CENTRE FOR
TECHNOLOGY &
GLOBAL AFFAIRS

www.ctga.ox.ac.uk

Workshop on Applied Artificial Intelligence: Privacy Ethics & Organizational Yield

A workshop organized by Oxford University's Centre for Technology and Global Affairs, r4 Technologies, and Palantir

Lady Margaret Hall, Oxford

On March 16th, 2018, Oxford University's Centre for Technology and Global Affairs held a workshop on Applied Artificial Intelligence (AI) in Lady Margaret Hall, Oxford. In attendance were academic and industry experts, business managers, and government representatives from the U.S. and Europe. r4 Technologies and Palantir – two industry leaders in the deployment of AI technology and data analytics respectively – co-organized the event with the Centre.

The main goals of the workshop were to:

- address privacy and other ethical concerns associated with the applied aspects of machine learning technology and Big Data Analytics; and
- explore the practical use of AI to improve organizational yield in the public and private sectors in the face of an accelerating pace of change in economic and social structures.

Too often, ethical concerns and organizational opportunities are viewed as conflicting goals. After extensive examination of real-world examples and hypothetical situations, however, participants strongly agreed that the bar of success for AI-related companies must be set high in terms of accomplishing *both* sets of goals simultaneously. Achieving this aim will require important changes in how companies design and implement AI and in how public authorities govern its use.

In particular, two key themes emerged from the workshop:

- the emergence of enterprise AI and its applications to government and industry; and
- new guidelines and policies to address ethical concerns in the use of AI.

Following are the key conclusions and recommendations based on the proceedings:

- **Organizational Yield is emerging as a new management discipline.** Enabled by AI, it seeks to connect the points across an organization's different functions and break down process and data silos to identify enterprise-wide opportunities for growth and major performance improvements.
- **Create a consensus-based ethical code of conduct for companies.** To ensure a level playing field and to ensure that AI is used for benefit and progress, rather than for exploitation, the development of an ethical code of conduct for companies is essential. In the UK, the establishment of the AI Council, which will gather leaders in the field from across academia and industry is an important first step in this direction.
- **Create an accountability framework.** As algorithms are increasingly relied upon for automated decisionmaking, a framework of accountability should be developed. Such a framework should take into account the stakeholders involved as well as the relationship between, on the one hand, algorithm developers and companies, and, on the other, the users of algorithms and decisionmakers.

The remainder of this report summarizes the above topics as they were discussed in the workshop.

1. ENTERPRISE AI CAN BE A DRIVER OF ORGANIZATIONAL YIELD

"Enterprise AI" has arrived. Enterprise AI is the application of machine learning and artificial intelligence techniques to run enterprises more efficiently.[1]

Businesses are successfully adopting the use of data and artificial intelligence to drive business performance. Academic and industry experts are aligned in describing AI as a fundamental technology shift that will enable and drive changes in competitive strategies in all industries – at a rate and pace never experienced before.

Ralf-Dieter Wagner of r4 discusses the benefits of AI technology

Technologies are adopted in phases. In the beginning we have early innovators, followed by a wave of rapid adoption as the technology becomes mainstream. Eventually the adoption of the technology saturates, before it is surpassed by another technological advancement. r4 provided context on the adoption of *Enterprise AI to drive Organizational Yield*, and noted that enterprise AI is at the tipping point between early innovators and a wave of rapid adoption.

In the past, the majority of the world's enterprises have achieved competitive advantage and sustained profitability by engaging in a management discipline focused on controlling costs by optimizing economies of scale – most of the time within functional silos. This discipline and its attendant IT infrastructure has delivered the advantages of tightly controlled costs, human productivity, and reliability and stability of transaction systems. More recently, most enterprises struggle with growth and face existential threats from highly agile digital disrupters. Value is trapped within silos.

Enterprise AI helps companies to break these silos and connect the dots across the different functions of an organization in order to enable organizations to continuously sense what people want, and to constantly align their value proposition accordingly – faster and faster, for increasingly micro-segmented markets, in order to identify perishable opportunities and to drive new growth and agility. A new cross-enterprise management discipline, "Organizational Yield", is emerging, enabled by AI.

In order to leverage the full benefits of Enterprise AI, organizations have to consider a new set of design principles for it. Organizations need to break out of the traditional platform and tool set paradigms and create ecosystems that integrate data seamlessly, that can integrate new algorithms and tools, and connect to operational systems to take immediate action.

r4 provided examples for new Enterprise AI design principles, which include:

Data as Fuel

Examine your data through the lens of your business; create a flexible, dynamic view of your organization and its ecosystem. Technical data integration such as creating data lakes alone is not the answer – creating a proprietary "market ontology" allows to break data silos and identify patterns and associations to discover opportunities across the organization and create competitive advantage.

