



University of
Salford
MANCHESTER

AI management an exploratory survey of the influence of GDPR and FAT principles

Addis, MC and Kutar, MS

<http://dx.doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00102>

Title	AI management an exploratory survey of the influence of GDPR and FAT principles
Authors	Addis, MC and Kutar, MS
Publication title	2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)
Publisher	IEEE
Type	Conference or Workshop Item
USIR URL	This version is available at: http://usir.salford.ac.uk/id/eprint/56950/
Published Date	2020

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: library-research@salford.ac.uk.

AI Management: An exploratory survey of the influence of GDPR and FAT principles

Chiara Addis

Salford Business School

Salford, UK

M.C.Addis@edu.salford.ac.uk

Maria Kutar

Salford Business School

Salford, UK

M.Kutar@salford.ac.uk

Abstract— As organisations increasingly adopt AI technologies, a number of ethical issues arise. Much research focuses on algorithmic bias, but there are other important concerns arising from the new uses of data and the introduction of technologies which may impact individuals. This paper examines the interplay between AI, Data Protection and FAT (Fairness, Accountability and Transparency) principles. We review the potential impact of the GDPR and consider the importance of the management of AI adoption. A survey of data protection experts is presented, the initial analysis of which provides some early insights into the praxis of AI in operational contexts. The findings indicate that organisations are not fully compliant with the GDPR, and that there is limited understanding of the relevance of FAT principles as AI is introduced. Those organisations which demonstrate greater GDPR compliance are likely to take a more cautious, risk-based approach to the introduction of AI.

Keywords— Artificial Intelligence, Machine Learning, Data Protection, Fairness, Accountability, Transparency, Management, GDPR

I. INTRODUCTION

The development of AI has sparked a debate on its potentials, and concerns and new questions on the protection of personal data have emerged. The EU's General Data Protection Regulation (GDPR) has imposed new requirements for organizations handling personal data, and the implications for businesses managing AI technologies are particularly important. This research aims to explore the praxis of data protection and AI Management in UK organizations, investigating how adopters and users of AI technologies perceive, understand and apply the Fairness, Accountability, and Transparency (FAT) principles, and how this affects organizations. Existing literature demonstrates that while the FAT principles are discussed within the AI community [1], a substantial engagement is absent from dominant practices in innovation management in UK organizations. It is notable that there is little reported research in this area and much of what is reported arises from grey literature and popular media rather than systematic scientific studies; this is expected to change as the area matures and the research reported here intends to contribute to the gap. We present the initial stage of a study, in which interviews designed to explore how AI and DP are addressed, providing an analysis of perceptions, understanding, and experiences of managers and AI experts that are implementing or using AI in their specific contexts. It is hoped that the research will lead to useful guidance for organizations in their AI Management.

II. AI AND MACHINE LEARNING (ML)

We're starting slowly but surely to employ machine learning in ways where the machine's actions actually have an impact

on the world, from which the machine then keeps learning.
(Jakob Uszkoreit, 2018) [2]

Generally defined as the capacity of a machine to act or think like a human being, AI has rapidly advanced in recent years due to the increase of available data, more powerful computing, and better algorithm techniques. Breakthroughs have been followed by a growing debate on its potential, applications and concerns [3]. More than other technologies, AI seems to have the potential to transform social, economic and political orders, impacting economic possibilities, and the rights and freedoms of individuals. AI has become a General-Purpose Technology used in varied applications and sectors to increase efficiency and production. For example: Google Search, Google translation, email spam filters, personal assistants, speech to text applications, image recognition, prediction of energy consumption and transports, and their optimization, robots, self-driving cars all utilize AI. Applications now extend to emotions and feelings, for example monitoring physical and emotional states of drivers by measuring faces, heartbeat and voices [4]. The advancements in AI are inspiring, in particular in healthcare; DeepMind has achieved breakthroughs that are milestones in the AI evolution, using deep learning in healthcare, to improve diagnosis of cancer, eye and kidney diseases [5, 6]. However, the deployment and use of the technology can produce other consequences, such as reduction of privacy, misuse of personal data, reinforcement of patterns of discrimination already existing in societies, and socio-economics changes. For example: DeepMind Health raised many questions for its use of patients' data [7, 8]. Some speech to text applications can clone voices [9], or generate new words from voice samples, as the case of Adobe VoCo, called the 'photoshop' for faking words. Some ML applications are able to learn from watching videos, or to generate new images from real ones, while others can be used to detect and measure other qualities, such as "morals" or "honesty" of travelers at airports [10]. There are concerns around biased algorithms leading to discriminatory decisions, the lack of transparency in opaque algorithms (particularly in deep learning) and the use of AI to manipulate public discourse (which has already emerged with Cambridge Analytica [11], and further malicious uses of AI are predictable [12].

