

Gender-Aspekte der Data Science

**Sitzung 2b: Themensuche; Daten extrahieren;
Forschungsdatenmanagement**

16.04.2026

Dozentin: Dr. Daria Tisch

Roadmap

	Einführung: Gender & Data Science
19.03.2026	Einführung in Wikipedia und Wikidata; Arbeiten mit R; Bildung von Projektteams
16.04.2026	Der männliche Prototyp Themensuche; Daten extrahieren; Forschungsdatenmanagement
	Datenfeminismus
30.04.2026	Versionskontrolle, Datenanalyse
	Geschlechterbezogene Verzerrungen im gesamten Datenprozess
28.05.2026	Projektvorstellungen: Zwischenstand; Datenanalyse; Peer Feedback
	Thematische Vertiefung: Gender Bias im Gesundheitsbereich
11.06.2026	Projektvorstellungen: Zwischenstand; Datenanalyse; Peer Feedback
18.06.2026	Online Fragestunde
	Thematische Vertiefung: Algorithmische Diskriminierung
25.06.2026	Gemeinsame Erarbeitung eines Posters zu Gender & Data Science
09.07.2026	Projektpräsentationen & Reflexion

Teamprojekt = Studienarbeit = Prüfungsleistung



- **Teamprojekte zu einem Gender-Bias oder einem Gender-Data-Gap auf Basis von Wikipedia oder Wikidata Daten**
- **Alle Mitglieder eines Teams sind für die Arbeit des Teams verantwortlich.**
- **Jedes Mitglied des Teams erhält dieselbe Note.**
- **Bei der mündlichen Präsentation muss jeder Studierende in der Lage sein, das gesamte Teamprojekt vorzustellen und Fragen dazu zu beantworten.**

Formalien

- **2.000 – 3.000 Wörter (ohne Literaturverzeichnis, ohne Code)**
- **Abgabe: 09.07.2026**
- **Abgabe als Link zu einem github repository**
 - **Code und Manuskript (PDF) oder**
 - **Quarto Dokument (Text und Code kombiniert)**
- **Teams: max. 4 Personen**
- **Deutsch oder Englisch**

Bewertung (100 Punkte gesamt)

Inhalt und wissenschaftliche Qualität (45 Punkte)

- Klare Fragestellung und Motivation (10)
- Roter Faden, Logische Struktur und Verständlichkeit (10)
- Richtiges Zitieren (5)
- Kritische Reflexion und Einordnung der Ergebnisse (20)

Code und Reproduzierbarkeit (25 Punkte)

- Strukturierter, gut kommentierter Code (5)
- Vollständige Reproduzierbarkeit (15)
- Sinnvolle Organisation des GitHub-Repositories (5)

Datenanalyse und Methode (25 Punkte)

- Angemessene Auswahl und Anwendung von Methoden (10)
- Nachvollziehbare Datenaufbereitung (5)
- Nachvollziehbarkeit der Analyseschritte (5)
- Visualisierungen und Tabellen (5)

Mündliche Präsentation (5 Punkte)

- Verständliche Darstellung des Projekts
- Fähigkeit, Fragen zum gesamten Projekt zu beantworten

Mindestanforderungen

Die Prüfung gilt als **nicht bestanden**, wenn:

- kein funktionsfähiger Code abgegeben wird
- wesentliche Teile des Projekts fehlen
- einzelne Gruppenmitglieder das Projekt nicht erklären können

Struktur

- **Einleitung**
 - Motivation
 - Fragestellung des Projekts
 - Theoretischer Hintergrund und bisherige Literatur
- **Daten und Methode**
 - Beschreibung der verwendeten Daten (Wikipedia/Wikidata + optional weitere Daten)
 - Stichprobe
 - Datenaufbereitung
 - Beschreibung der analytischen Vorgehensweise + Begründung der gewählten Methoden
- **Ergebnisse**
 - Darstellung zentraler Analyseergebnisse, Visualisierungen
- **Diskussion**
 - Interpretation der Ergebnisse, Einordnung im Kontext von Gender-Bias / Gender-Data-Gap
 - Reflexion von Limitationen
 - Fazit
- **Literaturverzeichnis**

KI Nutzung

- Siehe: **Wie dokumentiere ich KI-Nutzung?**
- https://leko.service.th-nuernberg.de/owl/artikel/ki_nutzung_in_wissenschaftlichen_arbeiten_dokumentieren/#wie-dokumentiere-ich-ki-nutzung
- **Variante 2: Detaillierte Dokumentation!**

Ideen für Projekte, die Gender-Bias auf Wikipedia untersuchen:

- Gender-Bias in der Darstellung von Berufen auf Wikipedia: Wie stark spiegeln Wikipedia-Artikel typische Geschlechterrollen in Berufsbildern wider?
- Gender-Bias in medizinischen Wikipedia-Artikeln: Spiegelt Wikipedia gender-spezifische Unterschiede in Gesundheitsinformationen korrekt wider?
- Gender-Bias in der politischen Repräsentation auf Wikipedia: Sind Politikerinnen auf Wikipedia unterrepräsentiert oder anders dargestellt als ihre männlichen Kollegen?
- Gender-Bias in der Sichtbarkeit von Wissenschaftlerinnen und Wissenschaftlern auf Wikipedia:
- Gender-Bias in historischen Artikeln: Werden Frauen und Männer in historischen Kontexten auf Wikipedia unterschiedlich dargestellt?
- Gender-Bias in Artikeln zu Filmen, Serien und Musiker*innen: Welche Geschlechterrollen werden in Wikipedia-Artikeln über Medien und Popkultur reproduziert?
- Gendergerechte Sprache in Wikipedia-Artikeln
- Gender-Bias in der Verlinkung von Personen auf Wikipedia: Werden männliche CEOs stärker verlinkt und sind damit sichtbarer auf Wikipedia?

Ideen für Projekte, in denen Wikipedia nur als Datenbasis genutzt wird:

- **Bias in Datenerhebung: alle relevanten Artikel oder nur eine Teilmenge (z. B. „Top 100“). Welche Gruppen werden durch die Auswahl über- oder unterrepräsentiert? Welche Verzerrung entsteht?**
- **Bias in Datenselektion: filtern nach Kriterien (z. B. Artikel mit Bildern, Artikel >500 Wörter, nur bestimmte Länder). Wie erzeugen Filter, Kategorien oder Auswahlkriterien Gender-Bias?**
- **Bias in der Datenauswertung: Wie können Analyse-Methoden unbewusst Bias erzeugen?**