie das Gefühl hatten, die Lernziele erreicht zu haben, wie sie die allgemeine Struktur der Veranstaltung und das Tempo erlebten und stellten Fragen zu den verwendeten Tools und Methoden. Nach der Transkription der Interviews führten wir eine Thematische Analyse durch. Dabei folgten wir dem Ansatz von Braun & Clarke [BC06] aufgrund ihrer Flexibilität, Zugänglichkeit und ihres Potenzials, unerwartete Erkenntnisse zu gewinnen. Wir analysierten die Daten vor allem dahingehend, inwieweit es uns gelungen ist, das kritische Verständnis der Studierenden für die unterrichteten Themen zu fördern. Die Transkripte wurden in Bezug auf die Lernziele kodiert und alle Codes und Zitate wurden schließlich zur Generierung der Themen zusammengetragen.⁴

2.2 Studienergebnisse

Innerhalb des Analyseprozesses konnten wir eine Reihe von Themen identifizieren, wovon wir nachfolgend drei zentrale vorstellen: „Erhöhtes Bewusstsein für die gesellschaftlichen Auswirkungen von Data Science“, „Zwiespältiges Interesse an den Themen des

³ Eine Übersicht zu den Themen gibt es auf der Veranstaltungsseite: https://www.mi.fu-berlin.de/en/inf/groups/hcc/teaching/Past-Terms/winter_term_2020_21/course_human_centered_data_science.html

⁴ Einen Überblick über die Fragen und die Transkripte teilen wir auf OSF: <https://doi.org/10.17605/OSF.IO/JD6TQ>

Kurses und seinem Bezug zur Informatik“ und „Beispiele aus der Praxis und Reflexion unterstützten das Verständnis der Studierenden“.

Obwohl die Studierende keine homogene Gruppe darstellen und eine Vielzahl unterschiedlicher Hintergründe und Vorerfahrungen besitzen, berichteten alle interviewten Studierenden von einem gesteigerten Bewusstsein dafür, wie sich aktuelle Entwicklungen in der Data Science auf die Gesellschaft auswirken. In einem Auszug, der dies betont, wird erwähnt, dass diese Veranstaltung den Studierenden die „Werkzeuge“ gab, um über diese Entwicklungen zu sprechen: *„[...] ich muss sagen, dass ich auch vorher schon in diese Richtung aufgepasst hab und auch eine ‚Awareness‘, also ein Bewusstsein dafür hatte, aber vielleicht jetzt noch nicht die Tools hatte. Es war eher ein diffuseres ‚Da muss man aber aufpassen!‘ und ich kannte auch schon ein paar Beispiele, wo eben Dinge passiert sind, die nicht so super sind [...] aber ich glaube, was mir auf jeden Fall geholfen hat ist, dass man bestimmte Sachen einfach besser definieren, oder bezeichnen kann, also Worte für bestimmte Dinge hat, wo man sagen kann so und so“* (S4). Studierende hatten also bereits ein latentes Gefühl für bestimmte Themen, sei es durch frühere Erfahrungen, Nachrichtenartikel oder ähnliches. Sie verfügten jedoch nicht über die richtigen Konzepte und Werkzeuge, um das Ausmaß der mit diesen Technologien verbundenen Probleme vollständig zu erfassen und auch nicht über die Sprache, um diese Themen mit ihren Kommiliton:innen oder potenziellen Interessengruppen zu diskutieren. Für das kritische Denken ist es jedoch unabdingbar, das richtige Vokabular zu besitzen, um solche Gedanken ausformulieren zu können. Die Teilnehmer:innen konnten am Ende teilweise sich selbst kritische Fragen zu ihren Arbeitsweisen stellen: *„Okay, wo können Fehler entstehen?“* (S4), *„Ist das [System] überhaupt gut, in jedem Zusammenhang einzusetzen?“* (S3). In Bezug auf die Existenz von Bias hat eine Person gut in Worte gefasst, dass sie nun weiß, *„[...] wie man [die Daten] analysiert, überwacht, vorbereitet und sich ihnen kritisch entgegenstellen kann“* (S4). Darüber hinaus sind sich die Studierenden nun bewusster, wie wichtig es ist, Blackbox-Modelle und Bias in den Daten interpretieren zu können: *„Ich sehe schon einen Wert darin, dass man auf jeden Fall eine Blackbox aufmachen und ein wenig reinschauen kann“* (S2). Weiterhin können sie sich auf einen menschenzentrierten Ansatz besinnen und die reale Welt sowie die Menschen, welche datengetriebene Systeme nutzen oder von ihnen betroffen sind, berücksichtigen: *„[...] praktisch hat das [System] in der realen Welt Leute beeinflusst und Auswirkungen auf sie gehabt und zwar ständig“* (S3).

