Digital Leadership Forum

AI for Good Report

The Ethics of Artificial Intelligence

October 2019

Technology Partner

DELL Technologies

Table of Contents

Technology Partner

DELL Technologies

What is Digital Leadership Forum?

At the Digital Leadership Forum, we help our members and their organisations to innovate and grow in the digital age. 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 damages 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

Session Videos

Watch video here

Using AI to Extend Human Cognitive Capabilities with Dr Karina Vold, University of Cambridge

What are the ethical issues that arise from using AI to extend human cognitive capabilities? Research Fellow/Postdoctoral Researcher, Dr Karina Vold challenges us to consider whether AI systems could be used to complement and extend our cognitive capabilities in more advanced ways and the ethical risks and implications of doing so.

Watch video here

Dell Technologies' Arash Ghazanfari Discusses Key Ethical Concerns with Artificial Intelligence

Field CTO at Dell Technologies, Arash Ghazanfari spoke with us about the ethical issues surrounding Artificial Intelligence.

Technology Partner

DELL Technologies

Session Videos

Watch video here

BPP's Stuart Kay Discusses How Employers Can Tackle Skill Gaps in AI and Automation

Chief Marketing Officer at BPP, Stuart Kay discusses how organisations can tackle skill gaps in AI and automation, and some of the most exciting possibilities surrounding AI technology.

Technology Partner

Executive Summary

The Ethics of Artificial Intelligence

23rd October 2019 | Barclays

In October 2019 the Digital Leadership Forum held the second in our series of quarterly AI for Good events, supported by our Technology Partner Dell Technologies.

The aim of the AI for Good programme is to encourage cross-industry collaboration on key ethical issues surrounding artificial intelligence and its implementation within organisations.

Representatives from leading organisations met at **The Ethics of Artificial Intelligence** session at Barclays Rise in Shoreditch to discuss the challenges of ethical AI, learn from academic and field experts, and work collectively towards developing professional best practices in a rapidly evolving technical and regulatory environment.

Attendees heard from academics working within AI. Dr Karina Vold, from the University of Cambridge, explained how AI could be used to extend human cognitive capabilities and the ethical issues that could arise from doing so. Dr Brent Mittelstadt, from the University of Oxford, outlined the variety and weaknesses in current AI ethical frameworks, presenting a pathway to a cohesive ethical framework.

We were also joined by a panel of industry experts from CMS, Dell Technologies, and Access Partnership, who discussed the practical steps that organisations can take to best prepare for new legal and regulatory challenges, particularly in light of the anticipated new European Commission regulations on AI in 2020.

Attendees discussed the importance of balancing accountability and responsibility with innovation, the need for diversity both within datasets and amongst stakeholders, and how it is essential to clearly communicate across teams to bring all members of an organisation along with you on the journey.

Technology Partner

DΦLL
Technologies

Using AI To Extend Human Cognitive Capabilities

Presented by Dr Karina Vold, Postdoctoral Research Associate at University of Cambridge

Dr Karina Vold challenged attendees to consider whether AI systems could be used to complement and extend our cognitive capabilities in more advanced and sophisticated ways than they are currently.

Embracing new technology

Vold began her presentation by arguing that while shifts in technology are generally positive, they have historically been met with suspicion. She gave the example of the Greek philosopher Socrates, who resisted the shift from the oral to written tradition as he thought that by writing things down we would become more forgetful and less social. "Those are exactly the same arguments that you hear against technology today," Vold said. "You hear that Google is making us more forgetful and Facebook is making us asocial. It's a story that's been happening for a very long time in philosophy and one that I'll argue against."

Redesigning the tasks

Studies show that when information is easily accessible we are less likely to remember the information itself, but instead how to access it. "One effect that technology is having is that it is redesigning the tasks. It's not making us more forgetful – we're still remembering stuff, it's just shifting the kind of stuff we have to remember," Vold explained. For example, we no longer need to remember phone numbers but instead just the passcode to our phones. "In a sense I've outsourced that task to the technology," Vold said.

Technology Partner

Using AI To Extend Human Cognitive Capabilities

Defining AI

Most AI definitions used today include a clause about autonomous agency. "The idea is that we should build a system that can do things entirely on its own without any kind of input or reliance on humans," Vold explained. However, Vold challenged this definition, suggesting that we should include non-autonomous systems in our definition of AI. These systems are built to interact with humans and become intimately coupled with us as we engage in an ongoing dialogue with them. Vold argued that these systems could know us better and have a more complete record of us than any human.

