|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-G-026 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | New Delhi, 13-15 November 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | Ministry of Communications (India) | | |
| **Title:** | | Ethical issues in AI for Health | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Deepa Tyagi, DDG(FN) Telecommunication Engineering Centre, Department of Telecommunication, Ministry of Communications, Government of India, India | | Tel: +91 11 23329089 Fax: +91 11 23329089 E-mail: ddgn.tec@gov.in |

|  |  |
| --- | --- |
| **Abstract:** | Artificial Intelligence (AI) - one of the truly life transforming technologies - has revealed an enormous potential in the field of Health. But for all of AI's potential in healthcare, it is an awesome responsibility to manage machines that can learn and make decisions without human direction. While Standardization Development Organizations like ITU, ISO/IEC, and IEEE SA have begun work on certain AI related areas however, it is imperative to lay down ethical foundation for using AI technology safely / effectively, creating open process for codifying rights / regulations around issues such as privacy, security, trustworthiness, robustness, transparency and above all, ethics. The rise of AI carries moral, legal and security questions. Development of a thoughtfully designed, high-quality and clinically validated AI technology for medical system is essential. Ultimately there will always be a human element in the practice of medicine. AI technology will most likely be a complementary tool and not a replacement for a physician. The essence should lie in becoming not overpowered but empowered by the technology. |

**1. Background:**

World leaders at the UN adopted the Sustainable Development Goals, also known as "Transforming our world: the 2030 Agenda for Sustainable Development." Therein, the UN made reference to technology as one of the major pillars of implementation of the 17 goals of the SDG as technology can help us achieve the development goals and have a massive impact on our world and our lives. *Chief among the life transforming technologies is Artificial Intelligence (AI).*

Advances in data collection and aggregation, algorithms, and computer processing power have enabled scientists and engineers to make great strides in developing AI.

Suddenly machines can perform tasks that once required human cognition. In the past, computers could execute only the rigidly defined tasks for which they were programmed. Now they can be given a general strategy for learning, enabling them to adapt to new data without being explicitly reprogrammed. Many such "machine learning "systems already have been put to commercial use.

AI gives innovators the tools and capacity to innovate and to use AI to help solve some of our greatest challenges, starting with the SDGs (poverty, hunger, health, gender equality, education, the environment and others).

AI refers to the ability of machines to exhibit humanlike intelligence e.g. the ability to solve a problem without requiring detailed human - developed software. By reviewing voluminous data sets for patterns, machines can "learn" to perform tasks, such as identifying images, recognizing speech, identifying relevant information in texts, synthesizing information, drawing conclusions, and forecasting. As AI's capabilities have dramatically expanded so has its utility in a growing number of fields, across sectors. Over the past decade, considerable resources have been allocated to exploring the use of AI for health, which has revealed an enormous potential.

But, for all of AI's potential, it is an awesome responsibility to manage machines that can learn and make decisions without human direction. These technologies are taking societies into uncharted territory. AI adoption touches on questions of social values that do not have any one –size –fits –all answers. It is vital to create an open process for codifying rights and regulations around issues such as privacy and security as well as trustworthiness, robustness, transparency, and above all ethics. ***Thus, in many ways , AI is just as much a new frontier for ethics and risk assessment as it is for emerging technology*** .

**2. Current Landscape for AI Standards:**

**ITU:** It has historically played a role in standards for information and communications technologies, particularly in telecommunications. It has a Focus Group on Machine Learning for Future Networks that falls within this telecommunications remit. Following the 2018 AI for Good Global Summit, it has also created a Focus Group on AI for Health, "which aims *inter alia* to create standardized benchmarks to evaluate Artificial Intelligence algorithms used in healthcare applications."

**ISO/IEC**: Recognising the importance of international standards harmonisation in addressing, managing and regulating new areas of technology, the ISO and the IEC Joint Technical Committee 1 (JTC 1) created Subcommittee 42 – Artificial Intelligence (SC42), in 2017 that has nine standards under development, focused variously on terminology, reference architecture and, more recently, trustworthiness. The Trustworthiness working group is currently drafting three technical reports on robustness of neural networks, bias in AI systems, and an overview of trustworthiness in AI. SC 42 committee is also driving work on the governance of AI within organizational settings, to ensure the responsible use of AI.

(ISO/IEC NP TR 24368 is an overview of ethical and societal concerns in AI)

**IEEE Standards Association**: Its AI standardization processes are part of a larger IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. IEEE has released a number of documents regarding the ethical development of AI through their Global Initiative on Ethics of Autonomous and Intelligent Systems, where they consulted across some areas of industry, academia, and government.

