

Krzysztof Dolega

Better Stewards of Our Nature: Can AI Foster Morality?

Debates on the ethics of AI have focused primarily on the dangers that artificial agents and machine learning systems pose for different aspects of human life, such as privacy, autonomy, and moral agency. Relatively little attention has been paid to the positive impact that AI could have on changing human behavior. This article argues that the known constraints present in current AI technologies, such as their susceptibility to perpetuating biases hidden in the training data, could be harnessed in service of individual and societal improvement. To support this claim, I investigate two possible scenarios in which available AI technologies could be applied to such ends.

On the individual level, I consider a system which builds a virtual persona of the user by collecting their behavioral data. Such profile could then be used for self evaluation, e.g. by revealing hidden biases inconsistent with one's self-image.

On the societal scale, I consider the use of search and recommendation algorithms, which have been shown to not only facilitate purchases, but also create and shape consumer demand. Given their ubiquity, recommendation systems could be used to bias consumer preferences towards more ethical choices, e.g. by promoting environmentally friendly products.

Although both of the proposed scenarios could lead to beneficial results, they are also weighted by negative implications, raising concerns about user privacy and autonomy. I address these concerns by pointing out that neither of the proposals goes beyond the already normalized and accepted dangers of AI technologies.