Dean Fleming

deanf1@umbc.edu

CMSC 471 – 01

Why the “AI” in “Dangerous AI” Should Stand for “As If”

The issue of artificial intelligence is one that has concerned human intelligence as long as computers have been around. The questions are endless: Is artificial intelligence possible? And if so, is it possible that an artificially intelligent machine could reprogram itself to do harm to humans or civilization? In my opinion, I do not know if artificial intelligence will ever get to or surpass human intelligence, but if it does, I welcome it with open arms. I do not think that artificial intelligence will ever become evil, and we can prevent this proactively.

The argument that some people say is that if a machine can reprogram itself to choose what it does next on its own, that it could choose something harmful and execute it. In my opinion, however, this is very preventable. The key to this is that in order for a machine to make motivated decisions, it needs to have some sort or motivation program it runs. All we have to do there is careful program that motivation program to only be motivated to do things that are helpful to humans or align with what we deem safe.

The counter argument to the previous argument would be: What if a machine decided, for example, that the best way to protect humans from harm would be to make sure no more humans were born, or if the best way to keep humans smiling is to implant electrodes in their faces? These are fallacies, however, along with being overly dramatic. The initial logic makes sense, but once we get to the “kill all humans” solution, this would only just feed back into the “protect humans” rule, which would cause an infinite loop of this logic. The solution contradicts the rule, so it could never execute.

Adding in a positive-outcome-only motivation algorithm would involve an important step. It would need to be designed and implemented before or at the same time as the rest of the artificially intelligent program. This way, the program cannot start executing and making bad decisions before it can be limited to making only good decisions. This would take a lot more preparation and forethought, but for something as significant as artificial intelligence, this would make it well worth the work and wait. Any hole needs to be closed before something that intelligent could be set free.

In conclusion, artificial intelligence would be a powerful tool, probably the most important tool mankind could ever invent, but it is nothing we should be afraid of. It is something we should embrace and work toward. It is truly the next step in human evolution; if we can invent it, we would never need to invent anything else ever again. The machines would just invent all new and useful things for us. We would become “transhuman”, and we could even start exploring the galaxy or even other galaxies. A real artificial intelligence can invent anything possible, and something like this cannot be turned away from. If anything, it is the safest bet for civilization’s future.