In Bezug auf die eigene Arbeit wurde festgestellt, dass die entscheidenden Anstrengungen vor der Implementierung und Programmierung der Systeme unternommen werden: *„[...] wenn ich jetzt daran [einem Data Science Projekt] sitzen würde, dass ich nicht nur ans Programmieren denke und wie kriege ich irgendwas umgesetzt, sondern dass das [die Analyse] nicht nebensächlich ist [...] ja [man überlegt], was man priorisieren sollte und das ist definitiv nicht nur das Programmieren“* (S4). Weiterhin wurde die Erkenntnis erlangt, dass Fairness für Informatiker:innen wichtig ist sowie dass der Einsatz von nicht fairen Modellen des maschinellen Lernens sogar nutzlos ist: *„Ich wusste nicht, warum ist das jetzt wichtig auf Fairness zu achten? ‚Ach Gott, dann ist ja jetzt diese eine Maschine ein bisschen diskriminierend‘ ist ja jetzt besser als gar nichts zu haben vielleicht. Aber dann hat es bei mir klick gemacht. Ja ne, eine falsche Maschine, das ist falsch‘. Es ist ja*

nicht nur ein bisschen falsch, es ist, wenn man es so sieht, einfach nur falsch und das kann man dann später für die Produktion oder so gar nicht benutzen“ (S2).

Es gab zwei eingesetzte Methoden, die von den Befragten als besonders hilfreich empfunden wurden, sowie eine, die auf weniger Resonanz stieß. Die meisten Studierenden verinnerlichten besonders Praxisbeispiele, die Probleme in realen sozio-technischen Kontexten beschrieben. Diese Ansicht wurde in allen Interviews geteilt, wie in diesem: *„[...] was ich immer schön fand war, dass man Praxisbeispiele hatte. Das fand ich immer sehr cool. Meistens waren das immer Sachen, die schief gelaufen sind und überhaupt nicht gut waren. Ich fand das, irgendwie sehr cool zu sehen, was dann auch passieren kann und was dann eben schon in der Welt passiert ist an absurden Dingen, die mit Data Science gemacht werden und welche sozialen Auswirkungen das auf Menschen hat“ (S4).* Überraschenderweise wurden Lehrmethoden, welche die Studierenden als aus anderen Disziplinen stammend ansahen, *„Ich weiß auch von anderen Studiengängen, dass da ganz viel reflektiert wird; immer über irgendwelche Hausaufgaben und alles Mögliche. Das ist da ein häufiger Aufgabenbestandteil“ (S3),* als wirklich hilfreich angesehen, da die Studierenden bemerkten: *„[...] okay, so eine Reflektion bringt wirklich was“ (S6).* Eine Person beschrieb, obwohl sie zuvor erwähnt hatte *„Programmieren mag ich immer mehr“ (S4),* dass das Aufschreiben ihr bei der Reflektion von Themen wirklich geholfen hat: *„[...] obwohl ich schreiben nicht ganz so gerne mag, aber es halt sein muss und man halt dadurch auch besser wird und irgendwann merke ich langsam auch, dass ich an einem Punkt bin, dass ich dann doch schreiben ganz gut finde und auch sehr sehr wichtig, weil man dabei auch viel reflektiert und beim Coding eigentlich nicht viel reflektieren muss zum Teil. Hauptsache es funktioniert halt“ (S4).* Wenig überraschend war, dass die große thematische Breite für die Teilnehmer:innen eine Herausforderung darstellte. Es war eine komplexe und schwierige Aufgabe, all die verschiedenen Aspekte des HCDS in einem einzigen Kurs zu behandeln. Insbesondere das Thema Bias wurde von den Studierenden wiederholt als Beispiel für diese Komplexität angesprochen. Häufig wurde beklagt, dass in der Vorlesung zu viele verschiedene Bias Formen behandelt wurden und dass die Studierenden nach einer Weile einfach den Anschluss verloren: *„[...] es ist so schwer für mich, der Vorlesung zu folgen, weil ich zum ersten Mal so viele verschiedene Arten von Vorurteilen höre“ (S1).*

