
ITU

Policy Considerations for AI Governance

Shailendra K. Hajela

ITU-T SG 3: Workshop on *Policies in relation to impact of Artificial Intelligence on ICT services,*

Geneva, 10 April 2018

(The views expressed are of the author and do not reflect the opinions of ITU and IAFI.)

General Understanding of AI (1)

Artificial Intelligence (AI) is the science and engineering of making intelligent machines, especially intelligent computer programs. - John McCarthy, father of AI, Dartmouth, 1956;

- AI refers to the ability of a computer or a computer-enabled robotic system to process information and produce outcomes in a manner similar to the thought process of humans in learning, decision making and solving problems. In a way, the goal of AI systems is to develop systems capable of tacking complex problems in ways similar to human logic and reasoning.
- A straightforward, consensus definition of AI is not yet there. It is best understood as a set of techniques aimed at approximating some aspect of human or animal cognition using machines.
- AI is a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason, and take action.

General Understanding of AI (2)

- While the rate of AI development has not grown steadily since its inception sixty years ago remaining confined mostly to the realm of scientific research in academia, science fictions and movies, the latter serving to increase general awareness of AI;
- But with the stupendous growth in computational power, Data analytics, leap-frogging of access to telecommunication the world over around the turn of the century, through sector reform leading to massive private sector investments in Network infrastructure and Services, Cellular mobile wireless technology 4G/5G, Broadband Internet, affordable smart phones, etc., AI is the buzz word now ushering in the 4th Industrial Revolution.
- Natural Language Processing (NLP) and knowledge representation and reasoning have enabled the AI-machines to beat the champions of Chess, Jeopardy and Go, and are bringing new power to searches on the Web, though great achievements, these technologies are highly tailored to particular tasks.

General Understanding of AI (3)

- Because AI has the ability to learn, think, reason out and as a result, provide outcomes that were not programmed or predicted by its creators, as it infiltrates our homes and the workplace, our corporate and government networks, the IT Service Management (ITSM) organizations will be responsible to keep these systems up and running. According to Gartner, notwithstanding that AI offers improved experience to customers at every point of interaction, without human governance, this may be squandered.
- As we look further into the future, advancements in AI technology will probably negate the need for human governance, but that could be addressed at appropriate time.
- AI and robotics will also be applied across the globe in industries struggling to attract younger workers, such as agriculture, healthcare, food processing, hospitality, and factories. They will facilitate delivery of online purchases through flying drones, self-driving trucks, or robots that can get up the stairs to the front door. [ai100 Stanford Report]

Concerns about AI (1)

- Alongside the growing interest in exploiting the potential of AI for human good, there are also concerns expressed about its unethical use. Entrepreneur/inventor Elon Musk of Tesla and SpaceX, who spends considerable time on the cutting edge of technology, is reported (Feb. 2015) to have warned that AI could be an existential threat to humanity.
- Bostrom and Yudowsky have studied and analyzed at length the ethics of AI, and emphasized that thinking machines are not as versatile as humans and can have only domain specific intelligence as per its design, suitable for the assigned task, but unsuitable for others. However, the possibility of developing super intelligent machines exists. They go further on to discuss the moral status of such machines themselves.
- Also, AIs with sufficiently advanced mental states may count as persons—though maybe persons very much unlike us and perhaps to be governed by different rules?

Concerns about AI (2)

- Artificial intelligence is providing beneficial tools for everyday use by people around the world. Its continued development, guided by certain principles, will offer amazing opportunities to help and empower people in the decades and centuries ahead.
- At the same time it can be a threat to humanity and not everyone, entrepreneurs, scientists, eminent persons and visionaries agree on the direction AI may take if left to itself.
- Asimov's Laws for conduct of Robots that hold human safety as the prime consideration are well known and are more relevant than ever before. There is evidence to suggest that AI may threaten humans simply by positive feedback leading to endless unbridled AI power.
- To keep humanity out of the harm's way it is necessary to put in place Policy guidelines for AI- governance to harness its huge potential for public good and standards including ethical from concept/design stage itself, in order that AI devices may think as humans and the AI is aligned to human values so as to build in adequate safeguards against threats that this new paradigm poses.