Distinguishing between externalised & extended cognition

Vold presented a scale of autonomous to non-autonomous systems, and explained two different ways in which we could engage with AI.

Externalised cognition is outsourcing a subprocess to a machine, for example using a system to translate language. This requires human input and output but the system can run an entire process on its own.
Extended cognition is when a system is tightly coupled with humans to add capacity, for example using GPS on a smartphone to navigate. The tool is always needed as the process is not internalised, and if you lose the system you lose the capacity entirely.

Learning from Move 37

In 2016 during a Go match in Seoul between world champion Lee Sedol and a computer program developed by Google DeepMind, called AlphaGo, AlphaGo played an unexpected and successful move that no human player would have played. This became known as Move 37, and is an example of systems creatively generating new concepts and ideas. Expert human Go players continue to train on Move 37, trying to work out why it was so unexpected and what we can learn from it. "One of the reasons that people think that the system came up with that move was that it wasn't being burdened by some of our own social norms, our own game-playing norms and our own human wisdom about what's good and what's not good," Vold said. "It's really interesting when you think about situations where the stakes are higher: scientific discoveries, drug discoveries, or healthcare. How do we think outside the box?"

Technology Partner

DELL Technologies

Using AI To Extend Human Cognitive Capabilities

Putting it into practice

Vold argued that by offloading cognitive labour to AI we free up our internal resources to do more interesting and creative tasks. "What are the kinds of things that we can use AI for to offload some of the tasks that either we're not good at, we're just not built for, or that we just don't like doing?" "Obvious weaknesses for us are easy tasks for some systems," Vold said, suggesting that memory processes, psychometrics, and quantitative and logical reasoning were all areas that could be offloaded. Vold also argued that we should consider offloading decision-making to systems in order to avoid bias. "We don't really make decisions in the way we think we do," Vold said. "A lot of times even though we think we're making judgments in a particular way, we're being informed by all sorts of built-in systematic biases." When evaluating our own human decisions, Vold said that we can find "lots of systematic deviations from what most of us would objectively call rational judgments."

Managing the risks

While AI offers exciting opportunities to extend human cognitive capacities, Vold identified three key risks and implications to be aware of:

1. Cognitive atrophy: By becoming too reliant on a system or technology we start to diminish our own internal capacities to perform that task independently. Depending on context this can lead to safety issues.

2. Responsibility: If your processes become reliant on technology but you don't know how it is reaching decisions and it is difficult to judge or override the system's decision then it may become a professional norm to implement those decisions without understanding them. The person implementing them will still be held responsible for any negative consequences, but without the ability to understand and rectify the problem.

3. Privacy: Devices and systems are susceptible to hacking, and as we put more information onto our devices, so that they become a part of us, we need measures to protect that data and our privacy.

Technology Partner

DELL Technologies

AI Ethics: Too Principled to Fail?

Presented by Dr Brent Mittelstadt, Research Fellow & British Academy Postdoctoral Fellow at University of Oxford

Dr Brent Mittelstadt reviewed the current landscape of AI ethical frameworks, identifying four recurring principles that are common to biomedical ethics, and arguing that using biomedicine as a baseline is an inadequate solution for AI. Mittelstadt then suggested how we can move forward towards a single cohesive framework for ethics in AI.

Competing frameworks

Dr Brent Mittelstadt opened his presentation with a overview of the current landscape of AI ethics. There are many different ethical frameworks, possibly as many as 110, and these strategies are generally western-focused, coming from governments, companies, and third-sector organisations in Europe and the US.

"The challenge - for businesses, for governments, for researchers – is that we have all these frameworks and depending on where you're operating potentially more than one of them applies to you," Mittelstadt said.

The question for organisations is how they can bring these frameworks down to earth and put them into practice in the best way possible.

Technology Partner

DELL Technologies

AI Ethics: Too Principled to Fail?

Recurring principles

In reviewing the AI ethical frameworks, Mittelstadt identified four recurring principles which are common to biomedical ethics:

1. Respect for autonomy
2. Beneficence
3. Nonmaleficence
4. Justice

Mittelstadt suggested that what we value in AI is not that different from what we value in medicine, and that these borrowed terms and principles give us a familiar vocabulary to talk about what is ethical or not and why in AI. "Can we reasonably expect it to be successfully translated in practice, and can we expect them to lead to any sort of unification or harmony in terms of how we are approaching AI ethics across the globe, at a company level, and at a sectoral level?" Mittelstadt asked. Mittelstadt said that unfortunately it is not that simple. Not all characteristics are mirrored and terms in AI are not as narrowly defined as in medicine, so these borrowed terms and principles can actually become barriers in working toward AI ethical frameworks.

