

Digital  
Leadership  
Forum

# AI for Good Report

AI for Good Steering Group

February 2020



Technology Partner



[www.digitalleadershipforum.co.uk](http://www.digitalleadershipforum.co.uk)

# Table of Contents

Background and Executive Summary	3
AI for Good Frameworks Discussion	5
Discussion Group Topics	7
Key Discussion Group Points	8
Supplementary AI for Good Insight Points/Resources	10
What is Digital Leadership Forum and AI for Good?	13

Technology Partner



# Background and Executive Summary

## AI for Good Steering Group

Tuesday 25th February 2020 | Clifford Chance

The purpose of the AI for Good Steering Group is to help leading organisations introduce AI into their organisations in a responsible way. The AI for Good Steering Group met for an introductory meeting at Clifford Chance in February.

The members of the AI for Good Steering Group reviewed and discussed a series of AI for Good frameworks. They considered how AI ethics could be applied in different markets globally; how data can be acquired appropriately; how to maximise the fairness of AI outputs, and how to guide analytics teams to use AI responsibly in their organisations.

The group recognised the ability of AI to sift through huge quantities of data to gain insights and to deliver new services; and to protect against cybersecurity and fraud.

However aligning AI programmes with human preferences and organisational brand values remains a challenge requiring new methods of training; and transparent and explainable models.

Technology Partner



# Background and Executive Summary

## AI for Good Steering Group

Tuesday 25th February 2020 | Clifford Chance

### The following people took part:

- Lee Baker, Commercial Leader, Seldon
- Alison Berthet, Human and Digital Rights Strategy Manager, BT
- Moniek Buiter, Digital Marketing Manager, Clifford Chance
- Martin Faux, Head of Digital Platforms, LV=
- Arash Ghazanfari, CTO, Dell Technologies
- Stuart Kay, Chief Marketing Officer, BPP Professional Education
- Antonia Lim, Head of Quantamental Investments, Schroders
- Paul Marsden, Digital Product Owner, LV=
- Sue McLean, Technology and Fintech Partner, Baker McKenzie
- Grant Millar, Head of Digital, First Rate Exchange Services
- David Moloney, Director: Innovation and Transformation, PwC
- Arup Paul, Deputy Chief Medical Officer, AXA PPP Healthcare
- Steve Pybus, Manager International Development, NATS
- James Rowland, Digital Performance Director, Refinitiv
- Margi Sheth, R&D Data Policy Director, AstraZeneca
- Russell Smith, Head of Enabling Units IT, AstraZeneca
- Ian Stevens, Partner: Specializing in Technology and Outsourcing, CMS
- Russell Tarr, Vice President: Global Sales, Artificial Solutions
- Mark Taylor, Product Director, International Compliance Association
- Luke Vilain, Responsible AI Product Owner, Lloyds Banking Group
- Abhijit Akerkar, Head of Applied Sciences: Business Integration, Lloyds Banking Group

Technology Partner



# AI for Good Frameworks Discussion

The members of the Steering Group reviewed and discussed seven algorithm questions created and developed by British scientist, David Spiegelhalter and reviewed their viability as a potential framework for their organisations:

## Seven Spiegelhalter Algorithm Questions:

1. Is it any good when tried in new parts of the world?
2. Would something simpler, and more transparent and robust, be just as good?
3. Could I explain how it works (in general) to anyone who is interested?
4. Could I explain to an individual how it reached its conclusion in their particular case?
5. Does it know when it is on shaky ground, and can it acknowledge uncertainty?
6. Do people use it appropriately, with the right level of scepticism?
7. Does it actually help in practice?

## Spiegelhalter: 7 Algorithm Questions Review

The steering group had comments and feedback on a number of the specific questions. Below is a summary of the key points in relation to particular Spiegelhalter questions.

### Question 1: Is it any good when tried in new parts of the world?

- Would it make more sense to say 'in different environments?'
- Does the model suffer from bias?
- Has it just been designed for one local territory? It is transferable?
- If your data is based on 'white males' it wouldn't work well in other territories and locations.

Technology Partner



# AI for Good Frameworks Discussion

## Question 3: Could I explain how it works (in general) to anyone who is interested?

- Democratise expert explanation. Customer wants to know WHY application was rejected. Give people tool-kits to interpret and explain model. Rely on the wisdom of the crowds online.
- Consider how can you translate embed your corporate values into AI processes and systems
- We don't need algorithms to make all decisions – human decision making still has a role
- People need to understand the overarching system so they have the ability to challenge it. Just looking at an accuracy score based on training data is not going to cut it.

