AI Ethics: The Next Big Thing In Government

Anticipating the impact of AI Ethics within the Public Sector

Ali Hashmi
February 2019

WORLD GOVERNMENT SUMMIT 2019

in partnership with

Deloitte.

Answering Tomorrow's Questions Today

The World Government Summit is a global platform dedicated to shaping the future of governments worldwide. Each year, the Summit sets the agenda for the next generation of governments with a focus on how they can harness innovation and technology to solve universal challenges facing humanity.

The World Government Summit is a knowledge exchange center at the intersection of government, futurism, technology, and innovation. It functions as a thought leadership platform and networking hub for policymakers, experts and pioneers in human development.

The Summit is a gateway to the future as it functions as the stage for analysis of future trends, concerns, and opportunities facing humanity. It is also an arena to showcase innovations, best practice, and smart solutions to inspire creativity to tackle these future challenges.

Table of Content

Topics

Despite remarkable achievements, the rapid development of AI has raised a host of ethical concerns. Governments face challenges and choices pertaining to how to apply AI technologies in the public sector and in governance strategies.

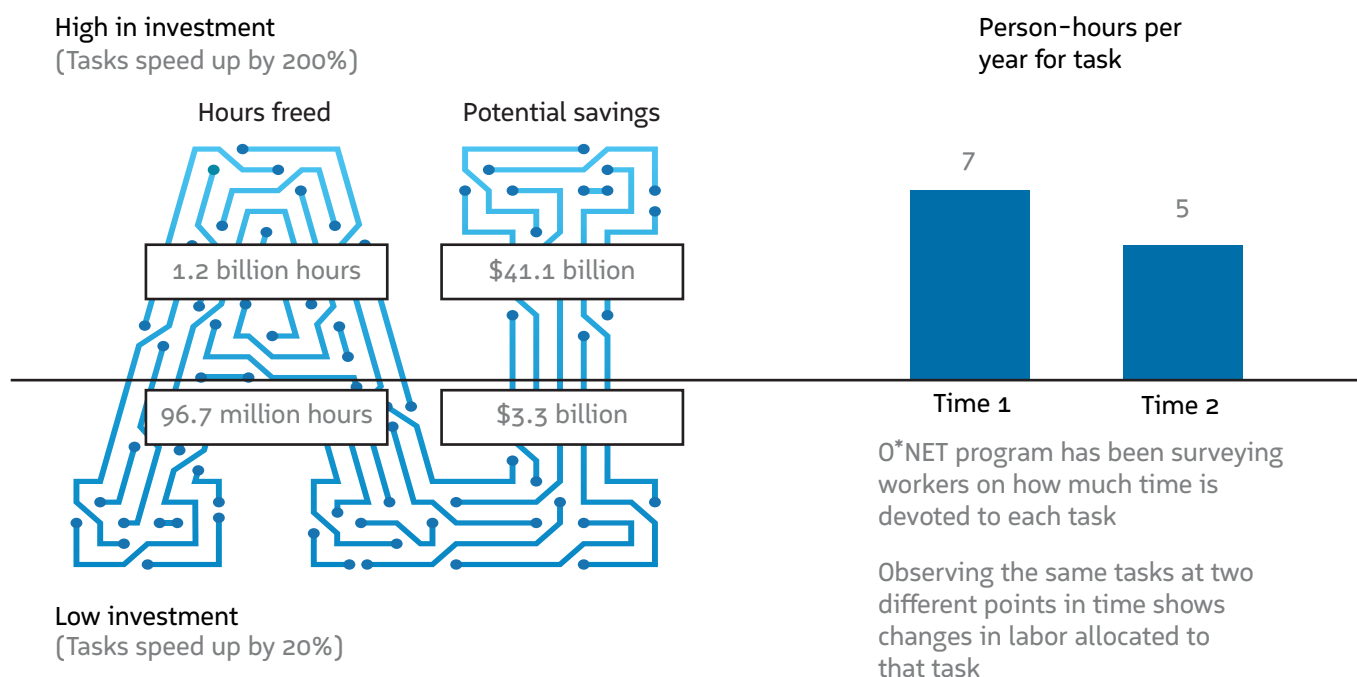# Setting the Context – Ethics at the heart of AI