Missing characteristics

Mittelstadt identified four common characteristics present in medicine that are lacking in AI:

1. Common aims: In medicine there is a single shared aim: the patient's well-being. This leads to co-operative ethical decision-making. There is no equivalent in AI, so different groups may have different competing aims which then leads to competitive ethical decision-making.

2. Professional history: Medical ethics have developed over a long time, and have been revised and tested many times. They are also supported by professional organisations, and are embedded within professional practice and culture. By contrast, there is no rich unified history to draw from in AI, and research shows that current ethical frameworks have no actual effect on day-to-day decision-making.

Technology Partner

D&LL Technologies

AI Ethics: Too Principled to Fail?

3. Translation methods: Medicine has mechanisms in place to anchor principles to their day-to-day implementation, such as licencing, ethics boards, and professional bodies. While AI ethical principles are generally agreed, their practical implementation is not, and so what appears to be agreement may in practice be hiding disagreement between parties. We are starting to introduce anchoring mechanisms, but this will take time to have an effect.

4. Accountability mechanisms: In medicine there are strong accountability mechanisms, such as licencing. Doctors can be sued, held liable, and have to explain their decisions. There is nothing similar in AI development.

Moving forward

"Principles aren't enough, the real work that we have to do is coming in the future." "We have to do our best in terms of putting ethics into place as early as possible in the development cycle, because that's where it does actually have the most impact," Mittelstadt explained. As software development is not a legally recognised profession there are no clear standards or professional practices against which individuals can be held liable. However, Mittelstadt suggested that we should reframe AI ethics away from the individual failings of developers, and instead towards a business ethics model. "We need to think not just about unethical individuals, but unethical business models," Mittelstadt said, adding that we're beginning to see this in the space of facial recognition.

Mittelstadt also encouraged the development of standards and sharing case studies. "We need more bottom-up cases coming from developers themselves to say 'here are the challenges we ran into in practice and here's how we dealt with them," Mittelstadt said. Yet it's not the case studies where companies said yes, nor the ones were they said 'definitely not' that need to be shared. Instead, they should share the case studies where they almost said no, and the changes that they made to the system to make it acceptable. "If we can get more of those cases out to be discussed then we're going to start learning about how to specify these very high level concepts and what they actually mean on the ground," Mittelstadt argued.

Technology Partner

DELL Technologies

Legal and Regulatory Challenges in Ethical AI

Panelists: Arash Ghazanfari, Dell Technologies; Rachel Free, Baker McKenzie; Matthew McDermott, Access Partnership.

Our panel of experts from across legal, regulatory, and technical fields discussed upcoming challenges in ethical AI, and how businesses can best prepare.

1. Review European Commission Trustworthy Guidelines

The European Commission recently published their ethics guidelines for trustworthy AI, with a view to piloting these within the European community before introducing regulations. Rachel Free noted that current Commissioner for Competition, Margrethe Vestager, stated she would bring AI ethics regulations by January 2020. While current Brexit uncertainty makes it difficult for businesses to plan for these regulations, the guidelines offer a framework for best practices.

2. Avoid parallel workstreams for innovation and risk teams

Matthew McDermott explained that while governments are excited about the possibilities of using AI for education and innovation, they're also thinking about regulation. There is a tendency to end up with a parallel track approach, with one group working on innovation and another working separately on risk. McDermott encouraged companies to join those two groups together so that the conversations are linked from the start.

Technology Partner

DELL
Technologies

Legal and Regulatory Challenges in Ethical AI

3. Know when regulation is necessary (and when it's not)

McDermott cautioned that as AI can be difficult to define it is important to acknowledge variation between systems, regulating where required and allowing innovation to flourish where it's not. If data is going to fundamentally affect peoples' lives then you need ethics boards and other mechanisms in place to catch things before they go wrong.

4. Beware compliance culture

Arash Ghazanfari noted that it is important to make sure that you are measuring the right things, and not simply regulating for the sake of it. Ghazanfari cited the example of GDPR, and specifically cookies warnings on websites. The warnings don't always clearly explain what exchange the user is being asked to consider, so the original intention of GDPR, that you own your own data and are in control of what you choose to share and exchange it with companies for, is ignored.