III. GDPR

The GDPR increased the protection of personal data of data subjects in the EU, created new obligations for organizations, and created a more uniform Data Protection regime across the EU. It expanded the definition of personal data, introduced the principle of Accountability, and provided organizations with a more flexible mechanism for operating their business in different EU countries (i.e. One-

stop-shop). The Regulation constitutes a crucial evolution in Data Protection legislation and has influenced new data protection legislation beyond the European Union, such as the California Consumer Privacy Act of 2018 (CCPA) and the Washington Privacy Bill [13]. The Regulation prescribes new requirements for the processing of personal data by AI systems, such as: the right of explanation or information (right to have meaningful information about the logic or an explanation of some type); the right to request human intervention (in the case of decisions which produce legal effects or similarly significant affects to the data subjects); the right to refuse (in specific cases) a decision made only via automated processing, and requires a Data Protection Impact Assessment in the case of high-risk processing for a greater range of organizations. There are, however, some areas where the protection is less clear:

- Personal and non-personal data. Aggregating non-personal data can identify data subjects.
- Afterlife data rights are excluded by the GDPR. However, there are projects which use personal data (such as voices) to recreate digital copy of the dead, or chatbots that are digital duplicates of the dead [13].
- Consciousness, emotions and feelings are other areas of interest. If human brains are scanned to recreate a virtual post-mortem replica, should that copy be considered as personal data? [14]. Can emotions and feelings be protected as biometric data?

A. Fairness, Accountability and Transparency in Algorithm Systems

Interest in the social implications of AI has increased in the last few years, and debates on Fairness, Transparency, and Accountability are becoming more frequent. Key concerns are unfairness resulting from biased algorithms leading to discriminatory decisions, transparency (mainly linked to lack of transparency arising from opaque algorithms), and accountability as AI systems become more autonomous. Research on AI and FAT is mostly focused on algorithms and on how to regulate them. Consideration of how AI is used in management, human factors, contexts, and FAT in operational contexts, are still underexplored.

IV. INNOVATION AND AI MANAGEMENT

Digital transformation has become a business necessity for organizations, and the number of organizations that are using or planning AI technologies is rising.

A new form of economy has emerged, a virtual economy [16], based less on production and more on the distribution of information and on sharing between people. Systems of constant and ubiquitous surveillance are increasing. Data, algorithms and business processes “continuously communicate with one another” [17, p3]. and create an autonomous “external intelligence” which use “huge libraries of intelligent functions...(that) bit by bit render human activities obsolete” [Ibidem, p34]. This process has already started, and Arthur identifies two main reactions from organizations: integrating new technologies into existing structures or creating new business models using new parts. Furthermore, traditional domains are merging, and new innovation strategies that encompass knowledge of different sectors are now required [18]. This entails different competences in understanding, creating and driving innovation, and different approaches to the management of