## Question 5: Does it know when it is on shaky ground and can it acknowledge uncertainty?

- Does this question mean – when should it hand the task over to a human or does it know when to check back with a human?
- If it exhibits biased behaviours, it can be stopped very quickly
- It's hard for humans to correct autonomous vehicles quickly enough in a range of changing contexts
- Autonomous vehicles have no in-built moral sense
- Set up barriers/introduce red flags which tell the software/AI to go back to a human to check
- Humans kill 14,000 people a year driving – why do we think autonomous vehicles have to be perfect?

Technology Partner



# Discussion Group Topics

The AI for Good Steering Group discussed a range of topics:

- How can you ensure artificial intelligence benefits your company?
- How can you guide your analytics team to build and use AI responsibly?
- Are you acquiring data appropriately?
- How can you maximise the fairness of AI outputs?
- Can your company AI models be explained to stakeholders?
- What regulations are applicable to the AI work in your organisation?



Technology Partner



# Key Discussion Group Points

## Are you acquiring data appropriately?

Acquiring data is a problematic area. Anyone involved needs to:

- Understand their own business and have empathy with their clients
- Understand regulations
- Be up to speed with discussions concerning ethics and conduct -it's a big challenge
- EU & UK have regulations such as: Consumer Protection and GDPR

By contrast, in places like China, it is a very different society and different views prevail about collection of data and protections about data. It is a challenge to identify and apply consistent values that work in and across different societies.

'AI can help unlock the potential of data. When you can see patterns in dynamic data, you can gain a substantial edge over the competition. Massive sets of unstructured data can be turned into valuable intelligence to help predict risks, enhance automated trading, create new services and protect against cybersecurity threats. The data you work with has to be appropriate.

The foundation has to be right to capture higher returns and lower risks' -

Arash Ghazanfari, CTO, Dell Technologies

## How can you maximise the fairness of AI outputs and how can you guide your analytics team to build and use AI responsibly?

- We are talking about company purpose and values; and how the purpose and values of the company is structured and how that can be embedded into AI processes.
- What needs to change is your compliance frameworks and processes to ensure that teams are actually following responsible processes.
- It is not just our internal teams developing, but also our suppliers and contractual relationships. So how do we identify who in our supply chain is using artificial intelligence? And how do we ensure that they are using it ethically, and responsibly, and teach people how to challenge on these grounds?

Technology Partner



# Key Discussion Group Points

- You need to guide and train your team to act fairly and responsibly through the development of the product into actual operation.
- We need new frameworks and infrastructure to overcome business silos to ensure responsible AI and fairness is a continuous element of working practice.
- In training, we should have a strategy for Responsible AI (fair AI) and this needs to have senior endorsement and it needs to be clearly communicated so teams can refer to it.
- From a practical point of view, it is not just a case explaining the AI system once you build the system, but why you doing it.
- A recommended responsible AI message is that we are looking to augment human capability in partnership with a machine. We are not looking to replace them.
- There is no way at this stage that an AI machine can take a huge amount of complexity in different fast changing contexts and accurately understand act upon it at pace like humans can do.
- A key question is how do you audit AI processes? How do you record and store the decisions that are made and also the data sets that were used?
- Data scientists are not yet trained to think about ethics as key parts of their job. So, it's a question of how much focus and time should we take out of development with these key expensive resources and force them to do so.



Technology Partner



# Supplementary AI for Good

## Insight points/resources

The Digital Leadership Forum has researched and identified a number of supplementary insights to support the thinking and the discussions of the AI for Good Steering Group which are highlighted below. These are sourced from other *Digital Leadership Forum AI meetings*; *McKinsey.com*; and *Human Compatible by Stuart Russell*; *PyTorch London – AI meetup*)

- Underlying data rather than the algorithm itself are often the main source of bias.
- Bias can be introduced into the data through how it is collected or selected for use.
- A machine learning algorithm may also pick up on statistical correlations that are societally unacceptable or illegal. For example, if a mortgage lending model finds that older individuals have a higher likelihood of defaulting and reduced lending based on age, society and legal institutions may consider this to be illegal age discrimination.
- In order to minimize bias, how do we define and measure fairness? Several approaches to enforcing fairness constraints on AI models have emerged. The first consists of pre-processing the data to maintain as much accuracy as possible. On the data side, researchers have made progress on text classification tasks by adding more data points to improve performance for protected groups.

