Artificial Intelligence: Global Risk and Long Term Strategy for Ending Fast Take off

Md. Faruk Hussain Khan, Shadman Sipar Ocean, and Wali Mohammad Abdullah

***Abstract*—**Artificial intelligence is advancing very fast in wide range. If it were to surpass that of humans significantly, there is no doubt in near future it would constitute a significant risk for humanity. This is the high time to think about it and consideration the issues that must include progress in AI as much as insights from the theory of AI. The effort in this paper tries to make cautious headway in finding the problem, evaluating predictions on the future of AI, proposing ways to ensure that Artificial Intelligence will be useful to humans – and logically evaluating such proposals.

***Index Terms*—**Artificial Intelligence, Human Vs Machine, Technological Risks, Robotics

I. Introduction

The field of artificial intelligence is growing very fast along a range of fronts. Recently it has seen dramatic improvements in AI applications like autonomous robotics, game playing and image and speech recognition; these applications have been driven in turn by advances in areas such as neural networks (i.e. deep learning techniques), search (i.e. Monte Carlo search and meta reasoning), and the scaling of existing techniques to modern computers and clusters.

In spite of the major scientific and economic impacts of AI and machine learning at present seen, the most of transformative impacts from this technology lie in the future. While the field shows lot of benefits, a growing body of researcher within and outside the field of AI has raised concerns that future developments may represent a major technological risk.

As AI algorithms become both more general and more powerful – capable of functioning in a wider range in different environments – their positive benefits and their potential for harm will increase rapidly. Even very simple algorithms, such as those implicated in the 2010 financial flash crash, shows the difficulty in designing safe goals and controls for AI that prevent unexpected catastrophic behaviors and interactions from occurring.

Technology has both positive and negative outcomes as well. Since computers are more capable of producing accurate results, they will probably replace humans in jobs that are better suited for them. It will create a situation that the workplace will no longer be man's domain. Unemployment problem will increase and people will lose the place as a dominant human being. Most drastic of possibilities is complete destruction of the human race. If AI at the level of Moravec's Fourth Generation Robots is created, these machines will have a "mind" of their own and could potentially destruct humanity.

At a more basic level, the use of artificial intelligence in household tasks might produce laziness on the part of humans. Mentality might become; "if the computer can do it why should we waste our time trying it ourselves?" Humans have an extraordinary ability to think, analyze, and use judgment. If AI is used for interpreting, then the human mind and its capabilities might go to waste.

Manuscript received May 8, 2016.

Md. Faruk Hussain Khan is an undergraduate student in the department of Computer Science & Engineering (CSE) of Military Institute of Science & Technology (MIST), Mirpur Cantonment, Dhaka-1216, Bangladesh

Shadman Sipar Ocean is also an undergraduate student in the department of CSE of MIST, Mirpur Cantonment, Dhaka-1216, Bangladesh

Wali Mohammad Abdullah is a Lecturer in the department of CSE of MIST, Mirpur Cantonment, Dhaka-1216, Bangladesh. He is the supervisor of Mr. Khan and Mr. Ocean in this research

II. Literature review

In [2], Omohundro introduced the problem of risk and the author emphasizes his points that even an simple artificial agent, like one program to win chess games, can very easily turn into a dangerous threat for humans. For example, if it starts acquiring resources to accomplish its goals: "The seemingly harmless chess goal therefore motivates harmful activities like breaking into computers and robbing banks." He suggested that human need formal methods that provide proofs of safe systems.

The two following works dealt with prediction of ensuing success in AI. Armstrong et. al proposed a decomposition schema to compare predictions on the future of AI and then test five famous predictions [3]. T. Goertzel argued in [4] that while most progress in AI so far has been 'narrow' technical AI, the next stage of development of AI, for at least the next decade and more likely for the next twenty-five years, will be increasingly dependent on contributions from strong-AI.

