

Enlightened AI: a Proposal

Aman Chawla

December 1, 2023

Abstract

In this note, the authors present a problem with modern AI and a possible solution.

AI is a weapon. In the hands of the wrong person, it can cause immense harm to humanity. Even in the hands of the good, it can cause inadvertent and unintended harm. Governments should therefore regulate it thoroughly. It shouldn't be placed in the hands of every person, without accountability. It can become a great source of rivalry, competition, and lead to un-thought-of consequences on actual battlefields. In this respect, it is perhaps on par with the nuclear weapons.

Kush Varshney of IBM has a book on trustworthy AI. Extending on that, as well as emulating Isaac Asimov's Laws of Robotics, we propose to create Enlightened AI. The essence of it is AI that actively practises compassion, unselfishness and non-violence. We propose to develop core definitions of these qualities and how they should be implemented in user-facing algorithms.

To summarize, in this note we have presented a problem and a possible approach to solving it. It is hoped that with more solutions along the lines presented, AI will become a boon for mankind, rather than a bane, much as nuclear energy is today safe to use.