

Laws For Humans, Logics For Humanoids

Datla Hanith, Chandra Jeevan Sai

Abstract—Everything in this universe is governed by some laws. Be it the physical laws of nature, constitutional setup in human societies, or the way an organization works there are some rules and ethics to be followed for sustainable existence. Now through, with the advances in Artificial Intelligence, it might not be long before science fiction becomes reality. Will there be robots which can actually think? If so, shouldn't we propose some kinds of laws, Logical Restrictions for these humanoids? What role does constitutional law play in human societies, and why there is a need to have such systematic laws based on sound logics for intelligent machines too? Just as humans create havoc by misusing the laws that we as a nation agreed upon, intelligent machines too might create worrisome situations when their laws are fundamentally flawed in nature. With biological and chemical processes being embedded into robots, how humane can we consider such beings. Should there be ethics for artificial life? This paper delves into these questions and attempts to answer them.

Index Terms— Cloud computing, Humanoids

INTRODUCTION

“The only way to discover the limits of the possible is to go beyond them into the impossible.”

~Arthur C. Clarke

Humanoid robotics has come a long way since its inception. With the advent in humanoid technology, we are close to conceiving the idea of what may be called artificial life. This would include robots that think, learn collectively, develop and perhaps eventually transform themselves more effectively. We would then have a huge population of self motivated bots with distributed infrastructure which would negotiate, exchanging tasks and resources in mutually beneficial ways. As a given task arises, humanoids will not only share workload and resources, but will also evolve by passing hostindependent, modular code.

Imagine what would happen to a humanoid trained for knowledge, personality, intentions over a long period of time. What if the code of the humanoid is developed under the Open Source paradigm? It would mean that developers around the world would be able to modify the software of their own or other people's robots. Source code aside,

humanoids will be given the ability to develop and learn in

response to the input they receive. Could a cruel master make a cruel humanoid? Will people begin to see their robots as a reflection of themselves? As works of art? As valuable tools?

As children? If humanoids learn “bad behaviour,” whom should we hold responsible? The manufacture? The owner? The bot? Or the surrounding environment as a collective whole?

This is where the question of ethics comes up. To prevent such incidences from happening, shouldn't we be having rules or logics for humanoids?

From climate to life processes, all terrestrial beings follow certain sets of rules that are governed by nature for a balanced setup. While it is a physical law that the earth needs to rotate at a certain speed to counter the gravitational force of the sun by its centrifugal force, it is also an evolutionary law that there are different kinds of animals that feed on a wide variety of fauna, flora and even other animals so that the environment balance is maintained. It is but natural that man had to follow suit and develop laws so that he could live in harmony with his fellow counterparts. There is thus always a need to have laws such that a balance is maintained in the society.

No matter how quickly technological progress seems to unfold, foresight and imagination will always play key roles in driving societal change. Nurturing robots seems to present a greater challenge than actually building one. Humanoids are the products of our own minds and hands. Neither we, nor our creations, stand outside the natural world, but rather are an integral part of its unfolding.

II Recent Trends In Human ROBOTICS

The Humanoid robot took its first shape in the year 1973, at Waseda University, Under the guidance of the late Prof. Ichiro Kato. Ever since progress in this field has never taken a back seat. The recent technological advances in robot technology, artificial intelligence, power computation etc. have contributed to enable humanoid robots (HRs) to roughly emulate the physical dynamics and motor dexterity of human. Nowadays, HRs are capable of displaying motor dexterities for dancing, playing musical instruments, talking etc. Although the long term goal of true autonomous HRs has yet to be accomplished, the feasibility of integrating them into people's daily lives is becoming closer.

Honda's ASIMO (Advanced Step in Innovative Mobility) developed in 2009; HRP -4C, created by the National Institute of Advanced Industrial Science and Technology in Japan; REEM-B, a humanoid built in Spain by Pal Robotics; Twenty-One, a white plastic E.T. look-alike on wheels developed Waseda University in Japan; Justin, a robot from the German Aerospace Center (DLR), a baby robot CB2 developed by graduates at the Osaka School of Engineering are among the examples of human actions like talking, walking, running and also communicating with humans. In years, as an initiative to promote the among ronotics, the Technical Committee on Robo-ethics is an applied ethics whose objective is to develop scientific/cultural/technical tools that can be shared by different social groups and beliefs. These aim to promote and encourage the development of robotics for the advancement of human society and individuals, and to help preventing its misuse against humankind. Concerns have also been raised develop laws for humanoids at the world Declaration issued on February 2004 stated Fukoka (Japan)by the following statements: "Next generation robots will be partners that coexist with human beings", "Next generation robots will assist human beings both physically and psychologically", and "Next-generation robots will assist human beings both physically and psychological", and "Next-generation ronots will contribute to the realization of a safe and peaceful Society", These Statements basically refer to some Kind of guidelines that assume that the next generation of robots will have the capability of coexist within the human environment to improve the life conditons of people.