Now this is our concern and we stepped into the proposals on how to achieve human friendly, positive, safer and ethical general AI. Brundage investigated the general limitations of the approach to supply an AI with a 'machine ethics', and found them both serious and deeply rooted in the nature of ethics itself [5]. In [6], Yampolskiy founded, which utility functions we might want to implement in artificial agents and particularly how we can prevent them from finding simple but counter-productive self-satisfaction solutions. B. Goertzel explained how his "Goal-Oriented Learning Meta-Architecture" (GOLEM) may be capable of preserving its initial – benevolent – goals while learning and improving its general intelligence [7]. Potapov and Rodinov demonstrated an approach in [8], to machine ethics in AIXI that is not based on 'rewards' (utility) but on learning 'values' from more 'mature' systems. Kornai argued in [9], that Alan Gewirth's dialectical argument, a version of classic Kantian ethical rationalism, showed how an artificial agent with a certain level of rationality and autonomy will necessarily come to understand what is

moral. Kornai thus denied what Bostrom called the 'orthogonality thesis', namely that ethical motivation and intelligence are independent or 'orthogonal'. Editorial: Risks of Artificial Intelligence. Sandberg in [10], looked at the special case of general AI via whole brain emulation, in particular, he considered the ethical status of such emulation: Would the emulation (e.g. of a lab animal's brain) have the ability to suffer, would it have rights?

Dewey investigated strategies to mitigate the risk from a fast takeoff to super intelligence in more detail [11]. In [12], Bishop argued and took a different line which shows that there is no good reason to worry about existential risk from AI, but that we should rather be concerned about risks that we aware about such as the military use of AI. Like many people working in AI, Bishop remained unimpressed by the discussion about risks of super intelligence because he thought there were principled reasons why machines will not reach these abilities: they would lack phenomenal consciousness, understanding and insight.

III. Possible Threats of AI

Many researchers have raised the issue that, by way of an "intelligence explosion" near future in the 21st century, a self-improving AI could become so gigantic more powerful than humans that we would not be able to stop it from achieving its goals.

As the robotics industry surges, so too are claims involving bodily injury, property damage, and financial loss. Robot-linked workplace deaths occur. While certain robots accumulate 'artificial intelligence,' this (as with human intelligence) is no immunity from errors and omissions in performance.

When a robot-related accident or incident takes place, complex liability questions are must: Is it the result of an error arising from the manufacturer of the robotics hardware, firmware, software, or artificial intelligence architecture or perhaps some incorporated subcomponents thereof? Is it a professional liability error or general liability loss? The opportunity for ambiguity is clear. So there is a need for a comprehensive risk exposure and management solution.

A. Robot Race

A strong AI seeks to be life like. Therefore in near future people may embody human intelligence in to machine i.e. robot which may be mobile and sustain its own existence. If many of these machines were built, they could be called a race of robots, and a race of human-like intelligences would likely be a competitor for natural resources.

Bill Joy, chief engineer at Sun Microsystems and author of the manifesto "Why the Future Doesn't Need Us," argued that this is a conflict we would surely lose. Intelligent robots could easily rebuild themselves. They would have no gestation period. A new robot would be "born" in the time that it takes to put the pieces together.

B. AI-Enabled Crime

AI-enabled crime is one important risk in the short term. Intelligent malware that mutates over time has not yet appeared but is feasible. This is particularly alarming because those who would activate such malware probably would not care about its potentially wide-ranging negative impact. Applications of such AI include:

1) Malware that could imitate users on, e.g., banking websites.
2) Intelligent bots that could spread misinformation via social networks in order to cause panic or to manipulate public opinion, with the goal of affecting stock movements. Such misinformation campaigns could be very effective.
3) AI-automated insider trading, which may already be occurring.

C. The Singularity and Machine Memory

Research is being done on "lifelong" machine learning wherein machines learn from their experiences at a steady rate. However, most intelligent systems cannot be run for unlimited time because of their memory limits. These limits would prevent their capabilities from increasing exponentially.