Reinvent the Step Between Insight and Action

Classical analytics capabilities (and some of today's AI tools) provide scores, reports and visualizations. Business executives are left alone to come up with the appropriate action to take. Today's state of the art AI capabilities however go far beyond reporting and visualization to automated generation of recommended actions – which dramatically cuts time (and effort) between analysis, insights and business outcomes. Machine Learning enables continuous validation and optimization of those recommended actions and targeted business key performance indicators (KPI).

Human Intuition Meets Science

Do not treat AI as just another technology, but as one critical element of a new way of enabling organizations to accelerate business performance. Business executives have to step up and take courage to embrace that technology and not leave this to the IT departments. It is a Board-level issue as opposed to a delegated set of "IT Requirements". Do not treat AI as a black box. *Let human judgement and creativity guide AI*: *let AI accelerate human decisionmaking.*

1 "The Road to Enterprise AI," Rage Frameworks and Gartner (2017).

2. NECESSITY FOR NEW GUIDELINES AND POLICIES TO ADDRESS PRIVACY AND OTHER ETHICAL CONCERNS

While AI has clear benefits for organizational yield, the use of data and personally identifiable information, raises important ethical issues which were discussed in the second theme of the workshop.

Understanding clients' motivations in applying AI to private data and ensuring that solutions are engineered to make proper use of these data sources is considered key by the workshop audience. In addition, organizations that use any kind of personal data – whether in conjunction with AI or not – have an ethical obligation to consider and protect the basic rights and dignity of the individuals who will be affected by the systems that they build and/or deploy.

Participants stated that different applications of AI have different levels of ethical and privacy risks. Public discussions and policy making need to take this into consideration and develop an appropriately differentiated view of AI risk and benefits. At the same time, while AI presents novel ethical challenges, participants also found that there remain significant ethical questions surrounding even the most basic uses of data that society continues to struggle to address. The intense interest in AI ethics discussions should be leveraged to reinvigorate other important technology ethics debates.

Points for consideration discussed include:

Value Selection and Priorities

Before talking about ethical considerations, we first have to define what those considerations are. This in itself is a complicated process – particularly in the technology sector, where new capabilities can present novel ethical questions without clearly established societal norms to guide behaviour. Consequently, organizations must attempt to define their own ethical principles (grounded to the greatest extent possible in established values), which becomes increasingly complex as these organizations scale to include more and more people with diverse backgrounds and viewpoints. Palantir have tried to resolve this by integrating core "top-down" values from management and the "bottom-up" values of employees. Creating such an internal knowledge base of ethical principles, which are ideally communicated to clients as well, is a much needed step the industry needs to take. Collaboration on ethical issues, especially by the key players in the industry will also go a long way in establishing ethical norms.

Education

The issue of ethics education, both of stakeholders and clients and the people who are going to create AI/ML software, cannot be ignored. It is not clear what this would entail, but the most effective focus would be on teaching a framework for informed ethical discussion versus attempting to develop and apply a specific code of conduct that would almost certainly be outdated the moment it was completed. Effectively teaching this could involve compulsory requirements for ethics courses in undergraduate studies, as well as industry requirements for ethics knowledge when applying for AI/ML jobs. Another aspect of education is that of the next generation who will grow up with these technologies that will pervade their lives. (See for example a Swedish study which engages teachers and students to identify fake news and misinformation[2].) Education at all levels to raise awareness of the issues that AI/ML will cause is essential.

Accountability and Liability

Accountability for mistakes arising out of processes involving algorithms is another issue that was discussed in the workshop. Certainly in the early stages of the development of these capabilities, if not always, any decision that might negatively affect the rights and privileges of an individual should be ultimately reviewed by a human being and not just left entirely to an algorithmic process. Oversight mechanisms must be in place to ensure that an actual review is undertaken and that the machine output is not simply "rubber stamped" by a human functionary. This not only prevents unfairness, but it also generates useful feedback that can be incorporated to improve future iterations of the algorithm.

While "human in the loop" scenarios can be a solution to preserve liability, regulations or codes should be in place such that the humans who are in charge of automated systems, as well as those affected by them, fully understand the failure modes of these systems and how these could be subverted.

Identify Human Bias

Both human bias and algorithmic bias, which typically arises out of initial bias in training datasets can cause feedback in complex, tiered systems. While machines can help with identification of human bias, it is more helpful to think of bias propagation as a complex systems issue. An algorithm is not run over a simple, single set of data. Rather, a data set is more than likely composed of data pulled from other data sets that may have been transformed in part by the use of other algorithms trained on yet more data sets potentially with their own inherent biases and so on *ad*

2 www.tii.se/projects/nyhetsverderaren

infinitum. In short, data exists not in discrete "sets" but as part of complex and potentially vast data ecosystems.