The IEEE sets out five core principles to consider in the design and implementation of AI and ethics. These include adherence to existing human rights frameworks, improving human wellbeing, ostensibly to ensure accountable and responsible design, transparent technology and the ability to track misuse. IEEE SA recently launched the development of an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS). ECPAIS seeks to develop three separate processes for certifications related to transparency, accountability, and algorithmic bias.

(P 2801 pertains to Recommended practice for the Quality Management of Datasets for Medical AI;

P 2049.4 pertains to Methodologies and Processes for Ethical considerations)

**3. Key Challenges**

Adoption of AI is not always a simple proposition and no single organization acting alone can solve all of the issues surrounding these technologies.

Creating robust data ecosystems for AI technologies must be fed a steady diet of reliable, actionable and secure data; it is how they "learn" and function. Everything that we do and the key differentiator in everything that is done with Analytics and Results starts and ends within the data.

**(i)** 5 Rs of data:

Ideally the Data should be:

**Rich** - sufficiently prescribed richness is required to enable monitoring, detection, prediction or inquiry of potential questions, patterns and trends and the scrutiny to understand. Dimensional richness, attribute richness and context richness around key items of interest /objects of inquiry is critical for speed and quality of pattern detection and problem resolution.

**Relevant** –with many "big data initiatives" working from the bottom up, there can be a tendency to have a lot of irrelevant data that serves no useful purpose. On the other hand, relevance is subjective, purpose and time based, depending on the quest of the inquirer. It must therefore be available "on tap" to serve its purpose on demand. Data must be context based so that it may become "relevant" "on demand" to pursue a line of inquiry or support a specific problem. Thus there is a need for broad and deep dataset around themes and inter-related themes to support multi-faceted aspects of relevance.

**Reliable** –dependability and ultimate trust; the more powerful the proposition /insights generated, the more impact they will have; the more contention there is, the more reliable they must be or they will be dismissed, used to distort other behavioural pattern as reliability will be brought into question –and an undesired course of action can ensue. "Reliability" is one key attribute, the user will home in on to either embrace or dismiss the insight.

**Robust** -to secure ongoing quality and utility of the solution, data robustness is necessary. Key to exploiting Robustness is to ensure a sufficient "operating framework" for analysis which cuts through users, technology, governance and process.

Not factoring these in or over emphasis on the wrong pillar will lead to non –robustness or fragility; the end- to- end solution must strive for 'anti-fragility.

**Reconcilable –** ability to reconcile drives trust and with trust we get action. "Acting on insights" are systematically addressed with reconciliation and validation procedure. Key to addressing this aspect is to work from the bottom up with the lowest common denominator possible.

**(ii)** While the data is key to all AI systems, how the data is collected and the quality of data is dependent on the quality of the sensors. There has to be a strong and fool proof system of testing and certification of the various sensors used for data collection.

**(iii)**Network bandwidth /network availability /latency /SLA for the transmission of the data collected at site to the central control room /AI platform is also critical. Advances in communications technologies such as 5G /network slicing etc. should be helpful in achieving these objectives in future.

**(iv)**Standardization of Data formats, data porting across AI platforms, assessment /grading of AI (ML) algorithms for various parameters such as trustworthiness /robustness etc. is yet another challenge.

**(v**) Methodologies may have to be arrived at for measuring the AIQ (Artificial Intelligence Quotient) of the Machine /Platform on similar lines as human Intelligence quotient (IQ).

**(vi)** It is well understood that with its robust ability to integrate and learn from large sets of clinical data, AI can serve roles in diagnosis, clinical decision making and personalized medicine, none-the- less, this powerful technology creates a novel set of ethical challenges that must be identified and mitigated since AI technology has tremendous capability to threaten patient preference, safety and privacy. Anticipating potential ethical pitfalls, identifying possible solutions and offering policy recommendations will be of benefit to physicians adopting AI technologies in their practice as well as the patients who receive their care.

**4. Some key ethical issues**

Considering the number of times our computers have crashed /got hacked /virus infected, a lot needs to be invested in AI Safety research because as we put AI in charge of even more decisions, we need to figure out how to ensure truly trustworthy robust AI systems that do not malfunction and harm us or get hacked and be turned against us. AI safety work has to include work on AI value alignment i.e. AI accomplishing goals that are aligned with ours; we need to figure out how to make machines understand our goals, adopt our goals and retain our goals.

