Interview mit Kiara Stempel (Stand 14.02.2023)

Wie sieht ein normaler Arbeitstag bei dir aus?

Wir alle arbeiten alle recht selbstständig, lesen verschiedene Paper, schauen uns an, was andere schon gemacht haben, überlegen, was man sonst machen, erweitern oder verbessern könnte, wie man Ideen zusammenführen kann und implementieren bzw. testen dann diese Ideen. Manchmal arbeitet auch zum Beispiel ein anderer im Büro, Nachbarbüro oder Forschungsprojekt mit überschneidenden Themen, womit man dann direkt einen Gesprächspartner zum Entwickeln neuer Ideen hat.

In welchem Fachbereich der Universität verortet sich denn dein Teilprojekt?

Im Fachbereich Informatik. Ich bearbeite ja das Teilprojekt Fairness und Transparenz von Machine Learning-Algorithmen. Trotzdem gehört da thematisch eigentlich auch ein bisschen Rechtswissenschaften dazu, durch die gesetzlichen Regulierungen, die gerade stattfinden, und auch Philosophie, also was bedeutet Verständlichkeit, was bedeutet Transparenz, was heißt es überhaupt zu verstehen.

Und was begeistert dich an der Informatik? Wie bist du dazu gekommen?

Ich habe das Abitur auf einem Wirtschaftsgymnasium gemacht, wo Informatik als Leistungskurs angeboten wurde, das habe ich gewählt und es hat mir Spaß gemacht. Ursprünglich wollte ich es nicht studieren, weil es sich für mich eher wie ein Hobby angefühlt hat. Aber dann habe ich gemerkt, dass mir einfach nichts anderes besser gefällt. Ich habe angefangen Informatik zu studieren und es hat mir sehr gut gefallen.

Und wieso hast du dich dann für die Promotion entschieden?

Als ich im Entscheidungsprozess war und auch nach Jobs in der Wirtschaft geschaut habe, habe ich gemerkt, dass ich immer nachgeschaut habe, ob die entsprechende Stelle einen Bezug zur Forschung hat. Und das ist der Punkt gewesen, der mir gezeigt hat, dass ich erstmal hier, also an der Uni, bleiben möchte, weil es in der Wirtschaft einfach nicht den gleichen Forschungsbezug gibt wie hier.

Und warum gerade in Mainz? Hast du deinen Master hier gemacht?

Ja, Master und Bachelor.

Und wie bist du dann zum Projekt TOPML gekommen?

Ich habe meine Masterarbeit bei Prof. Stefan Kramer geschrieben und er hat mich informiert, dass das Projekt bald starten wird und gefragt, ob ich Interesse hätte an einem der Themen. Und dann habe ich darüber nachgedacht und gleichzeitig gemerkt, dass ich eigentlich nur nach Forschungsstellen schaue und habe ich mich deshalb dazu entschieden, mich zu bewerben.

Was machst du genau in deinem Projekt? Was erzählst du beispielsweise einem Familienmitglied bei einem Familienfest darüber, was du Interessantes und Relevantes machst?

Ich bearbeite die Themen Fairness und Transparenz von Algorithmen des Maschinellen Lernens und es geht darum, diese beiden Eigenschaften von Algorithmen irgendwie zusammenzubringen oder zu schauen, ob es Zielkonflikte bzw. Trade-Offs gibt.

Bei Transparenz gibt es verschiedene Aspekte: beispielsweise gehe ich davon aus, dass man mehr Vertrauen in einen Algorithmus hat, wenn man versteht, was ein Algorithmus macht und wie er funktioniert, sodass man diesen dann auch eher verwenden möchte oder möchte, dass er von

Unternehmen oder in der Medizin verwendet wird, um Entscheidungen vorzunehmen, die einen selbst betreffen. Und zum anderen können Fehler ggf. besser identifiziert werden. Wenn beim autonomen Fahren oder bei einer medizinischen Diagnose Fehler passieren oder im Nachhinein festgestellt werden, kann man mithilfe eines transparenten Algorithmus besser herausfinden, wie diese entstanden oder an welcher Stelle in der Ausführung des Algorithmus sie passiert sind, um dann beispielsweise beim nächsten Mal entgegenzuwirken. Es ist aber schwer, abzuwiegen, wann ein Algorithmus wirklich als transparent angesehen werden kann.