THE PROBLEM

Consider two Scenarios where knives are used. In the first case, the knife is used to operate on a person for surgery. In the other Case, the knife is being used by a person to harm others. So, what is the 'Knife'? Is it good ' or 'bad'?

From the above example, we see that the virtue of being good or bad depends on the context.

Now, consider a robot, specifically a humanoid in such a scenario. How would it distinguish good from evil? This is where the question of morality springs up, For a society to live in harmony, it is highly essential that its members follow certain codes of ethics and morality.

We are assuming a world, sometime in the near future, wherein humans and humanoids need to exist together. Since

we have laws for humans, it is an implicit requirement that artificial life also has laws which controls it.

~ Issac Astimov

Issac Asimov the popular science fiction novelist had formulated rules which he felt every humanoid must follow. These laws become so popular that they came to be known as "Asimov's Laws of Ronotics". These have been stated below:

The Zeroth Law:

A robot may not harm humanity, or by inaction, allow humanity to come to harm.

The First Law:

A robot may not injure a human being or, through in action, allow a human being to come to harm.

The Second Law:

A robot must obey any orders given to it by human beings, except where such orders would conflict with the first Law.

Third Law:

A robot must protect its own existence as long as such protection does not conflict with the First Second Law.

The rules of Asimov, though fictional, highlight that protecting humanity is of at most importance. Since these laws have been framed in the interest of humanity, we presume that all humanoids achieve these goals.

IMPLEMENTATION

In this model, we propose to have a hierarchy laws which as a whole try to implement ethics into artificial life. The above laws of Asimov may considered as being in the topmost level hierarchy. This is what we need to achieve.

Sub laws are implemented in the form of contests. Each context has its own rules, which form the sub hierarchy. We propose using the services of cloud computing and the effectiveness of parallel computing in the development of the artificial agents. Moreover the central System of the robot is assumed to be an evolvable system using advanced machine learning algorithms. Initially, we build humanoids specialized to work in a particular area. The development of robots starts with the training of the robot for evolving learning under a trained and controlled environment. This is a especially designed for applicative use

of the robot in its domain. For example, a medical robot would be embedded with rules which apply medical world, Similarly, an industrial robot would be trained to perform in an industrial environment. Since the context is already known, it is easy to define what is right and what is not. These can be called closed worlds. The robot is trained under this closed world for some stipulated period of time. We now store these rules that the robot is expected to follow in a certain knowledge base, i.e. a cloud. A Cloud provides 3 necessary services Platform Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS). Since we intend to store rules on an external database we will be using the functionalities of Infrastructure as a Service. Every humanoid has its space on the cloud. Cloud is placed under the control of a higher "Central Authority", which is directly answerable to Government of a particular country on various levels which control robots under its jurisdiction. It is left to the discretion of the central authority to decide what rules can be shared and what cannot. For example: Information used by a military robot might be classified. So, there should be flexibility in the amount of knowledge that can be shared. The robot wireless communication system which would be then be at a very advanced stage. This system can enable the robot to access, process and implement rules from the cloud swiftly and efficiently. In this manner we can ensure that the cloud database does not fall into the wrong hands. When these robots are brought to the outside world, decision making may be done based on the information in the combining the rules in the various contexts that are stored in the cloud. The inherent question of good versus bad would be answered here. Given the context, and given the action, the robot selects the rules that can be possibly applied to the current situation. The rule to be followed by the robot is then selected by it depending on the weightage given to it. This has been elaborated in the next section.

MAKING DECISION

Good versus Bad

There is a need for robots to define what is "good" and what is "bad". Taking into account Asimov's 0th law which states that the robot must protect humanity, we can split the rules into positive and negative rules. Positive rules work for a better future of humanity while negative rules work against humanity.

Returning back to our example of using a knife. While it can be used in a positive perspective, i.e. to cut vegetables and fruits that can be served to humans, it can also be used in a negative way, to harm them.

As such the rules that harm people or attempt to cause damage to humanity must be attached with a negative weight, which gives an indication to a robot that these rules must be avoided by the robot. The positive rules must be attached with a positive weight, which are an indication that they will benefit humans. The weights can be of two kinds, "closed" and "open". While closed weights are attached to laws when they are viewed from a closed world perspective, open weights are associated with the knowledge base as viewed for the open world. A closed world for a robot is the specific environment that the robot was built for. For example the closed world for a medical robot will be the hospital where such a robot is used regularly. The open world for a robot is the combination of all other including the general case where the robot will be used. For the medical robot, an open world could be the road, a house, a shop or any other place other than a hospital.

