erschienen in der FIfF-Kommunikation,
herausgegeben von FIfF e.V. - ISSN 0938-3476
www.fiff.de

Claudia Müller-Birn

Human-Centered Data Science

Etablierung einer kritisch-reflexiven Praxis bei der Entwicklung von datengetriebener Software

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. 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 zu erfassen (Aragon et al., 2022). Aus diesen Herausforderungen hat sich das Gebiet Human-Centered Data Science *entwickelt, in welchem ein Verständnis für die komplexen Interaktionen zwischen Gesellschaft, Technologie und von Menschen erzeugten Daten vermittelt wird (Aragon et al., 2022). 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 können. Ein zentrales Anliegen vom 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.*

Human-Centered Data Science – Erweiterung von Data Science durch Qualitative Methoden

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 (Aragon et al., 2022). Human-Centered Data Science liegt dabei ein menschenzentrierter Ansatz bei der Technologiegestaltung zugrunde, das *human-centered design*, um die Praktiken des Data Science zu verbessern. Dieser Ansatz der menschenzentrierten Gestaltung basiert auf einer Reihe von Leitsätzen, die Kling und Star bereits vor über 20 Jahren aufgestellt haben (Kling & Star, 1998). Danach sollte datengetriebene Software menschliche Fähigkeiten sinnvoll ergänzen, aber diese nicht ersetzen oder automatisieren. Soziale Konstrukte (wie Fairness) sollten nicht in mathematische Konzepte übersetzt werden, da mathematische Operationalisierungen die Vielfältigkeit unserer 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 den 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 Probleme wie beispielsweise die der sozialen Gerechtigkeit lösen. Daher bildet ein Verständnis über die inkludierten Werte, die der zukünftigen direkt oder indirekt durch den Softwareeinsatz betroffenen Personengruppen und die der Datenwissenschaftler:innen selbst, eine wesentliche Grundlage der menschenzentrierten Gestaltung. Durch einen solchen menschenzentrierten Ansatz bezüglich Data Science können eine Vielzahl von Methoden aus dem Bereich der Mensch-Computer-Interaktion im Data Science verwendet werden, wie beispielsweise die partizipative 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 ergänzen.

Förderung kritisch-reflexiver Praktiken im 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. Gegenwärtige Ethikkurse 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 (Fiesler et al., 2020). Ethik wird in diesen Veranstaltungen aber häufig als statischer, antizipatorischer und formalisierter Prozess operationalisiert. Es werden ethische Theorien (wie Utilitarismus, Deontologie) erläutert und anhand

moralphilosophischer Gedankenexperimente (z. B. Trolley-Problem) diskutiert. Es werden darüber hinaus bestehende ethische Grundsätze (wie Privatsphäre, Rechenschaftspflicht, Transparenz und Erklärbarkeit, Fairness und Nichtdiskriminierung) eingeführt, aber deren konkrete Umsetzung in der jeweiligen Data Science-Praxis bleibt theoretisch. Das hat wahrscheinlich damit zu tun, dass die Ethik-Ausbildung häufig aus Sicht der Geistes- oder Sozialwissenschaft vermittelt wird. Das 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 *Ethics-in-Action* (Handlungsethik) wie sie von Frauenberger et al. formuliert wurde (Frauenberger et al., 2017). 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 der 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" (Schön, 1983, S. 79). Schön bezeichnet dieses Gespräch mit einer Situation als reflexiv (ebd.). Im Bereich Data Science kann eine solche *Situation* als konkrete Entscheidung bei der Datenauswahl, der Datenaufbereitung, dem Feature Engineering, etc. angesehen werden. Bei diesen Entscheidungen ist es wichtig, diesen *Dialog* herbeizuführen, der letztlich zu einem reflexiven Handeln befähigt.

Aus der Handlungsethik 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 (wie 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" (Frauenberger et al., 2017). 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 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 (beispielsweise 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 Aufgaben abzuleiten, die dann anhand verfügbarer *Lösungsansätze* umgesetzt und praktisch evaluiert werden. Als konzeptioneller Rahmen dienen dabei Greens Stufen einer *critical technical practice* (Green, 2021). Nachfolgend erläutern wir die 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 Übungsveranstaltung sind. Diese Phasen können innerhalb einer Lehrveranstaltung mehrmals unter Anwendung verschiedener ethischer Grundsätze durchlaufen werden.

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 Technologie-Einsatz auf die positive Beeinflussung der Gesellschaft zu lenken (Green, 2021). Hierfür verwenden wir Fallstudien, die eine bestimmte *reale Situation* beschreiben. Die Fallstudien dienen dazu, Studierende nach erfolgtem theoretischem Input (wie im Bereich der Diskriminierung) zu aktivieren. Geeignete Beispiele finden sich in der Sammlung der Gewissensbits (Kurz et al., 2009), die von Mitgliedern der Fachgruppe *Informatik und Ethik* der Gesellschaft für Informatik in den letzten zehn Jahren erstellt wurden.[1] Die Fallstudien decken ein breites Spektrum gesellschaftlicher Themen ab, wodurch Studierende leichter nachvollziehen können, inwieweit eine bestimmte Technologie auch ihr Leben unmittelbar beeinflussen kann.

