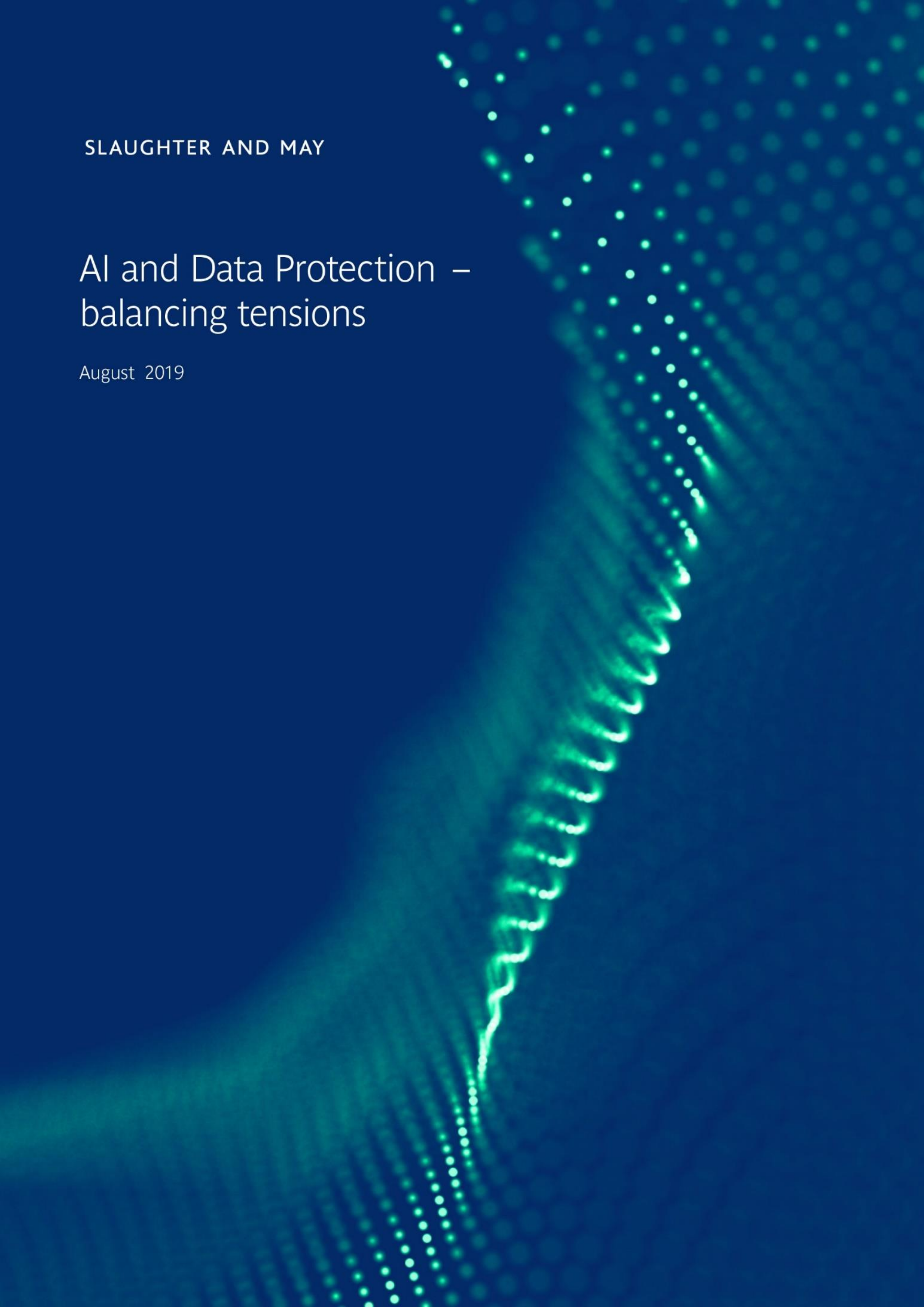
SLAUGHTER AND MAY

AI and Data Protection – balancing tensions

August 2019



AI and Data Protection

Balancing Tensions

In this paper Rob Sumroy and Natalie Donovan of Slaughter and May's Emerging Tech practice look at the some of the key data protection concerns relating to AI and how the UK's data regulator (the ICO) is responding to the new risks and opportunities it raises.

Artificial intelligence (AI) has "the potential to transform our world, from improving healthcare, reducing energy consumption and predicting climate change to credit scoring and fraud detection" (see EU Communication on AI for Europe). At a more commercial level, it can help organisations to profile, interact with, and sell to, their customers (see box "*What is artificial intelligence?*").

