The ICO's Consultation on Generative AI: plugging the regulatory gap

Andrew Kimble, Partner, Jenny Gibbs, Associate, Jess Mant, Solicitor, and Katie Simmonds, Managing Associate, highlight key challenges with the use of generative AI, and offer proactive steps organisations can take now to leverage the benefits of the technology whilst maintaining good compliance

In response to requests for further clarity on how data protection law applies to the development, testing and use of generative AI, the Information Commissioner's Office ('ICO') recently launched a consultation ('the Consultation'), which will be in the form of a series, on generative AI and data protection. Its stated aim is identifying and addressing the gaps between existing legislation and the data protection risks presented by generative AI technologies.

In this article, we consider:

- the key challenges highlighted in the Consultation;
- how the ICO hopes to address some of these challenges; and
- what proactive steps organisations can take now to leverage the benefits of these technologies while ensuring good data protection compliance.

Key challenges

1. How to identify an appropriate lawful basis (Article 6 UK GDPR)

The first part of the Consultation centres on whether it is lawful to use personal data that have been scraped from the internet to train generative AI models. More specifically, the first part of the Consultation focusses on the need for organisations to have a valid lawful basis under Article 6 UK GDPR for processing personal data. This is challenging in the context of web-scraping, which the ICO refers to as an 'invisible' processing activity, as individuals are unlikely to be aware of their personal data being processed in this way. This means that 'consent' can be easily discounted as a lawful basis.

The ICO has published a policy guidance note which indicates that 'legitimate interests' may be a valid lawful basis for training generative AI models on web-scraped data. Legitimate interests enable processing to be undertaken where it is necessary for the purposes of business, or other interests. The exception to this is where such interests are overridden by the interests or fundamental rights and freedoms of individuals.

As well as determining the lawful basis for the web-scraped personal data, organisations will need to consider what lawful basis is most appropriate when using generative AI to process individuals' personal data. For example, in a generative AI chatbot situation, will organisations be relying on legitimate interests, or seeking consent? This becomes more complex where sensitive or special category personal data are collected, as organisations need to identify both a lawful basis under Article 6 UK GDPR and a separate condition for the processing under Article 9 UK GDPR.

Where organisations are looking to rely on legitimate interests for generative AI activities, a Data Protection Impact Assessment ('DPIA') should be completed, as invisible processing and AI related processing are both seen to be high-risk processing activities.

2. The expectations in relation to compliance with the accuracy principle (Article 5(1)(d) UK GDPR)

This challenge was specifically identified by the ICO through its engagement with innovators as an area where organisations would welcome further clarity on how data protection law applies to the development and use of generative AI.

A key problem with AI technologies is that they operate in a 'black box', and are continually learning. It is therefore often extremely difficult to explain how the AI works. As AI is becoming increasingly complicated, it also follows that it is harder to identify erroneous outputs or decisions that relate to individuals. This causes difficulties in the context of the accuracy principle in Article 5(1)(d) UK GDPR, which requires organisations to take all reasonable steps to ensure that the personal data they hold is not incorrect or misleading as to any matter of fact.

Erroneous outputs can come in the form of false statements, fabricated references and discriminatory decisions. For example, last year, two

(Continued on page 12)

(Continued from page 11)

lawyers were fined in the US for using fake citations in Court that had been generated by ChatGPT. This emphasises how inaccurate outputs and decisions can often be tricky to spot and can result in organisations relying on false data to make decisions about individuals.

Why does AI produce erroneous output? Considering Chat GPT as an example, it is an excellent web scraper, but what it cannot do is challenge the accuracy of the underlying sources that it 'scrapes' data from. This can lead to misleading outputs. Examples of the potential consequences for organisations and individuals include:

- AI-enabled fraud. At the extreme end of the spectrum, customers could be exposed to significant levels of false and misleading information and AI-enabled fraud;
- discriminatory decisions. A real-life example of discrimination involved an advertising algorithm which showed more technical jobs to men and secretarial jobs to women. This was due to the algorithm having made decisions on what roles to advertise based on historic training data, that predominantly related to men. It is therefore easy to see how the algorithm perpetuated historic biases from the underlying dataset it was trained on;
- erroneous data. This could also result in organisations holding inaccurate data about individuals; and
- commercial issues. In addition to customer harm, it is easy to see how this could cause serious commercial issues for organisations. For example, if an organisation were to rely on inaccurate market analysis or competitor data to inform its buying practices.

