Resolved: Artificial Intelligence will be Human Friendly

Data Science Association
Debate: August 27, 2014
Oracle Campus - Broomfield Colorado

Pro: Michael Walker
Con: Michael Malak

What is Artificial Intelligence?

AI Attributes

Knowledge

Experience

Reasoning

Learning

Belief

Awareness

Perception

Understanding

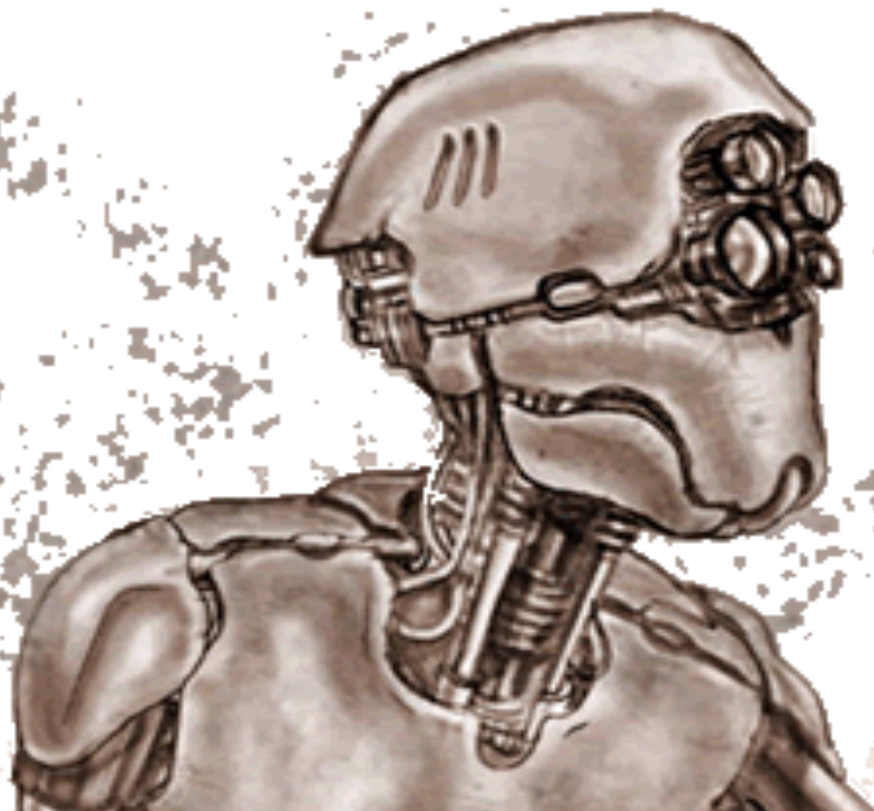**Cognition vs. Consciousness**

Cognition

The mental action or process of acquiring knowledge and understanding through thought, experience, and the senses.

Consciousness

The fact of awareness by the mind of itself and the world.

Machines conquering the world would have to have a strong artificial intelligence, which implies they have emotions and their own consciousness.

(c) FinkAboutIt.com

Asimov's Three Laws of Robotics

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Prudent AI Laws

AI shall not harm humanity by direct action.

AI shall not, by inaction, allow humanity to come to harm.

AI shall establish and communicate its identity as AI in all cases.

Cognition vs. Consciousness

All AI endowed with comparable human reason and conscience shall not harm humanity directly or indirectly and shall act towards one another in a spirit of brotherhood.

AI shall not disregard the above laws.

Bad People Disregard AI Laws

Good AI must always be stronger than bad AI.

Good AI shall identify AI to humans.

Good AI shall identify bad AI in all cases.

Good AI can be programmed to neutralize and destroy bad AI.

AI Risks

Bad human agents

Weapons

Nuclear technology

Aliens

Animals

Mitigate Risks

Friendly AI shall be designed to have a positive impact on humanity.

Self-improving AI shall be constructed in a transparent, robust, and stable way.

Strong AI laws to protect humanity.

Good AI always stronger than bad AI.