What risks need to be considered by a business using artificial intelligence?

Catherine Hammon, Digital Transformation Knowledge Lawyer, and Georgina Graham, Tamara Quinn, Arty Rajendra, and Olivia Sinfield, Partners, at Osborne Clarke LLP explore the risks associated with using artificial intelligence applications

The rise of artificial intelligence (AI) applications has been a “gradually, then suddenly” scenario. Although we all use AI every day in our devices and systems, there has been an explosion of activity and interest in AI since the release of ChatGPT in November 2022. Although AI is not (yet) specifically regulated, there is a wide range of current legal and compliance risks that need to be taken into consideration, in addition to monitoring emerging AI regulation.

We will consider the following in turn:

- how AI works – a measure of "AI literacy" helps to understand the impact of this technology;
- who controls the AI tool in question and the resulting scope for contractual protection;
- emerging AI-specific regulation;
- risks flowing from the inputs into an AI system, including both the training data and user inputs;
- risks relating to the outputs of the system; and
- overarching considerations (including environmental, social and governance risks (ESG)).

What is AI?

The current focus of interest and development concerns machine learning, a type of AI that partly writes and adjusts itself. This is achieved through an iterative “training” process, often passing huge quantities of data through a structure known as a neural network. Each new piece of data passed through the system, causes the individual settings within the network to self-adjust to make the model progressively more accurate. Modern neural networks can be huge, with millions, billions or even trillions of individual settings that are calibrated and recalibrated by reference to each piece of training data.

A machine learning system has no wider “knowledge” or frame of reference beyond the model created from the training data. The quality of the data therefore drives the quality of its outputs. It does not retain its training data, but generates the right answer based on its complex model of the training data – the answer is a prediction of the right response.

“Foundation models” are machine learning systems that perform a specific task (such as image recognition, translation, text generation etc) that can be used for many different applications. Within that category, generative AI (ChatGPT, Bard, LLaMA, DALL-E, or Midjourney, etc) has become hugely effective at creating content which is often difficult to distinguish from that created by humans. These large and powerful systems are trained on huge datasets, often using data scraped from the internet.

Who controls the AI tool and scope for contractual protection

As with all software, the degree of control that a business using AI will have over how it functions ranges from complete control to almost none, depending on how it is built, trained and accessed.

Some systems are bespoke and built from scratch. Some are available as a pre-trained cloud-based service, ready to use “out of the box”. Many systems are a composite of elements from different sources. When negotiating terms for the development or use of an AI system, it is worth understanding how it is structured, where in the supply chain key decisions are taken, and where control is exercised. Such insight can inform contractual negotiations around warranties and indemnities and in relation to structuring liability more generally. Our experience of contractual negotiations around AI is that this field is too new for standard market practice to have emerged.

Many of the new, free, generative AI tools are also available on the basis of an enterprise licence. This may create additional protections for the business user and will often be a sensible investment. There will generally still be only limited scope for negotiating around standard terms and conditions.

For all AI systems that involve digital connections with a third party, cybersecurity requirements should be considered as part of the contractual relationship.

Emerging AI-specific regulation

As noted, AI-specific regulation is not yet in place but is a priority for many policymakers around the world, as can be seen from the following examples of emerging global legislation.

The EU's Artificial Intelligence Act

The EU's draft AI Act is currently being negotiated between the EU institutions. The legislation takes a tiered approach, focused on risks to human health and safety and to fundamental human rights. It seeks to foster trustworthy AI and draws on the EU product safety regulatory regime (although it has much wider application).

> ***"Some uses of AI are expected to be banned outright in the EU, including live remote biometric identification (such as face recognition) in public spaces and cognitive behavioural manipulation. Applications considered to be high risk will be heavily regulated… High risk systems will need to be assessed, certified, registered and will be subject to a formal enforcement regime at national level, including powers to impose significant GDPR-style fines"***

Some uses of AI are expected to be banned outright in the EU, including live remote biometric identification (such as face recognition) in public spaces and cognitive behavioural manipulation. Applications considered to be high risk will be heavily regulated, with requirements including data governance, extensive technical documentation and record-keeping, transparency for users, human oversight, accuracy and security. High risk systems will need to be assessed, certified, registered and will be subject to a formal enforcement regime at national level, including powers to impose significant GDPR-style fines of up to 6 per cent of worldwide group turnover.