> ***"organisations buying AI tools should look to understand what data have gone into the tool to train it, and build an understanding of the possible weaknesses or gaps in the data."***

The two key takeaways for organisations are that firstly, an algorithm is only as good as the data it is trained on and provided with; and secondly, the outputs of an AI tool need to be put into a wider context and understood before any important decisions are made. This means that organisations buying AI tools should look to understand what data have gone into the tool to train it, and build an understanding of the possible weaknesses or gaps in the data. In addition, organisations will need to ensure that they have a process and team in place to understand how a tool is being used and provide oversight to any decisions.

3. How to ensure compliance with the purpose limitation principle (Article 5(1)(b) UK GDPR)

The purpose limitation principle requires organisations to be clear about what its purposes for processing are from the outset. This has been identified by the ICO as another area of focus in the context of generative AI.

Complying with the purpose limitation principle is challenging in a machine learning/neural networks context, as these technologies are continually learning. Further, generative AI tools are flexible and can be adapted to a number of different use cases. This means that an organisation may start using a tool for one purpose and find that several weeks/months later this purpose has evolved significantly.

Examples of ways that we are seeing organisations tackle this challenge include:

- trialing new tools to identify potential weaknesses. We are seeing organisations taking part in different trials of AI systems. This enables organisations to risk assess particular use cases for a tool, identify any weaknesses and put a mitigation plan in action, before adopting a tool on a long-term basis;
- using short term versus long term contracts. We are also seeing organisations enter into initial/short term contracts, which can then be adapted in future, once it is clearer how the AI tool will be used. The exercise of entering into short-term contracts (that likely need to be reviewed and renegotiated) might feel expensive from an internal time and legal cost perspective, but it is likely to be needed until the particular use cases for a tool have been determined and the legal landscape has settled — more on this second point to follow below; and
- using internal environments. We are seeing a lot of preferences towards using tools that allow personal data to stay within the user's own environment. This allows them to leverage the benefits of advance generative AI technologies, while ensuring that their personal data remains secure.

In addition to the above, completing a DPIA and ensuring it is kept up to date as the use of an AI tool progresses will be key to ensuring compliance with the purpose limitation principle.

Proactive steps

Steps that organisations can take now to leverage the benefits of these technologies while ensuring good data protection compliance are:

Conduct a DPIA for all generative AI processing activities: Crucially, this DPIA should be kept up to date as the particular use case for a tool develops and changes. The DPIA will help organisations to determine key aspects of compliance, including what lawful basis to rely on and how to ensure sufficient transparency is given to individuals. This will be particularly useful in the context of ensuring that the organisation is prepared to respond to requests from individuals about how it is making automated decisions about data subjects under Articles 13(2)(f) and 14(2)(g) UK GDPR.

Determine an internal AI strategy and the individual or team responsible for dealing with AI within the organisation: The UK government's AI strategy sets out five core principles aimed at encouraging innovation and utilising AI, while ensuring that public and fundamental rights are protected. One of these is 'accountability and governance', which refers to the expectation that organisations and individuals will adopt appropriate measures to ensure the proper functioning of AI systems throughout the entire project lifecycle. We are therefore seeing an increasing number of new AI-specific teams and roles being created, such as Chief AI Officer, to lead on AI compliance and strategies.

Expressly ask suppliers how they mitigate against the risks posed by their products: We recommend that customers of AI tools get curious about the solution being purchased and ask the supplier key questions about how the tool works. These could include asking whether the supplier has been able to identify any inherent biases in its tool and, if so, how the supplier plans to address any issues. Another question could be to focus on how you as an organisation will understand and interpret the outputs/decisions made by a tool and what support, if any, the supplier can provide here.

Andrew Kimble, Jenny Gibbs,
Jess Mant and Katie Simmonds
Womble Bond Dickinson (UK) LLP
andrew.kimble@wbd-uk.com
Jenny.gibbs@wbd-uk.com
Jess.mant@wbd-uk.com
Katie.simmonds@wbd-uk.com