Responsible AI

Modus: Einzelarbeit

Typ: Miniprojekt

Beschreibung

Für diese Arbeit vergleichen Sie die Ansätze zur gesetzlichen Regulierung von AI-Systemen in unterschiedlichen Legislaturen (jedenfalls USA, Europa, Österreich) und erarbeiten eigene Vorschläge dafür, wie eine ideale Regulierung Ihrer Meinung nach aussehen sollte, wenn das Ziel eine verantwortliche und gesellschaftsdienliche Verfügbarkeit von AI-Systemen sein soll.

Ablauf

Führen Sie während des gesamten Prozesses ein Forschungstagebuch (siehe Beschreibung im Anhang). Dokumentieren Sie darin die Aktivitäten, Ergebnisse, Hindernisse und Erfolge sämtlicher Schritte Ihrer Arbeit.

1. Recherchieren Sie zum Begriff »künstliche Intelligenz«. Beantworten Sie u.a. konkret die folgenden Fragen:

- Wo kommt der Begriff »artificial intelligence« her? Wer hat Ihn geprägt? Seit wann wird er verwendet?
- Welche Konzepte sind damit gemeint? Identifizieren Sie verschiedene Strömungen bzw. »Geschmacksrichtungen« von AI!
- Gab bzw. gibt es Kritik an diesem Begriff? Gibt es Vorschläge, welche anderen Begriffe stattdessen verwendet werden sollten?
- Suchen Sie eine kurze Timeline von AI, die die verschiedenen Dinge, die unter dem Begriff verstanden werden, in einen historischen Kontext stellt, und inkludieren Sie diese.

2. Verschaffen Sie sich einen Überblick über aktuelle und vergangenen Bestrebungen, AI bzw. AI-Systeme zu regulieren. Eine wesentlichen Quelle dafür ist

https://en.wikipedia.org/wiki/Regulation_of_artificial_intelligence

Ordnen Sie die unterschiedlichen historischen oder aktuellen Entwürfe in die Timeline ein. Vergleichen Sie: welche Fragen standen bei historischen Regulierungsversuchen jeweils im Mittelpunkt.

Hinweis: bitte übersehen Sie nicht die Regulierungen von AI in der DSGVO!

3. Im Anhang dieser Arbeit finden Sie eine Zusammenfassung des Buchs »Automating Inequality« von Virginia Eubanks. Lesen Sie diese Zusammenfassung, und suchen Sie mindestens zwei davon unabhängige Quellen, die mögliche andere gesellschaftliche Problem durch AI beschreiben, und fassen Sie die wesentlichen Probleme in einer Liste zusammen.

4. Überlegen Sie für jedes der Probleme, welche Formen staatlicher Policies (siehe Abschnitt »Mögliche Ansatzpunkte für Policy Thinking« im Slidebook) geeignet wären, diese Probleme zu vermeiden/verhindern/abfedern/etc.

Denken Sie (im Sinne der Diskussion im Panel »Policy Thinking) darüber nach, wie man eher generellere Regeln formulieren könnte, die einen Handlungsrahmen vorgeben, anstatt jede einzelne Missbrauchs-Möglichkeit einzeln zu regulieren.

5. Formulieren Sie für eine der Regelungen, die Sie vorgeschlagen haben, einen entsprechenden Gesetzestext. Schauen Sie sich dazu an, wie Gesetzestexte formuliert werden. Sie können sich auch Entwürfe mit einem LLM Ihrer Wahl produzieren lassen; sie sollen jedoch jedenfalls die finale Version überarbeiten und detailliert prüfen.

Machen Sie in diesem Text (gegebenenfalls) deutlich, welche Textteile von Ihnen verfasst wurden und welche Original durch ein anderes System generiert wurde. Halten Sie sich jedenfalls an die allgemeinen Regeln (aus dem PDF »Denkweisen der Informatik 2023«) zur Verwendung von generative AI in den Abgaben.

6. Diskutieren Sie die Grenzen und Schwächen Ihres Gesetzes. Wo kann das Gesetz umgangen werden? Welche Rolle spielt die Tatsache, dass viele Dinge im (internationalen) Internet passieren und in Österreich nicht reguliert werden können? Welche Probleme können durch das Gesetz entstehen, sowohl für Menschen, für Innovation, aber auch für Organisationen und Einrichtungen?

Abgabe

7. Ihre Abgabe besteht aus Ihrem Forschungstagebuch, eventuell bereinigt um persönliche Einträge, die Sie nicht preisgeben wollen, sowie den Teilen, die oben als Teile der Abgabe genannt sind. Gliedern Sie dieses Dokument bitte sinnvoll, und bemühen Sie sich, ein gut lesbares Layout zu gestalten. Erzeugen Sie dann daraus ein PDF[1] und geben Sie dieses im entsprechenden Abschnitt in TUWEL ab.