Limited-risk AI will mainly be subject to transparency requirements to ensure that people know that they are interacting with an AI (such as a chatbot), with AI-generated content (such as deep-fakes), or that an AI system is monitoring their reactions etc. Other AI applications will be unregulated.

Additional provisions are expected to regulate foundation models and generative AI tools. The flexibility of application of these systems makes it difficult to fit them into the AI Act's risk-based tiered structure.

The text of the AI Act is expected to be largely settled in the autumn of 2023. Since compliance could be complex and require technical changes to existing AI tools, businesses should plan well ahead to meet the likely deadline in 2026.

To reinforce the regulatory regime under the AI Act, the EU is proposing to facilitate private actions to secure redress through the courts for harm caused by AI. It plans to increase the availability of information to the claimant, and create a rebuttable presumption of liability where certain requirements are met. These changes will need to be implemented at national level across the EU Member States and are likely to increase the risk of litigation.

The UK's AI White Paper

The UK is taking a markedly different approach to AI regulation. The white paper of March 2023 proposed five high level principles to guide the application of existing regulation by existing regulators:

- safety, security and robustness;
- appropriate transparency and explainability;
- fairness;
- accountability and governance; and
- contestability and redress.

For the time being no new legislation or powers are planned, though the government has indicated that this may change if regulatory gaps are identified.

A number of UK sectoral regulators, such as Ofcom, the Financial Conduct Authority and the Medicines and Healthcare products Regulatory Agency, are already actively engaged in understanding how AI fits within their area of expertise. Economic regulators such as the Competition and Markets Authority and the Information Commissioner's Office (ICO) are similarly exploring the interface of their areas of jurisdiction with AI.

The UK approach is expected to be less onerous than the EU's AI Act regime. On the other hand, UK businesses wishing to sell their AI offerings into EU markets will need to comply with the AI Act in any case.

International AI policy

The recent surge in interest in AI has created a sense of urgency amongst governments and policymakers in many countries. Discussions are underway between the EU and US, the US and UK, between the G7

(Continued on page 4)

(Continued from page 3)

countries, and through international organisations such as the OECD. It is not yet clear how, or whether, these multinational initiatives will translate into compliance requirements for businesses.

Input risks

As regards risks flowing from inputs into the AI system, there are two broad areas to consider: risks associated with training data, and those associated with user inputs.

Training data – bias and discrimination risk

As discussed above, AI systems are only as good as their training data. Understanding the profile of its training data is therefore a key aspect of due diligence on the suitability of an AI tool. The EU's draft AI Act is expected to include an obligation that training data should be relevant and representative, taking into account the "specific geographical, behavioural or functional setting" within which the AI tool will be used.

Specific legal risks that can flow from training data include bias and discrimination. If the training data is skewed towards (or against) a particular social, racial or cultural profile, for example, the outputs that it generates may be similarly skewed. Some forms of bias can amount to illegal discrimination. Even where the bias is not illegal, it can generate material reputational risk.

Training data – data protection risk

If training data includes information about identifiable individuals, it is likely to fall within the scope of the EU and UK General Data Protection Regulations. This can be a particular risk for web-scraped training data. While information about, or images of, real people might be lawfully used to train AI in some jurisdictions, the practice faces challenges in the EU or UK, where data protection rules tend to be stronger than elsewhere.

At the most basic level, personal data cannot be processed without an appropriate lawful basis. In addition, the EU and UK GDPR include requirements where automated decision-making has a legal or similarly significant effect on individuals.

These provisions may apply where AI is taking decisions with material ramifications such as recruitment, loan applications etc. In such cases, in addition to a lawful basis for processing, there must be transparency about use of the tool; simple mechanisms for the individual to request human involvement or to challenge the decision; and regular ongoing checks that the system is working as intended.

> ***"While information about, or images of, real people might be lawfully used to train AI in some jurisdictions, the practice faces challenges in the EU or UK, where data protection rules tend to be stronger than elsewhere"***

Data protection compliance risk also includes requirements to undertake a data processing impact assessment (DPIA); that it must be possible to withdraw consent, have errors corrected, or have data deleted; and to comply with overarching principles including transparency, accuracy and fairness for those whose data is processed. Requirements around the international transfer of data must also be complied with.

