

Name: \_\_\_\_\_  
ID: \_\_\_\_\_

**ECS 089: Ethics in an Age of Technology**  
**Quiz 7**  
**November 13, 2024**

***1) In 2003, Nick Bostrom imagined a superintelligent robot, programmed with the seemingly innocuous goal of manufacturing:***

- a. Paper clips
- b. Scissors
- c. Pencils
- d. iPhones

***2) According to the article, research suggests that YouTube has been helping to:***

- a. Sell more Google Pixel phones
- b. Teach ethics to CS students
- c. Polarize and radicalize people
- d. Educate young people to the need to vote

***3) One of these principles is NOT part of Stuart Russell's three "principles of beneficial machines,":***

- a. The machine must not harm a human or allow a human to come to harm
- b. The machine's only objective is to maximize the realization of human preferences.
- c. The machine is initially uncertain about what those preferences are.
- d. The ultimate source of information about human preferences is human behavior.

***4) In standard inverse reinforcement learning, a machine tries to***

- a. Adapt its database to its reward function
- b. Learn the reward function that a human is pursuing
- c. Teach a human what he/she is expected to pursue
- d. Teach ethics to programmers

***5) The author of the article mentions an issue in Stuart Russell's principles that Stuart Russell himself did NOT consider:***

- a. Our behavior is so far from being rational that it could be very hard to reconstruct our true underlying preferences
- b. Human preferences change
- c. What about the preferences of bad people?