BIG INNOVATION CENTRE
مركز الإبتكار الكبير

REPORT
SEPTEMBER 2019
Zaabeel Hall 1

DATA GOVERNANCE AND ETHICS

AI EVERYTHING WORKSHOP 3

Data Governance and Ethics is a report based on the third workshop of AI Everything - held on 1 May 2019 at Zaabeel Hall, Dubai.

This meeting was chaired by Lord Clement-Jones, UK House of Lords, and Younus Al Nasser, Assistant Director General, Smart Dubai & CEO, Smart Dubai Data.

We would like to express our appreciation to the following people for their oral evidence: Dr Eva-Marie Muller-Stuler, IBM, Chief Data Scientist - Middle East & Africa; Ott Velsberg, Ministry of Economic Affairs and Communications - Republic of Estonia, Chief Data Officer; and Dr. Sid Ahmed Benraouane, Professor at the University of Minnesota and Co-Chair of the US/ISO Working Group on Innovation Management System.

The evidence presented in the report is not exhaustive but reflects what was discussed at the meeting, and the views and experiences put forward by the people giving evidence. Written submissions by individual expert advisors in relation to this meeting are also included.

www.biginnovationcentre.com | info–me@biginnovationcentre.com| @BigInnovArabia

© Big Innovation Centre 2019. All Rights Reserved

TABLE OF CONTENTS

WORKSHOP OVERVIEW

Details

- Date: 1st May 2019
- Time: 1:00 – 2:30 pm
- Location: Zaabeel Hall, Dubai.
- Participants: 40 registered attendees

Speakers

- Dr Eva-Marie Muller-Stuler, IBM, (Middle East & Africa) Chief Data Scientist
- Ott Velsberg, Chief Data Officer, Ministry of Economic Affairs and Communications, Republic of Estonia
- Dr. Sid Ahmed Benraouane, Professor University of Minnesota / Co-Chair of the US/ISO Working Group on Innovation Management System

Questions

- How should we govern data from individuals to the tech-giants and from the nation to the globe?
- Whose data is it anyway? Personal data, business data, government data, machine data?
- What ethical behaviours should we follow? Data user rights, algorithm-bias, global governance of data, right to be forgotten, and AI in business to customer relationships?
- What is good and bad data? Accurate and quality issues.

INTRODUCTION

As part of AI Everything in Dubai, Big Innovation Centre hosted a multi-stakeholder workshop to discuss the importance of data governance in a context where AI technologies are increasingly being developed and deployed across the world.

Lord Clement-Jones and Mr. Younus Al Nasser chaired the timely session and the panel included: Dr Eva-Marie Muller-Stuler, IBM, Chief Data Scientist - Middle East & Africa; Ott Velsberg, Ministry of Economic Affairs and Communications - Republic of Estonia, Chief Data Officer; and Dr. Sid Ahmed Benraouane, Professor at the University of Minnesota and Co-Chair of the US/ISO Working Group on Innovation Management System.

The audience of the workshop were attendees at AI Everything, representing all sectors of life from government to business and from academia to civil society.

Together, the participants had a fruitful discussion on data governance and its ethics, as well as suggestions on potential new forms of data governance fit for the AI era.

It was clear that AI and data are deeply interrelated. Access to accurate and quality data is critical to make AI work and, in many ways, data is useless without AI-driven analytics and insight. However, dealing with data in AI, is a complex and multifaceted process – particularly for stakeholders to obtain the desired results.

Smart data governance, in this context, can help manage both the process and outcome. Data governance can also help control data availability, usability, integrity, and security – as well as help public and private sectors around the world face complicated data challenges which revolve around ownership, control, transparency, data misuse, data bias, access, and data monopolization.

Ultimately, a robust data governance approach is essential for the implementation of AI in the economy and the society and for stakeholders to reap the benefits these AI technologies promise.

DATA GOVERNANCE

Data governance is how we manage data in terms of integrity, security, usability, and availability. To answer the question of how should we govern data, we first have to know what data can do to AI. The quality of data affects AI, therefore the best the quality of data would give us the best of AI and vice versa.

As AI and data thrive and feed from each other, problems arise when there are critical issues in the governance of either the former or the latter. Simply put, if the data being collected and used has ingrained issues so does the AI. In fact, rather than solving these issues, AI makes them more pronounced.
Therefore, given the codependent and symbiotic relationship of AI and data, AI governance and data governance must adapt a coherent strategy, addressing key areas of growing concern, including ownership, control, misuse, bias, access, and monopolisation.

Throughout history, data has always been collected, used, and managed by different actors for different purposes. However, we have now entered a period in which the volume and velocity of data has skyrocketed. The created use of data has completely transformed industries, routines at work and the way we live our lives. Social notions are with people, machines, and software.

In the realm of AI, individuals, businesses, and governments have to provide accurate data to attain the best products. Accuracy issues happen when there is a defect in providing or misuse of data, therefore data

governance helps in solving this issue. The low quality and accuracy of data could result in bad decisions. Therefore, companies and organizations have to maintain the quality of data to their own good, thus they can make better and knowledgeable decisions, and they could provide the best goods and services upon these decisions.

