SCIE300 Reading Assignment 5 (RA5)

80 points

Question 1 (20 points)

Describe at least two of Bostrom's "malignant failure modes" in your own words.

Question 2 (5 points)

Give a brief description of Bostrom's opinion about the ethical treatment of superintelligent computer programs.

Question 3 (10 points)

In your opinion, what is likely to be the most effective of the capability control methods described in Chapter 9?

Question 4 (10 points)

What is the difference between a motivation selection method and a capability control method?

Question 5 (10 points)

In section 1, Turing describes an "imitation game," which these days is called the Turing test. Turing actually describes two versions of his imitation game. The first version involves three humans. In this first version of the imitation game: (a) Who are the players? (b) What are the rules for winning the game? (c) Who do you think will win on average?

Question 6 (5 points)

The second version of the imitation game involves two humans and a computer (note that Turing often uses the word "machine" where we would use "computer"). This second version is what we commonly refer to as the Turing test today. What are the rules for winning the second version?

Question 7 (20 points)

In section 6, Turing rebuts nine possible objections to his belief that a computer program could, in principle, win the imitation game. The objections are numbered (1)-(9). From these nine objections, choose the one objection that you find *most* compelling. (a) Explain why you believe your chosen objection may be plausible. (b) Which part of Turing's rebuttal to your chosen objection do you believe may be incorrect?