According to Deloitte's 2018 Global Human Capital Trends report, 42 percent of surveyed executives expect Artificial Intelligence (AI) – intelligent machines that imitate human behavior – may be widely deployed in their organizations in the next three to five years. The public sector follows the same path seeking and adopting applications to improve public services and manage the growing difficulty of analysis and decision making by effectively exploiting increasingly available amounts of information. Through cognitive applications, AI already helps governments reduce backlogs and cut costs, predict fraudulent transactions and identify criminal suspects via facial recognition. By adopting AI for automation, governments can focus on more creative and complex aspects of service delivery to citizens.

Despite remarkable achievements, the rapid development of AI has raised some concerns being a subject of fear and skepticism in the media. Could AI long-term development lead to the end of humankind as Elon Musk, Bill Gates and numerous technologists have speculated? What is the role of ethics in the design, development and application of AI? How will ethics help maximize the benefits of AI to increase citizen well-being and common good?

AI- augmented government

Artificial intelligence looks to make all of government more efficient by automating and improving routine tasks allowing public sector employees to spend fewer hours on noncore tasks and more on innovation. For instance, finance ministries in the Gulf region are using machine learning for detecting complicated fraud scenarios; public welfare organizations are employing machine learning to disburse welfare payments in a more efficient and equitable manner; numerous public-facing entities have deployed Chatbots to interact with citizens; health organizations are employing AI for triaging health care cases; and police agencies are using facial recognition AI for improving its surveillance capabilities. In the 2017 Deloitte report AI-augmented government, our analysis of the US Public Sector found cognitive technologies could save up to 1.2 billion hours and potential annual savings of $41.1 billion[1] as indicated in figure 1 below.

FIGURE 1. HOW MUCH SAVINGS CAN AI IN GOVERNMENT GENERATE?

High in investment
(Tasks speed up by 200%)

Hours freed | Potential savings

1.2 billion hours | $41.1 billion

96.7 million hours | $3.3 billion

Low investment
(Tasks speed up by 20%)

Person-hours per year for task

O*NET program has been surveying workers on how much time is devoted to each task

Observing the same tasks at two different points in time shows changes in labor allocated to that task

Source : Deloitte analysis

AI Ethical Dilemmas

While there is an increasing interaction between AI technologies and our socio-political and economic institutions, consequences are not well defined. The advent of AI raises a host of ethical issues, related to moral, legal, economic and social aspects of our societies and government officials face challenges and choices pertaining to how to apply AI technologies in the public sector and in governance strategies. From Uber's self-driving car fatality to Amazon's gender biased recruitment tool, examples of AI ethical concerns abound and reinforce the idea that they should be taken into account before an AI system is deployed. In this perspective, "ethics" can be defined by the pursuit of "good" actions based on "good" decision-making —decisions and actions that lead to the least possible amount of unnecessary harm or suffering.[2] It implies that our government and business leaders understand and define what "good" means for AI systems. Gaining societal consensus on the ethics of AI is one of the key tasks of the government.

As per the Deloitte survey, 32% of respondents ranked ethical issues as one of the top three risks of AI while most organizations do not yet have specific approaches to deal with AI ethics.

