

Controlling the Jekyll and Hyde of Artificial-Intelligence Assisted Information Architectures: Accepting Customization while Resisting micro-targeted Manipulation

Initiative: Künstliche Intelligenz – Ihre Auswirkungen auf die Gesellschaft von morgen

Ausschreibung: Künstliche Intelligenz – Ihre Auswirkungen auf die Gesellschaft von morgen - Planning Grant

Bewilligung: 06.02.2019

Laufzeit: 1 Jahre

People's experience on the Internet is shaped by artificial intelligence (AI). Search engines customize results based on geographical location, entertainment sites recommend books based on presumed preferences, and ads on social media are tailored to users' inferred personalities. Intelligent customization assists with our daily chores, but AI also opens the door to manipulation and subterfuge: Micro-targeted political messages can exploit people's personal vulnerabilities without their knowledge and the opportunity for rebuttal. There is much concern that microtargeting affected the Brexit referendum and the U.S. election in 2016. How do we harness AI's power for the public good while protecting the public from manipulation? The cognitive, ethical, and legal implications of AI-assisted information architectures (AIAIA) are poorly understood, and scholarly opinions range from fears about the demise of democracies to assurances that online political fractionation is harmless. The key idea is to "reverse-engineer" manipulative customization to provide users with the cognitive and technological tools - an approach we call technocognition - to resist manipulation while remaining open to useful customization. The planning grant allows us to pursue 4 steps to guide future research: (1) A conceptual scoping of ethical issues of AIAIA. (2) An assessment of the public's knowledge of, and attitudes towards AIAIA. (3) Cognitive and technological scoping research to establish the technocognition approach's feasibility. (4) Consolidation of the team during preparation of the full application.

Projektbeteiligte

Prof. Dr. Ralph Hertwig

Max-Planck-Institut für
Bildungsforschung
Zentrum für Adaptive Rationalität
Direktor
Berlin

Prof. Dr. Tina Eliassi-Rad

Northeastern University
Khoury College of Computer Sciences
Network Science Institute
Boston, MA
USA

Prof. Stephan Lewandowsky

University of Bristol
School of Experimental Psychology
Bristol
Grossbritannien

Dr. Stefan Herzog

Max-Planck-Institut für
Bildungsforschung
Zentrum für Adaptive Rationalität
Berlin

Prof. Dr. Awais Rashid

University of Bristol
Faculty of Engineering
Computer Science
Bristol
Grossbritannien

Open Access-Publikationen

How behavioural sciences can promote truth, autonomy and democratic discourse online

Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools

Artificial intelligence in online environments: Representative survey of public attitudes in germany