

## Einführung kritische Data Science: Übersicht einschlägiger Terminologie

---

- *privilege hazard*: Diskriminierung findet häufig nicht bewusst, sondern basierend auf Ignoranz/ Unfähigkeit Diskriminierungen zu erkennen statt. → fehlende Sensibilisierung (siehe auch *plainness*)
- *plainness* (nichts-aussagende Art) der (weißen) Mehrheitsgesellschaft, welche Diskriminierung nicht erfährt, und kein Problembewusstsein hat, sich aber auch aus der Verantwortung entzieht.
- *gender data gap*: Das Konzept richtet sich gegen die fehlende Datenerhebung von Diskriminierung oder Schlechterstellung von Frauen (siehe auch *The Library of Missing Datasets*), aber auch die Unterrepräsentation von Frauen in Datensätzen. Beispielsweise beim Aufbau von Autosicherheit (Airbag, Gutposition...) nach Größe und Gewicht von Männern.
- *New Jim Code* (nach Benjamin Ruha): Technologie wird vermeintlich als neutral und faktenbasiert wahrgenommen: "[T]he employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era" - Ruha, *Race after Technology*  
Wortherkunft: *Jim Ceow*: Katoon Figur (1832) welche PoC verspottete und marginalisierte → Mit dem black power movement findet Aneignung und Ausweitung des Begriffs statt, um rassistisch motivierte Unterdrückung, Marginalisierung oder Segregation zu benennen, welche durch Institutionen, Verhalten.. "Weiße Vorherrschaft" stützt.
- *Data Sharing/ Data Fusion*: Austausch persönlicher Daten zwischen (staatlichen) Institutionen. Negatives Beispiel: "[If someone is] denied a bank loan despite having a high income and no debt, because the lender had access to her health file, which showed that [he/she] had a tumor." - Ruha, *Race after Technology*
- *tech addiction*: psychologische Gefahren durch Mediennutzung: aktives Forschungsfeld von Sucht, über Persönlichkeitsentwicklung bis Einfluss auf die Selbstwahrnehmung
- *Targeting/ personalized content*:
  - 1. Problem - Illusion von Repräsentation: "[A]s algorithms become more tailored [(personalized)], the public will be given the illusion of progress." - Ruha, *Race after Technology*
  - 2. Problem - Undurchsichtigkeit der Algorithmen (siehe auch *blackbox*, Anwendung von Targeting ist Rezipient\*innen nicht bewusst)
- *black box*: Undurchsichtigkeit sowohl über die Entstehung von Algorithmen, als auch die Funktionsweise von neuronalen Netzen (siehe auch *understanding AI* oder *bias training*)
- *oversampled*: Beschreibt eine nicht repräsentative Datenlage, auf Basis derer das Training des Algorithmus stattfindet. Beispiel: Amazon hat in der Vergangenheit einen Algorithmus zur Unterstützung von Einstellungen verwendet, welcher Bewerberinnen benachteiligt hat. Grund dafür war, dass in den Datensätzen, bestehend aus vergangenen Einstellungen Männer gegenüber Frauen überrepräsentiert (oversampled) waren.
- *scarcity bias*: Die Überzeugung, dass fehlende Arbeitskraft durch Automatisierung kompensiert werden kann, wobei die Gefahr besteht, dass Fehler oder Benachteiligung durch die Automatisierung nicht auffallen.
- *racial coding*: Kontinuität und Reproduktion von Rassismus durch Algorithmen, kann auch indirekt stattfinden, beispielsweise über den Parameter Postleitzahl

D'Ignazio, Catherine und Klein, Laure. Data Feminism. 2020. Data Feminism. Einführung und Kapitel 1.

Ruha, Benjamin (2019): *Race After Technology: Abolitionist Tools for the New Jim Code*.