A recent survey[3] by Deloitte of 1,400 U.S. executives knowledgeable about AI identified that one of the biggest challenges facing AI is around the ethical domain. As per the survey, 32% of respondents ranked ethical issues as one of the top three risks of AI while most organizations do not yet have specific approaches to deal with AI ethics. For instance, how do we ensure that AI systems serve the public good rather than exacerbate existing inequalities? There is a big gap between how AI can be used and how it should be used. The regulatory environment has to progress along with AI which is rapidly transforming our world. Governments and public institutions need to start identifying the ethical issues and possible repercussions of AI and other related technologies before they arrive. Objective is twofold:

- First to properly manage risks and benefits of AI within the government for an **AI augmented public sector**;
- Second to develop smart policies to **regulate AI intelligently** and secure it benefits for the society and economy.
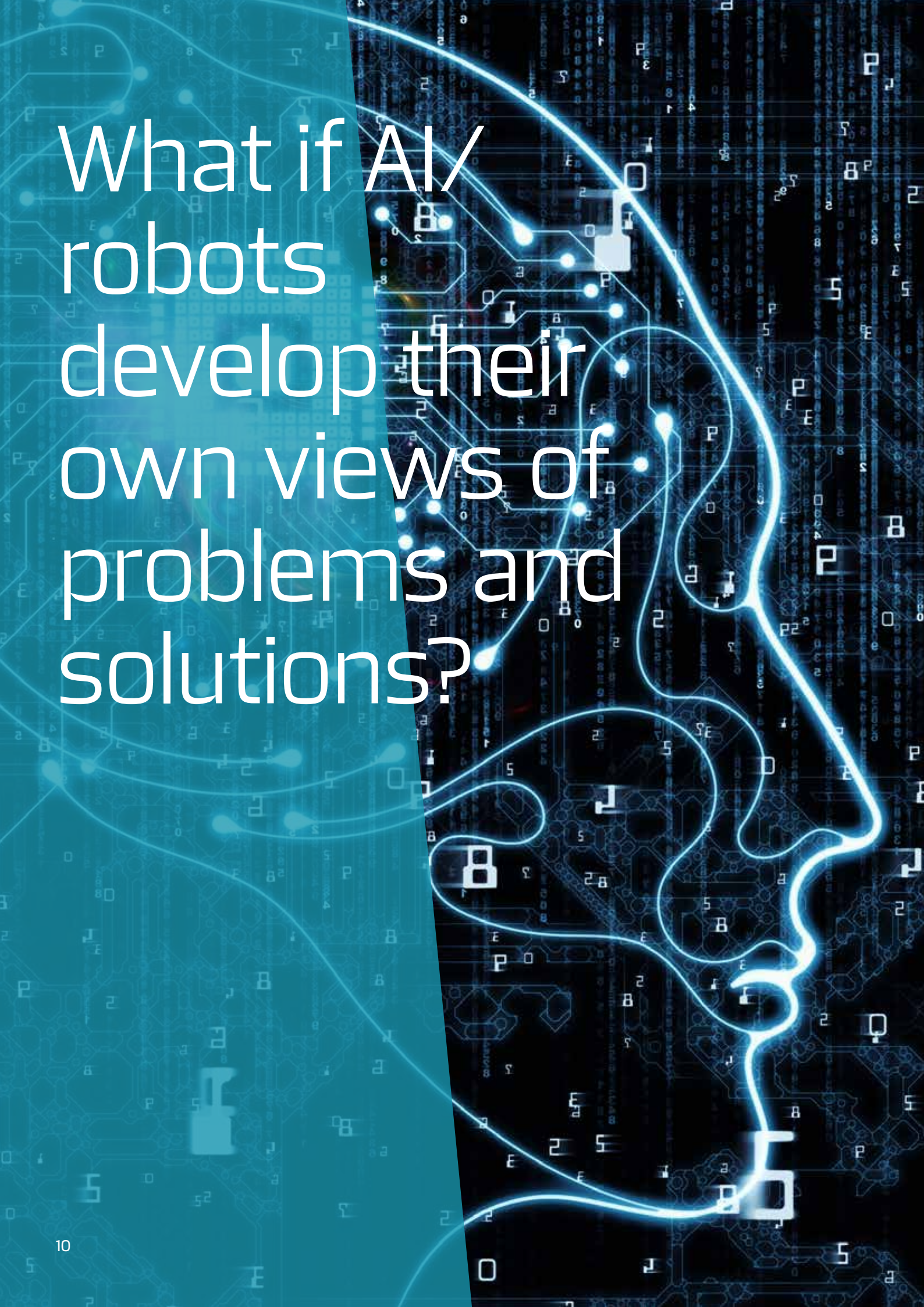
AI systems' behavior should reflect societal values. Gaining societal consensus on the ethics of AI is one of the key tasks of the government.