Bitte beachten Sie, dass Aufgaben dieses Typs **nach spätestens 2 Wochen abgegeben** werden müssen (ab der Verfügbarkeit dieser Beschreibung), und dann noch eine Review-Phase (1 Woche) durchlaufen. **Ihr selbst gewählter Termin gilt erst für die Endabgabe!**

Zusatz für Endabgabe

Ein wesentlicher Teil Ihrer Endabgabe ist der Abschnitt *Reflexion & Feedback*. Beantworten Sie dabei die folgenden Fragen für die finale Abgabe, also nachdem Sie die Reviews geschrieben/bekommen haben, und ergänzen Sie Ihr PDF um einen entsprechenden Abschnitt:

- Wurde Ihr Verständnis der gewählten Denkweise durch diese Übungsarbeit verändert?
- Glauben Sie, ein nachhaltiges Verständnis der gewählten Denkweise wird Ihnen im Studium oder danach im Beruf helfen?
- Welche Teile dieser Arbeit fanden Sie besonders schwer, welche zu einfach?
- Welche Aspekte dieser Arbeit haben Ihnen gut gefallen, welche würden Sie ändern?
- Was haben Sie bei dieser Arbeit gelernt? Ist diese Art von Übungsformat Ihrer Meinung nach sinnvoll?

---

[1] Beachten Sie bitte, dass inzwischen alle aktuellen Betriebssysteme die Erzeugung von PDFs ohne zusätzliche Software erlauben. Geben Sie keine PDFs ab, bei denen Werbung oder Wasserzeichen von Gratis-Software eingebettet ist. Für Unterstützung befragen Sie bitte die allwissende Müllhalde (das Internet) bzw. https://www.wikihow.com/Convert-a-File-Into-PDF

- Hat das Schreiben der Reviews geholfen, Ihre eigene Arbeit zu verbessern? Falls ja: wie?
- Haben die Reviews, die sie bekommen haben geholfen, Ihre eigene Arbeit zu verbessern? Falls ja: wie?
- Sind Sie mit Ihrer Arbeit zufrieden?

Beachten Sie: Die Antworten auf die Fragen im Abschnitt *Reflexion und Feedback* gehen **nicht** in die Beurteilung Ihrer Arbeit ein!

Beachten Sie bitte die Richtlinie zur Verwendung von generativer AI, die im PDF »Denkweisen der Informatik 2023« zu finden ist. Wesentliche Teile der Arbeit dürfen nicht durch generative AI-Systeme verfasst werden!

Anhang: Forschungstagebuch

Ein Forschungstagebuch ist ein (physisches oder digitales) Medium, in dem Sie den Fortschritt Ihrer Arbeit und Ihre Gedanken dazu bzw. Probleme damit schriftlich festhalten. Damit Ihr Forschungstagebuch dabei helfen kann, zufällige Ideen oder plötzliche Inspirationen notieren können, sollten Sie es immer bei sich haben (das spricht stark für ein digitales Forschungstagebuch). Für die Zwecke dieser Arbeit genügt eine einfache Text-Datei. Jeder Eintrag ist mit Datum und Uhrzeit versehen.

Einträge im Forschungstagebuch werden zB. zu folgenden Anlässen gemacht:

- Artikel gelesen (mit kurzer Anmerkung der Relevanz für Ihr Thema, Auflistung für SIe wesentlicher Punkte)
- Gute Suchbegriffe für Ihr Thema
- In einem Gespräch etwas relevantes gehört, mit Ideen, wie Sie das weiterverfolgen könnten
- Teil der Arbeit geschrieben, mit Einschätzung der Qualität

Sie können auch persönliche Dinge im Forschungstagebuch festhalten, also erfreuliche (zB. Gute Quelle gefunden!) wie unerfreuliche (zB. heute gar nichts weitergegangen, sehr frustrierend). Für die Abgabe des Forschungstagebuchs können Sie Teile, die Sie nicht preisgeben wollen, entfernen.

Anhang: Qualität von Quellen

Ein wesentlicher Teil der Recheche im Internet ist die Einschätzung der Qualität von Quellen. Dazu gibt es, nicht ganz unironisch, viele Hilfestellungen im Internet. Wir haben einige davon für Sie zusammengestellt, denen wir vertrauen:

- Saferinternet, Quellen richtig beurteilen – https://www.saferinternet.at/news-detail/online-quellen-richtig-beurteilen-aber-wie
- Lehrerfortbildung Baden-Württemberg, Arbeitstechnik 2: Überprüfung von Quellen im Internet – https://lehrerfortbildung-bw.de/u_gewi/gk/gym/bp2016/fb5/2_komp/6_vorlagen/3_methode/02_technik2/
- Wer es ganz genau will: Qualitätskriterien für wissenschaftliches Arbeiten – https://soztheo.de/forschung/qualitaetskriterien-fuer-wissenschaftliches-arbeiten/

Anhang: wie man einen wissenschaftlichen Artikel liest

Wissenschaftliche Artikel sind meistens nicht dafür geschrieben, von vorne bis hinten gelesen zu werden. In Ihrem Studium werden Sie aber viele wiss. Publikationen lesen. Da hilft es oft, eine klare Strategie zu haben, wie man das angeht.