2.3 Einordnung der Ergebnisse

Die erste Version des Kurses wurde aufgrund mangelnder bestehender Lehrbeispiele und nicht bestehender Textbücher recht intuitiv entworfen, wobei wir uns weitgehend von der bestehenden Forschung und unseren eigenen Interessen leiten ließen. Aufgrund unserer eigenen Erfahrungen und der Ergebnisse der Interviewstudie haben wir uns jedoch entschlossen, genauer über unseren Vermittlungsansatz nachzudenken.

Bei der Erörterung des Feedbacks der Studierenden stellte sich heraus, dass das größte Bedürfnis darin bestand, die Werte und grundlegenden Thesen zu vermitteln, die das Verständnis und die Problemlösung von Informatiker:innen in Human-Centered Data Science-Projekten prägen (vgl. [OV21]). Wir stellten fest, dass wir zwar die verschiedenen

Methoden, Konzepte und Techniken, die zugrundeliegenden Werte und grundlegenden Thesen, die die Praxis der Human-Centered Data Science prägen, vorgestellt hatten, diese aber nicht ausreichend an praktischen Kenntnissen und Fähigkeiten ausgerichtet hatten. Infolgedessen hatten die Studierenden zwar das Gefühl, ein Vokabular für die Diskussion bestehender Probleme in ihrer Data-Science-Praxis gelernt zu haben, wir hatten aber den Eindruck, dass sie nicht über einen ausreichenden „Werkzeugkasten“ verfügten, diese Probleme in ihrer Praxis dann auch nachhaltig anzugehen.

Dieser „Werkzeugkasten“ sollte gerade nicht aus Metriken, Techniken, Algorithmen oder Software bestehen, sondern aus einer „Haltung“ bzw. „Einstellung“, die es Studierenden erlaubt, über diese Probleme nachzudenken. Nachdem wir diesen Eindruck diskutiert hatten, wollten wir das Bewusstsein für eine ethische Verantwortung in Human-Centered Data Science enger mit der Praxis verknüpfen. Daher haben wir uns insbesondere mit Lehrensätzen aus dem Bereich der Ethik auseinandergesetzt, wie wir nachfolgend vorstellen werden.

