ICT20-079 - Humanized Algorithms: Identifying and Mitigating Algorithmic Biases in Social Networks

Zusammenfassung

Viele Ranking- und Social-Recommender-Algorithmen basieren auf Daten aus sozialen Netzwerken. Zum Beispiel, Social-Media-Plattformen wie Twitter oder LinkedIn nutzen die Informationen in sozialen Netzwerken, um Personen zu ranken und empfehlen den Benutzern neue soziale Links. Diese Netzwerke, die von Menschen generiert werden, werden angetrieben von grundlegende gesellschaftliche Mechanismen wie Popularität und Homophilie und enthalten oft unterschiedliche soziodemografische Merkmale von Menschen. Diese Attribute spielen eine wichtige Rolle bei der Art und Weise, wie Individuen interagieren mit anderen und bestimmen damit die Struktur von Netzwerken. Noch wichtiger ist die Struktur von Netzwerke spielen eine entscheidende Rolle bei dynamischen Prozessen in Netzwerken wie Informationsverbreitung, Bildung und Entwicklung von Vorurteilen, Normen und Kultur. Über die Wirkung von Netzwerkenstrukturen auf Algorithmen und inwieweit Methoden des maschinellen Lernens soziale Vorurteile verstärken und praktische Ansätze zur Minderung algorithmischer Verzerrungen ist jedoch nur sehr wenig bekannt. Das übergeordnete Ziel dieses Projekts ist es, die Rolle der Empfehlungs- und Rankingalgorithmen zur Verstärkung von Verzerrungen in sozialen Netzwerken mit besonderem Fokus auf Minderheiten. Die Ergebnisse dieses Projekts werden dazu beitragen, bessere Algorithmen zu entwickeln, die Minderheiten gegenüber gerecht werden.

Wissenschaftliche Disziplinen:
Statistical physics (20%) | Social psychology (40%) | Information technology (20%) | Data science (20%)

Schlagworte:
Algorithmic Bias, Social Networks, Minorities, Ranking algorithms, Recommender Systems

---

Principal Investigator:	Fariba Karimi
Institution:	Complexity Science Hub Vienna
Co-Principal Investigator(s):	Markus Strohmaier (RWTH Aachen) Julia Anna Koltai (Center for Social Sciences)

Status: Abgeschlossen (01.01.2022 – 28.02.2026)
GrantID: 10.47379/ICT20079

Weiterführende Links zu den beteiligten Personen und zum Projekt finden Sie unter
https://www.wwtf.at/funding/programmes/ict/ICT20-079/