Ich habe hier für Sie die Ultrakurzversion zusammengeschrieben. Sie finden nach diesem kurzen Guide einige Links zu längeren Versionen. Dieser Guide gilt für »typische« wissenschaftliche Texte, also solche, die dem üblichen Aufbau folgen.

1. Überfliegen Sie das Abstract. Sie werden dann verstehen, um was es im Artikel geht, warum die Arbeit verfasst wurde, und in wenigen Worten üblicherweise auch, was das Ergebnis der Arbeit war. Das hilft Ihnen, den Rest besser einordnen zu können.

2. Lesen Sie jetzt den letzten Abschnitt des Papers, üblicherweise »Conclusions« oder »Discussion« genannt. Damit sollten Sie jetzt wissen, was die Autor_innen gemacht haben, und warum Sie es gemacht haben. Sie wissen auch, was dabei herausgekommen ist.

3. Der Abschnitt vor den Schlussfolgerungen sind üblicherweise »Results«. Überfliegen Sie diesen Teil, um zu sehen, wie relevant er für Sie ist.

4. Sehen Sie sich die Abbildungen an. In groben Zügen können Sie jetzt verstehen, um was es in diesem Paper geht, und was die Autor_innen gemacht haben. Zugegeben, das wird einfacher, je öfter Sie es machen.

5. Es sollte einen Abschnitt geben, der die Methodologie beschreibt, meistens »Methods« o.ä. Versuchen Sie grob zu verstehen, wie die Autor_innen gearbeitet haben (qualitativ, quantitativ, etc.).

Sie haben jetzt ein gutes Bild davon, um was es geht, und können entscheiden, ob Sie den Rest des Papers auch lesen wollen (zB. weil es relevant oder interessant ist). Eventuell ist aber auch nur noch der Abschnitt »Related Work« (o.ä.) für Sie spannend, weil Sie dort weitere Papers finden, die sich mit derselben oder einer ähnlichen Fragestellung beschäftigen – und vielleicht suchen Sie ja genau solche Arbeiten.

Weitere Guides:

- https://drewdennis.medium.com/how-to-read-scientific-papers-quickly-efficiently-e7030c4018fa
- https://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper
- https://paperpile.com/g/read-scientific-paper/

Anhang: Zusammenfassung von »Automating Inequality« von Virginia Eubanks[2]

Dieses Buch ist eine kritische Untersuchung der Auswirkungen von AI- und algorithmischen Systemen auf soziale Ungleichheit in den USA. Das Buch analysiert, wie Algorithmen und AI-Systeme in verschiedenen Bereichen wie Sozialhilfe, Bildung und Strafjustiz eingesetzt werden und dabei bestehende soziale Ungerechtigkeiten verstärken können.

Eubanks präsentiert Fallstudien, um zu zeigen, wie algorithmische Entscheidungsfindungssysteme dazu neigen, diskriminierende Muster zu reproduzieren, insbesondere gegenüber marginalisierten Bevölkerungsgruppen. Sie argumentiert, dass diese Technologien oft auf bestehenden Vorurteilen basieren und soziale Ungleichheiten weiter zementieren.

Das Buch hebt auch die Tatsache hervor, dass diejenigen, die von diesen Systemen betroffen sind, oft wenig Einfluss auf die Entscheidungsprozesse haben. Menschen, die auf staatliche Unterstützung angewiesen sind, werden beispielsweise durch AI-Systeme oft mit unfairen Sanktionen konfrontiert, ohne angemessene Möglichkeiten zur Überprüfung.

Ein bemerkenswertes Beispiel, das im Buch dokumentiert ist, ist die Untersuchung des Systems für die Vergabe von Sozialleistungen in Indiana, das als "Noch mal überprüfen" (englisch: "Check 'n Go") bezeichnet wird. Hierbei handelt es sich um ein algorithmisches System zur Überprüfung von Anträgen auf staatliche Unterstützung.

In diesem Fall zeigt Virginia Eubanks, wie das System dazu neigt, extrem restriktiv und fehleranfällig zu sein. Viele Menschen, die legitimen Anspruch auf staatliche Hilfe hatten, wurden fälschlicherweise als nicht berechtigt eingestuft und erhielten keine Unterstützung. Das System basierte auf Algorithmen, die oft zu scharf formuliert waren und dazu führten, dass Menschen ungerechtfertigterweise ihrer dringend benötigten Leistungen beraubt wurden.