AI [19, 20]. This has considerable implications for Data Protection. Data are being used in all innovation processes: algorithms are created and trained with data, use data to make predictions, and use feedback data to continuously learn and improve from the environments where they are deployed. Data are becoming the most important asset for organizations, that are using varied internal and external data: data in the organization, from partners, the data industry, and from interactions on social media. Furthermore, the pace of digital innovation in digital business is fast, as a new idea can be realized, tested, and changed in a short time. Those who experiment first (and those who follow fast) gain a huge competitive advantage [21]. Therefore, organizations are using more data, are creating more products and services in shorter times, and are accepting bigger risks, as inactivity could be too risky for their business continuity. There is a pressing need to embrace digital innovation, and the process is even faster with Artificial Intelligence systems, which poses specific risks to organizations:

- Relationship with vendors. Vendors are bringing AI to organizations, are selling the technology and educating clients on AI specificities and compliance. However, they are also co-deciding on how to process data and using clients’ data for their own innovation, sparking intense debate on their role as mere processors, joint-controllers or controllers [22].
- Use of “out of the box” systems. They can create a false sense of security in managers and untrained staff and could potentially lead to a lack of control which impacts on risk. Moreover, algorithms may be created in and use data taken from other contexts, and may not work properly in the new context, leading to discrimination and non-compliance in those application environments.
- The probabilistic nature of ML can be mistaken for a more deterministic one. Predictions bear a level of uncertainty, but they can be mistaken for certain forecasts.
- ML and new taxonomies; the ML capacity to identify patterns in datasets is used to identify correlations that can be mistaken for causations. When this is used with knowledge representation, the organization of knowledge into categories, subcategories, subclasses and their relations [23] can produce implications for individuals, who are ranked according to different criteria, such as facial features or moral qualities, and for business and business models. For example, a new range of personalized products based on consumers’ DNA and other biometrics data are being offered without much scrutiny or deep considerations in relation to data protection and fairness and potential risks for the data subjects rights [24].

Therefore, while many issues related to AI/ML, data protection and FAT are specifically related to data, others are linked to their praxis within organizations, involving management and leadership’s responsibility.

A. An Ambivalent Technology

Adopting a wide and holistic approach (which includes human factors) and looking at the use of innovation in different contexts, and at power dynamics amongst

stakeholders, the research aims to explore limitations of people's agency within organizations, how an ambivalent technology is socially shaped and constructed) [Ibidem], and how it reflects the values existing in that space. A critical review of the existing literature has revealed that research on the relationship between Artificial Intelligence and Data Protection is in its infancy, whilst research on ethics and the social effects of algorithms is mainly focused on data and on technical aspects. Contexts, internal processes, people and stakeholders' dynamics are still underexplored. Research is mainly concentrated on the consequences of the technology, on remedies, and on mechanisms for redressing violations of rights resulting from decisions made by AI systems. Research on "preventive" Data Protection is still scarce. Substantial engagement with the FAT principles is absent from dominant practices in innovation management in UK organizations, with little research focused on FAT in the identification and application of AI systems, Operational FAT.

V. THE PROJECT

This project aims to provide insights into AI Management, to understand the impact of Data Protection and AI technologies on organizations, exploring how leaders and managers adopting and using AI technologies perceive, understand and apply the FAT principles, and how this may affect organizations. The project is ongoing, and the results presented and discussed below represent emerging findings from the initial survey.

A. Study Overview: Participants and Question Themes

Interviews were used to understand perceptions, understandings and experiences of experts in data protection and AI. A semi structured interview approach was employed, with 6 participants including Privacy lawyers, Data Protection consultants, Technology Business, and ML experts. Interviews were conducted in March and April 2019. Participants have experience in both private and public organizations across a range of sectors. Participants jointly have considerable expertise in data protection, AI, ML, and data technology management, and comprise individuals who are on national and international expert groups on AI.

They provided overviews of trends and information from their industries, and insights from the experience in their current assignments. Questions were based on 4 main themes which were identified from the review of the literature:

1. AI Management and Strategy. This focused on organizations and leadership (reasons for and means of acquiring AI, level of AI understanding by leadership, the involvement of vendors), and data and risks (data type and quality checking).
2. GDPR and Processes. This theme aimed to understand what organizations had done with regards to new requirements, AI, and relationship to internal processes.
3. FAT. This focused the principles, and explored the involvement of stakeholders, AI specific challenges, and AI limits and exclusions.
4. Future. Participants vision of the future of AI and FAT were discussed.