Die andere Eigenschaft ist Fairness. Hierzu vielleicht am Beispiel: wenn Kredite vergeben werden und man eine Person hat, bei der die Kreditwürdigkeit geprüft wird, möchte man nicht, dass diese Kreditwürdigkeit von Eigenschaften abhängt, von denen sie nicht abhängen sollte. Die Kreditwürdigkeit darf natürlich vom Einkommen abhängen, aber die Herkunft oder die soziale Herkunft sollen keine Rolle spielen. Man würde auch nicht wollen, dass von Menschen getroffene Entscheidungen von solchen Aspekten abhängen, und das will man dann genauso vermeiden, wenn ein Algorithmus die Entscheidungen trifft. Auch zum Beispiel bei der Jobvergabe gibt es solche Faktoren, denn auch dort sollten Geschlecht und Herkunft keine Rolle spielen.

Schwierig ist dabei, dass ein Algorithmus immer von echten Daten lernt, d.h. von Entscheidungen, die durch Menschen getroffen wurden, und die echten Daten sind häufig so, dass die Herkunft oder das Geschlecht bei der Entscheidung eine Rolle gespielt haben. Dann ist es schwierig, den Algorithmus so zu trainieren, dass er das aus den Daten nicht mitlernt, dass die Menschen so entschieden haben.

Gerade baue ich zum Beispiel auf eine Methode auf, die schon fair ist und versuche dieses faire Modell, das noch nicht vollständig transparent ist, jetzt auch noch transparenter zu machen. Dann prüfe ich beispielsweise, ob das Modell noch genauso gut funktioniert oder ob es schlechter wird, wenn es zusätzlich transparent ist.

Du entwickelst ja gerade ein Modell weiter. Welche Daten gehen in dieses Modell rein und welche kommen raus?

Das sind unterschiedliche Arten von Daten. Es können beispielsweise Informationen über eine einzelne Person sein, wie Alter, Geschlecht, Bildung, ob jemand verheiratet ist, Herkunft oder die Herkunft eben auch nicht, wenn man das eben aufgrund der Sensibilität der Information nicht möchte. Diese Informationen über eine Person gehen rein und der Algorithmus gibt eine Entscheidung aus, bekommt diese Person einen Kredit oder bekommt sie keinen Kredit, bekommt die Person den Job oder nicht oder bekommt sie ihn zu einer gewissen Wahrscheinlichkeit. Und zusätzlich gibt der Algorithmus, mit dem ich aktuell arbeite, eine veränderte Darstellung dieser Person raus. Diese soll möglichst fair sein, sodass sie sensible Informationen, wie die Herkunft oder das Geschlecht, verschleiert.

Was ist eigentlich KI?

Ich finde es sehr schwierig, das zu definieren, weil es so viele Gebiete umfasst, so viele verschiedene Verfahren, aber ich würde sagen, dass es Programme oder Algorithmen sind, die versuchen zu imitieren, wie der Mensch lernt und wie der Mensch Entscheidungen trifft, diese Entscheidungen nachbilden und sich anhand von Erfahrungen verbessern. Es gibt Varianten, in denen eine KI etwas ausprobiert und dann aus ihren Fehlern lernt. Oder der Algorithmus lernt aus bekannten Daten basierend auf der Grundlage, wie Menschen in der Vergangenheit entschieden haben. Das machen wir ja auch in vielen Fällen so. Beispielsweise Anwälte, die sich, soweit ich weiß, vergangene Fälle anschauen und darauf aufbauend versuchen, diese früheren Entscheidungen auf einen aktuellen Fall zu übertragen. Und ich würde sagen, das versucht die Künstliche Intelligenz auch, aber es ist so

allumfassend, es ist schwer auf einen Satz herunterzubrechen. Da kann man sicherlich auch drüber diskutieren, wie man das korrekt darstellt.