Guiding Principles for AI development

- Like in 1975, a group of geneticists gathered in Asilomar, a small central Californian coastal town to decide if their work on genetic engineering, manipulating DNA to create organisms that didn't exist in nature would bring about the end of the world, and laid down guidelines comprising a strict ethical framework, in January 2017 the world's top AI researchers discussed this rapidly accelerating field and the role it will play in the fate of humanity.
- The Group established certain principles for collaboration among researchers and linkage with policy makers;
- Goal of research should not be to create undirected intelligence but beneficial intelligence, and safety, transparency, responsibility, value alignment, privacy, liberty, shared benefits and shared prosperity, avoidance of AI arms race, should be paramount considerations.
- Super-intelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than of one state or organization. Concentration of digital power should be avoided for upholding freedom and democratic ideals. These principles among others are useful for policy makers.

Key Policy Issues (1)

- **Checks & Balances:** Justice AI, like an AI ombudsman to resolve undesirable AI activity before human intervention may be required?
- **Explainability:** Currently, model predictions, especially with Deep Learning, Reinforcement Learning with associated complex architectures with large number of layers, millions of parameters, make predictions in a way that is hard for humans to understand. DARPA – Explainable AI (XAI) project - understanding how each of the ML layers are generalizing knowledge of increasing more complex understanding.
- **Fairness and bias:** The AI will be trained on people's behaviors so it will learn to amplify people's reactions and biases. This is not necessarily good. For example if there is a subconscious bias based on skin color the AI will pick that up from the data and amplify it unless there are controls. In a world filled with biases based on color, race, gender, religion, sexual orientation, we do not want to have industrialized inference engines that can promote biases.

Key AI Policy Issues (2)

- **Privacy and data sharing**: The current success of ML owes to 3 factors coming together - the algorithms, the hardware and the data. The more data is used to train the models, the better the accuracy. Given that the algorithms are readily available, the GPU hardware is equally available (especially with public cloud), the key differentiator is data.
- This is the main reason for companies such as Google, Facebook to offer free photo storage. We are freely giving up our data to large scale web companies, in exchange for ease of use. This data is acting as fuel for the incredibly powerful models being created - some of which are at superhuman level accuracy.
- **With image recognition capable to recognize up to 10m people in a given instance, we run the risk of having stalker-as-a-service and paparazzi-as-a-service**.

Key AI Policy Issues (3)

- **Security**: Wearables and other IOT devices reveal locations and other facts about objects they are embedded in. Reverse engineering allows deduction of information unavailable. For example data from wearables emitting from a military base can provide troup strength or equipment inventory.
- **(Non)Repudiation – impersonation**: AI makes it possible to take a few snippets of someone's speech and do text-to-speech using the tone, voice, accent of that person [LyreBird with oversight from ML visionary Yoshua Bengio]. There are examples of Obama and Trump on their website. The idea of such startup is to highlight the power of AI to impersonate people. What would happen if a prank call (which has actually happened) where a politician or a noted personality is impersonated and causes a national security issue? We could include markers similar to the invisible dots in printouts (to prevent currency from being printed) to have the appropriate distinction between AI generated and human generated data.

Key AI Policy Issues (4)

- **Employment:** As with all disruptive technologies, there is bound to be impact. This is a case where white collar jobs including high income professions such as lawyers, doctors, computer engineers can find themselves replaced by AI for routine work initially and over time for very complex work that cannot be done by humans.
- **AI and automation** replacing humans from jobs ranging from truck-driving to parts of medical and legal practices, will cause large scale unemployment, unless there is a wave of new jobs that we don't yet know about (similar to when the industrial revolution eliminated a lot of agriculture and manual jobs).
- Many years ago, when computers came on the scene, many office and factory workers thought that they were going to lose their jobs and threatened strikes, but soon realized that it made their lives more comfortable.

Key AI Policy Issues (5)

- **Jobs in the future** might be completely different but any unemployment as a consequence of AI would need to be carefully addressed in the Policy for AI Governance.
- **IT & ITeS job cuts forecast:** With the rise of AI, preparing itself for scaling up manufacturing sector is essential for countries like India to meet the demand for jobs from a dramatic surge in working age population, according to Paul Krugman, who won the Nobel Prize in economics in 2008 for his work on international trade theory.
- While speaking recently at a News18 event, Krugman has warned that **India could end up with huge mass unemployment if it does not grow its manufacturing sector**, just like its service sector. India's IT & ITeS services annual earnings of about USD 60 billion will be severely eroded by such jobs enabled by AI getting done domestically that outsourced such jobs.
- In future, while diagnosis may be outsourced to a doctor in India, it could also go to a firm based on artificial intelligence.