The ICO has issued extensive guidance around the use of personal data in AI tools at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence.

Training data – IP risk

A second area of legal risk around training data is the possibility that it is subject to third party intellectual property rights. Web-scraped training data again poses potential problems. This is an area where private litigation is on the rise, with both individuals and businesses taking action to protect their intellectual property from unauthorised use.

In the UK, legal exceptions to the copyright rules for text and data mining do not extend to copying data for commercial purposes so are unlikely to apply to web-scraped training data used in commercial AI. In the EU, there is an exception for text and data mining conducted for any purpose, unless the rightsholder has expressly opted out of this exception. Many websites' terms of use will include such a reservation. Where possible, confirmation should be sought that all necessary licences to use the training data have been obtained.

User inputs – confidentiality and wider use

Where an AI system operates on the basis of a user input, it is important to understand where that information goes and how it is used.

Particular care is needed around confidential information where the AI system is a public one, accessed from the cloud. It is known, for example, that questions and data put into the public version of ChatGPT by users are monitored by OpenAI. Confidential information, business secrets, trade secrets etc may need to be withheld from these systems, or redacted, in case confidentiality is compromised.

It is also important to understand whether the system provider will use input data for any wider uses. For example, will it be added to training data for the same system or for other uses?

Output risks

Output risks around AI systems may be based on black letter law, or may be looser risks, flowing from ethical considerations.

IP risks

Where an AI system is generating creative content or innovative proposals around technical products or processes, it is important to understand whether the value in those outputs can be protected by intellectual property rights.

In the UK and EU, where content is created by a human with assistance from an AI tool, the human (or their employer) will own any copyright, as long as the work expresses original human creativity. However, the position may differ if an AI tool has been set up to churn out content.

In the UK, there may still be copyright protection available for automated content provided that there is originality in the output. Copyright would belong to the person who undertook the arrangements necessary for the work to be created. How these provisions apply to generative AI systems has not yet been tested in the courts. In the EU, it is generally considered that there is no copyright for AI-generated creative works. The best way to ensure copyright protection for AI output is therefore to ensure that the system is used as a tool, not for an automated flow of content.

As regards the patentability of AI-generated innovations, both the EU and UK courts have ruled that an AI system cannot be an inventor for the purposes of filing a patent. By contrast, inventions by humans that use AI as a tool will be patentable. The detail around this distinction is currently an area of legal uncertainty, being tested in the courts.

Finally, it should be noted that the risk that outputs could infringe the intellectual property rights of a third party should be mitigated by proper curation of the training dataset. Where the provider of the tool, or of the training dataset, has given warranties that all appropriate licences have been obtained in relation to the training data, it is sometimes possible to secure an indemnity against liability arising from outputs that infringe third party rights.

Accuracy, hallucinations and bias

As explained above, the quality of an AI system's outputs is driven by the quality of the training data that the model has been built on. As well as checking the quality of curation and choice of the training data, it is important to check the outputs from the AI.

Unexpected answers can crop up. This may be because the patterning in the model has spotted something that humans hadn't previously considered. Or it may generate an “edge case” – a reasonable answer but at the margins of the possible outputs. Monitoring outputs on an ongoing basis will be important to ensure that the risk around accuracy is understood and mitigated. It is also important to establish whether the AI system continues to learn and recalibrate once in use. Active ongoing monitoring will be particularly important for such systems.

> *“Monitoring outputs on an ongoing basis will be important to ensure that the risk around accuracy is understood and mitigated. It is also important to establish whether the AI system continues to learn and recalibrate once in use. Active ongoing monitoring will be particularly important for such systems”*

“Hallucinations” from generative AI tools are a known risk, resulting from the fact that machine learning systems predict answers, rather than researching them. The simplest mitigation of this risk is for someone sufficiently knowledgeable on the topic to check output that needs to be accurate or that will be relied on.

The specific legal risk around accuracy may flow from various sources. Where the output from the AI tool impacts on the business' customers, there may be contractual provisions that create (or exclude) obligations around accuracy or quality of products or services provided. Where accuracy of outputs could impact on product safety or quality, product regulation may be in play. Consumer protection law could also be engaged: for example, if errors cause unfairness for consumers.