Welche Herausforderungen und Chancen bringt die fortschreitende Entwicklung der künstlichen Intelligenz mit sich und wie sieht die Zukunft der künstlichen Intelligenz aus und welchen Einfluss wird sie auf die Gesellschaft haben?

Ein Vorteil der KI ist sicherlich, dass sie sich in kürzerer Zeit mehr Daten anschauen kann als ein Mensch das kann. Das heißt, wenn man zum Beispiel ein Ultraschallbild hat, dann kann eine KI in kürzerer Zeit und effizienter einen Vergleich mit bisherigen Ultraschallbildern, zu denen man schon Ergebnisse hat, anstellen und ein Mensch würde deutlich länger brauchen, um sich all diese Bilder vergleichend anzuschauen. Letztendlich kann die KI mehr Informationen mit mehr Genauigkeit aufnehmen, was von Vorteil ist. Dafür könnten Details zu einzelnen Bildern übersehen werden, die vielleicht von einem Menschen wahrgenommen werden würden. Es ist schwer zu sagen, ob das tatsächlich passiert, aber ohne Transparenz hat man keine direkte Überprüfbarkeit, ob die Ergebnisse korrekt sind und wie sie genau zustande gekommen sind. Eine Option ist, dass die KI nur einen Teil der Aufgabe übernimmt. Beispielsweise in der Medizin, wenn der Algorithmus digitale Daten analysiert, Ergebnisse oder Teilergebnisse wie Diagnosen bereitstellt und der Arzt oder die Ärztin aber trotzdem noch einmal einen abschließenden Blick darauf wirft, vor allem wenn es sich um Ausnahmen handelt oder die Ergebnisse der KI merkwürdig erscheinen. Die Zeitersparnis könnte dann für Dinge genutzt werden, die die KI nicht abnehmen kann, wie Patient:innengespräche oder Untersuchungen. Der Algorithmus ist dann mehr eine Hilfestellung, soll aber nicht ganz allein eine Entscheidung treffen. Da soll immer noch ein Mensch dahinter sitzen, welcher dann die finale Entscheidung trifft. Bei beispielsweise Netflix-Empfehlungen steckt ebenfalls ein Algorithmus dahinter, den wir alle nutzen. Bei so etwas würde sich aber kaum jemand beschweren, weil das Situationen sind, die nicht so viel Risiko mit sich bringen. Aber wenn es um deine Gesundheit, um eine Stelle auf dem Arbeitsmarkt oder Geld geht, dann ist das was anderes. Dann braucht man mehr Vertrauen in die KI.

Ich glaube schon, dass KIs in der Zukunft mehr verwendet werden. Aber ich denke auch, dass mehr Regeln und Gesetze erfüllt werden müssen, was gut ist. Natürlich muss man dann auch klären können, wo die Verantwortung liegt und wer für irgendwelche Schäden aufkommen muss.

Was sind die größten ethischen Herausforderungen bei der Entwicklung und Anwendung von KI und wie gehst du damit um?