Data must be governed by the public and private sectors, from all sources, through the collection and management of accurate and true data. That is how data risks such as bias, privacy, transparency, and safety can be mitigated.

The participants of the workshop recognized the urgency to rethink data governance. Ultimately, they identified two different approaches for the governance of how data is collected, used, and managed. The first approach – the hard approach – calls for new legislation and regulation. The second approach – the soft approach – calls for the use of soft-structures (new and old) to address the data issues of the AI revolution. This approach includes new norms, standards, and ethical guidelines.

The first step to changing data governance is to analyse the landscape currently in place and gather evidence on the existing regulations, policies, and institutions existing. Using the data gathered from the landscape analysis, the second step is to provide society with a framework, guiding them on how data should be collected, used, and manged.

During the workshop, Dr. Eva-Marie Muller-Stuler discussed the need for an updated framework which will provide a new model of data governance for AI systems.

Dr Eva-Marie Muller-Stuler
Chief Data Scientist, IBM

Talking points from Dr.Muller-Stuler's presentation

- A personal story about a quite disturbing interaction I had about 1 year ago with a robot. When it saw me, it came down to my level and immediately started flirting with me. To change the topic, I asked it to tell me a joke. But it didn't react until my male colleague in his deeper voice repeated my question. The robot immediately reacted and told a very insulting joke about sexual harassment of women.
- The main problem explaining why we are in this situation now is that we have a pacing problem.
- The technology that we develop currently changes on an exponential speed, while our social, economic and legal system only change incrementally. This

leads to our technical innovation increasingly outpacing the ability of laws and regulations to keep up.

- The social consequences of a technology cannot be predicted early in the life of the technology.

- We have seen the same with our Data Science and AI solutions that we implemented over the last few years. We realized that there are many different issues that our solutions can inherit from data ownership, user rights, biases to accountability.

- We have implemented credit scoring or loan application systems in the past that strongly discriminated again postcodes, race or other input data.

- Just because we know how the software and hardware work does not mean that combining data sources and patterns will lead to a fair, right and trusted solution.

Most of the concerns around AI and DS can be categories into one of the following topics:

- Privacy
- Security
- Safety
- IP protection
- Economic disruption
- Fairness & Equal right

And be split in 4 main areas:

- Individual
- Organization
- National
- Aociety

And we would need all 4 bodies to implement AI governance.

- Personal Responsibility
    - What Data do I own and show?
- Organizational Responsibility
    - Make sure the solutions we buy/sell/ build/ use are safe and not against our laws and goals as a society
- Inter/Intra national Responsibility
    - Build rules and frameworks about what we need and how we monitor things.

- Build an audit body that is trained in AI and DS and for example ensure that the solutions used in the country have been audited and approved

When the solution and the risk inherent in the solution falls into one of the following categories ,it becomes a national interest and we should take precautionary measures:

- Highly probable
- Tangible
- Immediate
- Irreversible
- Catastrophic

- The computational speed we see today, is far greater than it has ever been, to rely so heavily on our own machines to make decisions for us, we should insist that the highest possible governance is in place, with the best possible measures ensuring the **INDIVIDUAL**, the **ORGANIZATION** and **SOCIETY** as a whole.

- If we ensure these three pillars, we are able to begin to create a framework for how to secure AI and ensure sustainable innovation in the field of AI.

DATA OWNERSHIP

Linked closely to data governance, a second point that was raised during the workshop revolved around the controversial questions: who owns what data?

This is particularly problematic in a context where data is now noted as one of the most valuable resources.

However, unlike most other goods people can own, data is non-rivalrous. One player having access and using data does not prevent, in theory, another from doing so.

Data used by AI is ultimately everyone's data. It is a combination of personal data, business data, government data, and machine data. For instance, big-tech companies are often criticized for making extreme profits from individual data. This is partly true as they rely on consumer data to offer a personalized product or service, but this personal data is often collated with billions of other types of data including commercial and public. With the aide of AI systems, data about 'me' can now also provide insight about 'others.'

Furthermore, data is now collected, used, and managed in revolutionary ways. Organisations no longer rely on their own, independent processes of collecting data but rather use the data that is generated in interconnected open networks. As a result, this means access to a lot more data and, hence, a lot more opportunity. Data can be used by different organisations in the same industry but also across sectors to increase the value further. Most companies, at the moment, take personal data and anonymise it in order to subsequently apply it for various value-generating processes. This also means that the task of figuring out ownership for a specific data set becomes more complex. It becomes a true challenge to differentiate between well-founded, clean data and the reverse.

For these, the concept of data ownership is complicated to solve. Many suggest that the argument around data ownership should shift to one of control. Lack of data control and

agency are two of the key problems underlying the need for new data governance models. Increasingly so, people (and, also, organisations and communities) are feeling they are unable to control how their data is being collected and used by AI systems in both the private and public sectors.