B. Emerging findings from interviews

A primary thematic analysis of 6 interview transcript was carried out considering the 4 themes above.

1) AI Management and Strategy.

Organizations are acquiring AI technology mainly due to pressure, interest, and curiosity coming from various subjects. There is pressure originating from boards, senior management, or innovation areas within the business that is driven by the desire to improve operational efficiency, reduce costs, and maintain competitiveness. In some cases, this includes awareness of precarious business continuity, often intensified by the difficulty of embedding and operating AI in more traditional organizations

"Organizations that have been around for more than 8 years, they have a problem...an industrial core, how do they embed AI into their business? How do they change their companies? A lot of organizations born industrial 15 years ago still have industrial activities they have to do" (Participant 1).

Similarly, individuals and teams on the ground are becoming informed about AI, see the potentials for improving their activities, and are suggesting it to management.

"They suddenly realize they are sitting on lots of data, and that could do something, like getting more insights using machine learning and that might be useful for the business or they may create systems that could monetized" (Participant 1).

The availability of open-source technology and cloud-based frameworks makes AI easily and quickly accessible. Of relevance is the role of start-ups, believed to be less aware of the GDPR requirements and to be taking more risks to react quickly to market. This supports the literature on the pace of digital innovation, and its impact on risks and compliance.

In bigger organizations, many are developing their own systems, without fully understanding the implication for risks and compliance, and without central governance oversight of what is being done in different business areas. This is intensified in organizations with strong silo mentalities.

A general low knowledge of AI technology was reported in executives and management. Leaders are reported to be signing off systems they do not properly understand, including the implications for security and data protection compliance. Some have high expectations from AI, seen as a tool able to do everything better. As AI can radically change the way organizations work, they are facing big decisions, and some are becoming aware of future big challenges for their businesses (such as job cuts) while others are concerned about the impact on employees. The low level of AI knowledge translates into uncertain AI strategies. The shortage of specific reskill training for staff seems to suggest a lack of interest or knowledge in adopting AI augmented strategic models. "People are not being liberated to do other things, most employers are just liberating the people full stop" (Participant 1). Vendors provide off the shelf technology when organizations face uncommon challenges and need tailored solutions that generic solutions cannot provide. They are more aware of security risks and obligatory compliance and are informing clients. They are also using clients' data for their innovation, thanks to

contracts that allow data capture, which confirms their role as joint-controllers and not as mere processors.

Specific risks for ML, such as data manipulation, reverse engineering or adversarial data able to trick the models, are viewed as very low, or are not considered at all in organizations' agenda, with some differences due to data sensitivity and contexts.

Other areas of concerns related to data sources including the collection of uncontrolled data from internet services. ML is often trained with real data in a very controlled environment, and its access by various developers, and security in case of system failure were mentioned as specific risks for data protection.

GDPR is identified as having increased the awareness of risks coming from data location, and of access by different people. As the GDPR allows a risk-based approach, organizations are minimizing risks in some areas, and taking higher risks in others where compliance is more challenging, for example, uncertainty on consent withdrawal for data used to create AI systems. Other specific ML risks are related to training, deployment and loss of accuracy. The efficiency of ML algorithms can reduce over time. Systems developed in the lab with training data can show a high accuracy rate, which then drops when deployed it into the real world. "Some systems might be able to retrain it, others would continue to degrade performance and accuracy...because they are not being retrained with real data" (Participant 4). Not many in organizations are believed to be aware of this reduction of efficiency, which was compared by Participant 4 to old Intrusion Detection Systems, good for detecting only known threats. Similarly, ML systems that do not adapt "are just looking to match the past things we were aware of in the past...this is a serious problem for senior people if they are not educated, as the name artificial intelligence makes you think that that is already happening" (Participant 4).