3 Förderung kritisch-reflexiver Praktiken in der Data Science

Eine Erkenntnis aus den Herausforderungen des Einsatzes von datengetriebener Software im gesellschaftlichen Kontext war, dass Informatiker:innen oder Datenwissenschaftler:innen neben der Vermittlung von technischen Fähigkeiten auch im ethischen Denken geschult werden sollten. Bestehende Ethik-Lehrveranstaltungen verfolgen das Ziel, den Studierenden beizubringen, ethische Probleme in der Welt zu erkennen, diese Probleme kritisch zu beurteilen und Technologien in Bezug auf diese Probleme zu bewerten sowie gut begründete Argumente auf der Grundlage der Kritik zu formulieren [FGB20]. Ethik wird in diesen Veranstaltungen aber eher selten aus der Entwicklungspraxis heraus operationalisiert. Es werden vielmehr ethische Theorien (z.B. Utilitarismus, Deontologie) erläutert und anhand moralphilosophischer Gedankenexperimente (z.B. Trolley-Problem) diskutiert. Es werden darüber hinaus bestehende ethische Grundsätze (z.B. Privatsphäre, Rechenschaftspflicht, Transparenz und Erklärbarkeit, Fairness und Nichtdiskriminierung) eingeführt, aber deren konkrete Umsetzung in der jeweiligen Data Science-Praxis bleibt un spezifiziert. Dies hat damit zu tun, dass die Ethik-Ausbildung in fachlich losgelöste Veranstaltungen ausgelagert und häufig aus Sicht der Geistes- oder Sozialwissenschaft vermittelt wird. Dies ist fachlich sicherlich sinnvoll, aber damit ethische Überlegungen auch über den jeweiligen Lehrkontext hinaus nachhaltig wirken können, sollten diese stattdessen in Data Science-Praktiken eingebettet sein.

Es bedarf in der Data Science-Ausbildung einer Handlungsethik (Ethics-in-Action), wie sie von Frauenberger et al. formuliert wurde [FRF17]. Die Handlungsethik ergänzt dabei den durch bestehende ethische Grundsätze (formuliert durch Regierungen, Unternehmen, Fachgesellschaften und Nichtregierungsorganisationen) formalisierten institutionalisierten Ethik-Rahmen, verbindet diesen aber gleichzeitig mit der praktischen Datenarbeit. Die Handlungsethik baut auf Donald Schöns Reflective Practice auf, nach dem eine Person

durch einen gewählten methodischen Ansatz in einen Dialog mit einer Situation tritt. Dieser Dialog „*should create a dynamic in which the situation 'talks back', and [the designer] responds to the situation's back-talk*“ [Sc83]. Schön bezeichnet dieses Gespräch mit einer Situation als reflexiv [Sc83]. Darüber hinaus unterscheidet Schön zwischen einer Reflection-in-Action (Handlungsreflexion) und Reflection-on-Action (Reflexion über die Handlung). Erstere findet während der Handlung statt und ergibt sich aus unerwarteten Ergebnissen einer Handlung. Letztere beschreibt eine Reflexion, die auf der Grundlage einer früheren Handlung erfolgt und mit einer Phase der Rückschau beginnt, in der diejenigen, die reflektieren, die frühere Handlung überprüfen. Im Bereich Data Science kann eine solche „Situation“, in der eine Handlungsreflexion erfolgt, als konkrete Entscheidung bei der Datenauswahl, der Datenaufbereitung, dem Feature Engineering etc. angesehen werden. Bei diesen Entscheidungen ist es wichtig, diesen „Dialog“ zur Handlungsreflexion herbeizuführen, der letztlich Auswirkungen auf die Entscheidungen hat.

Aus der Handlungsethik und der Handlungsreflexion ergibt sich somit die Notwendigkeit, ethische Überlegungen eng in die Daten- und Programmierpraxis von Datenwissenschaftler:innen einzubetten. Der Dialog kann nur durch bestehende ethische Grundsätze (z.B. Fairness und Nichtdiskriminierung) initiiert werden, aber das reflexive Handeln selbst muss zu einem inhärenten Teil der Data Science-Praxis werden. Um diese Verbindung herzustellen, wird *ēthos* (altgriechisch „Charakter“) benötigt, *“a moral commitment or stance, a moral attitude that underlies a particular practice”* [FRF17]. Im Gegensatz zu formalen Ethikgrundsätzen ist *ēthos* also intrinsisch und personifiziert. Somit ist bei der akademischen Ausbildung ein wichtiges Ziel, dass sich Studierende ihrer Werte bewusst werden. Diese Werte bilden letztlich die Grundlage für die Anwendung der ethischen Grundsätze. Es ist somit von zentraler Bedeutung, dass Studierende diese eigenen Werte formulieren und danach handeln können.