Below are some important questions to consider by the government and public sector at large to approach AI ethical challenges and opportunities. These are not easy questions to answer. In the pages that follow, we will explore the interplay between ethics and emerging technologies like AI. We hope this paper will help trigger conversations that lead to action to tackle these important issues.

What if AI/ robots develop their own views of problems and solutions?

AI Ethical considerations for Public Sector

- **Regulatory and Governance:** What are the principles of governance that governments should adopt as part of anticipatory regulation? How do we allow the development of AI applications for the public good? What is the moral status of AI machines? What properties must a machine have if it is seen as a moral agent? Who is liable for decisions that AI and robots make? How do we bring transparency in the implementation of AI algorithms to prevent encoding of bias in machine decisions?

- **Legitimacy and non-repudiation:** How do we ensure the AI we are interacting with is legitimate? How do we know that training data are legitimate? Are we sure decisions are made by the proper AI agent? (Principle of nonrepudiation).

- **Safety and Security:** Does AI warrant a new science of safety engineering for AI agents? How do we ensure that machines do not harm other humans? Who will cover in case of damage? Will an accident caused by our robot make me responsible?

- **Socio-economic Impact:** How do we prevent job losses caused by AI intrusion in work place? What are the social and moral hazards of predictive profiling? Will humans reach a point where there is no work for us due to AI? Will humans do different type of jobs? Will the society become a jobless society, as described by many authors and how will the governments tackle it?

- **Morality:** Do we have the right to destroy a robot? Is a robot the property of a human or belongs to public wealth? How could we control a system that has gone beyond our understanding of complexity? What if AI/robots develop their own views of problems and solutions? Should a robot/AI have its own digital identity that allows it to own decisions, assets and other things? Should they have legal status and rights?

AI pushes us to ask bigger questions such as what is the future of humanity, society and work.

At the 2016 World Economic Forum, top technology/ AI experts have highlighted key AI related ethical challenges that should frame any AI conversations.[4] Some of the points are related to the redefinition of our humanity vs. AI rights: How do we anticipate and control the impact of machines on our behavior and interaction? How do we define rights of AI vs. rights of humans? AI pushes us to ask bigger questions such as what is the future of humanity, society and work.

The surge of AI Ethics across the world

There has been an increasing interest in the global academic, corporate and government community to develop ethical frameworks for maximizing AI benefits while minimizing its risks. A few examples are listed below:

Academic institutions
Launched by Harvard Law School's Berkman Klein Center, together with the MIT Media Lab, the $27 million – Ethics and Governance of AI initiative – aims at developing new legal and moral rules for artificial intelligence and other technologies built on complex algorithms.

Corporate Organizations
Many technology companies have also designed programs that support AI as a tool to create a better society. For instance, Google initiative called "AI for Social Good" and Microsoft's $115m "AI for Good" grant aims to fund artificial intelligence programs that support humanitarian, accessibility and environmental projects. Recently, Microsoft committed $50 million to its "AI for Earth" program to fight climate change.

Public Sector
Over a short period, an increasing number of countries have announced the release of AI ethical guidelines. In December 2018, the European Commission, supported by the High-Level Expert Group on AI released the first draft of its Ethics Guidelines for the development and use of artificial intelligence.[5] At the same time, Canada recently released the Montreal Declaration of Responsible AI, which is a document to guide individuals, organizations and governments in making responsible and ethical choices when building and utilizing AI technology.[6] The effects of AI are almost certain to be very far-reaching; hence, there is a need for governments to delineate the legal, ethical, and regulatory implications of AI through guidelines and code of ethics. Ideally, a global consortium or institution should develop global standards for AI ethics.