As AI becomes ever more popular, organisations are starting to grapple with the reality of how to balance AI design and deployment with data protection compliance. At the same time, regulators and governments are faced with their own AI balancing act: ensuring the safe and ethical deployment of AI without stifling innovation.

Some commentators consider that the EU has failed to strike this balance. For example, in March 2018, the Centre for Data Innovation argued that failing to amend the General Data Protection Regulation (*2016/679/ EU*) (*GDPR*) to reduce its impact on AI will all but consign Europe to second-tier status in the emerging AI economy.

While it is hard to measure the impact of EU data protection legislation on AI development, the European Commission (the Commission) did recognise, in its communication on AI for Europe published in April 2018, that Europe is behind Asia and North America in terms of private investments in AI.

In the UK, the Information Commissioner, Elizabeth Denham, gave a speech on AI and privacy in March 2018, during which she acknowledged the excitement about how AI is already enriching daily life. However, she also noted that it is one of the top three priorities of the Information Commissioner's Office (ICO), given AI's ability to intrude into private life and have an impact on human behaviour through the manipulation of personal data.

When looking at the ways in which AI can work, it is easy to see where this tension arises.

The processing of large quantities of data, sometimes for new purposes, to produce outcomes where it can be unclear why or how that decision was reached, can bring transformative benefits to those adopting and benefiting from AI. However, it does seem at odds with many of the key principles underpinning data protection regulation. It is therefore vital, when advising on AI and its privacy risk profile, to understand how and when personal data is used, and how this use fits with the requirements of the GDPR.

This paper looks at:

- The rise of AI and why it poses particular privacy concerns.
- How AI fits with some of the key principles of the GDPR.
- How the ICO is responding to the new challenges that AI raises.

AI privacy concerns

While not all potential applications for AI use personal data, a significant number do. Personal data can be processed both when training an AI algorithm and when deploying the AI. AI can even determine whether information falls within the definition of personal data, as the ability of AI to recognise patterns in data, or link data sets, can potentially enable data that would not normally be considered personal data to become "identifiable".

The challenge for organisations using AI, and which are within the scope of the GDPR, is that a number of the typical characteristics of AI seem, at least at first glance, to be at odds with the principles of data protection law. In its March 2017 guidance on big data, artificial intelligence, machine learning and data protection (the ICO guidance on big data and AI), the ICO defines big data analytics as the combination of AI, big data and machine learning. It lists three distinctive aspects of big data analytics that can raise data protection implications: the use of algorithms in a new way; the opacity of the processing; and the tendency to collect "all the data", often for new purposes.

New ways of using algorithms

Big data analytics typically does not start with a predefined query to test a particular hypothesis. Instead, it often runs large numbers of algorithms against data to find correlations in a discovery phase. The uncertainty of the outcome of this phase of processing has been called "unpredictability by design".

Opaque processing

Some AI uses deep learning, a type of machine learning that feeds large amounts of data through non-linear neural networks which classify the data based on the outputs from each successive layer. The sheer volume of data involved and complexity of the processing creates a "black-box" effect, meaning that it is difficult to understand the reasons for the decisions made. Possibly the best-known example of this was when AlphaGo, a computer programme developed by Google's DeepMind, won a game of Go against the (human) world champion. In a game renowned for its complexity, AlphaGo's winning move was so unusual, or incomprehensible to humans, that it prompted match commentators to assume that AlphaGo had malfunctioned.

The use of "all the data"

Big data analytics tend to collect and analyse all of the data that are available. This can include new types of data, such as observed, derived and inferred data. Those data are then often used for new purposes. For example, a retailer may use loyalty card data of all purchases made to find correlations rather than merely to invite a sample of shoppers to take a survey.

AI and the GDPR

The GDPR contains seven principles relating to the processing of personal data that must be followed (*Article 5*). These are:

- Lawfulness, fairness and transparency.
- Purpose limitation.
- Data minimisation.
- Accuracy.
- Storage limitation.
- Integrity and confidentiality.
- Accountability.

When looking at these key principles, it is easy to see how the distinctive characteristics of AI, as identified by the ICO, can cause tensions with GDPR compliance in practice.

What is AI?

The concept of artificial intelligence (AI) has existed since the 1950s but rapidly increasing computational power, and reducing costs for processing and storing data, mean that it is now a practical reality.

The European Commission, in its April 2018 Communication on Artificial Intelligence for Europe, said that AI refers to systems that display intelligent behaviour by analysing their environment and taking actions, with some degree of autonomy, to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world; for example, voice assistants, image analysis software, search engines, and speech and face recognition systems. Alternatively, AI can be embedded in hardware devices; for example, advanced robots, autonomous cars, drones or internet of things applications. It is used today in a variety of sectors (see box *"Current examples of AI"*).