D. Autonomous Weapons Systems

It is feasible that autonomous weapons systems, such as intelligent drones, will be developed. Autonomous drones would be more consistent in following military rules of engagement than drones operated by humans. However, there is certain risk of such technology falling into the hands of enemies or criminals, in which case, it would follow whatever rules they chose.

There has been a recent outcry with regards to the engineering of artificial-intelligence weapons and within short time AI is going to take over the place of human being. AI weapons do present a type of danger different than that of human controlled weapons. Many governments have begun to fund programs to develop AI weaponry. Developing counter system and robots when there is weaponization of artificial intelligence. Some individual or country will try to develop a weapon of Mass Destruction with Artificial Intelligence.

IV. Proposed System

Basing on the threats discussed above we propose some principle and system to reduce the risk of AI. Those are as follows:

1) Strict use of roboethics which dictate to the morality of how humans design, construct, use and treat robots and other artificially intelligent beings. It considers both how artificially intelligent beings may be used to harm humans and how they may be used to benefit humans. So a central body may be formed by any concerned institution it may be any recognized university or IEEE who may look after and set the standard to ensure the roboethics issues.
2) Forming of a Specialized Risk Management Services that Support policy holders in navigating the new and shifting terrain of robotics-related risks – and helps to mitigate exposures.
    - A robotics-focused workplace safety evaluation with an aim to support emerging safety standards issued by ISO, ANSI, and the Robotic Industries Association.
    - Robotics Workplace Audit covering the spectrum of workplace laws impacted by the introduction of specific robotic systems or services introduced to

the workplace.

- Review of state- and federally-mandated Illness and Injury Prevention Programs as they relate to the use of robotics.
- Ongoing updates on changes in regulations and legislation regarding robotics, including updates from the Congressional Robotics Caucus and the Littler-created Workplace Policy Institute.

3) Dedicated Robotics Claims Team. Enhancing claims management and supporting continuity of coverage with general, products, and professional liability robotics-related claims processed through a single point of contact.
4) In case of any weapon system with AI is built we suggest a system or robots which we call counter robot system with following capacity which can control the destructive weapon system.
   - Our counter robot/system will be able to control the destructive robot.
   - It will be able to judge the capacity of the destructive system and make a program from its intelligence which can instantly take the decision and destroy its counterpart.
   - It should not destroy the human being and this decision should be taken from its artificial intelligence.
5) The best step at providing adequate safeguards would be controlling and regulating the AI itself: requiring the development of testing protocols for the design of AI algorithms, improved cybersecurity protections, and input validation standards—at the very least. Those protections would need to be carefully prepared to each industry or individual application, requiring countless AI experts who understand the technologies, the regulatory environment, and the specific industry or application. At the same time, regulatory proposals should be crafted to avoid stifling development and innovation.

V. Proposed Concept for Mitigating AI Risk

As we discuss above necessary steps to mitigate the risk of AI we have proposed a system which is shown with a flow chart to detect the possible risk and required preventive measures in Figure 1.

Figure 1 explains how to monitor suspicious activities and take necessary action analyzing the perceived risk. Thereby risk factor from AI can be minimized.