Basierend auf diesen Überlegungen stellen wir nachfolgend unseren Vorschlag für einen methodischen Ansatz für die Vermittlung von Human-Centered Data Science in der akademischen Ausbildung vor, welche das Ziel verfolgt, die kritisch-reflexive Praxis der Data Science-Studierenden zu fördern, indem die technischen und ethischen Kompetenzen gestärkt werden. Wir verwenden dazu bestehende ethische Grundsätze (z.B. Fairness und Nichtdiskriminierung, Transparenz und Erklärbarkeit) und verknüpfen sie mit konkreten Implementierungsaufgaben. Jeder ethische Grundsatz wird anhand einer Fallstudie eingeführt, um Interesse zu wecken. Dieser Fall wird durch Studierende anhand ihrer Werte reflektiert, um potenzielle Lösungsansätze abzuleiten, die dann anhand verfügbarer „Lösungsansätze“ umgesetzt und praktisch evaluiert werden. Als konzeptioneller Rahmen dienen dabei Greens vier Phasen einer Critical Reflexive Technical Practice [Gr21]. Nachfolgend erläutern wir diese vier Phasen (Interesse, Reflexion, Anwendungen und Praxis) exemplarisch anhand des ethischen Grundsatzes der Nichtdiskriminierung.

Die ersten zwei Phasen werden in der Vorlesung durchlaufen, während die letzten beiden Phasen Teil der Übung sind. Diese Phasen können innerhalb einer Lehrveranstaltung mehrmals unter Anwendung verschiedener ethischer Grundsätze durchlaufen werden.

4 Beispiel einer kritisch-reflexiven Praxis anhand des Grundsatzes der Nichtdiskriminierung

In der ersten Phase geht es vor allem darum, Interesse für das Problem (abgeleitet aus dem ethischen Grundsatz) zu wecken. Dazu sollten gesellschaftlich relevante Daten anstelle von banalen Beispieldaten verwendet und konkrete gesellschaftliche Probleme diskutiert werden. Dies zielt darauf ab, das Denken der Studierenden vom einfachen Technologieeinsatz auf die positive Beeinflussung der Gesellschaft zu lenken [Gr21]. Hierfür verwenden wir Fallstudien, die eine bestimmte „reale“ Situation beschreiben. Die Fallstudien dienen dazu, Studierende nach erfolgtem theoretischem Input (z.B. im Bereich der Diskriminierung) zu aktivieren. Geeignete Beispiele finden sich in der Sammlung der Wissensbits [We09], die von Mitgliedern der Fachgruppe „Informatik und Ethik“ der Gesellschaft für Informatik in den letzten zehn Jahren erstellt wurden⁵. Die Fallstudien decken ein breites Spektrum gesellschaftlicher Themen ab, wodurch Studierende leichter nachvollziehen können, inwieweit eine bestimmte Technologie auch ihr unmittelbares Leben beeinflussen kann.

In der zweiten Phase soll die Reflexion gefördert werden, indem Erkenntnisse und Theorien aus Bereichen wie der Wissenschafts- und Technikforschung, Technikphilosophie, Soziologie und Politikwissenschaft bei der Beurteilung einer „Situation“ einbezogen werden. Die Situation wird dabei in Daten, Modell, Zielgruppe, Zweck, Kontext etc. zerlegt, während die erörterten Theorien dazu anregen sollen, über bestehende Annahmen in diesen Bereichen nachzudenken. Durch die Diskussion der Fallstudie wird es Studierenden ermöglicht, über ethische Herausforderungen in der Data Science nachzudenken und zu erkennen, dass es oft keine eindeutigen Lösungen gibt. Ziel ist es, Studierende zu befähigen, die eigenen Argumente nachvollziehbar zu formulieren und darzustellen sowie gegenteilige Argumente aufzunehmen und zu bewerten. Ein wesentlicher Ansatz in der Diskussion ist es daher, auf andere Meinungen einzugehen und gemeinsame Lösungen zu suchen. Eine wichtige Erkenntnis in diesen Diskussionen soll es sein, dass bestehende Zielkonflikte beispielsweise bezüglich des Ressourcenverbrauchs, des Sicherheitsniveaus und der Zuverlässigkeit nicht eindeutig lösbar sind. Darüber hinaus sollen Studierende verstehen, dass datengetriebene Software innerhalb eines soziotechnischen Kontexts eingebettet ist und Zielkonflikte auch dadurch häufig nicht lösbar sind, da es nicht nur um technische Lösungen geht.