AI can be achieved using a number of different technologies, from machine learning to natural language processing. Machine learning is a subset of AI, and is a set of techniques and tools that allow computers to "think" by creating self-learning mathematical algorithms based on accumulated data.

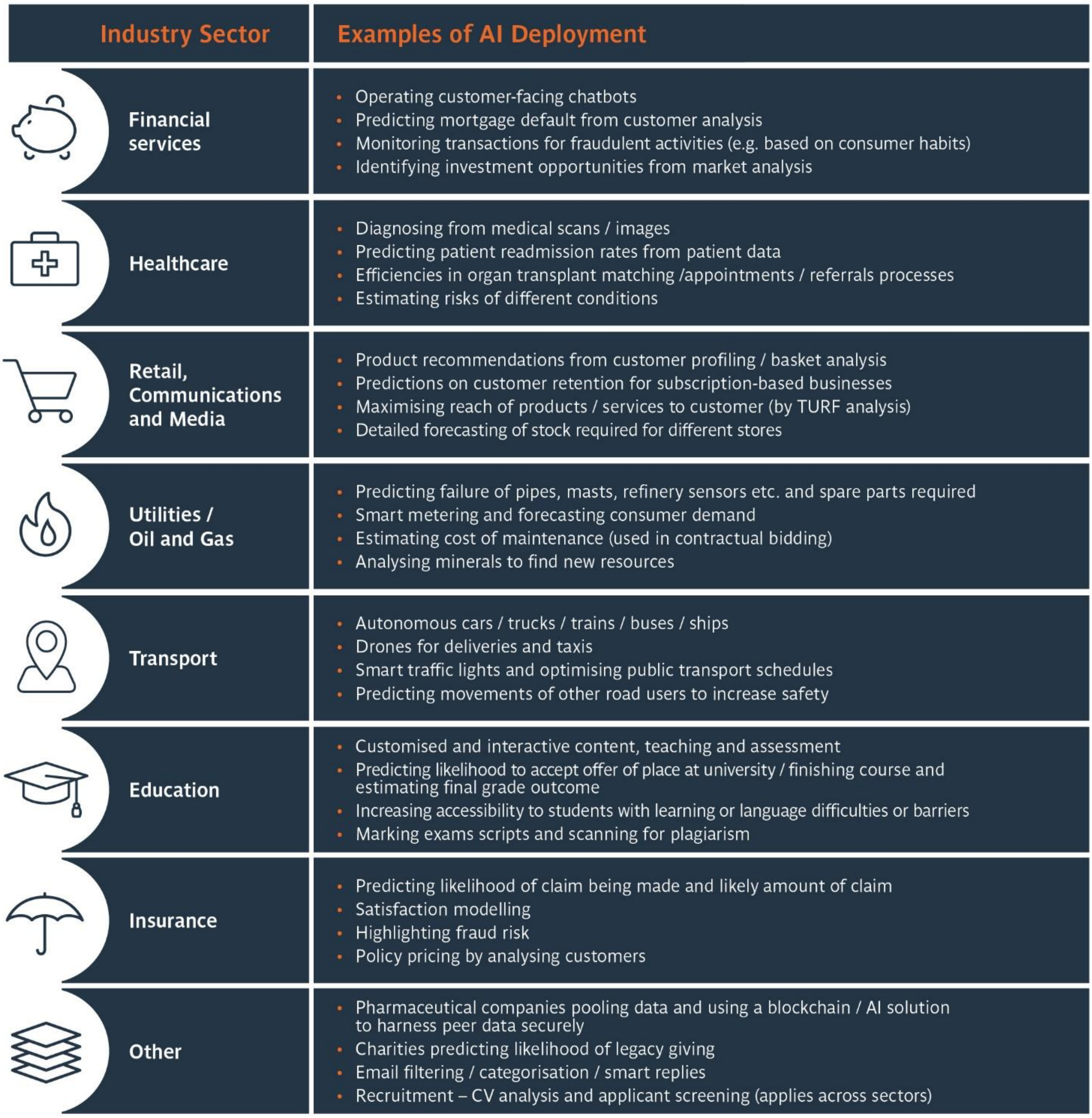
AI is often linked to the use of big data. In its March 2017 guidance on big data, artificial intelligence, machine learning and data protection, the Information Commissioner's Office describes big data as an "asset that is difficult to exploit" and AI as "the key to unlocking" its value. A key component of many AI examples is the identification of patterns in large data sets. The National Cyber Security Centre explains it by saying that "AI does not learn like a human. People can learn a fact by simply being told a few times. An AI has to "see" this fact in the data at a high enough frequency to detect a pattern. This is the reason why you need such high quantities of data to train an AI. It is also why it's difficult to correct a mistake".