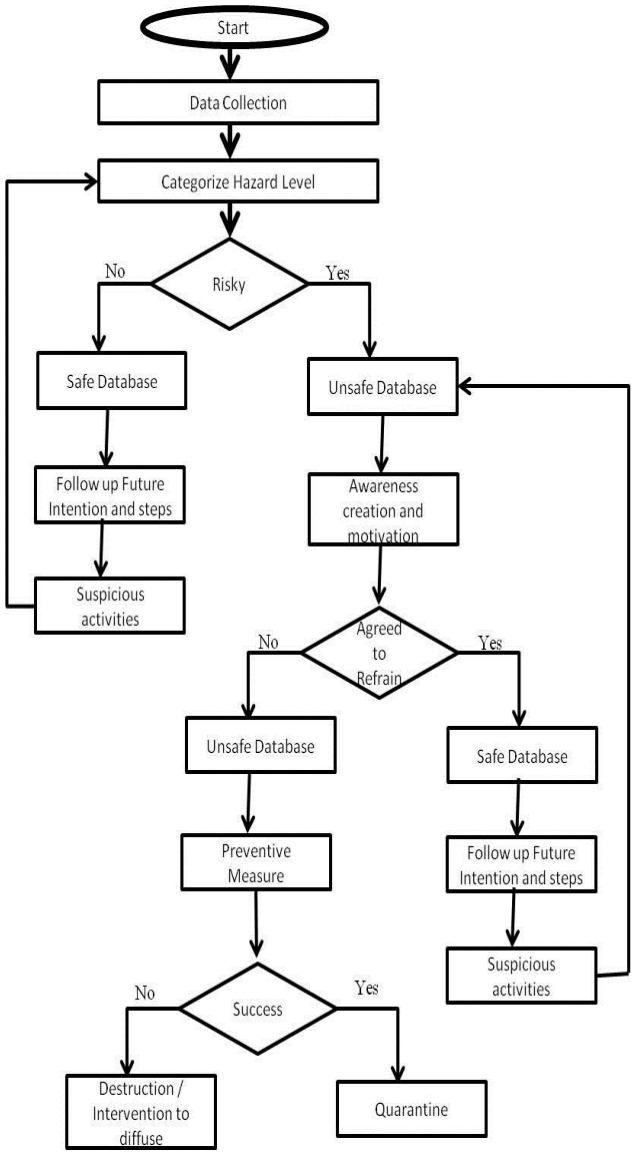

Fig. 1. Flow chart of possible risk and required preventive measures.

VI. Conclusion

The threats enumerated by many scholars and researchers are real and worthy of our immediate attention, in spite of the immense benefits artificial intelligence can potentially bring to humanity. As robot technology increases steadily toward the advancements it is necessary to facilitate widespread implementation. In near future robots are going to be in situations that pose a number of courses of action. The ethical dilemma of bestowing moral responsibilities on robots calls for rigorous safety and preventative measures that are fail-safe, or the threats are too significant to risk.

It is not only the concern of the researcher and scholars rather of all. There is no doubt that the risk is realized by all. Now this is the time to take necessary action on it. We should not allow any organization or any country to produce destructive weapon using AI. All developer of AI and researcher should come to an agreement that only the positive side of AI will be explored. But as we are human being there will be competition and one of the groups will try to down the others by destructive use of AI. To

curb the negative race of AI we need to bring booth scholar and researcher under same umbrella. But one area of concern is the level of uncertainty associated with both the speed of development and the potential risks of artificial intelligence. This is due to several factors that include conceptual barriers in AI research, a perceived lack of communication between experts in different subfields of AI, and a paucity of research on appropriate safety and control mechanisms for AI development.

The end goal is both to significantly advance both the state of research on AI safety protocol and risk, and to inform industry leaders and policymakers on appropriate strategies and regulations to allow the benefits of AI advances to be safely realized.

REFERENCES

[1] Bostrom, Nick. "The superintelligent will: Motivation and instrumental rationality in advanced artificial agents." Minds and Machines 22.2 (2012): 71-85.
http://dx.doi.org/10.1007/s11023-012-9281-3

[2] Bostrom, N. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press, 2014.

[3] Chalmers, David. "The singularity: A philosophical analysis." Journal of Consciousness Studies 17.9-10 (2010): 7-65.

[4] Clark, Andy. Supersizing the mind: Embodiment, action, and cognitive extension. Oxford University Press, 2008.
http://dx.doi.org/10.1093/acprof:oso/9780195333213.001.0001

[5] Ditterich, T., and E. Horowitz. "Benefits and risks of artificial intelligence. medium. com. Retrieved 23.01. 2015." (2015).

[6] Dreyfus, Hubert L. What computers still can't do: a critique of artificial reason. MIT press, 1992.

[7] Dreyfus, Hubert L. "A history of first step fallacies." Minds and Machines22.2 (2012): 87-99.
http://dx.doi.org/10.1007/s11023-012-9276-0

[8] Good, Irving John. "Speculations concerning the first ultraintelligent machine." Advances in computers 6.99 (1965): 31-83.

