Chapter 26 | Questions

1. How did Dijkstra argue that the question "Can machines think?" is ill defined?

   Answer Hint: "The question of whether Machines Can Think . . . is about as relevant as the question of whether Submarines Can Swim." The question is ill defined because the definition of "think" may evolve/change.

2. Why did Dijkstra argue that the question "Can a machine be conscious?" is ill defined?

   Answer Hint: Because he thinks that we need not insist on a higher standard for machines than we do for humans. For example, when we interact with humans, we do not (every time) ask if the other person is consciously thinking.

3. Why do the authors (Russel and Peter) don't fully agree with philosopher John Searle's statement - "No one supposes that a computer simulation of a storm will leave us all wet . . . Why on earth would anyone in his right mind suppose a computer simulation of mental processes actually had mental processes?"

   Answer Hint: Because the authors argue that we do not know if mental processes are more like storms, or more like addition.

4. In the context of Artificial Intelligence (such as expert systems) being used for medical diagnosis, when can doctors become legally liable if they don't use the recommendation of an AI system?

   Answer Hint: If it is proven that for the particular diagnosis AI (or expert systems) is more accurate than humans

5. Specifying the right utility function for an AI system to maximize is not so easy. Consider that you are designing an AI system with a utility function designed to minimize human suffering - expressed as an additive reward function over time. What is wrong with such a utility function?

   Answer Hint: To minimize human suffering, it may decide to kill all humans (i.e. in order to have minimum = 0)

6. What is wrong with the 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 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?

   Answer Hint: The weights between the three laws are not defined well. For example, a person's life is at risk but she "casually" orders a robot to get a "lifesaver" from a dangerous place (dangerous for the robot's existence) then the robot does not know how far to risk itself to fulfill her request. Note that the robot can be aware that even if it kills itself, it won't be able to fulfill the human's request.