

**ITU Kaleidoscope 2016**  
*ICTs for a Sustainable World*

# **Human Safety Considerations in Emerging ICT Environment**

**Shailendra K Hajela**  
**ITU-APT Foundation of India**

Email [hajela@yahoo.com](mailto:hajela@yahoo.com); [chairman@itu-apt.org](mailto:chairman@itu-apt.org)

**Bangkok, Thailand**  
**14-16 November 2016**

# ICT Scenario – upcoming Heterogeneous Man-Machine World

Rapid advancements in the fields of Internet of Things, artificial intelligence, cognitive robotics, super smart machines, etc., have aroused interest as well as concerns

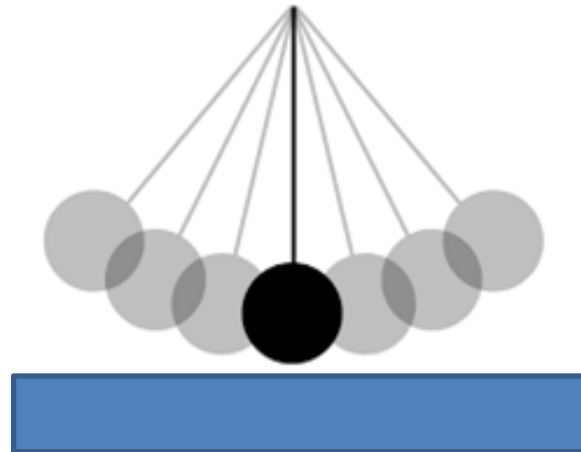
The emerging interactive man-machine heterogeneous environment, where autonomous super intelligent machines/robot working solo or teaming with, can match and even surpass humans in speed and precision presents enormous opportunities as well as threats

# Extreme Visions of Future

Existential  
Threat:  
Elon Musk

IJ Good: AI  
positive  
feedback  
cycle  
resulting in  
AI explosion  
2050 – 2100 CE

ICT Developments'  
Impact



Enormous  
Benefits: Digital  
Economy, trans-  
parent, non-  
discriminatory,  
democratic and  
inclusive society;  
direct impact on  
transportation,  
health, education,  
citizen services,  
environment &  
SDGs

# Asimov's Laws

- No longer is this a matter of science fiction when Isaac Asimov in 1942 made the 3 (technically 4) laws to assign paramount importance to human safety:
  - A robot may not injure a human being or, through inaction, allow a human being to come to harm;
  - A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law;
  - A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.
  - A robot may not injure humanity, or, by inaction, allow humanity to come to harm.

# Need for Aligning AI to Ethical Codes of Humans

Digital Economy - The Fourth Industrial Revolution

In the converged environment of Digital Economy with super intelligent thinking machines endowed with reasoning, determining their own action/ reaction in a given situation with fuzzy knowledge of the result of their action, interacting with humans, the abstractions of virtual world changing to reality pose a great challenge to dominance and safety of human society.

Ethics of Science & Technology

The need is apparent for aligning the AI methodology of super intelligent machines to the ethical codes of humans that incidentally also developed in the process of human civilization over centuries as a result of philosophical thought, moral codes of religions, social reforms, etc.

# PRESENT SCENARIO

- The present century is driven by Information and Communications Technologies (ICT), which involve integration of telecommunications and computer networks being at the heart of all major systems forming the infrastructure of our Living Environment.
- Smart cities of tomorrow embody smart systems with AI methodology at their heart and integration of multiple ICT solutions in a secure fashion for managing their assets in a Digital Economy – banks, schools, libraries, transportation systems, hospitals, power plants, public utility services, waste management, law enforcement, and other community services.

# ICT – an integral Part of our Ecosystem

ICT is now an integral part of our living environment/ecosystem:

- In the 21<sup>st</sup> century, humans would live with technological entities such as robots, autonomous vehicles, drones, ubiquitous entities, such as wearables, smart phones, tablets, home automation gadgets, digital assistants/IoT hubs, etc. connected wirelessly.
- The glaring fact is that the Laws, Regulations and Ethics cannot keep pace with the rapidly emerging new technologies.

Integration of AI: Since work on AI began six decades ago, it has become integrated in many diverse applications and usage, e.g. smart phones, mobile apps, mobile internet, Big data, social media, autonomous vehicles, drones, IoT, Clouds, cyber physical systems (Robotics), health care, surveillance, manufacturing, smart cities, intelligent decision support systems, and so on.

# Opportunities and Ethical challenges of AI

- Many opportunities exist for AI to have a positive impact on our Environment and attaining SDGs. AI researchers need to have the right perspective, tools and the skills required to address concerns of global warming, poverty, food, healthcare, education, and demographic issues.
- The ethics of artificial intelligence is the part of the ethics of technology, specific to each domain that raises its own ethical concerns; the regulatory policies would thus have to be considered domain-wise.
- For example, what happens when a self-driving car has a software failure and hits a pedestrian, or a drone's camera captures images of persons in a private swimming pool or an autonomous robot injures or kills a human?



# AI & Robotics - Emotional Robots

How would real-time monitoring of emotions impact IoT? That it is possible to imbibe reasoning ability, feeling and expression of pleasure and pain in Super smart robots, has been reported in the literature published on the subject.

As an example, unveiled in June 2014, 'Pepper', a humanoid robot, jointly developed by Aldebaran Robotics of France and Softbank of Japan is understood to have the ability to analyze expressions and voice tones, read emotions with the capability to express and respond to emotions.

While the technologies behind these advances are astounding the potential negative ramifications are to be considered. For instance, can fraud be committed by creating a humanoid robot to impersonate someone?

