# AI ethics in the post-GDPR world: Part 1

***Sandy Tsakiridi, Senior Legal Counsel at HSBC's Group Data Privacy & Digital Legal team, discusses the current landscape of AI ethics in the EU, as well as the main ethical challenges of AI from a GDPR perspective***

We live in an era where artificial intelligence ('AI') is emerging as one of the most widely used technologies that increasingly permeates different aspects of our daily life. Algorithms are used in a variety of sectors, including healthcare, transport, finance and leisure, to engage in conversations, recommend products, improve collaboration and make decisions in areas, such as lending and recruitment. In light of its impact on society through all these avenues, AI has sparked ample debate about the ethical principles and values that should guide its development and use.

The proliferation of AI-driven applications carry great potential benefits and can improve social and economic welfare. However, in order to harness AI's full potential, and avoid unintended, negative consequences and risks that may arise from the implementation of AI systems, it is necessary to explore the ethical aspects. This will emphasise the benefits of AI technology while safeguarding ethical values defined by fundamental rights and basic constitutional principles.

In the last years, the advancement of AI has prompted discussions at global level on how to design, implement and govern ethical AI. In the EU, such discussions have led public sector organisations, research institutions and the industry to issue principles, guidelines, processes, codes of practice and research papers around ethics in AI. Importantly, as highlighted by the European Data Protection Supervisor, ethics in the EU is not conceived as alternative to compliance with the law, but as the underpinning for genuine compliance in order to avoid box-ticking approaches which undermine trust in digital services.

The General Data Protection Regulation ('GDPR') does not refer to AI specifically, but rather regulates the processing of personal data regardless of the technology used. As pointed out by the European Data Protection Board ('EDPB') in a recent letter to a Member of the European Parliament, "the GDPR is built in a technologically neutral manner in order to be able to face any technological change or revolution". That said, there are GDPR provisions that specifically allude to technologies or methods of processing that incorporate aspects of AI, notably those on automated decision-making.

This article aims to provide an overview of the interplay between data ethics and certain key GDPR principles when implementing AI technology. After looking at the current landscape of AI ethics in the EU, the main ethical challenges of AI from a GDPR perspective in relation to the principles of fairness, transparency, data minimisation and accountability will be analysed, and future developments together with possible solutions will be discussed. This article appears across two editions (Part 2 will be published in *Volume 12, Issue 7*).

## What is AI?

While the dynamic between ethics and AI is vividly discussed, there is still a lack of widespread agreement on the definition of AI. The European Commission ('EC') defines AI in its 2018 Communication on AI as "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 (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)".

AI holds great promise from an economic, social, medical, security, and environmental perspective. Depending on its application, it can benefit people (e.g. by improving healthcare through more precise diagnoses), businesses (e.g. by enabling them to better understand their customers and develop products tailored to their needs), and the public interest (e.g. by contributing to climate change mitigation through improved sustainability of products).

(Continued on page 14)

Sandy Tsakiridi is speaking on 'Data Protection Solutions for Emerging AI Technologies' at the 19th Annual Data Protection Practical Compliance Conference, taking place in London on 8th October 2020. See www.pdpconferences.com for details.

(Continued from page 13)

## What is ethics?

Ethics is generally defined as the moral rules and values that govern human behaviour or the conduct of an activity, as well as principles for evaluating those rules. Depending on the circumstances, such principles can form part of concepts that are fundamental to human nature (e.g. the protection of human life, freedom and human dignity) or be of individual character (e.g. an employer's expectation that employees accept the company code of conduct).

It is worth noting that AI ethics is not a novelty, but rather a form of applied ethics. In the context of AI, ethics is concerned with the important issue of how the design, production and application of AI-driven solutions should be approached in order to minimise the ethical harms that can arise in society. The scope of AI ethics extends from more imminent concerns about data privacy risks, to medium-term issues around the impact of AI on employment, and more long-term considerations about the so-called 'superintelligence', namely the possibility of AI systems exceeding human capabilities.

## Where does Europe currently stand?

During the last few years, AI ethics has shifted from a topic of academic discussion to a point of debate by international organisations, national governments and the industry. This has led to a number of initiatives around ethical principles, standards and strategies for AI, the most notable of which are highlighted below.

***"A number of organisations and professional communities across different industries that rely on AI for all or part of their business have developed AI guidelines and principles with analogous values and principles. This proliferation of soft-law efforts can be interpreted as a governance response…"***

### EU institutions

The EU expressed early on its view that "[e]thical AI is a win-win proposition that can become a competitive advantage for Europe: being a leader of human-centric AI that people can trust". In April 2019, the High-Level Expert Group on AI appointed by the EC presented its Ethics Guidelines for Trustworthy AI ('the Guidelines'). According to the Guidelines, trustworthy AI should be "(i) lawful — respecting all applicable laws and regulations; (ii) ethical — respecting ethical principles and values; and (iii) robust — from a technical perspective".

Furthermore, the Guidelines put forward a set of key requirements that AI systems should meet in order to be deemed trustworthy, namely human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability.

In February 2020, the EC published a White Paper on AI following Ursula von der Leyen's announcement that as President of the EC, she intended to put forward legislation for a coordinated European approach on the human and ethical implications of AI within her first 100 days in office, which commenced on 1st December 2019. The White Paper, which is aligned with the key principles set out in the Guidelines, details policy options and calls for debate on issues pertinent to AI and data. In particular, the EC proposes to opt for a risk-based approach and to make proportional regulatory intervention in order to address mainly 'high-risk' AI applications depending on the sectors where they are deployed (e.g. healthcare) and their intended use.