In der dritten Phase konzentrieren wir uns darauf, Studierenden dabei zu unterstützen, die erlernten Konzepte/Perspektiven in ihre Data Science-Praxis einzubringen. Daher stellen wir Studierenden Datensätze zur Verfügung (z.B. Kreditdaten⁶), welche sie zur Umsetzung einer konkreten Anwendung (z.B. Kreditberatungssoftware) verwenden sollen. Im

⁵ Wissensbits werden regelmäßig in einer Rubrik der Fachzeitschrift Informatik Spektrum von Springer veröffentlicht. Weitere Informationen finden Sie unter <http://gewissensbits.gi.de>

⁶ Der Datensatz ist verfügbar unter <https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29>

Themenbereich der Diskriminierung erarbeiten Studierende durch den Einsatz der „Datasheets for Datasets“ [Ge21], was sie über den Datensatz wissen und was nicht. Wir diskutieren anschließend, wie durch diese Dokumentation die Nutzbarkeit solcher Datensätze erhöht werden könnte. Anschließend untersuchen die Studierenden mögliche Verzerrungen im Datensatz mithilfe des AI360 Toolkit⁷. Das Ziel dieser praktischen Übungen ist es immer wieder, dass Studierende lernen, ihre Entscheidungen zu hinterfragen, um so eine kritische Denkweise aufzubauen. Dabei geht es auch darum, bestehende Techniken, Ansätze und Hilfsmittel zu kritisieren, d.h. deren Grenzen zu verstehen und Veränderungsvorschläge zu erarbeiten.

In der letzten Phase sollen Studierenden in die Lage versetzt werden, eine Human-Centered Data Science-Praxis zu realisieren. Dabei lernen Studierende Methoden aus der menschenzentrierten Gestaltung kennen und wenden diese innerhalb eines praktischen Beispiels an. Beispielsweise geht es bei der Kreditberatungssoftware nicht nur darum, dass Studierende ein entsprechendes Vorhersagemodell entwickeln, sondern auch mit den potenziellen Gruppen, die vom Einsatz dieser Anwendung (direkt oder indirekt) betroffen sind, in Kontakt treten und die Auswirkungen ihrer Software verstehen lernen. Dazu werden zunächst Evaluationsstudien durchgeführt, indem die Studierenden, die in den vorhergehenden Phasen realisierten Anwendungen (vor allem in Form von Jupyter Notebooks) kennenlernen und gegenseitig beurteilen. Durch die gegenseitige Evaluation der unterschiedlichen Anwendungen werden die Studierenden ermutigt, die „beste“ Anwendung zu küren. Auch diese Aufgabe soll Studierende befähigen, verschiedene Perspektiven auf das Problemfeld einzunehmen und auch potenzielle Konsequenzen zu antizipieren.

5 Zusammenfassung und Ausblick

Human-Centered Data Science ist ein interdisziplinäres Forschungsgebiet, welches sich vor allem auf einen menschenzentrierten Ansatz der Technologiegestaltung fokussiert und dabei Erkenntnisse aus den Bereichen Mensch-Computer-Interaktion, Sozialwissenschaften, Statistik und maschinelles Lernen integriert. Bestehende Lehrkonzepte zur Vermittlung dieses Forschungsgebiet existieren bisher noch nicht. Mit dem vorliegenden Artikel möchten wir diese Lücke schließen und bauen auf Schöns Handlungsreflexion [Sc83] und Greens Phasen einer Critical Reflexive Technical Practice [Gr21] auf, um eine kritisch-reflexive Praxis von Studierenden zu fördern.