The field of AI is generally subdivided into two categories:

- General AI; that is, AI that has such broad applicability that it could successfully perform any tasks or solve any problem requiring human intelligence.
- Narrow AI; that is, algorithms that are designed to solve a particular problem, such as playing a game.

The distinction between the two is a continuous spectrum but, to date, no truly general AI (or, more accurately, artificial general intelligence) has been created. Expert estimates for achieving general AI still differ widely, with estimates ranging from between 2029 and 2200, and some suggesting that it may still not be possible (see theverge.com).

Current examples of AI



Industry Sector	Examples of AI Deployment
Financial services	• Operating customer-facing chatbots • Predicting mortgage default from customer analysis • Monitoring transactions for fraudulent activities (e.g. based on consumer habits) • Identifying investment opportunities from market analysis
Healthcare	• Diagnosing from medical scans / images • Predicting patient readmission rates from patient data • Efficiencies in organ transplant matching /appointments / referrals processes • Estimating risks of different conditions
Retail, Communications and Media	• Product recommendations from customer profiling / basket analysis • Predictions on customer retention for subscription-based businesses • Maximising reach of products / services to customer (by TURF analysis) • Detailed forecasting of stock required for different stores
Utilities / Oil and Gas	• Predicting failure of pipes, masts, refinery sensors etc. and spare parts required • Smart metering and forecasting consumer demand • Estimating cost of maintenance (used in contractual bidding) • Analysing minerals to find new resources
Transport	• Autonomous cars / trucks / trains / buses / ships • Drones for deliveries and taxis • Smart traffic lights and optimising public transport schedules • Predicting movements of other road users to increase safety
Education	• Customised and interactive content, teaching and assessment • Predicting likelihood to accept offer of place at university / finishing course and estimating final grade outcome • Increasing accessibility to students with learning or language difficulties or barriers • Marking exams scripts and scanning for plagiarism
Insurance	• Predicting likelihood of claim being made and likely amount of claim • Satisfaction modelling • Highlighting fraud risk • Policy pricing by analysing customers
Other	• Pharmaceutical companies pooling data and using a blockchain / AI solution to harness peer data securely • Charities predicting likelihood of legacy giving • Email filtering / categorisation / smart replies • Recruitment – CV analysis and applicant screening (applies across sectors)

Lawfulness, fairness and transparency

The GDPR states that the processing of personal data must be lawful, fair and transparent; this includes considering the effects of the processing on the individuals involved with AI. Some data are processed to find general trends and correlations while other processing, such as where profiling is used to determine credit references, can have significant legal effects on individuals and even perpetuate bias or discrimination.

Perhaps unsurprisingly, automated decision making and profiling are specifically regulated under the GDPR (see box "*Profiling and automated decision making*").

Fairness. Fairness is also about what is in the reasonable expectations of the relevant data subjects. This includes considering if, for example:

- The processing is naturally connected to the purpose for which the data were collected, which is linked to the purpose limitation principle (see

“*Purpose limitation, data minimisation and storage limitation*” below).

- Whether data subjects were aware of how their data would be used, which is linked to transparency.

Transparency. The Cambridge Analytica scandal, and the recent backlash against the use of personal data by some social media companies, show the importance of keeping data processing transparent and within expectations. However, AI use can be difficult to explain using the traditional privacy notice model. It is technically complicated and, when the data are collected, it can sometimes be hard to know how that data will be used and for what purpose. There are also concerns in some areas that too much transparency may allow individuals to manipulate a system in areas such as fraud detection, raise security issues by making it easier to infer private information about the individuals used to train the AI model, or create commercial sensitivities such as intellectual property infringement. Just in time notifications may help increase transparency; these are focused privacy notices that appear when a person provides an organisation with particular information and give the person a brief message about how the information will be used. In addition, new methods can be used to help “interpretability” such as “local interpretable model-agnostic explanation”, known as LIME, which explains a specific output rather than the AI model generally (see the ICO’s blog on automated decision making for more information).

Lawfulness. Processing must also be lawful, which includes ensuring that one of the legitimising conditions under Article 6 of the GDPR applies. Two of the most commonly used conditions are consent and legitimate interests.