2) *GDPR and Processes*

There was a consistent view of participants that many organizations are not GDPR compliant, as many were reported to have done the minimum, or nothing. "Most of them had done nothing, they are waiting for their competitors to have a major data breach...the GDPR has not even scratched the surface" (Participant 1). Organizations seem to be more concerned around commercial confidentiality of information, than personal confidentiality of information (Participant 5). Differences exist across sectors, with organizations in the regulated sector being more mature, confirming the gap that emerged in pre GDPR research [25]. The GDPR requirements related to AI are not seen as easily achievable at this early stage of the use of the technology. Similarly, not many organizations were reported performing DPIA or privacy by design as these preventive activities were seen as "luxury" and barriers to innovation. "Many are thinking a little bit about it, but it is not essentially privacy by design...it is more we need to think about this now we are going to deal with it." (Participant 5). Organizations are said to be focusing on "getting intelligence out of the data to be able to monetize" (Participant 4), and on creating systems to respond quickly to market needs. Furthermore, while some organizations are aware of taking risks, others do not have the necessary knowledge or information, and many assumptions are being made. For example, the expectation that security measures are all done in the cloud, or that AI systems self-learn as they are "intelligent". The sector, size

and maturity of organizations are key factors in having data protection and data teams that work together, such as in the public sector, where organizations already had the obligation to perform DPIAs. However, it was not clear if organizations starting to deploy or to use AI/ML were performing DPIAs, or if vendors were doing them (something considered particularly important for the Privacy Lawyer), or if fairness and social impact were already taken into consideration by more mature companies.

Of interest, is the creation of working groups "to make sure everybody is connecting on the same page" (Participant 4). This is happening particularly in Finance and Big Tech (especially after Cambridge Analytica), and is less evident within medium and smaller companies. This multidisciplinary activity is not happening amongst start-ups, where there are fewer data scientists. "They themselves don't fully understand what they are doing, and they probably have a very high level of abstraction. They don't know which questions they should be asking, and do not understand the deeper level to be worried about, like biased algorithms". (Participant 4). Yet, they are using open-source technology to create ML systems, and they are doing it very quickly.

3) *FAT*

Fairness is usually discussed in relation to discriminatory outcomes derived from biased data. Other sources of unfairness are not generally considered by companies. However, other relevant situations were identified by participants. For example, systems created in other countries could be brought into the UK, leading to potential unfair results. The issue would not be caused by biased data but by their use, as data used to train models are usually specific to the domains where they are going to be applied. "If you're trying to train a model with data from somewhere else, to me this always raises a red flag...because you are making the assumption that customers are always the same...how do you enforce that?" (Participant 3). Human oversight, with the power to stop the process, audit, and internal control were suggested as potential tools to increase fair outcomes. In contrast however, one participant considered algorithms as inherently biased as being created by humans, "the machine will always have biases, always being created by a programmer, and the programmer has prejudices" (Participant 1).

Another potential issue for fairness was identified in the request by organizations to use ML to automate existing decision processes. Starting by analyzing the process was said by Participant 3 to replicate possible existing biases. Instead, they suggested to start by looking at the data, "see what the data is telling us", and taking the truth from the data, "proactively building something that is less likely to come back in court because you can say this is the rigor we put into the algorithm". Ownership of data in organizations was another reported issue: "In my experience the most part of the majority of executive there is no thought about that, at all. If you think about the data in the company, if there is a problem, who owns the data? IT says Business, and business says IT." (Participant 1).

Accountability is acknowledged to be a big issue. Management's responsibility and lack of understanding were raised by Participant 3, who believed many were delegating responsibility or signing things they did not fully understand. The progressive autonomy of AI systems and lack of scrutiny of some ML systems were the biggest concerns. For

example: autonomous systems that continuously learn without human oversight; non-transparent autonomous systems making very key decisions or left for a long time without scrutiny; autonomous algorithms used by different organizations which collectively lead to a unique adversarial decision. Decisions can be based on a reduced number of factors, without human awareness or intervention.

