Can morality by taught to an AI?

Abhi Agarwal

How do we define morality?

- Distinction between right and wrong. It is the determination of what should be done and what should not be done.[1]

- Biblically, morals are derived from God's character and revealed to us through the Scriptures.[2]

- Law or a legal system is distinguished from morality or a moral system by having explicit written rules, penalties, and officials who interpret the laws and apply the penalties.[3]

- The term comes from both Latin and Greek. In Latin it is "mores", and in Greek it is "ethos". Each derives their meaning from the idea of custom.[4] Therefore morality can refer to:

  - Customs
  - Precepts
  - Practices of people and cultures
  - Virtues, values, and principles of people

- The concept of being moral seeks to establish principles of right behavior. It can be used to serve as a guide for individuals and groups.

What is it to be a moral person?
What is the nature of morality?
Why do we need morality?
What function does morality play?
How do I know what is good?
What do morals depend upon?

---

[1] http://carm.org/dictionary-morality
[2] http://carm.org/dictionary-morality
[3] http://plato.stanford.edu/entries/morality-definition/
[4] http://www.slideshare.net/dborcoman/chapter1-9042561

- Morals differ among cultures, and there are morals that are relative, i.e., dependent upon situations and context.[5]

What can the term morality be used for?

- Descriptively to refer to some codes of conduct put forward by a society (some other group, such as a religion or accepted by an individual for her own behavior)[6]

- Normatively to refer to a code of conduct that, given specified conditions, would be put forward by all rational persons.[7]

What are the moral characteristics?

- Being Honest, Truthful, Trustworthy

- Having Integrity

- Being Caring/Compassionate/Benevolent

- Doing One's Civic Duty

- Having Courage

- Being Willing to Sacrifice

- Maintaining Self-Control

- Being just and fair

- Being Cooperative

- Being Persevering/ Diligent

- Keeping Promises

- Doing no harm

- Pursuing excellence/takes pride in work

- Taking personal responsibility

- Having Empathy

- Benefiting others

---

[5]http://carm.org/dictionary-morality
[6]http://plato.stanford.edu/entries/morality-definition/
[7]http://plato.stanford.edu/entries/morality-definition/

- Having Respect for others

- Having Patience

- Being Forgiving

- Making Peace

- Having Fidelity/Loyal

- Respecting Autonomy

- Being Tolerant

- Having Self-respect

- Competitiveness

- Valuing Life

Observations

- When the set of morals we follow are written down do they become rules? Is law different to morality because it is written down?

- Do we need a sense of morality to make a judgement?

- Are moral principles absolute?

- How do we investigate which values and virtues are important for a worthwhile life in society?