

# Algorithmic Bias

**Modus:** Einzelarbeit

**Typ:** Guided Research

## Beschreibung

In dieser Guided Research Aufgabe beschäftigen Sie sich intensiv mit Vorurteilen, die in Datenbanken und Softwaretools unbeabsichtigt codiert sind. Dazu lesen Sie zuerst über die Schnittmenge zwischen human und machine Bias, Arbeiten dann mit einem Simulations-System, und Kategorisieren schließlich reale Vorfälle.

## 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. Lesen Sie als Grundlage das folgende Buchkapitel:

[https://dwi.lva.tuwien.ac.at/buch/GR\\_critical\\_kap6.pdf](https://dwi.lva.tuwien.ac.at/buch/GR_critical_kap6.pdf)

Vergleichen Sie den Inhalt des Buchkapitels mit den Inhalten der Intro-Vorlesung und des Slidebooks zu diesem Thema. Verfassen Sie dann eine Email an [peter.purgathofer@tuwien.ac.at](mailto:peter.purgathofer@tuwien.ac.at) mit dem Subject »Ergänzung Critical Thinking: AlgoBias«, in der Sie eine oder mehrere Ergänzungen oder Erweiterungen vorschlagen. Diese Email ist Teil Ihrer Abgabe.

2. Arbeiten Sie dann die *Most Likely Machine* durch

<https://mostlikelymachine.artefactgroup.com/>

und dokumentieren Sie Ihren Prozess mit mehreren Screenshots (Step 6 - alle drei Reihungen; Step 7 - Ihre Vorhersage & das Ergebnis) sowie einer kurzen schriftlichen Ausarbeitung der 6 Fragen in Step 9.

Hinweis: beim Test ist die *Most Likels Machine* auf allen Browsern gut gelaufen, aber es gab Konflikte mit Script- und Adblockern. Deaktivieren Sie diese für diese Übung temporär, um Probleme zu vermeiden.

3. Denken Sie zurück: welche Kriterien waren für Sie wichtig, als Sie die Eigenschaften den drei Entscheidungspunkten zugeordnet haben? Beschreiben Sie Ihre Entscheidungsmuster und suchen Sie eventuelle Fehler in Ihren Überlegungen bzw. Ihrem Vorgehen.

4. Durchlaufen Sie die *Most Likely Machine* ein zweites Mal, und versuchen Sie, die Fehler vom ersten mal zu vermeiden. Recherchieren Sie zudem über Personen, die Sie nicht kennen. Dokumentieren Sie Ihren zweiten Durchlauf mit Screenshots von denselben Schritten wie beim ersten Mal.

5. Reflektieren Sie noch einmal über Erfolg oder Misserfolg Ihres Versuchs. Welche Informationen (vom ersten Durchlauf, von der Recherche, etc.) haben Ihnen geholfen? Welche haben Sie gehindert?

6. Beschäftigen Sie sich mit Abschnitt 3 der *Grundlagen zu Bias & Fairness in KI-Systemen*:

<https://bias-and-fairness-in-ai-systems.de/grundlagen/>

Überlegen Sie, ob Ihre Erfahrung mit der *Most Likely Machine* sich als einer der 11 Bias-Arten einordnen lässt. Begründen Sie für jede der Bias-Arten kurz, warum Sie hier zutrifft oder nicht zutrifft.

7. Sehen Sie sich die folgenden drei Beispiele von algorithmischer Diskriminierung an:

- <https://gizmodo.com/why-cant-this-soap-dispenser-identify-dark-skin-1797931773>
- <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>
- <https://www.science.org/doi/10.1126/science.aax2342>

Identifizieren Sie, um welche Bias-Kategorien aus *Grundlagen zu Bias & Fairness in KI-Systemen* es sich in diesen drei Fällen handelt, und begründen Sie kurz.

## Abgabe

8. 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<sup>1</sup> und geben Sie dieses im entsprechenden Abschnitt in TUWEL ab.

Bitte beachten Sie, dass Aufgaben dieses Typs spätestens **2 Wochen nach der Verfügbarkeit** dieser Beschreibung abgegeben werden müssen, 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:

- Wie 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?
- 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?

---

<sup>1</sup> 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>

**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.

Bitte führen Sie das Forschungstagebuch in digitaler Form; handschriftliche Abgaben werden nicht akzeptiert.

## **Anhang: Qualität von Quellen**

Ein wesentlicher Teil der Recherche 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/](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/>