Transparency. Even though some AI models can provide confidence, explainability was recognized as a big issue in Deep Learning, and contexts, purpose of processing and people involved in the process identified as important elements in providing more clarity and information, as Computer-Aided Detection (CADe) can do with radiographic images. Ongoing tests to identify discrepancies in the model were also seen as an expression of rigor around training, as they could for example facilitate the identification of discriminatory outcomes for a specific ethnic group, even when the ethnic background data are not fed into it but resulted from other factors. By looking at the data, via reverse-engineering, those factors could be identified. This would constitute, and demonstrate, good rigor in face of poor explainability. However, this comes with a cost for the organization.

Furthermore, some expressed some concerns over AI systems able to provide an explanation:

“The way that AI works it's completely different to the way a rational human brain sees thinks...AI has not contextual understanding of what's going on so to explain decision-making. It's very difficult to say what AI is actually doing, in what counts in human terms as an explanation.” (Participant 5).

In general, people are not very familiar with FAT and GDPR principles. “In more mature organization, people are asking the right things in the right way, for example, at some high level of risk management, but they are not doing it consistently and not in relation to the GDPR” (Participant 4). In the interplay between ML, Data Protection and FAT there are a lot of unknowns. While some people are just avoiding scrutiny, which could impact their ability to innovate (something allegedly happening a lot), others need specific guidance. “We do not have yet a strong governance around the ethics side... We want to be better at understanding if what we are doing is ethical...in Finance you have ML algorithms that are doing auto-trading...how you know that the algorithm is ethical when what it is doing is completely autonomous...there are morality questions, and there is a lot of psychology coming into play in terms of interdisciplinary research. That is a lot of unknown around what is considered to be good practice in this space” (Participant 4). Transparency, auditing, human supervision and power to overwrite decisions made by AI, are all starting to be discussed inside organizations. However, there is little enthusiasm in engaging in this debate due to the complexity of the topics and concerns that this can become a barrier for their activities. “They want to develop the technology and use it and engaging in that debate might stop them get the benefits” (Participant 4).

4) *Future of AI and Data Protection*

More regulations guaranteeing fair competition, the power of big tech, and a future role for the UK in Ethics of AI were all envisaged. Participants have visions for a fair,

and accountable AI, regulated to guarantee standardized access and data sharing, processing and competition, and potentially regulated by international agreements (similarly to nuclear power). FAANG groups (Facebook, Apple, Amazon, Netflix and Alphabet's Google), are expected to grow, but some surprising innovations from smaller companies were also not excluded. A potential unintended consequence of the GDPR is its impact on innovation, which may lead other countries to take the lead on the development of ML. However, this was seen positively, as fair AI is needed, more research is required into this area, and the UK is working on becoming an international player in Ethical AI, with legislators viewed as challenging on legality, morality, and ethics.

C. *Discussion*

Many UK organizations are choosing AI and ML for their digital strategies, with differences according to size and sector. They are looking to improve efficiency, maintain competitiveness and respond to market requests. This is mainly due to the senior management's input or sometimes the desire coming from specific teams. AI is acquired either via vendors, who then implement it, support and inform clients, or directly via open-source technology. Vendors are using clients' data for their innovation.

Specific AI/ML risks are not known and are sometimes underestimated or just avoided by organizations. They are not well informed, although it is unclear whether this is due to a lack of specific resources, or avoidance the issues that might compromise their innovation. Not many organizations were reported as being GDPR compliant or having done a great deal to become compliant. Those which are engaged and informed reported an increased level of awareness in their activity, with more preventive thinking in their activity, indicate the positive effect of GDPR on organisational awareness. However, this arguably translates into a cost that other organisations may be unwilling or unable to pay.

FAT principles are mainly discussed in relation to biased data, and risks coming from non-transparent and autonomous systems. Various potential solutions for a more fair, transparent, and accountable AI were envisaged (such as testing, auditing, human intervention). However, organizations are currently focusing on getting, implementing and using AI, and discussing FAT principles while understanding what AI can do for their needs is rarely done. Those who do understand potentials and risks are aware of the many unknowns, and of the lack of specific guidance on AI Ethics, and are asking for more future guidance.