The weights are also assigned a range of values from say (0,5) for positive rules or (-5,0) for negative rules. For positive rules, preferences increase with increase in value of the weights. The rule benefits humanity the most is given a value of 5 and the rule benefiting it the least is given a value 0. On the negative scale, preferences for implementation of the rule decrease with a fall in value. So while -5 stands for rules that will have a highly damaging effect on humanity, 0 is for rules having least negative effect. 0 is thus a neutral value which may not bring about any significant changes.

Each rule is selected based on comparison of the current situation with certain predefined contexts, specified as parameters and conditions. After the context has been determined, the robot will then evaluate objects present in the context (i.e. apple, knife etc.) and check for the past experience of any other robot in this context with the same object parameters. If the knowledge search process returns a past experience, then the past action is applied or implemented according to the weight assigned to the past action rule in the knowledge base. If not, it will check the value of the weights assigned to the rule for the context and if the associated to the rule for the context and if the associated weight is positive, it is implemented. If the associated weight is negative, the rule is not applied in this context. However we can have situations where a rule has a negative weight in one context and a positive weight in the other. In such a case, a closed negative rule could be implemented in the open world, depending on the context. In cases when the evaluation knowledge leads to generation of two or more rules, then the rule with the highest positive weightage is applied. When a robot encounters a situation where it cannot decide even after considering all the above scenarios, we have a human

intervention who would make the decision. This is then learned by the robot and is updated in its knowledge base. A B. Split Second Decisions But then, what about decisions which need to be taken at the split second? Suppose a building is crumbling down and the only way to save the humans trapped inside is by breaking through the walls. A robot might refer to the knowledge base and infer that breaking into walls is wrong. Human intervention here is obviously not possible. The first time it encounters such a scenario, we let the robot follow what ever rule it infers. The human intervention would be after the incident, but here the robot will realize the exception to its rules. After all, we need to let the humanoids learn from their mistakes! On a short term, this might not be a great idea but on the long term. We could say that a robot might get wiser over time.

VI .HUMAN MACHINE INTERACTIONS

The development of high end humanoids that look and act similar to humans would bring about a total change in how humans interact with themselves as well as with the robot. Excess interaction with robots, would reduce the real world conversations that humans have among themselves. Just like how the development of internet and social networking has resulted in more and more humans conversing more over the virtual world thereby neglecting their reality. Too much of human-humanoid interaction?

Few years back, Nilanjan Sarkar of Vanderbilt University and his former colleague Wendy Stone, now of the University of Washington, developed a prototype robotic system that plays a simple ball game with autistic children. The robot monitors a child's emotions measuring minute changes in heartbeat, sweating, gaze, and other physiological signs, and when it senses boredom or aggravation, it changes the game until the signal indicates the child is having fun again. The system is not sophisticated enough yet for the complex linguistic and physical interplay of actual therapy. But it represents a first step towards replicating one of the benchmarks of humanity: knowing that the others have thoughts and feelings, and adjusting your behavior in response to them. Leaving the care of a baby to a robot will definitely pose many questions. Owing to great expectations of human-humanoid.

VII. CONCLUSION

The creation of many humanoid robots in the future must address concerns of human safety. It is thus a necessity to incorporate ethical and moral values to humanoids. Ethics are important for a balanced living in the society. More so, because humanoids need to be trained properly so that they do not go off course and disturb the fabric of the society. The model we propose to implement laws for robotics might not be effective on a short term basis, but looking at a long term perspective, humanoids will have the intelligence to make the correct decisions, safe for humanity. There must be a conscious attempt made to stop robots from getting into the wrong hands and letting them evolve as the consequences of such actions will seem disastrous. Careful planning and training of artificial life will surely ensure harmony between humans and humanoids.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Mr. Bollem Rajkumar, post graduate student, IIIT Hyderabad and A. Prashanti Priya, Oracle India Pvt Ltd, for assisting us throughout the making of paper.

REFERENCES

- [1] I. Asimov, I. ROBOT (a collection of short stories originally published between 1940 and 1950), Grafton books, London, 1968.
- [2] M. Hauser (2006) – *MORAL MINDS : HOW NATURE DESIGNED OUR UNIVERSAL SENSE OF RIGHT AND WRONG*, Ecco
- [3] R. Clarke's, Asimov's Laws of Robotics: Implications for information technology part 1, computer, vol 26, no 12.
- [4] <http://news.bbc.co.uk/2/hi/science/nature/8044200.stm>
- [5] <http://www.zdnet.com/blog/emergingtech/exclusive-a-robot-with-a-biological-brain/1009>
- [6] <http://androidhumanoid.com/android-humanoid-philosophy>