Zum Beispiel zur Erzeugung von Fairness werden die Eingabedaten verändert, um eine fairere Darstellung von einer Person zu bekommen. Diese Veränderung von Daten kann eine ethische Herausforderung sein. Auf der einen Seite ist es nachvollziehbar das zu machen, wenn man Merkmale wie Herkunft oder Geschlecht verschleiern möchte, aber auf der anderen Seite wird eine Person verändert, bevor eine Entscheidung über sie getroffen wird, wozu es sicherlich verschiedene Meinungen gibt. Wenn meine Daten in einen Algorithmus gegeben werden und da wird beispielsweise mein Alter verschoben oder die Anzahl meiner Berufserfahrungsjahre wird heruntergesetzt, um irgendwas auszugleichen, dann würde ich das hinterfragen und würde wissen wollen, warum diese Dinge verändert wurden und ob das wirklich fair ist, dass das so verändert wurde. Ich denke, dass man darauf achten muss, dass nachvollziehbar ist, warum Daten angepasst worden sind, was oft gar nicht so einfach ist. Generell könnte es schwierig sein, dass Einzelfälle nicht so stark herauskommen oder nicht so differenziert betrachtet werden. Der Algorithmus arbeitet mit Regeln oder Gewichtungen, nach denen er sich entscheidet und der Mensch würde sich Details vielleicht nochmal ganz anders angucken als der Algorithmus, bei dem man nicht ganz genau weiß, ob der das macht. Gleichzeitig kann der Mensch auch Fehler machen. Das ist eine schwierige Frage.

Es wäre aber sehr schön, sie zu lösen.

Und wenn du es jetzt nicht auf dein Teilprojekt beziehst, sondern auf KI im Allgemeinen, wo siehst du da ethische Herausforderungen?

Algorithmen könnten Entscheidungen über Menschen treffen, die ihr Leben verändern können. Das kann ein Job sein, es kann eine Diagnose sein oder auch eine Kreditvergabe, wenn jemand ein Haus bauen möchte oder eine Firma eröffnen, dann würde ein Programm automatisiert darüber entscheiden, wie das weitere Leben von diesem Menschen verläuft, zumindest im Extremfall. Das ist schon etwas, das man im Hinterkopf behalten muss.

Dazu zählt auch die Verwendung der Daten generell. Die sollten auf jeden Fall sensibel behandelt werden. Vielleicht sollten Menschen auch mitentscheiden können, ob sie möchten, dass ein Algorithmus die Entscheidung trifft, denn die Menschen müssen diesem vertrauen.

Lehrst du? Und wenn ja, macht es dir Spaß?

Wir unterstützen das Seminar der Arbeitsgruppe, das in jedem Semester angeboten wird, haben aber sonst keine Lehraufgabe. Wir betreuen die Studierenden und geben Feedback zu den Präsentationen und den Ausarbeitungen, die sie schreiben. Sie können Fragen stellen, man schaut sich das Thema mit ihnen gemeinsam an und redet darüber.

Mir hat es während des Studiums in Seminaren sehr geholfen, wenn mir jemand mit etwas mehr Erfahrung vor der Präsentation Feedback gegeben hat und deshalb macht es mir Spaß, das jetzt ein bisschen zurückgeben zu können.

Um welches Thema geht es in dem Seminar?

Die beiden Seminare, die sich abwechseln, beziehen sich jeweils auf Maschinelles Lernen und Data-Mining, es sind also Themen der Arbeitsgruppe. Teilweise sind es Themen, die ich selbst im Studium gelernt habe, teilweise auch Themen, mit denen ich mich noch nicht so viel beschäftigt habe, und dadurch lernt man dann wieder etwas Neues kennen.

Ein aktuelles Thema im Seminar ist beispielsweise Repräsentationslernen, wo es darum geht, eine neue Darstellung der Eingabedaten zu finden, die kompakter oder informativer ist als die originale. Die Themen bauen aufeinander auf und man entdeckt die Zusammenhänge, wenn sie vorgetragen werden, dennoch hat jedes Thema seinen abgeschlossenen Bereich.

Und wie schätzt du das Potenzial des Projekts für die Forschung an KI ein?

Es sind schon aktuelle Themen, die wir bearbeiten, also Transparenz sowie Fairness wegen der Gesetze und Regulierungen, die aktuell entworfen werden. Ressourceneffizienz, bezogen auf Green IT, und Datenschutz sind auch sehr gefragte Themen. Deshalb denke ich, es ist relevant, dass hier geforscht wird, dass man diese Eigenschaften bei der Verwendung von Algorithmen verbessert.