Governments are encouraged to develop and implement regulatory and ethical frameworks for AI.

How should governments respond?

Considering there are risks, benefits, and uncertainty associated with AI, there is a particular need for governments to develop and implement regulatory and ethical frameworks for AI. From a regulatory and ethical point of view, public sector organizations should look at developing and implementing a code of ethics that includes injunctions and guidelines to govern AI. Before we discuss the principles that should be the basis of such a code of ethics, we should first outline the philosophical foundations that inform the ethics of AI.

Foundation 1: AI and consciousness

Any questions pertaining to the ethics of AI should first address the ontology of a machine or a program as an object of ethical concern. How do brains work? What are the ethical implications of intelligent programs? Can we say that the program is conscious of what it is doing? Could AI be considered as a human extension?

There is broad consensus among the AI community on the following two assertions:
• AI does not have mental states
• AI does not have personhood

Alan Turing[7] in 1950 developed a simple test to assess the ability of a machine to exhibit intelligent behavior. If the behavior of the machine is indistinguishable from that of a human, then one can argue that machines are intelligent. Many philosophers claim that if a machine passes Turing test, it merely simulates thinking.

John R. Searle,[8] an AI philosopher, argues that AI programs can simulate intelligence through symbol manipulation; however, this symbol manipulation lacks awareness of what the program is doing through this symbol manipulation. Consciousness is a subjective experience caused by physical processes of the brain. We do not know, given what we know of the brain, what is the seat of consciousness in brain; hence, we cannot simulate a mental state in an AI agent except in the form of data that keeps record of program's state.

It is, however, possible that cognitive behavioral capabilities of a machine become indistinguishable from that of a human in certain dimensions. We are already seeing such examples in Chatbot technologies deployed today. When interacting with these Chatbots, we may perceive these Chatbots as human beings but from a cognition point of view, these Chatbots are ontologically different things and we cannot just see them as human extensions. The AI that we see around us optimizes one or many aspects of our intelligence. We do not have AI that exhibits general, holistic intelligence that is capable of integrating various dimensions of our thinking into one. However, it is possible that in near future we can 'simulate' general human intelligence. There are numerous recent examples where AI machines have proven to be better than humans in certain aspects of human intelligence. Google's AlphaGo Zero, a Go program player created without using data from human games, has achieved superhuman performance which is unmatched by any Go player in known human history.[9]

What does it mean from a moral or ethics point of view? From an ethics point of view, we have to think about these entities in terms of personhood if we want to treat them as human beings and bring them under the locus of morality or ethics. A simple litmus test for answering the question whether AI agent has personhood is whether an AI agent can suffer like human beings. Machines do not suffer; they are incapable of suffering. Machine, like a stone or a rock, is not capable of experiencing either pleasure or pain. Hence, machines are not legitimate subject of moral concern. Any ethics around AI machines will be linked with people who create, use and deploy AI technologies.

Foundation 2: AI and Moral Agency

In order to answer the question whether AI agents are moral agents, we need to see if AI agents are free agents.

AI agents and computer systems do not have moral agency because they do not have mental states or intentionality to act freely. Current state of AI replicates human intentionality but this intentionality is encoded. Hence, AI agents cannot be held accountable for their actions. For an AI agent, all laws, under which the agent behaves, are ultimately provided or encoded by the designers. Since AI agents do not possess the ability to take decisions other than what the designers have restricted them to, i.e. restricted locus of operation, these agents are not free agents. Similarly, AI agents have intentionality, i.e. they have the ability to infer how others will respond to new information and what others likely want from the AI agent. However, this intentionality is coded into the AI agent by intentional acts of the designers.

If AI takes over most human tasks and intelligence work, would human brain still have opportunities to learn?