Key AI Policy Issues (6)

- **Lack of interpretability**: When the AI is trained on the data it is not exactly interpretable - people cannot inspect the AI and understand how it will behave and decide in hard situations. For example if a self driving car can either run over a child or cause an accident that can kill all the car occupants how should it decide?
- **Equitability:** AI is the next technology frontier which has the potential to further widen the gap between the haves and have-nots. Compared to the stupendous national and international efforts to bridge the digital divide, far more intense efforts may be warranted to minimize the disparity and that would need to be addressed seriously at the policy level.
- **Performance for Real Time inference**: Especially in Real-time use life-and-death cases such as autonomous driving, robotic surgery: there have to be policies and guidelines in place to minimize injury and death.

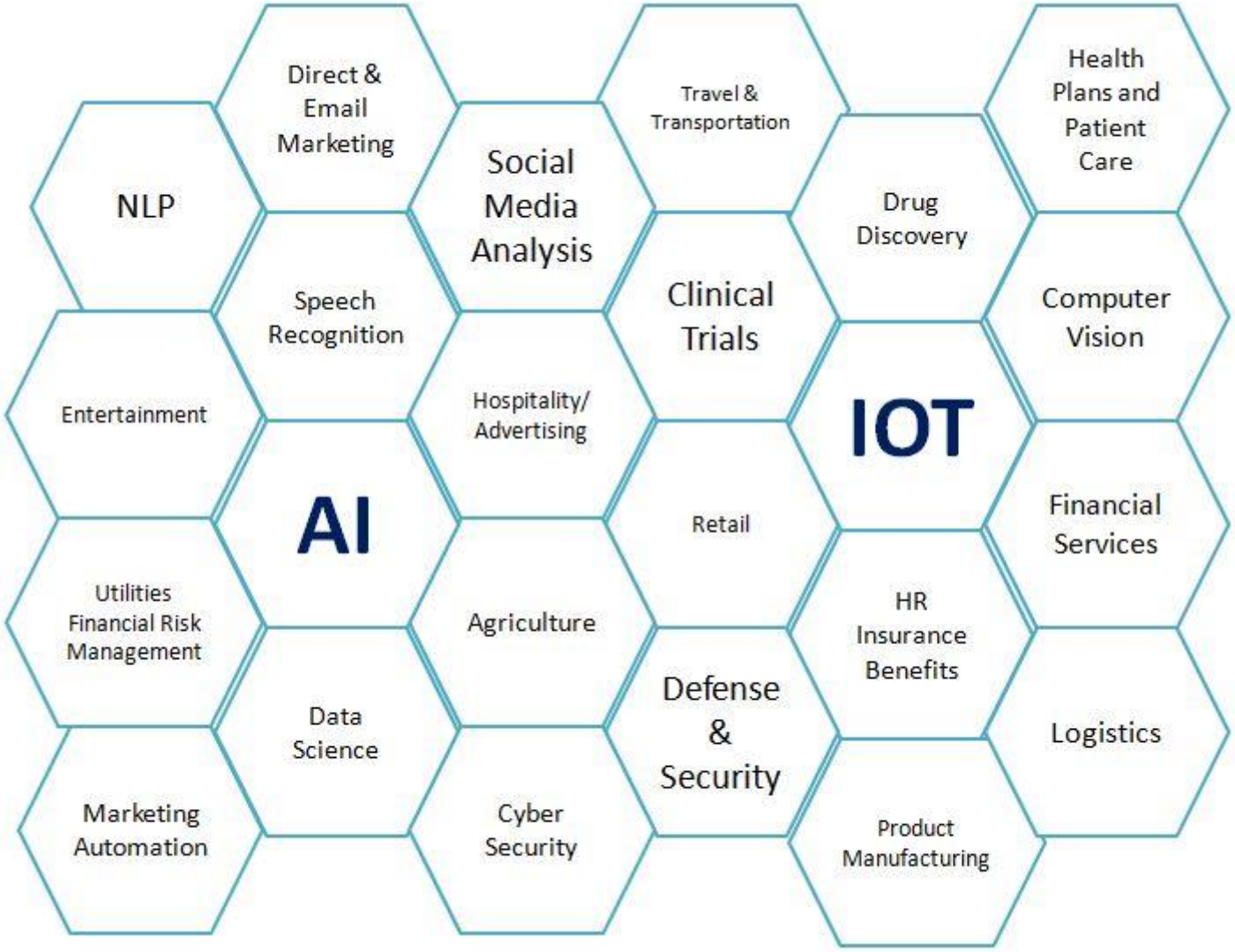
Key AI Policy Issues (7)