Obtaining meaningful consent, which must be freely given, specific, informed and unambiguous, can be difficult in an AI context, as can the fact that there must be an opportunity for the data subject to withdraw their consent at any point (Article 7, GDPR). The very nature of machine learning, where the way in which (and the purpose for which) data are processed and analysed can evolve without human intervention, makes obtaining accurate, specific and detailed consent difficult. The ICO guidance on big data and AI suggests that novel and innovative approaches to consent, which go beyond the simple notice and consent model, may be helpful. Examples include graduated consent and just in time notices, as well as more automation both in the collection and withdrawal of consent. However, issues still remain in practice.

Profiling and automated decision making

The General Data Protection Regulation (*2016/679/EU*) (GDPR) regulates:

- Automated individual decision making, which involves making a decision solely by automated means without any human involvement.
- Profiling, which is the automated processing of personal data to evaluate certain things about an individual. Profiling can be part of an automated decision-making process.

Additional rules, set out in Article 22 of the GDPR (Article 22), apply to protect individuals where decisions based solely on automated processing, including profiling, are made which have a legal or similarly significant effect on the individual; for example, the automatic refusal of an online loan application or the selection of preferred job candidates without any human intervention (see *“Accuracy”* in the main text).

This type of decision making can only be carried out where the decision is:

- Necessary to enter into or perform a contract.
- Authorised by law (which must be an EU law or EU member state law that applies to the data controller).
- Based on the individual’s explicit consent (*Article 22(2)*).

In these situations, organisations should give the relevant data subjects information about the processing and introduce simple ways for them to request human intervention or challenge a decision (*Article 22(3)*). This should be planned at the design phase, bearing in mind that the more complex and opaque the AI model, the harder it may be for individuals to understand, and challenge, any decisions made. Organisations must also carry out regular checks to ensure their systems are working as intended, complete a data protection impact assessment and document any requests for human intervention or challenges to decisions made. The Information Commissioner’s Office has produced detailed information relating to automated decision making in both guidance and a blog.

The legitimate interests condition may be more appropriate to use, although the ICO is keen to stress that it is not a soft option. Organisations are responsible for balancing their own interests, or those of a third party, against the interests of the individuals concerned. This may require them to have an ethics review board or a framework of values against which proposed processing can be tested, and to document their “legitimate interest assessment”. The processing must also be necessary, which includes looking at whether there is another way of meeting the legitimate interest.

In addition, the legitimate interests condition has limitations; for example, it cannot be used for special categories of data such as health data, nor in a public sector context, and is more complex to use where children are involved. The breadth, and ever-changing nature, of data processing that is required for an AI solution to function effectively also poses problems from a GDPR transparency perspective, as specific information around the data controller’s legitimate interests must be included and kept up-to-date in privacy notices provided to affected individuals.

Purpose limitation, data minimisation and storage limitation

Under the purpose limitation principle, personal data must be collected for specified and legitimate purposes and not be further processed for incompatible purposes. It must also be adequate, relevant and limited to what is necessary (the data minimisation principle) and, subject to some exemptions such as scientific research, only stored as long as is necessary (the storage limitation principle).

In the ICO guidance on big data and AI, the ICO emphasises that while the purpose limitation principle does not create a barrier for AI models, it does mean that organisations must carry out an assessment of compatibility of processing purposes.

Fairness is a key factor in determining compatibility (see “*Lawfulness, fairness and transparency*” above). The GDPR states that, when assessing compatibility (and having met the requirements for lawfulness), controllers should consider:

- Any link between the original and new processing.
- The context in which the personal data were collected, in particular the reasonable expectations of the data subjects.
- The nature of the personal data.
- The consequences of the intended further processing for the data subjects.
- The existence of appropriate safeguards (recital 50).

Where the new purpose is unexpected and involves making decisions about a person, such as where information placed on a social media platform is used to assess creditworthiness, specific consent from that person will often be required.

AI models can also encourage organisations to collect personal data that are excessive, and to retain those personal data for longer than necessary, in conflict with the data minimisation and storage limitation principles. However, even where this allows organisations to find unexpected correlations in the data they process, the ICO has said that this does not retrospectively justify using the data in the first place (ICO guidance on big data and AI). It can also create tension with other rights, such as the right for individuals to have their data erased or corrected, as it can be difficult in practice to find and erase someone’s data when they are spread across several different systems. Removing data from a model may also affect its results.

While recognising these issues, the ICO still considers that organisations should be able to articulate, at the outset, why they need to collect, process and retain certain datasets and be clear about what they expect to learn from them (*ICO guidance on big data and AI*). The challenge for organisations is therefore to define the purpose of the processing and to ensure that the data are relevant and not excessive.