Shaping the governance guidelines for AI

Given we have established that any ethics around AI should focus on the AI creators and the organizations that use AI, the development of a code of ethics for AI could be divided into two broad areas:
A. Macro considerations for deploying AI in an organization
B. Ethical AI systems design

A. Macro considerations for deploying AI in an organization

Since AI challenges the foundations of standard codes of ethics of IT systems because of its evolving nature and manifold applications, there is an urgent need to design and develop codes of ethics to govern it. Macro level concerns that impact the socio-economic fabric should be taken into consideration.

Jobs lost due to AI, robots and automation: There is no consensus on the prediction of jobs created and destroyed by automation and AI; however, overall, there are more studies that predict a net jobs loss than a net jobs gain as a result of AI. The World Economic Forum's Future of Jobs Report 2018,[10] estimates "a net employment impact of more than 5.1 million jobs lost to disruptive labor market changes over the period 2015-2020" due to technology disruptions, including AI and machine learning. In the near term, our analysis suggests, large government job losses are unlikely. But cognitive technologies will change the nature of many jobs—both what gets done and how workers go about doing it—freeing up to one quarter of many workers' time to focus on other activities.[11]

One of the most direct consequences of AI could be the creation of a "useless class" of millions of human beings. When undertaking an AI project, an organization should consider the following questions guidelines:
• What will be the economic impact of a project that will result in job loss on the society as a whole?
• Can we accommodate people who have lost jobs in new roles?

Jobs gained due to AI, robots and automation: Automation has resulted in job loss but on the other hand, it has created newer job opportunities. For example, when we deploy robotic process automation in an organization, new roles are created - bot

manager is needed to supervise and manage the soft robot; a service support prime is needed to provide first line of support for robotic process automation operation; and a change manager is required for managing process change via automation. Most organizations do not take into account the new roles structure that is required for supporting AI and automation operation. As cognitive technologies progress, government agencies will need to bring more creativity to workforce planning and work design. Some of the key questions regarding the new job structure for AI that the organizations should consider are as follows:
• What kind of governance and change management roles are needed for supporting the AI operation?
• What sort of service support structure is required?

The use of AI systems might result in loss of accountability: Legal liability is an important issue for AI systems, especially when it comes to public sector where the liability has no caps. When a medical practitioner in a public hospital uses judgment of an AI-based system for diagnosis, who is liable if the diagnosis is incorrect? Questions pertaining to liability should be answered from the outset to assess the viability of an AI project in a public organization. Some of the key questions regarding accountability are as follows:
• What could be the consequences of an incorrect AI decision, prediction or profiling?
• Who is liable when AI results in faulty behavior?
• If we cannot establish clear lines of liability, should we deploy AI in the organization?
• Shall AI have a digital identity therefore a legal status? Can AI own things? Assets?
• Shall AI pay for its own mistakes?

People might lose their sense of individuality, human-ness: Modern advancements in AI and machine learning allow us to distinguish humor from regular speech; to classify human emotional states using simple webcam images; to generate language like a journalist; and to produce music and art. Such capabilities can have a devastating impact on human self-esteem. When people see human aspects such as humor, speech, and emotional handling, as replaceable by machine, the may feel useless and expendable. It is therefore imperative that organizations should consider the psychological effect of deploying AI in an organization. Some key

questions that organizations should consider when dealing with such issues are as follows:

• What are the psychological impacts of deploying AI in an organization?
• How do we create a symbiotic relationship between humans and AI in the organization?

Organizations should anticipate the birth of artificial general intelligence: The natural evolution of current AI is the development of artificial consciousness or artificial general intelligence (AGI) – the intelligence of a machine that could successfully perform any intellectual task that a human being can do. With the development of such forms of AI, machine itself could become an object of moral concern. The anticipation of such forms of AI should also be part of the formulation of ethical codes for AI. In particular, a risk-centric approach is needed to anticipate the impact of such forms of AI on human societies. Organizations should develop risk-assessment frameworks that should evaluate the sophistication of intelligence in constantly evolving AI systems to minimize unknown risks associated with such evolution. If AI takes over most human tasks and intelligence work, would the human brain still have opportunities to learn? How will this affect the nature of human intelligence?

