

Exploring moral behavior of thinking machines

Abhi Agarwal

The question that arises from this is - what

My proposal is to begin by defining what a moral agent and an artificial moral agent is - the general idea of Moral Machines.

What does it mean for morality to evolve?

Is morality learned? Does evolution of morality mean we take principles we have learnt of morality and teach/embed that within computers?

Since we could be able to program a computer to make certain deliberations, what should the deliberations be?

What goals would we program a computer to fulfill?

References

- [1] Bostrom, Nick. *Superintelligence: The Coming Machine Intelligence Revolution*. Oxford: Oxford UP, 2013. Print.