Accuracy

Accuracy is one of the key principles of data protection and any incorrect or misleading personal data should be corrected or deleted without undue delay. Accuracy issues relating to AI may apply to matters of fact; for example, whether or not a data subject is a parent. However, AI outputs may generate personal data when there is no current matter of fact; for example, an AI system could predict when a person is likely to become a parent. The AI system may therefore be more or less accurate as a matter of statistics. Guidance from the European Data Protection Board suggests that, in these cases, individuals still have the right to challenge the accuracy of the predictions made about them on the basis of the input data or models used, and (under Article 16 of the GDPR) to provide supplementary information (see the ICO blog on accuracy of AI systems for more information).

Accuracy requirements are more stringent for solely automated AI systems which make decisions that have a legal or similar effect on the data subjects (see box “*Profiling and automated decision making*”). Organisations should use appropriate mathematical or statistical procedures for the profiling, and implement technical and organisational measures that are appropriate to ensure, in particular, that inaccuracies in personal data are corrected and the risk of errors is minimised (*recital 71, GDPR*). Data controllers should therefore consider whether it is appropriate to automate any prediction or decision-making process, which includes assessing if acceptable levels of accuracy can be achieved.

It is important, when building and deploying AI systems, to adopt appropriate accuracy measures, recognising that trade-offs may need to be made, for example between accuracy and fairness and accuracy and transparency. It is also important to understand the different consequences of different errors (such as false positives and false negatives) and to recognise that accuracy is not a static measure. AI systems may become more or less accurate over time: this is sometimes called “concept drift”. Potential accuracy risks and trade-offs can be considered as part of a data protection impact assessment (DPIA). Organisations can also take steps to help mitigate those risks, such as adopting common terminology that staff can use to discuss accuracy performance measures, using mathematical techniques to minimise trade-offs and regularly reviewing any trade-offs which are made (see the ICO AI blogs on accuracy of AI systems and trade-offs for more information).

Integrity and confidentiality

Personal data must be processed in a manner that ensures the appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage. Using AI may have implications for an organisation’s security risk profile, which must be assessed and managed. Some may be new risks, such as adversarial attacks on machine learning models, and some may be known risks that are exacerbated by the use of AI, such as the risk of outages.

AI systems are complex, often rely on third party (or sometimes open-source) code or relationships and generally need to be integrated with several new and existing IT components. This can also raise security risks, as can the fact that AI systems sometimes involve sharing and copying large data sets, which increases the risk of a breach. In addition, the complexity of AI systems can make security issues more difficult to identity and manage.

As yet, there is no established market practice around AI security, and the people designing and deploying AI tend to have a wider range of backgrounds than those deploying traditional IT, meaning that security practices, expertise and expectations can vary significantly. In addition, some languages that are commonly used by machine learning developers are not the most secure languages available. For example, in January 2019, a security vulnerability was discovered in a popular library for the Python programming language. However, the ICO suggests (in its AI blog on security risks) that one solution to this may be to develop a model in one language and convert to another language, such as Java, before deployment.

It is therefore vital for organisations to review risk management practices to ensure that appropriate security is in place; for example, organisations should:

- Review their information governance and security policies to check that they are fit for AI.
- Establish clear audit trails and, where possible, separate development environments from the rest of the IT infrastructure, especially when less secure tools and languages are used.
- Follow up-to-date security guidance.

The ICO is planning to update its security guidance to manage the new GDPR requirements and, while this will not be specific to AI, it will cover topics that are relevant to AI such as software supply chain security and the increasing use of open-source software. The ICO has also previously produced guidance on managing the security of internal and external code in relation to online services, which includes external code security measures, and similar measures will apply to AI applications. In addition, it is working on GDPR certifications, which should enable organisations to demonstrate compliance with the GDPR. Once established, certifications and kitemarks could help increase market confidence around AI security.

Accountability

Data controllers are responsible for, and must be able to demonstrate compliance with, the GDPR principles. There are a variety of ways that organisations can demonstrate accountability, some of which are best practice and some are explicitly specified in the GDPR; for example, record keeping, appointing a data protection officer and completing a DPIA. Up-to-date, accurate and rigorous document and assessment processes are therefore key to developing AI solutions that do not contradict the word or spirit of the GDPR.

Algorithmic accountability should also be considered; that is, the ability to check that the algorithms used and developed by machine learning systems are doing what they should be and are not producing discriminatory, erroneous or unjustified results. Bias is a key risk area for AI and, to comply with the accountability principle, the ICO states that detecting discriminatory decisions in hindsight will not be sufficient: discrimination detection must be built into the machine learning systems to prevent these decisions being made (ICO guidance on big data and AI). It has also confirmed, when discussing human bias and discrimination in AI systems, that an organisation's governing body will be accountable for the approach taken to manage discrimination risk (see ICO blog on human bias and discrimination in AI).