B. Ethical AI systems design

AI systems should be designed on principles that allow systems to be assessed objectively for transparency and accountability. The following seven principles highlight key areas that should be built into any code of ethics that govern design and implementation of AI systems:

• AI systems have to be explainable
• AI systems have to be transparent
• AI systems have be designed on human-first design principles
• AI systems have to be interpretable
• AI systems have to be designed on common-sense principles
• AI systems have to be auditable-accountable
• AI systems have to be built on unbiased data

Ethical AI systems design: explainable, transparent, human-interpretable
AI algorithms, which are the core of AI agents and

engines have to be intelligible to common users of the AI. If the AI is not explainable, we cannot understand the actions produced by the AI agent or system. This has given rise to the concept of "right to explanation" which is a right to be given an explanation for an output of an algorithm. For instance, if a person has been profiled in a certain category by an AI program, there should be a clear explanation available for such a classification. Currently, there are very limited legal rights that specifically address the notion of right to explanation. However, as we move towards a more AI-based ubiquitous computing environment, the right to explanation will become increasingly important in the legal and social domains. In principle, an end user should be able to determine what the AI program does and how it reaches its outputs. This in turn requires that AI designers have to:

• Produce more explainable modes;
• Enable human users to understand, manage and trust the AI agent.

As we move towards a more AI-based ubiquitous computing environment, the right to explanation will become increasingly important in the legal and social domains.

Typically, end users approach AI as a black box, with limited insight into what the AI agent does. With the evolution of deep neural networks, it has become increasingly difficult to understand how the hidden layers of a deep learning system function. If the systems is not transparent, it is very difficult to understand the rationale behind the decisions or outputs of an AI agent. In order to address this, AI agents and systems have to become more transparent in terms of functionality and processing. To solve the problem of opaqueness in the AI programs, methods have to be built in the machine learning process to delineate the processing inside the AI programs-for example, define what is happening in the hidden layers of a deep learning algorithm. In addition, human users should be able to determine through design documentation the inputs, the outputs and the formal logic behind the AI system. Finally, given that an AI system is built on explainability and transparency principles, the AI system should be human-interpretable.

Ethical AI systems design: common sense design
Common sense based design principle dictates that the context and outcome of an AI system has to be designed in common sense terms. While AI programs are remarkably good at executing complex tasks at extremely fast speeds, most AI programs lack basic understanding of common sense objects and actions. For example, a translation engine confuses "having someone for dinner" with "eating someone for dinner"; or, an object recognition program is unable to differentiate between a child and a doll. While building AI programs, AI designers have to be cognizant of common sense scenarios that will be encountered by AI during its operation.

Ethical AI systems design: human-first
AI systems have be designed on human-first design principles. Isaac Asimov,[12] in 1942, proposed the following three ethical laws on artificial agents:
• First Law – A robot may not injure a human being or, through inaction, allow a human being to come to harm.
• Second Law – A robot must obey the 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 or Second Laws.

These in turn have given rise to human-first principle in the ethics of AI. When designing an AI system, detailed feedback has to be solicited from multiple stakeholders, which include sponsors, users and designers. The design should also support human-first ethics that ensure that AI is not contravening basic human rights.

Ethical AI systems design: unbiased data
AI programs are only as good as the data we feed into them. If the data is biased then the decisions taken by the AI programs are also biased. Bad data results in codifying our implicit racial or gender biases into AI programs. It is therefore imperative that data used for building AI systems should be unbiased and unconscious preferences of the AI designers should not seep into training data.

Ethical AI systems design: auditable, accountable
Finally, AI systems have to be auditable and accountable. There should be a clear accountability structure that governs who is accountable for AI decisions in case of liability issues. For example, if an AI trading program executes illegal trades that result in loss of millions of dollars, an accountable entity has to be there to take the responsibility of the action. Similarly, an AI system should be auditable in terms of accountability, transparency, explainability and interpretability.

Is a global AI Ethics framework the solution?

