

ICT20-058 - Interpretability and Explainability as Drivers to Democracy

Abstract

An increasing number of decisions with significant societal impact is supported by intelligent and complex models (ICMs). For instance, the government's strategy for fighting the Covid pandemic is informed by complex models forecasting the spread of the virus. As these models thereby significantly impact our lives, there should be room to debate their usage and merits broadly within a democratic society. However, enabling this is non-trivial because (a) the precise workings of ICMs are typically hard to comprehend – for laymen as well as experts, and (b) it is often not made public how these systems are used within a political decision making process. This hinders the electorate's evaluation of the political decision making and the made decisions with respect to an ICM's usage: Whether and how to use an ICM? Which ICM to use? We believe that the electorate can be better integrated into the democratic decision processes by building interpretable and explainable ICMs and informing the public about how these systems are used. Therefore, we propose to study the following research questions: (i) How can we build interpretable and explainable ICMs to improve transparency of decision making? How does this affect trust in made decisions? (ii) How to communicate decisions supported by ICMs to the public and how to integrate the public in the decision making process? By answering these questions we innovate our understanding of ICMs with the goal of supporting democratic processes.

Scientific disciplines:

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: Ongoing (01.10.2021 - 30.09.2026)

GrantID: 10.47379/ICT20058

Further links to the persons involved and to the project can be found under
<https://www.gmbh.wwtf.at/funding/programmes/ict/ICT20-058/>