VI. CONCLUSION

Artificial intelligence is rapidly advancing, and it is deemed to become a General-Purpose Technology, whose application will produce profound changes in organizations and in societies. The GDPR has increased personal data protection and created new obligations for organizations using AI. The adoption of AI is growing and will be central in many digital transformation strategies. Potentials and concerns are fueling the debate, and discourse on Fairness, Accountability and Transparency are starting to appear in relation to the application of AI.

This project is looking at the praxis of AI/ML, data protection and FAT principles in UK organizations. The

interviews presented here have highlighted a number of issues: the market pressure to acquire and use AI, combined with the roles of different internal and external stakeholders in making strategic decisions. Participants reported a low understanding of AI/ML in Senior Management, with scarce awareness of specific risks, together with inadequate GDPR compliance across a range of sectors and sizes of organisations (it is notable that those more advanced are said to be adopting a risk-based approach). Findings suggest there is inadequate or non-existent internal debate on FATs, with some avoiding it for its potential impact on innovation, but with others willing to learn more about ethics and compliance.

The findings illustrate the gap between expert understanding of Data Protection and FAT principles and their relationship to the introduction of AI, and the expert participants view of the praxis in organisations. Leaders and managers must balance the need for digital innovation against both the legislative requirements of GDPR and the FAT principles it enshrines. The pressures of the commercial environment do not naturally allow for a considered approach, and this is compounded by weak understanding of the data and technologies. However, the fact that those organisations which have more effectively managed the introduction of GDPR are taking a more cautious, risk-based approach to the introduction of AI technologies, indicates that the legislation is having the desired effect where it is understood. There are however clearly many challenges arising from the lack of consistency of organisations in embedding the enhanced requirements of GDPR across their organisations and the consequences of this as they embrace AI. Having identified clear areas of interest in the application of FAT principles to the introduction of AI, the next stage of this research will be conducted employing case studies to provide greater depth, which will help us to understand the interrelations between innovation, regulations, contexts, and industry specificities.

ACKNOWLEDGMENT

C. Addis thanks Christian Klesse for his feedback on a draft of this paper, and for stimulating conversations on technology, classifications and fairness.