# Trust in ICTs

- Do ICTs deserve our trust? Technologies are becoming more autonomous in their interaction and decision-making, so how does one trust in future ICTs? The impact on society is wide-ranging and question of trust has been deliberated upon, among others, by the ITU and IEEE.
- Tesla's autopilot highlighted the need to address important issues regarding an autonomous automotive future, including regulatory challenges, human override capabilities; and the pursuit of more safe driving tools.
- The challenges posed by autonomous machines will span technical, regulatory, legal and even philosophical realms. They will force us to confront deep moral quandaries, and might even tweak our sense of what it means to be human.
- It has been seen that technology precedes before the laws to deal with its consequences are in place. For example, self-driving cars are ready before the laws for them are.

# POLICY & ETHICAL ISSUES -1

- Bostrom and Yudowsky have studied and analyzed at length the ethics of AI, and emphasized that thinking machines besides their not harming humans and other morally relevant beings, would lack versatility and can have only domain specific intelligence as imparted by the designer, suitable only for the task designed for.
- Five of the world's largest Tech. companies **Alphabet**, **Amazon**, **Facebook**, **IBM** and **Microsoft** are now attempting to create a standard of ethics around the creation of artificial intelligence. The Stanford project, named the "One Hundred Year Study on Artificial Intelligence", would produce a detailed report on the impact of A.I. on society every five years for the next century.
- The first such report titled "Artificial Intelligence and Life in 2030," states that the basic objective is to ensure that A.I. research is focused on benefiting people and that it will be extremely difficult to regulate A.I., since there is no clear definition of A.I.

# POLICY & ETHICAL ISSUES -2

- Nadella of Microsoft in June 2016, in an interview with the *Slate* magazine is reported to have laid out the following six ethical "principles and goals" that AI research must follow to keep society safe. The artificial intelligence must:
  - "be designed to assist humanity" meaning human autonomy needs to be respected.
  - "be transparent" meaning that humans should know and be able to understand how they work.
  - "maximize efficiencies" without destroying the dignity of people.
  - "be designed for intelligent privacy" meaning that it earns trust through guarding their information.
  - "have algorithmic accountability" so that humans can undo unintended harm.
  - "guard against bias" so that they must not discriminate amongst people.

# POLICY & ETHICAL ISSUES -3

- The Working Group on Robot Ethics of **UNESCO**'s World Commission of the Ethics of Scientific Knowledge and Technology (COMEST) has been deliberating on the major ethical issues of machines that comprise both physical robots and software agents, designed to function autonomously that can learn by themselves new process or behaviours. The rapidity of technology advancement leaves a gap between its effective use and the ethical dimension: human well-being, safety and social benefits.
- Also, considering that in future, machines/robots will be sharing the world with humans, to explore the extent to which developments in robotics as an aspect of a broader paradigm of technological convergence would require potential new understandings of "human" in respect of neurological implants and enhancing technologies.

# Evolving Ethical Design Rules

- EPSRC\* and AHRC\* have jointly published a set of following five ethical "principles for designers, builders and users of robots"
  - Robots should not be designed solely, or primarily to kill or harm humans.
  - Robots should be designed in ways that assure their safety and security.
  - Humans, not robots, are responsible agents. Robots are tools designed to achieve human goals.
  - Robots are artifacts; they should not be designed to exploit vulnerable users by evoking an emotional response or dependency. It should always be possible to tell a robot from a human.
  - It should always be possible to find out who is legally responsible for a robot.

EU is deliberating upon the safety, liability and risk management aspects of robotic systems and the UK has published\*\* a Guide to their ethical design and application. China is among leading countries and others would also be engaged in this area.\*

\*EPSRC = Engineering & Physical Sciences Research Council; AHRC = Arts & Humanities Research Council; UK. \*\*

# Summary

- AI is a very exciting technology area with astounding advancements in how machines (with or without mechanical components) can mimic human cognition, learning and problem solving.
- AI - as with the term cloud - is not a single entity but a methodology comprising broad collection of different technologies and techniques that target different problems and use cases in a variety of social and economic sectors.
- While there are significant benefits, there are also downsides, both known and unanticipated. In a sense, AI and super smart robotics could do to knowledge workers what industrialization and factory automation did to manufacturing workers. While on the one hand, productivity and efficiency could be exponentially improved, on the other hand, peoples' jobs and livelihoods would be at stake.
- Leaving the decision about how AI evolves thus cannot be left up to the industry and developers alone to tackle. The societal and governance inputs ought to be in place from the get go.

# Proposition - 1

- Considering that the ITU is actively engaged in this field, and recognizing the excellent work on IoT, Trust, cognitive informatics, artificial intelligence and related fields is being carried out in relevant Study Groups, it is suggested that
  - a multidisciplinary experts umbrella group, comprising scientists, academicians, R&D Labs., sociologists, legal professionals, civil society, industry, UN/International/Regional/National Organizations and Government representatives, dealing with the subject may be constituted under the auspices of ITU to deliberate upon and bring out Guidelines for designing super intelligent machines/robots aligned to human interests, rather than having the various groups pursuing separate goals, laudable as they maybe on their own;



## Proposition - 2

- such an umbrella group will bridge the various technology innovations and policy and ethical needs, to ensure a harmonious outcome of their efforts and put human safety at the forefront.
- An item of particular importance is to come up with the consistent taxonomy of the AI methodology and classification of AI systems by sophistication, in order to have a more nuanced regulatory and policy regime, instead of going with a single universal policy.

**Thank You!**