AI will radically transform and disrupt our world, but right ethical choices for AI can make it a force of good for humanity. Until governments, business sector and academics start thinking about bringing codes of ethics into the AI discussion there is no anchor for the AI disruption. We think there is a need for setting up global AI ethics standards. Codes of ethics for expert bodies have broader national or global context. An international regulatory model is essential for the responsible design, development and deployment of AI. For instance, there are global health standards like Health Level Seven that provide a wider context for policies around health standards. AI posits challenges that have the potential and breadth to affect the lives of billions of people around the world. The current challenge is to build a code of ethics for AI that has global reach and is acceptable internationally. The complexity of such a task goes without saying.

Currently, we do not have a mature, global-standards body to help shape global governance of AI. Given that public sector organizations are aligned on the "common good" principle, these entities are best placed to come up with standards of ethics for AI that are beneficial for all. At the same time, no single organization or institution is capable of shaping the governance guidelines for AI, given that we also have to consider ethics that are sensitive to local variations. This is why the role of a global consortium, comprising multiple government entities will be essential to provide a global reference for AI ethics.

Finally, there are three principal policy recommendations for developing an effective global code of ethics for AI:

Building relationships with the AI stakeholder community
No single organization or policymaking entity can address issues around AI ethics. Governments and public sector organizations have to reach out to external AI stakeholders— i.e. other governments, institutions— to build partnerships for developing effective codes of ethics.

Utilizing existing governance levers
Governments and public sector organizations are well advised to acknowledge the fact that standard professional ethical codes are limited to address matters around AI governance. Public sector policymakers have a range of strategic tools available to integrate AI ethics into existing governance structure including explicitly making AI code of ethics and standard setting part of business process improvement and extending governance platforms by including AI stakeholders and practitioners in the governance bodies.

Creating AI awareness at institutional level
There is a general lack of awareness at all levels about how AI will affect our lives and work. Governments have to play an active role in creating institutional awareness around AI, focusing on technology, governance, legal aspects and value at stake for AI.

References

1. Deloitte study – AI-augmented government. www2.deloitte.com/insights/us/en/focus/cognitive-technologies/artificial-intelligence-government.html
2. Herschel, Richard and Virginia M. Miori. "Ethics and Big Data." Technology in Society 49 (2017): 34.
3. State of AI in the Enterprise, 2nd Edition, October 2018
4. www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/
5. ec.europa.eu/digital-single-market/en/news/have-your-say-european-expert-group-seeks-feedback-draft-ethics-guidelines-trustworthy
6. globalnews.ca/news/4729450/montreals-booming-artificial-intelligence-sector-companies/
7. Machinery, Computing. "Computing machinery and intelligence-AM Turing." Mind 59.236 (1950)
8. Searle, John R. The rediscovery of the mind. MIT press, 1992
9. Silver, David, et al. "Mastering the game of Go without human knowledge." Nature
10. The Future of Jobs Report 2018, World Economic Forum, September 2018
11. Deloitte study – AI-augmented government.https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/artificialintelligence-government.html
12. Asimov, Isaac. "Runaround." Astounding Science Fiction

Report
Authors

Ali Hashmi

Artificial Intelligence Leader,
Deloitte Middle East
Email: alhashmi@deloitte.com

Rajeev Lalwani

Technology Consulting Leader,
Deloitte Middle East
Email: rlalwani@deloitte.com

Antonio Senatore

Chief Technology Officer,
Deloitte EMEA Blockchain Lab,
Deloitte Ireland
Email: asenatore@deloitte.ie

Costi Perricos

Global and UK Lead Partner
for Analytics & Information
Management, Deloitte UK
Email: cperricos@deloitte.co.uk

Virginie Clemancon

Chief of Staff, Public Sector,
Deloitte Middle East
Email: vclemancon@deloitte.com

WORLD
GOVERNMENT
SUMMIT

Join the conversation
worldgovernmentsummit.org

@WorldGovSummit
#WorldGovSummit