5. Market forces will punish bad behaviour

When asked how AI technologies would eventually become safer, Ghazanfari said that this would be driven by market forces working within a healthy regulatory environment. Individual expectations will shape the services that are offered, and bad behaviour will be punished by market forces.

6. Be clear on responsibility and accountability

While the panel agreed that ultimately it is the service provider who is responsible for their products and the data that they collect and use, attendees noted the nuanced nature of responsibility and accountability when using AI. For example, when a "black box" AI system in medicine suggests a treatment plan that we might not have predicted nor understand, how should we proceed? What are the risks of following it, and what are the risks of ignoring it?

Technology Partner

D∕LL Technologies

Contributors

Thank you to our speakers from University of Oxford, University of Cambridge, Dell Technologies, CMS and Access Partnership.

UNIVERSITY OF OXFORD

Dr Brent Mittelstadt
Research Fellow & British Academy Postdoctoral Fellow

Brent Mittelstadt is a Research Fellow and British Academy Postdoctoral Fellow in data ethics at the Oxford Internet Institute, University of Oxford, as well as a Turing Fellow and member of the Data Ethics Group at the Alan Turing Institute, and a member of the UK National Statistician's Data Ethics Advisory Committee. He is a philosopher focusing on ethical auditing, interpretability, and governance of complex algorithmic systems.

His research concerns primarily digital ethics in relation to algorithms, machine learning, artificial intelligence, predictive analytics, Big Data, and medical expert systems. Currently, he is examining the feasibility of ethical auditing of decision-making algorithms, and the development of standards and methods to ensure fairness, accountability, transparency, interpretability and group privacy in 'black box' algorithmic systems.

Technology Partner

DELL Technologies

Contributors

UNIVERSITY OF CAMBRIDGE

Dr Karina Vold

Research Fellow/Postdoctoral Researcher

Dr Karina Vold is a philosopher of mind, cognition, and artificial intelligence. She works on the ethical and societal impacts of emerging technologies and their effects on human cognition. Dr Vold is a postdoctoral Research Associate at the Leverhulme Centre for the Future of Intelligence, a Research Fellow at the Faculty of Philosophy, and a CanadaUK Fellow for Innovation and Entrepreneurship.

Technology Partner

DELL Technologies

Panelists

Arash Ghazanfari
Field CTO

Arash represents the office of the CTO in the UK & Ireland Region for Dell Technologies.

Arash serves as a Chief Technology Officer, a recognised Principal Technologist, supporting the overall go to market strategy across the full breadth of the Dell Technologies Ecosystem.

Prior to joining Dell Technologies, Arash held senior roles at Intel Security, VMware and other leading technology vendors in the hi-tech sector.

Technology Partner

Panelists

C/M/S/
Law.Tax

Rachel Free
Partner

Dr Rachel Free is a partner and UK / European patent attorney in the London Office Intellectual Property department.

Rachel uses her technical background in the fields of Artificial Intelligence, software and telecommunications to assist clients protect their technology through patents in this complex and changing area of patent law.

Technology Partner

D&LL Technologies

Panelists

Access Partnership

Matthew McDermott
Director, Policy & Regulation

International policy professional with a proficiency in government affairs.

Experience of global economic and commercial policy advocacy and analysis, especially ICT policy, including cloud computing, cybersecurity, spectrum acquisition and Internet governance.

Frequent interactions with United Nations and United Nations agencies, such as the International Telecommunication Union (ITU) at national and multilateral levels.

Technology Partner

DELL Technologies

Companies in Attendance

BARCLAYS

BT

eDF

DELL Technologies

AXA PPP HEALTHCARE

LLOYDS BANKING GROUP

Schroders

Pfizer

UNIVERSITY OF OXFORD

LV=

pwc

SLAUGHTER AND MAY

BDO

Pinsent Masons

NOKIA

CMS Law.Tax

UNIVERSITY OF CAMBRIDGE

BPP

FIRST RATE

Access Partnership

DataArt

BMI

SMARTER HUMAN

Technology Partner

DELL Technologies

![Dell Technologies]

Putting Data to Work

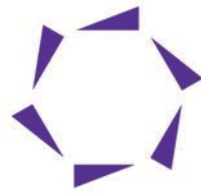How can you harness your data to give you that competitive edge?

🌐 www.DellTechnologies.com

🐦 @DellTech

Digital Leadership Forum

Join the next session

Applying AI for Social Good

3rd March 2020

Technology Partner

DELL Technologies

www.digitalleadershipforum.co.uk