Konrad Bartlett

0580964

COIS 3560

Assignment 2

The topic of computer superintelligence is actually what got me very motivated to go into computers in the first place so I think it’s a very exciting idea. My field of study is data analytics and AI which roughly translate to making computers do things that humans do but better, in this way I am sort of on the bleeding edge of this field, and am personally developing super intelligence in computers.

What makes computers so super intelligent is there incredible ability to take input, perform an equation (any equation), and produce a mathematical output that models something. This equation in the middle though is where it gets cool, because while traditional practices allow for the hard coding of these equations, modern advances in the power of computers has made it possible to make computers generate and analyze these equations themselves. This leads to incredible opportunities because instead of being limited to equations that the human mind can comprehend, we are limited by the equations that computers can create. What equations can computers create? Anything, the smarter a computer gets, the smarter equations it can make which in turn generate “smarter” computers.

After reading chapter 12 of the textbook my opinion has not changed very much. My field of computers is surrounded by vagueness and people telling me that this is the dooms day. While I see the outcome that the textbook provides (a super intelligence that has been left alone for so long that it becomes resistant to humans) as a slight threat to my survival, this is ultimately my personal end goal. AI and super intelligence are always unfairly being judged and put under a microscope, but really I just see letting AI take over the world as a future career goal.

The value loading problem. The textbook brings up the issue of how computers represent ideas vs. how humans represent ideas and calls it the value loading problem. Essentially what this is, is that computers operate on that a computer can understand and may produce values that are not legible by human brains. This is perfectly fine in my opinion and does not change my viewpoint in any way, simply because it just makes computers as much as a black box as anything else with conscious thought. If we’re going to say that a computer is dangerous because we can’t understand what is happening behind the scene then we might as well hold other sentient beings that we don’t understand to the same standards, are they a threat? Not in my world view, so the idea that an AI operates with a black box is not a significant worry to me. On top of this there are plenty of visualization tools such as PCA plots, WEKA, or Decision tress that can be used to describe how a computer comes to the conclusion it has.

It then brings up evolutionary learning and how computer AI will generate and increment on the performance of it’s agents by selected and “breeding” the best agents to produce better AI. I don’t find AI improving itself as a threat as much as I see it as a way to make my job easier.

Overall the chapters asks if we want a computer to want, to have a goal and I think if we ever want to further intelligence and ultimately progress as a society then we do need computers to eventually have a goal. It’s just not possible to teach computers to perform better unless we give them a measure and the goal of improving themselves.