Organisations must therefore determine and document their approach to bias and discrimination mitigation at the outset of any AI project so that appropriate safeguards can be put in place, and ensure that board members and other senior staff in oversight functions (including data protection officers) have sufficient understanding of the different approaches that exist. For example, various approaches and mathematical techniques are being developed to understand and manage imbalanced or biased training data. Interestingly, anonymisation and pseudonymisation, which are often seen as effective ways of managing privacy risk when using AI, can make bias more difficult to monitor and address.

Algorithmic accountability is also linked to accuracy, and the quality and reliability of the data. This is relevant to many different types of algorithm: from those relating to profiling decisions, to association algorithms, such as Google's autocomplete functionality, which has faced scrutiny in the German courts. It is therefore important that organisations using AI in the form of machine learning algorithms adopt measures such as:

- Auditing techniques, in order to identify the factors that influence an algorithmic decision.
- Visualisation systems, which help individuals to understand why a recommendation was made.

It is also important that organisations understand the distinction between correlation and causation, and the potential accuracy or inaccuracy of any resulting decisions as correlational associations can often be misinterpreted as causal associations.

Other issues

Focusing on the data protection principles only gives part of the picture; for example, AI also has an impact on the rights of those whose data are being processed, such as their right to have access to their data or to have their data deleted, and various supply chain issues. This is an area that the ICO is likely to look at more closely (see, for example, the ICO AI blog on developing an AI auditing framework and on trade-offs, which discusses outsourcing and third party AI in relation to managing trade-offs when designing AI models). However, it does demonstrate some of the data protection compliance issues raised by the use of AI. The use of compliance tools, such as using a DPIA and addressing the risks identified in the ICO's AI auditing framework, should go some way to ensuring that all relevant issues are addressed (see "*AI auditing framework*" below).

The ICO's response

The ICO guidance on big data and AI was published in 2014 and updated in 2017 to refer to the GDPR. It provides detailed guidance, running to 99 pages, on the various data protection implications of big data, AI and machine learning. It also lists a variety of compliance tools that can be used, from anonymisation and new approaches to privacy notices, to DPIAs, privacy seals, ethical approaches, algorithmic transparency and personal data stores. The latter is an area that the government has been working on with Innovate UK and the Open Data Initiative, as reported in its May 2019 report "AI Sector deal: one year on". The ICO guidance on big data and AI also includes an annex dedicated to helping organisations answer DPIAs in an AI context.

The ICO has listed AI as one of its three strategic priorities, and is currently taking a number of steps to try to help organisations manage AI risk. Three recent examples include:

- The development of an AI auditing framework.
- The ICO's new regulatory sandbox (see "*ICO sandbox*" below).
- Research that the ICO is currently carrying out with the Alan Turing Institute (the Turing), which will inform new AI guidance.

AI auditing framework

The ICO is currently developing a new auditing framework for AI which will have two key components:

- Governance and accountability, which will discuss the measures that an organisation must have in place to be compliant with data protection requirements.
- Eight AI-specific risk areas, which the ICO is examining in detail in a series of AI auditing framework blogs, not all of which have been published at the time of writing (see box "*AI-specific risk areas*").

The ICO launched its blog in March 2019 to provide regular updates on the framework's development and to encourage organisations to engage with the process: it has described it as "an informal approach to consultation". The ICO's plan is to conclude this initial consultation phase in October 2019 and publish a formal consultation paper on the AI auditing framework no later than January 2020.

ICO sandbox

As a part of its Technology Strategy 2018 - 2021, the ICO is developing a regulatory sandbox that will allow different organisations to develop a bespoke plan to receive support from the ICO when tackling complex data issues such as interpreting the GDPR for AI. A trial of the sandbox will run between July 2019 and September 2020. Ten participants will receive a bespoke plan using a variety of mechanisms to aid compliance, including a letter of negative assurance which will provide information about a product or service's compliance with data protection legislation.

Project ExplAIn

In 2018, the government tasked the ICO and the Turing to produce practical guidance to help organisations explain AI decisions to the individuals affected. To do this, they carried out research, using a "citizen's jury" method to find out public perception on the issues and held roundtables with industry stakeholders. Interim findings from the project, known as Project ExplAIn, were published in June 2019. Three key themes were identified:

- The importance of context in explaining AI decisions. Explaining an AI decision will be more important in some areas, such as recruitment and healthcare, than others. People who took part in citizens' juries also preferred to know that a decision was accurate rather than why it was made. They expected AI explanations when they would also expect a human to explain a decision and wanted the explanations to be similar, although there were some discussions at industry roundtables around whether AI should be held to higher standards.

