AI, HUMAN VALUES AND MEANINGFUL HUMAN CONTROL

22-23 June | CST Bonn

CONFERENCE PROGRAM

ZERTIFIZIERTE KI
Qualität sichern. Fortschritt gestalten.

CENTER FOR
SCIENCE & THOUGHT

UNIVERSITÄT BONN

June 22

9:30 AM – INTRODUCTORY REMARKS

10:00 AM – PROF. JOCELYN MACLURE, McGill University
AI, Cognitive Scaffolding and Social Reasoning

11:00 AM – COFFEE BREAK

11:15 AM – THEORETICAL FRAMEWORK

Claudia Schon, Werner Moskopp
Towards Benchmarking Moral Decisions in AI Systems

José Antonio Pérez-Escobar, Deniz Sarikaya
Wittgensteinian precepts for AI safety

Carsten Ochs
Negotiating Hypernormativity: A Situational Analysis of the AI Arena

12:45 AM – LUNCH BREAK

2:00 PM – DIGITAL NUDGING

Philippe Verreault-Julien
Ethical Nudging with Opaque Recommender Systems?

Marius Bartmann
Digital Nudging and Personal Autonomy

Aaron Schultz
Distraction and Attentional Freedom

Charlie Kurth
AI-Driven Emotion Nudges: Inviting a Wicked Problem?

3:40 PM – COFFEE BREAK

4:00 PM – PROF. BEATE ROESSLER, University of Amsterdam
Why Robots Can't Get Ill: On the Concept of the Digital Human Being

5:00 PM – AI & VALUE ALIGNMENT

Leonard Dung
Current cases of AI misalignment and their implications for future risks

Michael Cannon
Two Kinds of Control Problems

9:30 AM – PROF. JOERN LAMLA, University of Kassel
Artificial Intelligence as a Hybrid Life Form. On the Critique of Cybernetic Expansion

10:30 AM – DEMOCRACY & PARTICIPATION

Blair Peruniak
Artificial Intelligence and Workplace Democracy

Bjorn Fasterling, Gianclaudio Malgieri, Geert Demuijnck
Participating in – before contesting – data-driven activities? Addressing underrepresentation and power imbalances of vulnerable people

11:30 AM – COFFEE BREAK

11:45 AM – POLICY & REGULATION

Juan-Pablo Bermúdez
Ethics in the Grey Zone: Guidelines for Autonomy – Supportive Automated Influence

Monique Munarini
How soft compliance mechanisms aligned with due diligence principles can be a tool in eliminating societal harms from the deployment of AI driven systems, and at the same time promoting the inheritance of human values?

Matteo Fabbri
An ethical perspective on the new transparency requirements for recommender systems set by the Digital Services Act

1:00 PM – LUNCH BREAK

2:15 PM – DILETTA HUYSKES, University of Milan
Enforce ethics & rights, control digital technologies: some empirical insights from civil society, auditing and research

3:15 PM – AUTONOMY & TRANSPARENCY

Jose Luis Guerrero
AI opacity vs patient's autonomy in decision-making

Marten Kaas
A Perfect Technological Storm: Artificial Intelligence and Moral Complacency

4:15 PM – COFFEE BREAK

4:30 PM – MEANINGFUL HUMAN CONTROL

Alessio Tartaro
Transfer-of-control requirements for meaningful human control: reaffirming human autonomy against AI-induced heteronomy

Sabine Thürmel
First steps towards Meaningful Human Control of Generative AI

Martina Philippi
Talking to robots: The role of human value in AI-driven communication systems

6:00 PM – CONCLUSIONS & WINE RECEPTION