### EU Member States

In a declaration in 2018, 25 EU Member States expressed their will to ensure 'an adequate legal and ethical framework, building on EU fundamental rights and values' and that "humans remain at the centre of the development, deployment and decision-making of AI". Since then, several EU Member States have appointed committees or relied on their national Supervisory Authorities ('SAs') to produce reports and guidance documents on AI ethics.

In the UK, the Information Commissioner's Office ('ICO') is working closely with The Alan Turing Institute to develop an ethical framework for explaining the processes, services and decisions delivered by AI. In 2017, the ICO published its amended guidance on big data, AI, machine learning and data protection where it stressed the importance of ensuring fair, accurate and non-discriminatory use of personal data, and set out rules to ensure an ethical approach. The French SA (the CNIL) was also amongst the first regulators in the EU to issue a report on the ethical issues around algorithms following a public debate launched in 2017. Furthermore, Germany has instated a Federal Data Ethics Committee (Datenethikkommission) which released its final report on data and algorithmic systems in March 2020. The Spanish SA (Agencia Española de Protección de Datos) published in February 2020 a guide for those looking to make use of AI, including developers, setting out the privacy and quality guarantees that should be applied.

### Industry

In tandem, a number of organisations and professional communities across different industries that rely on AI for all or part of their business have developed AI guidelines and principles with analogous values and principles, intending to ensure the positive as-

pects and diminish any risks involved in AI development. This proliferation of soft-law efforts can be interpreted as a governance response from the private sector which must navigate novel risks and calls for using AI in a responsible manner that goes beyond strict legal compliance.

## Ethical challenges of AI and GDPR principles — fairness

Since entry into force of the GDPR, compliance with privacy rules has been on the forefront for organisations that process personal data. As stated by the EDPB, "[a]ny processing of personal data through an algorithm falls within the scope of the GDPR". Whenever AI systems process personal data, they are a vital component for their full lifecycle and all the standard provisions of the GDPR may apply.

The GDPR sets out that data processing activities must comply with the principles of lawfulness, fairness and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality (security); and accountability. While these principles apply to all processing of personal data, some of them are particularly relevant to the ethical challenges that can be raised by AI systems.

The GDPR provides that personal data should be processed fairly, which implies an analysis of whether the processing is unduly detrimental, discriminatory, unexpected or misleading to the individuals concerned. In the context of AI, this means that the controller should consider the likely impact of its use of AI on individuals and continuously reassess it. This has been reiterated by SAs in a number of EU Member States. For instance, according to the Dutch SA, "[a] controller must actively account for and justify why an algorithm is fair and the use of the chosen algorithm does not lead to inappropriate results".

Defining what is fair, though, is an ongoing challenge, since fairness is a subjective and contextual concept, influenced by several social, cultural and legal factors. At the same time, machine learning systems can entrench, reinforce or amplify existing bias in decision-making systems because they gain their insights from the structures and dynamics of the societies they analyse. Similarly, since the features, metrics, and analytic structures of the models that enable data mining are chosen to a large extent by their designers, these technologies can potentially replicate their preconceptions and biases. Finally, the data samples used to train and test algorithmic systems can often be insufficiently representative of the populations from which they are drawing inferences. As a result, the relevant outcomes may be biased and discriminatory given that the data being fed into the systems is flawed from the start.

As AI-based technology is used to make impactful decisions (e.g. filtering through CVs as part of the recruitment process for a vacancy, carrying out credit ratings for loans, assessing the risk of someone reoffending, etc.), it is very important to ensure that these outcomes are fair and non-discriminatory. However, avoiding bias produced by algorithms in relation to data sets used by AI systems both for training and operation purposes is much more complex in practice than it may appear at first sight.

Some researchers suggest that before devising a fair algorithm, it must be determined what constitutes a fair outcome. However, there are two issues with this approach. First, it assumes that it is always possible to predict what a fair outcome looks like from the outset whereas in practice, this is assessed on a case-by-case basis taking into account the specific circumstances at hand. Secondly, some of the pre-defined requirements for algorithmic fairness may contradict each other in a particular context but work well together in another.

Another practice which can arguably mitigate bias and enhance the fairness of algorithmic decisions is to use combinations of learning types, including unsupervised learning. This is based on the hypothesis that the labels of the data used for supervised learning can often create bias and more generally, to the fact that humans bring their own biases to machine-learning scenarios. To that effect, software solutions launched in the market offer the capability of including validation testing for algorithmic intent with regression testing that involve both real and synthetic data.

In order to tackle the challenge of mitigating potential bias and discrimination by AI solutions, the most effective strategy would be to adopt a holistic approach. This entails that both designers and users would ensure that the AI systems they are developing and deploying respectively are trained and tested on properly representative, relevant and accurate datasets (input fairness); employ model architectures that do not include correlations, interactions, and inferences which are unreasonable, morally objectionable, or unjustifiable (design fairness); do not impact in a discriminatory way the individuals they relate to (outcome fairness); and are deployed by users sufficiently trained to implement them responsibly and without bias (implementation fairness).

Meanwhile, organisations that use AI-driven solutions should consider how to instill mechanisms to collect feedback from their users in order to correct any unintentional bias in design or decision-making, and to periodically monitor training and results in order to quickly respond to any identified issues.

Fairness is closely linked with transparency as in order to conduct an assessment of whether a particular decision made by an AI solution is fair or not, it is necessary to know the reasoning behind it. For instance, an individual who believes that their job application has been unfairly rejected has the right to know the reasons that led the AI system to this decision. Therefore, if an AI system is not sufficiently transparent, it may be impossible for those overseeing its use to identify bias in its reasoning and output.

However, as we will see in Part 2, transparency in AI systems comes with its own challenges.

**Sandy Tsakiridi**
HSBC
sandy.tsakiridi@hsbc.com