Besonders besorgniserregend war, dass die betroffenen Personen wenig Möglichkeit hatten, gegen diese Entscheidungen vorzugehen oder Fehler zu korrigieren. Das System erschwerte den Zugang zu Ressourcen und verstärkte somit die bereits bestehende Armut und Unsicherheit.

Dieses Beispiel illustriert eindrücklich, wie algorithmische Technologien in sozialen Diensten dazu neigen können, Menschen in Not unverhältnismäßig zu benachteiligen und bestehende Ungleichheiten zu vertiefen. Es unterstreicht die dringende Notwendigkeit, solche Systeme zu überdenken, um sicherzustellen, dass sie fair, transparent und gerecht sind, insbesondere wenn es um lebenswichtige soziale Unterstützung geht.

Ein weiteres Beispiel im Buch ist ein System, das auf Algorithmen basiert, um das Risiko von Kindesmisshandlung vorherzusagen. Eubanks zeigt, dass diese Art von algorithmischen Systemen oft dazu neigt, stereotype und voreingenommene Modelle zu verwenden, was zu falschen Vorhersagen

---

[2] Dieser Text wurde mit Hilfe von ChatGPT 3.5 am 13.12.2023 erstellt; Die relevanten Prompts waren »zusammenfassung des buchs automating inequality auf rund einer a4-seite«, » zusammenfassung des stärksten beispiels aus dem buch«, » schlussfolgerungen des buchs«, » zusammenfassung des buchs automating inequality, wobei zwei gute beispiele detailliert erklärt sind, und mit einem fokus auf die schlussfolgerungen des buchs«. Die hier inkludierte Zusammenfassung ist eine Kombination aus den Ergebnissen der Prompts und selbst geschriebener Ergänzungen und Übergänge.

führen kann. Dies führt zu einer Überwachung von Familien, die möglicherweise nicht gerechtfertigt ist, und verstärkt das Misstrauen zwischen staatlichen Institutionen und den Gemeinschaften, die sie unterstützen sollen.

Die Schlussfolgerungen von »Automating Inequality« betonen die kritische Notwendigkeit einer sorgfältigen Reflexion über den Einsatz von AI- und algorithmischen Systemen in sozialen Diensten. Hier sind einige zentrale Schlussfolgerungen aus dem Buch:

- Reproduktion von Ungleichheit: Eubanks argumentiert, dass viele AI-Systeme dazu neigen, bestehende soziale Ungleichheiten zu reproduzieren und sogar zu verstärken. Algorithmen können auf Vorurteilen basieren und Diskriminierung perpetuieren, insbesondere gegenüber bereits benachteiligten Bevölkerungsgruppen.
- Mangelnde Transparenz und Rechenschaftspflicht: Das Buch hebt die Gefahr mangelnder Transparenz und Rechenschaftspflicht bei algorithmischen Entscheidungsprozessen hervor. Oft verstehen die betroffenen Personen nicht, wie Entscheidungen getroffen werden, und haben nur begrenzte Möglichkeiten, gegen ungerechtfertigte Entscheidungen vorzugehen.
- Geringe Partizipation der Betroffenen: Eubanks betont, dass Menschen, die von algorithmischen Systemen betroffen sind, oft wenig oder gar keine Möglichkeit zur Teilnahme am Entscheidungsprozess haben. Dies führt zu einer Entmachtung der Betroffenen und einer geringen Kontrolle über ihre eigenen Lebensbedingungen.
- Notwendigkeit für eine ethische Neugestaltung: Die Autorin schlägt vor, dass eine umfassende ethische Neugestaltung von AI-Technologien erforderlich ist. Dies sollte den Schutz der Privatsphäre, die Verhinderung von Diskriminierung und die Gewährleistung einer angemessenen Rechenschaftspflicht umfassen.
- Empowerment und soziale Gerechtigkeit: Das Buch plädiert dafür, dass Technologien so gestaltet werden sollten, dass sie die soziale Gerechtigkeit fördern und die Menschen stärken. Dies erfordert eine verstärkte Beteiligung der Betroffenen, Transparenz in Entscheidungsprozessen und eine ständige Überprüfung der Auswirkungen auf die Gesellschaft.

Insgesamt ruft »Automating Inequality« dazu auf, die sozialen Auswirkungen von AI und algorithmischen Systemen ernst zu nehmen und sicherzustellen, dass Technologie dazu beiträgt, bestehende Ungleichheiten zu verringern, anstatt sie zu verstärken. Es appelliert an Entscheidungsträger, Technologien ethisch zu gestalten und sicherzustellen, dass sie dem Wohl der Gesellschaft dienen.