The Ethical Gravity Thesis: Marrian Levels and the Persistence of Bias in Automated Decision-making Systems

Atoosa Kasirzadeh (Australian National University & University of Toronto; atoosa.kasirzadeh@anu.edu.au)

Colin Klein (Australian National University; colin.klein@anu.edu.au)

Full paper: https://doi.org/10.1145/3461702.3462606

Australian National University

AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

Our contribution

Offering a groundwork for a systematic analysis of the social, political, and ethical implications of automated decision-making. How?

1. Adapting David Marr's framework to discuss different levels of analysis of automated decision-making

2. Use our adapted framework to argue for the *Ethical Gravity Thesis:*
   The ethical considerations that appear at a higher Marrian level of analysis cannot be overcome, in any robust way, by interventions at lower levels.

Two arguments for the Ethical Gravity Thesis

The Realization Argument: realization relationships between levels (the arrows in the figure above) require lower levels to faithfully perform the task as specified at higher levels. This means that an ethical problem that can be identified at one level should be expected to be preserved by all lower levels.

The Institutional Argument: active social and political forces make even unsuccessful realizations---that is, realizations that only fulfill some of the conditions set by higher levels---relatively infrequent and difficult to maintain.

Adapting Marr's framework to discuss different levels of analysis of automated decision-making

Marr	Level	Algorithmic Decision-Making
"Everyday experience, coarse psychophysical demonstrations"	Functional	Overall business or political goal
"Computational problem"	Computational	Mathematical formulation of the problem
"Specific representation" and "Specific algorithm"	Algorithmic	Specific classification/ learning/ etc algorithm and representation of data
"Specific neural mechanism"	Implementational	Details of programs, data collection, auditing, etc

Case Studies

• Criminal justice decisions	Functional
• Healthcare expenditure	Computational
• Facial recognition	Implementational

Impact

- Framework for locating social and ethical problems with automated decision-making at the right level of analysis
- Movement away from mere technological fixes—some problems have no technological solution!
- Further research on the role of upward constraints as well as downward realization