ICT20-058 - Interpretability and Explainability as Drivers to Democracy

Zusammenfassung

Immer häufiger basieren Entscheidungen mit beträchtlichen Auswirkungen auf uns als Individuen und unsere Gesellschaft auf Vorhersagen komplexer mathematischer Modelle. Beispielsweise plant das österreichische Arbeitsmarktservice die Verwendung von intelligenten Modellen („AMS-Algorithmus“) zur Einstufung der Jobchancen von Arbeitslosen; und die Strategie der Bundesregierung im Umgang mit der Coronavirus-Pandemie basiert unter anderem auf komplexen Simulationsmodellen, die die Verbreitung des Virus vorhersagen. Daher wäre es in einer demokratischen Gesellschaft von großer Bedeutung, die Verwendung und Details dieser Modelle öffentlich zu diskutieren und zu beurteilen. Dies ist derzeit allerdings herausfordernd - die genaue Funktionsweise solcher Modelle ist sowohl für Laien als auch Experten oft schwer greifbar und die Art und Weise der Verwendung in Entscheidungsfindungsprozessen wird oft nicht offengelegt. In diesem Projekt verfolgen wir die Idee, dass Modelle mit speziellen Eigenschaften (Interpretierbarkeit und Erklärbarkeit) und geeignete Kommunikation dieser, sowie die Offenlegung deren Rolle zur Entscheidungsfindung, einen breiteren und qualitativeren öffentlichen Diskurs bezüglich der Nutzung dieser Modelle ermöglichen. Insbesondere entwickeln wir interpretierbare und erklärbare Modelle, Leitfäden für die Kommunikation von durch solche Modelle unterstützten Entscheidungsfindungsprozessen und reflektieren die Implikationen für eine (partizipative) Demokratie.

Wissenschaftliche Disziplinen:
Machine learning (30%) | Visualisation (20%) | Philosophy of technology (50%)

Keywords:
models; data-driven modeling and decision making; intelligent and complex models; ethics; transparency; explainability; interpretability; democracy; visual data analysis

Principal Investigator:	Sebastian Tschiatschek
Institution:	University of Vienna
Co-Principal Investigator(s):	Torsten Möller (University of Vienna) Mark Coeckelbergh (University of Vienna)

v.l.n.r. Sebastian Tschiatschek, Mark Coeckelbergh, Torsten Moeller ©BarbaraMair

Status: Laufend (01.10.2021 - 30.09.2025)

Weiterführende Links zu den beteiligten Personen und zum Projekt finden Sie unter
https://wwtf.at/funding/programmes/ict/ICT20-058/