

AI & ethics

Simon Westerlind
swesterl@kth.se

September 2016

Introduction

As we embark upon the quest of creating ever more advanced AI, not only for narrow compartmentalized tasks, but also more general AI, we must ensure that there are sufficient guidelines to prevent undesirable ethical ramifications. In this short essay we will be taking a brief look at the four major areas in which I see cause for concern. Namely, intelligence explosion, immoral intelligence, consideration for consciousness and human redundancy.

Intelligence explosion refers to a scenario in which we have managed to create computer programs which can optimize their own source code. Such an AI could conceivably increase its efficiency exponentially, and within a rather short period of time reach a level of intelligence far exceeding our own. Unless such a program is properly contained, it could have dire consequences. While portrayed in modern fiction as the primary cause for concern regarding AI, I highly doubt that we be faced with such problems* within a foreseeable* future. Nonetheless, I believe it is time that we start looking to create guidelines and safety protocol for how to prevent harm, should the problem arise. These guidelines would doubtlessly include having these program properly contained from our global IT-infrastructure.

Immoral intelligence is where we as programmers are faced with writing protocols for AI which will have ethical consequences. An example of which would be an AI which guides a self driving car. In the unfortunate event that this car happened to be in a situation where it was forced to decide between steering either left and into a motorcyclist with a helmet or to the right and into a motorcyclist without one, how should the AI act? Should it aim to minimize damage by colliding with the motorcyclist which was wearing a helmet? If that is the case, the lawful, helmet-wearing motorcyclist would indirectly be penalized for taking proper safety precautions. Furthermore, wouldn't you prefer an AI which puts you and your family's safety first? While I do think it is important that we create a set of rules which dictates how AI are allowed to prioritize in such situation, I do not know whom I would want to grant the authority

of dictating those rules to. My recommendations would be a broad coalition of experts appointed by the UN, consisting primarily of people within the the fields of artificial intelligence and moral philosophy.

By consideration for consciousness I refer to a scenario in which machines would gain some resemblance of a conscious mind, and whether or not they should be granted certain rights. The origin on consciousness is still a mystery, but there seems to be general consensus regarding the claim that consciousness correlates heavily with an entity's intelligence and awareness. Were we to be able to create a general AI which was intelligent, was self aware and had ambitions for the future, I would argue that these AIs might have to be granted certain rights. I would not be surprised if there laws forbidding AI abuse similar to the animal abuse regulations. I am unfortunately not enough educated on the subject to be able to propose any suitable solution. I can however elucidate the problem and plead to experts and politicians to process the topic.

If we were to create a general AI, we would essentially posses the ultimate labour saving device, which would essentially render humans redundant as producers of goods and services. In such a scenario a few people would likely reap all of the benefits from this invention whilst the rest of humanity would be unable to provide for themselves. I do not think today's capitalist market will adequate be able to handle this situation and distribute wealth fairly. My solution would be to develop a plan for the entire globe to transfer into a more socialistic market structure.

Wordcount: 631