- **Adversarial challenges**: Recently, image recognition models have been fooled by adversarial patches - by simply putting such a patch on a known object, such as STOP sign, the machine model can be fooled into thinking it's a banana. Similarly, as many would have heard of the latest Apple iPhone X's face recognition being fooled using a mask. While there's no 100% fool proof detection of such adversarial challenges, this shows the need to keep sight of such issues (which can be considered similar to virus or malicious attacks) and have approaches (including model ensembles - with multiple models making predictions - to reduce the possibility of such attacks).
- **Critical Evaluation:** AI is being used for monitoring verification of the nuclear test treaty to distinguish between natural seismic tremors and shocks triggered by nuclear tests.[UC Berkeley];
- **Democratization** – meaning accessibility and consummability of AI to all.

Key AI Policy Issues (8)

- **Massive interdisciplinary collaboration:** AI societal impact extends over multiple disciplines in cyber as well as physical space needing massive interdisciplinary collaboration.
- **Focus Areas:**

  Natural Language Processing; Speech Recognition; Direct & Email Marketing Social Media Analysis; Marketing Automation; Entertainment; Hospitality; Advertising; Retail; Clinical Trials & Drug Discovery; Recruiting; Computer Vision; Health Plans & Patient Care; Aerospace & Defense Business Intelligence; HR Insurance Benefits; Utilities Financial Risk Management; Data Science ; Cyber Security; Product manufacturing; Defence; Disaster management and recovery; Logistics; Financial services; Travel and transportation; Agriculture


NLP
Direct & Email Marketing
Social Media Analysis
Travel & Transportation
Health Plans and Patient Care
Drug Discovery
Speech Recognition
Clinical Trials
Computer Vision
Entertainment
Hospitality/ Advertising
AI
IOT
Retail
Financial Services
Utilities Financial Risk Management
Agriculture
HR Insurance Benefits
Data Science
Defense & Security
Logistics
Marketing Automation
Cyber Security
Product Manufacturing

Fostering Innovation

- Fostering a culture of innovation and research and to encourage innovation in AI research efforts and initiatives to build user communities in the field of AI will go a long way. Examples from around the globe include the DARPA's Cyber Grand Challenge which attracts a large share of AI research funding in the US, the European Union's technology funding programme, FP7, and the BRAIN initiative, a 10-year, multi-billion dollar funding initiative for AI research in the US.
- Also, the role of an AI system, as in the case of a driverless car, could be to assist the user. In such a situation, deciding liability for what the AI system has done will be difficult and need to be discussed and delved into deeply before arriving at any conclusion. The country level policies and plans for digital transformation have created data which is readable by machines. At the same time, technologies have also reached a level of maturity where they can think like humans in real time and, at times, in a cost-effective way.

Planning for future skill demands and readiness of workforce

- The national policy needs to define standards and benchmarks that can be effectively used to gauge progress in AI innovation and commercialization in a host of application domains. By nature, the AI space has no direct traceability of returns from investment in innovation and capability building. This makes it all the more important for intermediate tangible progress to be measured against set targets from time to time.
- A strong presence in AI R&D is a prerequisite for a nation to gain a lead in an ai-automation-driven future. The national policy needs to take into account the current and future demand forecast for AI experts. For building expertise, on the other hand, will require the evaluation of the current educational institutions and curricula and overhaul the same, if necessary, to provide skill up-gradation initiatives for a workforce in order that it stays relevant in a fast-evolving technology landscape.

Thank you for your attention