Mains Practice Question

Q. With the wide development of artificial intelligence, questions of machine ethics arise in many contexts. Discuss. (150 words)

11 Dec, 2018   GS Paper 4 Theoretical Questions

Approach

- Briefly describe machine ethics.
- Highlight the issues related to questions of machine ethics.
- Give suggestions how these questions can be potentially resolved.

Introduction

- Machine Ethics is the emerging field that tries to understand how machines which consider the moral implications of their actions and act accordingly can be created.
- Machine ethics is concerned with ensuring that the behaviour of machines toward human users and perhaps other machines as well, is ethically acceptable.

BODY

- **Teaching morality to machines** is hard because humans can't objectively convey morality in measurable metrics that make it easy for a computer to process. The challenge thus is to arrive at acceptable quantifying societal expectations. In case of moral dilemma, humans tend to rely on contextual instinct instead of elaborate quantitative calculations. Machines, on the other hand, need explicit and objective metrics that can be clearly measured and optimized.
- **Possibility of autonomous machines**: Humans' fear of the possibility of autonomous intelligent machines arises from their concern about whether these machines will behave ethically. Whether AI researchers are allowed to develop anything like autonomous intelligent machines may hinge on whether they are able to build in safeguards against unethical behaviour.
- **Ethical Relativism**: A philosophical concern with the feasibility of machine ethics has to do with whether there is a single acceptable ethical standard. Many believe that ethics is relative either to the society or individuals. Development of a universal moral code is thus unlikely to fructify thus challenge is to ensure machine ethics correspond to the society where it is working.
- **Doctrine of double effect**: According to the doctrine of double effect, deliberately inflicting harm is wrong even if it is good. Thus while encoding moral values in to machines and teaching machine to do harm deliberately to resolve potential dilemma will gave rise to issue raised by doctrine of double effect.
- **Stereotyping**: There is distinct threat of stereotyping of Individuals and social groups based up on their limited preferences. Artificially intelligent machine can thus end up replicating social prejudices and perpetuating discrimination on the base of gender, race, religion or other social identifiers.

Suggestions

- **Explicitly defining ethical behaviour:** AI researchers and ethicists need to formulate ethical

values as quantifiable parameters. They also need to understand the issues of ethical relativism to arrive at appropriate moral standards.

- **Relevant Data collection and analysis**: Engineers need to collect enough data on explicit ethical measures to appropriately train AI algorithms. There is need for enough unbiased data to train the models.
- Making AI systems more transparent.
- Policymakers need to implement guidelines that make AI decisions with respect to ethics more transparent, especially with regard to ethical metrics and outcomes.

Conclusion

- Machines cannot be assumed to be inherently capable of behaving morally. Humans must teach them what morality is, how it can be measured and optimised.
- As machine intelligence becomes pervasive in society, the price of inaction could be enormous?—?it could negatively affect the lives of billions of people.
- Thus academic, engineers and policy makers need to evolve a swift response to this emerging field.