[9] Haugeland, J. (1995). Mind embodied and embedded. Acta Philosophica Fennica, 58, 233– 267.

[10] Hawking, S. "Transcendence looks at the implications of artificial intelligence—but are we taking AI seriously enough? The Independent." (2014).

[11] Kurzweil, Ray. The singularity is near: When humans transcend biology. Penguin, 2005.

[12] Kurzweil, Ray. How to create a mind: The secret of human thought revealed. Penguin, 2012.

[13] McCarthy, John, et al. "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955." AI Magazine 27.4 (2006): 12.

[14] Muehlhauser, Luke, and Nick Bostrom. "Why we need friendly AI." Think 13.36 (2014): 41-47.
http://dx.doi.org/10.1017/S1477175613000316

[15] Müller, Vincent C. "Editorial: Risks of artificial intelligence." (2015).

[16] Müller, Vincent C., and Nick Bostrom. "Future progress in artificial intelligence: A survey of expert opinion." (2016).

[17] Pfeifer, Rolf, and Josh Bongard. How the body shapes the way we think: a new view of intelligence. MIT press, 2006.

[18] Russell, Stuart, et al. "Research priorities for robust and beneficial artificial intelligence." Future of Life Institute (2015).

[19] Searle, John R. "Minds, brains, and programs." Behavioral and brain sciences 3.03 (1980): 417-424.
http://dx.doi.org/10.1017/S0140525X00005756

[20] Soares, Nate, and Benja Fallenstein. "Aligning superintelligence with human interests: A technical research agenda." Machine Intelligence Research Institute (MIRI) technical report 8 (2014).

[21] Sotala, Kaj, and Roman V. Yampolskiy. "Responses to catastrophic AGI risk: a survey." Physica Scripta 90.1 (2014): 018001.

[22] Varela, F. J., and E. Thompson. "&E. Rosch. 1991. The embodied mind: Cognitive science and human experience."

[23] Yudkowsky, Eliezer. "Friendly artificial intelligence." Singularity Hypotheses. Springer Berlin Heidelberg, 2012. 181-195.
http://dx.doi.org/10.1007/978-3-642-32560-1_10

[24] Eden, Amnon H., et al. Singularity hypotheses: A scientific and philosophical assessment. Springer Science & Business Media, 2013.

Md Faruk Hussain Khan was born in 1983 in Chandpur, Bangladesh. He is pursuing his under graduate degree in the department of Computer Science & Engineering (CSE) of Military Institute of Science & Technology (MIST), Mirpur Cantonment, Dhaka..

He has keen interest in networking and artificial intelligence.

Mr. Faruk contributed in the 18th International Conference on Computer and Information Technology (ICCIT 2015).

Shadman Sipar Ocean was born in November 07, 1983 in Cox's Bazar, Bangladesh. He is pursuing his under graduate degree in the department of Computer Science & Engineering (CSE) of Military Institute of Science & Technology (MIST), Mirpur Cantonment, Dhaka.

He has keen interest in programming, networking and artificial intelligence.

Mr. Shadman actively contributed in the 18th International Conference on Computer and Information Technology (ICCIT 2015).

Wali Mohammad Abdullah was born on July 21, 1989 in Dhaka, Bangladesh. He received his M.Sc. Engg. degree from the Institute of Information and Communication Technology (IICT), Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh in 2014 and the B.Sc. degree from the Department of Computer Science and Engineering (CSE), Military Institute of Science and Technology (MIST), Dhaka, Bangladesh, in 2010.

He has been working as a lecturer in the Department of CSE, MIST, Dhaka-1216, Bangladesh since 2011. He has published number of journal and conference papers in renowned international journal and conferences. His research interests include Meta-heuristics, Networks, Bioinformatics and Artificial Intelligence.

Mr. Abdullah contributed in the 18th International Conference on Computer and Information Technology (ICCIT 2015) as a Joint Secretary of organizing committee.