Die zweite Phase soll die Reflexion fördern, 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 im 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 in einen soziotechnischen Kontexts eingebettet ist und Zielkonflikte auch dadurch häufig nicht lösbar sind, weil es nicht nur um technische Lösungen geht.

In der dritten Phase konzentrieren wir uns darauf, die Studierenden dabei zu unterstützen, die erlernten Konzepte/Perspek-

tiven in ihre Data Science-Praxis einzubringen. Daher stellen wir Studierenden Datensätze zur Verfügung (wie Kreditdaten[2]), die sie zur Umsetzung einer konkreten Anwendung (beispielsweise Kreditberatungs-Software) verwenden sollen. Im Themenbereich der Diskriminierung erarbeiten Studierende durch den Einsatz der *Datasheets for Datasets* (Gebru et al., 2021), was sie über den Datensatz wissen und was nicht. Wir diskutieren anschließend, wie durch diese Dokumentation die Nutzbarkeit solcher Datensätze erhöht werden kann. Anschließend untersuchen die Studierenden mögliche Verzerrungen im Datensatz mithilfe des AI360-Toolkit[3]. 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 sollten die Studierenden in die Lage versetzt werden, eine partizipative Data-Science-Praxis zu realisieren. Dabei lernen sie Methoden (wie partizipatives Design) aus der menschenzentrierten Gestaltung kennen und wenden diese an. Beispielsweise geht es bei der Kreditberatungs-Software nicht nur darum, dass sie 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 der vorherigen Phase realisierten Anwendungen (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.

Zusammenfassung und Ausblick

Die Veranstaltung *Human-Centered Data Science* ist Teil des Bildungsprogramms *Verantwortungsvolle Informatik*, das an unserem Institut für Informatik der Freien Universität Berlin gerade angelaufen ist. Wir haben die Veranstaltung *Human-Centered Data Science* im Wintersemester 2022/21 erstmalig für Informatik-Studierende an der Freien Universität angeboten. Im Anschluss an diese Lehrveranstaltung führten wir eine Interviewstudie durch, deren Ergebnisse das beschriebene Konzept maßgeblich geleitet haben. Aktuell wird das vorgestellte Konzept 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. Wir möchten aber diesen Artikel mit einem Zitat von Bates et al. (2020) beschließen: „students' critical and ethical thinking is clearly not a cure-all; unethical and socially irresponsible data practice will continue as long as it goes rewarded and poorly regulated."

Referenzen

Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff (2022) Human-Centered Data Science: An Introduction. MIT Press.

Jo Bates et al. (2020) Integrating FATE/critical data studies into data science curricula: where are we going and how do we get there? In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, 425–435.

Casey Fiesler, Natalie Garrett, and Nathan Beard (2020) 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, 289–295.

Christopher Frauenberger, Marjo Rauhala, and Geraldine Fitzpatrick (2017) In-Action Ethics. Interacting with Computers 29, 2 (Mar 2017), 220–236.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford (2021) Datasheets for datasets. Commun. ACM 64, 12 (Nov 2021), 86–92.

Ben Green (2021) Data Science as Political Action: Grounding Data Science in a Politics of Justice. Journal of Social Computing 2, 3 (Sep 2021), 249–265.

Rob Kling and Susan L. Star (1998) Human Centered Systems in the Perspective of Organizational and Social Informatics. SIGCAS Comput. Soc. 28, 1 (March 1998), 22–29.

Constanze Kurz, David Zellhöfer, Debora Weber-Wulff, Christina Class and Wolfgang Coy (2009) Gewissensbisse: Ethische Probleme der Informatik. Biometrie – Datenschutz – geistiges Eigentum.

Donald A. Schön (1983) The reflective practitioner: How professionals think in action. Basic books.

Anmerkungen

1 *Gewissensbits werden regelmäßig in einer Rubrik der Fachzeitschrift Informatik Spektrum von Springer veröffentlicht. Weitere Informationen finden Sie unter http://gewissensbits.gi.de.*

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

3 *Weitere Informationen sind hier verfügbar: https://github.com/Trusted-AI/AIF360.*

Claudia Müller-Birn

Claudia Müller-Birn leitet die Forschungsgruppe *Human-Centered Computing* am Institut für Informatik der Freien Universität Berlin. In ihrer Forschung untersucht sie die Verflechtung von Menschen, Daten, Algorithmen und Kontext, um eine Human-Computer-Collaboration zu ermöglichen. Ihr aktueller Anwendungsschwerpunkt liegt auf Technologien des maschinellen Lernens im Zusammenhang mit Privatsphäre, Reflexion und Interpretierbarkeit.