The rise of AI carries moral, legal and security questions some old, some new: could intelligent machines attack vital power and health or act as a channel for others to do so?

Although it is comforting to imagine AI algorithms as completely emotionless and neutral, but that is simply not true. AI Programs are made up of algorithms that follow rules. They need to be taught those rules and this occurs by feeding the algorithm with data, which the algorithm then uses to infer hidden patterns and irregularities. If the training data is inaccurately collected , an error or unjust rule can become part of the algorithm- which can lead to biased outcomes .If even a few of our data sources were biased ,if they contained information on sex ,race , colour or ethnicity ,or the data was collected that did not equally represent all the stakeholders ,we would unwittingly build bias into our AI .Poorly collected data may lead to erroneous conclusions that may not be valid and individuals may be disadvantaged as a result .There have been instances in recent years that showed even algorithms can be prejudiced as biases sometimes creep in on programs because human bias influenced those algorithms when they were being written; there could be instances where the machine learning algorithms might not provide equally accurate predictions of outcomes across race ,gender or socio economic status .There could also be instances where 2 different ML Algorithms , using the same set of data ,come out with differing outcomes. Occasionally biases can creep in without intention also. There will be a need to take action to mitigate and prevent them through awareness, motivation, transparency and ensuring we use the right data.

In the much touted robotic surgery and use of AI robot during surgery, the importance of proper informed consent of the patient cannot be overlooked. It is absolutely important that the potential harms related to the use of AI technology must be transparent to all involved including the patient.

Another key area of attention is in the reframing of the medical education from a focus on knowledge recall to a focus on training students to interact with and manage AI machines. This reframing would also require diligent attention to the ethical and clinical complexities that arise among patients, caregivers and machines. The use of Artificially Intelligent virtual patients (VPs) in medical education needs to be explored too. With their exciting applications in teaching medical history taking, such as in psychiatric intake evaluation, VPs offer a readily accessible platform with several benefits over traditional standardized patients; however, the disadvantages and shortcomings are equally important, emphasizing the need for clarity about the role of VPs in medical education.

Further, another area of attention is the legal and health policy conflicts that arise with the use of AI in health care. There is an immense likelihood of medical malpractice and product liability that arises with the use of "black – box" algorithms because users cannot provide a logical explanation of how the algorithm arrived at its given output. There is a policy gap governing the protection of patient photographic images as they apply to facial recognition technology, which could threaten proper informed consent, reporting of incidental findings, and data security. Considering all these, the American Medical Association's recently adopted Policy on AI in health care calls for the *development of thoughtfully designed, high-quality, and clinically validated AI technology, which can serve as a prototypical policy for the medical system.*

The ASILOMAR Conference on Beneficial AI, in 2017 organized by the Future of Life Institute where more than 100 thought leaders and researchers met to address and formulate principles of beneficial AI came out with a set of guidelines for AI research – the 23 Asilomar AI principles. Beneficial AI works on the shared benefit principle that AI technologies should benefit and empower as many people as possible; and the essence should lie in becoming not overpowered but empowered by the technology.

There is no doubt that AI will have widespread ramifications that revolutionize the practice of medicine, transforming the patient experience and physicians' daily routines. Nonetheless, there is much work to do in order to lay down the proper ethical foundation for using AI technology safely and effectively in health care. Ultimately, patients will still be treated by physicians no matter how much AI changes the delivery of care, and there will always be a human element in the practice of medicine. AI technology will most likely be a complementary tool and not a replacement for a physician.

Quoting Max Tegmark, founder of Future of Life Institute," just as with rocketry ,it is not enough to make our technology powerful ,we also have to figure out how to steer it and where we want to go with it".

**5. Proposal**

In many ways, AI is just as much a new frontier for ethics and risk assessment as it is for emerging technology. It is important to foster a broad and ongoing public dialogue across the Governments/‌Corporate Health and Social sectors to build consensus around AI governance and to ensure that it is used to deliver long –term social and health benefits. This debate is critical for ensuring that public investment in technology yields the kind of outcome that the public actually supports. It is also vital to create an open process for codifying rights and regulations around issues such as ethics, privacy and security.

This contribution is being submitted for the information and benefit of member states .It may be taken as an input while discussing the various issues related to regulatory and ethical considerations on AI for health under the standardization work plan of the new ITU-T Focus Group on AI for Health.

**6. References**

* journalofethics.ama-assn.org
* Asilomar conference on Beneficial AI ,2017
* Future of Life Institute
* World Economic Forum –ASEAN Summit 2018
* McKinsey Insights

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_