Autonomic identification of bias criteria may very well remain out of the province of AI/ML systems, but it is an area that requires further exploration alongside the development of human assisted AI/ML systems. If the point of AI is to further assist in the understanding of the "truth" of the world, then AI compromised by bias fails in that endeavour and AI developers therefore ignore this issue at their peril.

Ethical Code for Companies

Some of the ethical issues arise because there is no enforceable code of conduct for companies or for coders. This can lead to a "race to the bottom" whereby ethically dubious organizations will seek out equally unethical – or, at best, ethically agnostic – technology partners. Some of the points raised in the discussion were about how other professions have a guild which enforces and defines the norms which should be followed. As an example, lawyers are expected to follow a code of ethics, otherwise they will be prevented from executing their profession. The group discussed the challenges of trying to define computer scientists as a profession and create a code of conduct that was similarly binding.

Principle of Least Information

Data-focused companies tend to default to collecting as much data as possible, arguing that the potential future value of such data justifies the costs of collection and storage. But such value is highly speculative and often inhibits effective data analysis by contributing unnecessary "noise." Thus, when designing algorithmic systems which require information that infringes privacy (in particular personally identifiable information), designers should ensure that the least amount of information is collected.

Reproducibility and Transparency

Especially for public facing and government services, producibility and transparency of machine learning algorithms is vital. At any given point in time, an individual can – and should – demand an explanation as to how a decision was reached that affects him or her in a significant way. A human understandable explanation for that decision should be readily available, which may be particularly challenging when an algorithm is constantly evolving due to ongoing feedback loops.

Future-Proofing Privacy Concerns

Given the current pace of technological development, the utility of a single data set can rapidly evolve as new capabilities in turn allow organizations to derive new insights. Seemingly innocuous data can suddenly present significant unanticipated risks as new ways to exploit this information are developed. Technologists must, to the greatest extent possible, anticipate and guard against these risks, and where such developments are completely unpredictable develop protocols for obtaining renewed informed consent from individuals before deriving value from these capabilities.

3. CONCLUSION

Like other technological advancements before it, machine learning and other artificial intelligence techniques can be used for the benefit of industry, government, and society – but they carry negative implications as well. In contrast to previous technologies, where humans were a key part of the production, dissemination, and use of new technology, artificial intelligence and machine learning techniques can replace or augment some of the roles of humans in these stages. This reality has immense benefits and consequences. By augmenting human capabilities, AI and Big Data Analytics promise to provide a more accurate, holistic picture of human reality and business activities, which can help businesses improve their efficiency and drive organizational yield. On the other hand, automated insights from widespread data collection have potentially important implications for privacy, anonymity, and bias propagated by algorithms. The workshop's discussion of the ethical use of data and algorithms emphasized these dangers. It affirmed the necessity for the development of industry-wide guidelines and ethical codes of conduct.

The challenge is that as ethical guidelines tend to coalesce around societal norms, technological development dramatically outpaces the establishment of such norms (see for example the uptake of social media platforms vis-à-vis the development of an understanding of appropriate behaviour on such platforms). Establishing a workable code of conduct that is not out of date from the moment it is published may ultimately be a Sisyphean ordeal. Instead, organizations – with the help of institutions such as this Centre – should focus on the development of more sophisticated discussion frameworks that enable a more informed analysis of the ethical implications of technology and point the way towards reasonable solutions tailored for a broad spectrum of positive societal outcomes. Particularly important is the elaboration of a classification of different AI applications as well as their associated value and risk potentials to guide the governance and policymaking process.

CENTRE FOR
TECHNOLOGY &
GLOBAL AFFAIRS

UNIVERSITY OF
OXFORD

ABOUT THE CENTRE FOR TECHNOLOGY AND GLOBAL AFFAIRS

The Centre for Technology and Global Affairs at Oxford University is a global research and policy-building initiative focusing on the impact of technology on international relations, government, and society. The Centre's experts use their research findings to develop policy and regulatory recommendations addressing the transformative power of technological change.

The Centre serves as a bridge between researchers and the worlds of technology and policymaking to impact policy in the resolution of pressing problems across six technological dimensions: Artificial Intelligence, Robotics, Cyber Issues, Blockchain, Outer Space, and Nuclear Issues.

The Centre's mission is (a) to provide leadership in creating new knowledge on practical problems affecting the security and welfare of governments, citizens, and private enterprises; (b) to influence major policy decisions and opinions in these arenas; and (c) to guide the work of leading technology developers and policymakers.

The Centre is based in the Department of Politics and International Relations at Oxford University. It is supported by core funding from Kluz Ventures.

This worskshop was supported by funding from Kluz Ventures and r4 Technologies.

Centre for Technology and Global Affairs
Department of Politics and International Relations
University of Oxford
Manor Road
Oxford OX1 3UQ
United Kingdom

UNIVERSITY OF
OXFORD

DPIR
DEPARTMENT OF POLITICS & INTERNATIONAL RELATIONS