Technology Partner



# Supplementary AI for Good

## Insight points/resources

- Human judgement is still needed to ensure AI supported decision making is fair. In which situations should fully automated decision making be permissible at all? It requires human judgement and processes, drawing on disciplines including social sciences, law, and ethics, to develop standards so that humans can deploy AI with bias and fairness in mind. The work is just beginning.
- It is important to be aware of the contexts in which AI can help correct for bias as there is a high risk that AI could exacerbate bias.
- When deploying AI, it is important to anticipate domains potentially prone to unfair bias, such as those with previous examples of biased systems or with skewed data.
- Organisations will need to stay up to date to see how and where AI can improve fairness – and where AI systems have struggled.
- Organisations need to explore how humans and machines can work best together. This includes considering situations and use-cases when automated decision making is acceptable versus when humans should always be involved. Some systems use a combination of machines and humans to reduce bias. Techniques include ‘human-in-the-loop’ decision making where algorithms provide recommendations or options, which humans double-check or choose from.
- A more diverse AI community will be better equipped to anticipate, spot, and review issues of unfair bias and better able to engage communities likely affected by bias. This will require investment on multiple fronts, but especially in AI education and access to tools and opportunities.

Technology Partner



# Supplementary AI for Good

## Insight points/resources

### Appropriate data acquisition

Data serves as the fuel for AI. In general, the more data used to train systems, the more accurate and insightful the predictions. However, pressure on analytics teams to innovate can lead to the use of third party data or the re-purposing of existing customer data in ways that, while not yet covered by regulations, are considered inappropriate by consumers. As a result, leaders must be vigilant in asking data-science teams where they acquire data from and how the data will be used, and challenge them to consider how customer and society might react to their approach.

### Data-set suitability

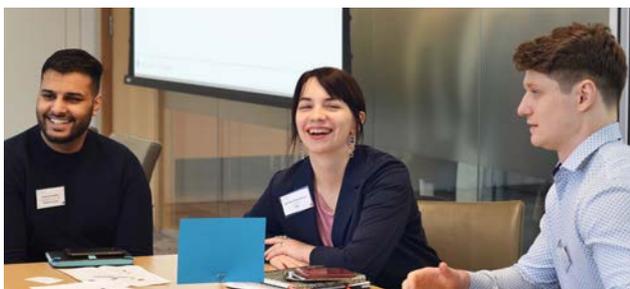
As history has shown, hard-working data scientists in the thick of deadlines and drowning in data can think they have covered all bases, when in fact they have not. As a result leaders must ask data science teams fairly granular questions to understand how they sampled the data to train their models. Do data sets reflect real world populations? Have they included data that are relevant to minority groups?

### Explainability

To ensure model outputs can easily be explained to stakeholders, leaders must probe their data science teams on the types of models they use by, for example, challenging teams to show the model they have chosen and demanding the use of explainability techniques for naturally opaque techniques.

Technology Partner





## What is the Digital Leadership Forum?

At the Digital Leadership Forum, we help our members and their organisations to innovate and grow in the Age of AI. We design and run highly interactive and collaborative sessions where our members can share experiences and gain valuable insights to help support their digital strategies. We tackle a range of core strategic business issues including digital transformation, AI & automation, digital marketing strategies, the future of work, and diversity in the workplace.

## What is AI for Good?

The membership of the Digital Leadership Forum has asked for more knowledge and insight on the rise of AI technologies and their applications. Following a well attended AI Ethics meeting in 2018, we decided to launch a dedicated AI for Good membership community project. This initiative involves members from world leading organisations, academics, regulators, policy advisors and AI experts. AI can lead to better business outcomes such as increased revenue and enhanced customer experience. But deployment of AI requires careful management to prevent unintended damage to your brand, work colleagues, and society as a whole. With the support of Dell Technologies, AI for Good is a quarterly series of sessions which are designed to help members use AI in a responsible way.

Technology Partner



**DELL**Technologies

# Putting Data to Work

How can you harness your data to  
give you that competitive edge?

Scan this code to get  
access to Glenn Mallon's  
presentation

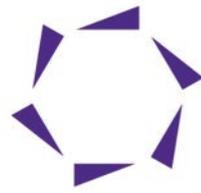


[www.DellTechnologies.com](http://www.DellTechnologies.com)



@DellTech





Digital  
Leadership  
Forum

# Check out our upcoming sessions

[www.digitalleadershipforum.co.uk/events](http://www.digitalleadershipforum.co.uk/events)



Technology Partner



[www.digitalleadershipforum.co.uk](http://www.digitalleadershipforum.co.uk)