- The need for education and awareness around AI. A broad range of voices need to engage and inform the public in the use, benefits and risks of AI decision making. The interim findings also discuss the need for board-level buy-in on explaining AI decisions.

- The various challenges to providing explanations. While industry felt confident that they could technically explain decisions, other issues were raised including cost, commercial sensitivities (such as intellectual property infringement), the potential for abuse of systems and the lack of a standard approach to establishing internal accountability for explainable AI decision systems.

AI-specific risk areas

The Information Commissioner's Office (ICO) has identified the following eight artificial intelligence (AI) risk areas as part of its AI auditing framework:

- Fairness and transparency in profiling, including issues of bias and discrimination, the interpretation of AI applications and the ability to explain AI decisions to data subjects.

- Accuracy, including both the accuracy of data used in AI applications (input data) and the accuracy of data derived from them (AI outputs). The ICO suggests that any potential accuracy risks can be considered and addressed as part of a data protection impact assessment.

- Fully automated decision-making models, including the classification of AI solutions (that is, fully automated or non-fully automated decision-making models) based on the degree of human intervention, and issues around human review of fully automated decision-making models (see box "*Profiling and automated decision making*").

- Security and cyber risk, including testing and verification challenges, outsourcing risks and re-identification risks.

- Trade-offs, covering the challenges of balancing different constraints when optimising AI models; for example, accuracy versus privacy.

- Data minimisation and purpose limitation.

- The exercise of rights, including individuals' right to be forgotten, data portability and the right to access personal data.

The impact on broader public interests and rights as they pertain to data protection legislation, such as freedom of association and freedom of speech.

These interim findings will feed into guidance that will be published in autumn 2019 following consultation. The ICO has already concluded three possible implications for the development of the guidance: the lack of a one-size-fits-all approach to explanations, including the potential for a list of explanation types to support organisations in making appropriate choices;

the need for board-level buy-in on explaining AI decisions; and the value of a standardised approach to internal accountability to help assign responsibility for explainable AI decision-systems and foster an organisational culture of responsible innovation.

AI regulation is wider than GDPR

While many of the issues relating to AI and the GDPR are clear, the solutions have been less obvious. To date, much of the regulatory response has focused on how to use existing compliance tools or has acknowledged, but not necessarily resolved, conflicts. There has been some forward- looking guidance, for example around the use of ethical approaches, novel consent methods and data stores. However, it is only recently that the ICO has developed new frameworks and models, such as the AI auditing framework, to help organisations manage (and the ICO regulate) the privacy risks associated with AI.

That said, the ICO is not the only source of guidance for organisations. A number of international guidelines and guidance exists. For example, at EU level, the Commission is pushing the AI agenda with a strong focus on ethics and data protection, and on 8 April 2019 its High-Level Expert Group on Artificial Intelligence published ethics guidelines for trustworthy AI, which list seven requirements that AI systems should meet, one of which relates to privacy and data governance.

The government has also set up a number of AI-related bodies as part of its AI strategy. These include the Centre for Data Ethics and Innovation (CDEI), which will take a central role in guiding organisations in their use of AI. It recently published interim reports on the reviews it carried out regarding online targeting and bias in algorithmic decision-making following a call for evidence.

Much of this work on the ethical deployment of AI is closely interlinked with the obligations that organisations face under the GDPR, and the government has said that it expects the CDEI to work closely with other regulators such as the ICO. As the government recently said in its response to the March 2019 House of Lords Select Committee on Communications report on regulating in a digital world, "the increased use of data and AI is giving rise to complex, fast-moving and far-reaching ethical and economic issues that cannot be addressed by data protection laws alone".

This article was written by Rob Sumroy (partner) and Natalie Donovan (professional support lawyer). The authors would also like to thank Charles MacRae for his help. It first appeared in the August 2019 edition of PLC Magazine.

If you have any AI related queries, please contact Rob, Natalie or your usual Slaughter and May contact.

Contacts

Rob Sumroy
t: +44 (0)20 7090 4032
E: rob.sumroy@slaughterandmay.com

Natalie Donovan
t: +44 (0)20 7090 4058
e: natalie.donovan@slaughterandmay.com

This article first appeared in the August 2019 issue of PLC Magazine
(http://uk.practicallaw.com/resources/uk-publications/plc-magazine)

This material is for general information only and is not intended to provide legal advice.
For further information, please speak to your usual Slaughter and May contact.

August 2019