REFERENCES

- [1] FAT/ML Conference, "FAT/ML Conference - 2018 Papers", Retrieved from <https://www.fatml.org/schedule/2018/page/papers-2018>, 2018.
- [2] B. Reese, "A Conversation with Jakob Uszkoreit.", *Voices in AI*, Episode 70, Retrieved from <https://gigaom.com/2018/10/04/voices-in-ai-episode-70-a-conversation-with-jakob-uszkoreit>, 2018.
- [3] S. Bird, S. Barocas, F. Diaz, K. Crawford, and H. Wallach, "Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI. Workshop on Fairness, Accountability, and Transparency in Machine Learning". Retrieved from <https://ssrn.com/abstract=2846909>, 2016.
- [4] Affectiva, "Emotion AI Overview What is it and how does it work?", Retrieved from <https://www.affectiva.com/emotion-ai-overview/>, 2018.
- [5] DeepMind, "Helping clinicians get patients from test to treatment, faster. Retrieved from <https://deepmind.com/applied/deepmind-health>, 2017.
- [6] E. Harwich, and K. Laycock, "Thinking on its own: AI in the NHS | Reform, (January). Retrieved from <https://www.wiltonpark.org.uk/wp-content/uploads/Thinking-on-its-own-AI-in-the-NHS.pdf>, 2018.
- [7] J. Powles, and H. Hodson, "Google DeepMind and healthcare in an age of algorithms." *Health and Technology*, 351–367. <https://doi.org/10.1007/s12553-017-0179-1>, 2017.
- [8] S. Shead, Google DeepMind is giving the NHS free access to its patient monitoring app. Retrieved April 7, 2018, from <http://uk.businessinsider.com/nhs-discloses-how-much-its-paying-google-deepmind-2017-6>, 2017.
- [9] Lyrebird. Retrieved from <https://lyrebird.ai>, 2018.
- [10] The UK Parliament, "Dr James O'Shea – Written evidence (AIC0226)." Retrieved from <https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/publications>, 2018.
- [11] P. Greenfield, "The Cambridge Analytica files: the story so far.", *The Guardian*, Retrieved from <https://www.theguardian.com/news/2018/mar/26/the-cambridge-analytica-files-the-story-so-far>, 2018.
- [12] Z. Tufekci, "We're building a dystopia just to make people click on ads." Retrieved from https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads/up-next, 2017.
- [13] B. G. Cesaratto, "Washington State Considers Comprehensive Data Privacy Act to Protect Personal Information.", Retrieved from <https://www.natlawreview.com/article/washington-state-considers-comprehensive-data-privacy-act-to-protect-personal>, 2019.
- [14] A. Matei, "New technology is forcing us to confront the ethics of bringing people back from the dead.", Retrieved May 15, 2018, from <https://qz.com/896207/death-technology-will-allow-grieving-people-to-bring-back-their-loved-ones-from-the-dead-digitally>, 2018.
- [15] M. Graziano, "Why You Should Believe in the Digital Afterlife.", Retrieved May 15, 2018, from <https://www.theatlantic.com/science/archive/2016/07/what-a-digital-afterlife-would-be-like/491105>, 2016.
- [16] W. B. Arthur, Where is technology taking the economy? *McKinsey Quarterly* 697, 2017.
- [17] A. Edmondson, and B. Saxberg, "Artificial intelligence takes shape." *McKinsey Quarterly*, (4), 56–62. Retrieved from <https://www.mckinsey.com/~/media/McKinsey/McKinsey%20Quarterly/Digital%20Newsstand/2017%20Issues%20McKinsey%20Quarterly/Q4-2017-McKQuarterly-Full-Issue.ashx>, 2017.
- [18] The Boston Consulting Group, "The Most Innovative Companies 2018: Innovators Go All In On Digital." Retrieved from http://image-src.bcg.com/Images/BCG-Most-Innovative-Companies-Jan-2018_tcm9-180700.pdf, 2018.
- [19] M. Luca, J. Kleinberg, and S. Mullainathan, "Algorithms need managers, too." *Harvard Business Review*, 94(1), 20. 2016.
- [20] Deloitte. "State of AI in the Enterprise, 2nd Edition Early adopters combine bullish enthusiasm with strategic investments." Retrieved from https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf, 2018.
- [21] J. Bughin, T. Catlin, M. Hirt, and P. Willmott, "Why digital strategies fail." *McKinsey Quarterly*. Retrieved from <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/why-digital-strategies-fail>, 2018.
- [22] A. Roosendaal, "DPIA Diagnostic Data in Microsoft Office ProPlus, Commissioned by the Ministry of Justice and Security for the benefit of SLM Rijk Vendor Management (Strategic Microsoft Dutch Government), Retrieved from <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rappor-ten/2018/11/07/data-protection-impact-assessment-op-microsoft-office/DPIA+Microsoft+Office+2016+and+365+-+20191105.pdf>, 2018.
- [23] P. Norvig and S. J. Russell, "Artificial intelligence: a modern approach." Harlow, Essex, England: Pearson. <https://doi.org/9780136042594>, 2016.
- [24] K. Fischer, "DNA Test to Pick Roommates? What's Next?," Retrieved from <https://www.healthline.com/health-news/dna-test-to-pick-roommates#3>, 2018.
- [25] M.C. Addis, and M. Kutar, "The General Data Protection Regulation (GDPR), Emerging Technologies and UK Organisations: Awareness, Implementation and Readiness." In *Proceedings of the UK Academy for Information Systems Conference Proceedings*, Oxford, UK, pp. 20-21. 2018.