Aktuell wird das hier vorgestellte Lehrkonzept praktisch in einer laufenden Lehrveranstaltung evaluiert. Für die Zukunft planen wir, alle Materialien als frei zugängliches Lehrmaterial unter einer offenen Lizenz zur Verfügung zu stellen.

Danksagung

Die Autor:innen bedanken sich für die finanzielle Unterstützung beim Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Vorhaben "ENKIS – Etablierung

⁷ Weitere Informationen zu dieser Software sind unter <https://github.com/Trusted-AI/AIF360> verfügbar.

nachhaltiger KI-bezogener Studienangebote für eine Verantwortungsvolle Künstliche Intelligenz an der Freien Universität Berlin" (Förderkennzeichen: 16DHBKI025) und bei allen Reviewer:innen für die hilfreichen Verbesserungsvorschläge.

Literaturverzeichnis

- [Ar22] Aragon, Cecilia; Guha, Shion; Kogan, Marina; Muller, Michael; Neff, Gina: *Human-Centered Data Science: An Introduction*. MIT Press, Cambridge, MA, USA, 2022.
- [Ba17] Baumer, Eric P. S.: *Toward human-centered algorithm design*. *Big Data & Society*, 4(2):2053951717718854, 2017.
- [BC06] Braun, Virginia; Clarke, Victoria: *Using Thematic Analysis in Psychology*. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [FGB20] Fiesler, Casey; Garrett, Natalie; Beard, Nathan: *What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis*. In: *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, S. 289–295, 2020.
- [FRF17] Frauenberger, Christopher; Rauhala, Marjo; Fitzpatrick, Geraldine: *In-Action Ethics*. *Interacting with Computers*, 29(2):220–236, Mar 2017.
- [Ge21] Gebru, Timnit; Morgenstern, Jamie; Vecchione, Briana; Vaughan, Jennifer Wortman; Wallach, Hanna; III, Hal Daumé; Crawford, Kate: *Datasheets for datasets*. *Communications of the ACM*, 64(12):86–92, Nov 2021.
- [Gr21] Green, Ben: *Data Science as Political Action: Grounding Data Science in a Politics of Justice*. *Journal of Social Computing*, 2(3):249–265, Sep 2021.
- [Ko20] Kogan, Marina; Halfaker, Aaron; Guha, Shion; Aragon, Cecilia; Muller, Michael; Geiger, Stuart: *Mapping Out Human-Centered Data Science: Methods, Approaches, and Best Practices*. In: *Companion of the 2020 ACM International Conference on Supporting Group Work*. GROUP '20, ACM, New York, NY, USA, S. 151–156, 2020.
- [KS98] Kling, Rob; Star, Susan Leigh: *Human Centered Systems in the Perspective of Organizational and Social Informatics*. *SIGCAS Comput. Soc.*, 28(1):22–29, 1998.
- [OV21] Olsson, Thomas; Väänänen, Kaisa: *How Does AI Challenge Design Practice?* *Interactions*, 28(4):62–64, 2021.
- [Sc83] Schön, Donald A.: *The reflective practitioner: how professionals think in action*. Basic Books, New York, 1983.
- [Va17] Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia: *Attention is all you need*. *Advances in neural information processing systems*, 30, 2017.
- [We09] Weber-Wulff, Dora; Class, Christina; Coy, Wolfgang; Kurz, Constanze; Zellhöfer, David: *Gewissensbisse: Ethische Probleme der Informatik*. *Biometrie - Datenschutz - geistiges Eigentum*. transcript Verlag, 2009.