During the workshop, there was a question to Mr. Ott Velsberg about how the Estonian government deals with these complicated data issues. – particularly around privacy. Ott Velsberg explained that the government of Estonia tries to make everything public and transparent 'if possible' and gives a great effort to provide personalized services to citizens.

In regards to the transparency of data use, the government of Estonia tries to notify citizens via different mechanisms such as press; and people have a trustworthy approach on how the government uses data. Moreover, Mr. Velsberg elaborated on the government using the EU ethics standards and how the government is eager to collaborate with other countries to solve issues associated with data governance.

Mr. Younus Al Nasser, further addressed the topic, illustrating Smart Dubai's standardization of addressing data. They launched a mandate for how to govern data and how data exchange can happen, how to open data, how to share data, and who owns data monitoring compliance to data standards.

It was argued that governance on data ownership and adopting ethical standards are the potential solutions moving forward.

Ott Velsberg,
Chief Data Officer, Republic of Estonia

Talking points from Mr. Veisberg's presentation

- I believe in the transparency of data use.
- In Estonia, citizens are already able to see which data, for what purposes and by whom have been used.
- In case of healthcare data, citizens are by default "opt in" sharing their health data, but if required, can "opt out".
- In some circumstances the government would still be able to use the data, mostly for statistical, research and security purposes.

- One of the first use-cases is in the healthcare sector, more precisely concerned with genome data (today over 150 000 have donated genome data) and prescription data.

- This allows citizens to decide how and by whom their data will be used. However, we have identified (and still continue with the impact analysis) that in some circumstances we should not allow access to individual data as it might have substantial negative effect on citizens, for instance: crime record, substantial healthcare information, political information etc.

- There is a great chance that giving such access could make it required by default, thus in the end negatively affecting an individual.

- However, such transparency and access to data has allowed the creation of services otherwise unimaginable, i.e. diet according to your DNA.

- At the core of transparent data management, data integrity should be protected, access to data should be controlled, impact analysis carried out (too much data on the individual could hurt them?), and there must exist the right to be forgotten, among other things.

ETHICAL BEHAVIOURS

The data governance challenges mentioned have created public backlash and erosion of trust and confidence across the world.

Individuals and organisations are starting to question how data is collected and used; and, ultimately, whether our current data governance models are fit for purpose. Data and AI have already changed our lives tremendously, helping generate great economic and social benefits. Simultaneously, however, reoccurring data breaches, lack of self-regulation by data-driven organisations, and little ethical judgment in the way day is collected and used are making the need for data governance one of high policy concern.

In the workshop, some argued that soft-structures such as ethical standards could be the appropriate channel to govern the trends of AI technologies – given their exponential speed. Most stakeholders in the ecosystem are already engaging with the ethical issues related to data and data-driven technologies. For example, companies are already proposing solutions to how to address ethical issues around privacy and transparency.

However, these approaches are fragmented and there is little monitoring or accountability for whether companies are to follow them. Ethical standards of data using and sharing, as a critical part of the soft governance model, must ultimately follow the rules of law. This means the involvement of the public sector and the civic society in their development and in their implementation.

Ethical standards of data using and sharing, as a critical part of the soft governance model, must ultimately follow the rules of law. Every country should follow the best standards available for ethical behaviors and develop these standards to fit the uniqueness of the society looking for adopting the technology. In this way, the ethical standards can easily be implemented within all governments' entities, private sector institutions, universities, etc.

Individuals or users have their own responsibility in data governance. When one signs up for a platform, for example, he/she often gives consent to give away his/her data in exchange for a specific service. Individuals must realize there is collective responsibility in every transaction including that of data.

Ethical implications could prove a significant dilemma on AI responses and might lead to severe problems. Data user rights, algorithm-bias, the right to be forgotten, or AI in business to customer relationship are just a few of the ethical issues that nations worldwide must now address.

Dr Sid Ahmed Benraouane,
Professor at the University of Minnesota

Talking points from Mr. Benraouane's presentation

- With AI, Big Data has acquired a new meaning and a new function. It is now used to teach machines how to make decision and train algorithm to create AI decision making models. The issue is that once a bias seep into the datasets used to build the AI training model, the decision the algorithm makes becomes itself biased.
- It is crucial for a credible AI system that we work very hard on training data scientists, engineers and programmer building AI systems, on how to collect, prepare and use data to train their AI models.
- According to Karen Hao, from MIT Technology Review, argues that data become corrupt in three ways: During the data collection stage, during the preparation stage, and during the stage in which you frame the problem.
- The way you collect data is an important step in removing bias from your data and from your AI decisions
- The way you prepare the data. Data preparation is an important step in which subconscious bias can be introduced into the AI decision making process.
- The third way that might create bias in your AI decision making is the way you frame the problem you want AI to solve. This is another way through which bias creeps into the AI decision making without intention. How you ask the

question, and what type of language you use to describe the issue, is in itself indicative of the subconscious bias you carry.

- As we say, machines don't discriminate. It is the dataset that discriminates. And dataset by itself does not discriminate. It is the way we handle dataset that can create bias for AI decision making.

www.biginnovationcentre.com