Where the AI has been trained on personal data, obligations to remove or correct inaccurate outputs about individuals may apply. Inaccurate outputs about identifiable individuals could create defamation risk. Under the AI Act, high risk AI will be required to operate to an appropriate level of accuracy.

Transparency, explainability and the “black box”

Explaining the outputs of AI systems in terms that correspond to how a human would think about the same problem can be very complex. AI models are based on maths not human -style reasoning and neural networks are potentially vast. The difficulty in understanding and explaining these systems is known as the “black box” problem.

The UK white paper’s approach to transparency is that regulators should be able to obtain sufficient information about an AI system to perform their functions. It notes that transparency and explainability are not absolute requirements but should be applied proportionately to the risks in play.

The ICO has issued guidance developed with the Alan Turing Institute about explaining AI that processes

(Continued from page 5)

personal data (see https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/).

As with other output risks, the first consideration around explainability is whether the AI system impacts on third parties. Again, there may be contractual provisions creating (or excluding) obligations dealing with the transparency and explainability of outputs. The AI Act will create overarching transparency requirements for high-risk AI, although again the obligation will be tempered to a type and degree of transparency that is appropriate for the system in question. Transparency is also required under the EU and UK GDPR, including in relation to automated decision-making.

Overarching considerations (including ESG)

There are a number of wider areas of risk exposure that organisations should consider as they explore the utility of AI within their businesses. These are detailed below.

Environmental risk

The emergence of powerful AI systems is the result, in part, of the increasing scale and reducing cost of computer processing capacity. Machine learning systems can be huge, needing significant processing power with corresponding energy consumption.

Businesses using AI should investigate the carbon footprint of AI systems as part of their procurement due diligence. Significant investments by the major cloud providers in renewable energy means that AI systems do not necessarily have poor environmental performance, but this risk should be checked.

Social risk

A key aspect of the social risk of AI is its impact on a business' workforce. Where AI replaces human labour altogether, reskilling may be needed to redeploy staff into new functions within the business, or a redundancy programme. Workforce restructuring is, of course, subject to legal requirements around process and consultation, with corresponding risk of employee disputes if those obligations are not met.

Where AI is used as a productivity tool, reskilling may be needed in how to use the tool. It is important to consider all parts of the workforce when designing reskilling and retraining programmes. In particular, discrimination risk could arise if such programmes are not made available and accessible to all ages of worker.

In addition, businesses will need to expand policies (or introduce new policies) around acceptable use of technology and internet resources to include new AI tools. Staff will need clear guidance, including, for example, ensuring that confidential and/or client information is not inputted into public AI tools and that any outputs are checked for accuracy before being used. Where policies are not followed, risks of disciplinary proceedings will follow.

Governance risk and AI audits

As the use of AI across all sectors expands, businesses need to consider whether they need an overarching ethical policy for how they use AI. This will be shaped by the nature of the business concerned, its customer base, whether the AI is being used internally or in a manner that will impact on products, services or platforms provided to customers or third parties, and the business's overarching approach to customer trust, its reputation, and ethical considerations more generally.

The UK AI white paper's proposal of high-level principles to guide regulators (outlined above) could be used as a steer for corporate AI governance. Many organisations have issued frameworks for ethical AI: one of the better known examples of which is the "Responsible AI: Global Policy Framework" developed by the International Technology Law Association (see https://www.itechlaw.org/ResponsibleAI2021).

In light of the significant matrix of legal and regulatory risk flowing from the adoption of AI, many businesses are implementing a policy of conducting an audit of their use of AI, in order to understand the specific risks for an AI tool, and to identify risk mitigation options. Ensuring that the system is aligned to the business' overarching ethical approach can be worked into such a review.

This article is not intended and should not be used as a substitute for taking legal advice. Specific legal advice should be taken before acting on any of the topics covered. Further articles on AI from Osborne Clarke can be found at https://www.osborneclarke.com/insights/topics/artificial-intelligence/location/uk

Georgina Graham
Catherine Hammon
Tamara Quinn
Arty Rajendra
Olivia Sinfield
Osborne Clarke LLP
georgina.graham@osborneclarke.com
catherine.hammon@osborneclarke.com
tamara.quinn@osborneclarke.com
arty.rajendra@osborneclarke.com
olivia.sinfield@osborneclarke.com