1. Do you agree with the above description of how AI may use copyright works and databases, when infringement takes place and which exceptions apply? Are there other technical and legal aspects that need to be considered?

Answer: Yes, this is an accurate and comprehensive summary of many of the issues faced. We would add that there are also issues around the incentives faced by companies using AI, because of copyright, particularly with regards to transparency and bias.

Firstly, with regards to bias, companies that follow copyright law to the letter by trying to train AI on databases that are either open or easily licensed may face problems of bias according to the kinds of data that are easily available to them. Certainly when compared to if they had trained the AI on the public web. This issue does not seem to have been explored at all.

Secondly, with regards to transparency, the law as it currently stands incentivises many companies to be reticent about sharing any details of how they train their algorithms. They rightly fear that revealing any details of how they have trained the AI may open them to copyright infringement suits — whether justified, accidental, or spurious. This is the case in the UK, certainly, where we can probably assume that much infringement is already taking place behind closed doors, and in the US also, where the law is still unclear. Although AI companies in the US have claimed fair use for data mining, and there are some indications that the law will uphold this, there is still a great amount of uncertainty. The overall situation is thus that many companies keep quiet about their methods and sources. Yet this also provides the UK with an opportunity to distinguish its copyright law with regards to AI, providing greater certainty than in the US.

2. Is there a need for greater clarity about who is liable when an AI infringes copyright?

Answer: Absolutely, yes. The greatest problem with copyright is that it is routinely infringed, often by people who are unaware of their infringement. So if AI can infringe on someone's behalf, people should be aware of the risks involved - especially small business owners and entrepreneurs, who are still relatively inexperienced, or who lack the resources to consult expert legal advice. We should not have a situation where people are later punished for their successes in implementing AI, due to ignorance of a poorly understood law.

3. Is there a need to clarify existing exceptions, to create new ones, or to promote licensing, in order to support the use of copyright works by AI systems? Please provide any evidence to justify this.

Answer: There are strong reasons to create new exceptions, in particular by extending the fair dealing exception for text-and-data mining to commercial, as well as non-commercial uses.

From the point of view of national economic competitiveness, this would be a clear win for the UK, in that it would provide clarity of a sort that is not yet found in the USA, where such issues

are still being litigated, and that it would provide a clear edge compared to the EU, where such an exception is not yet permitted.

With regards to licensing arrangements, licensed datasets that AI can be trained on are already protected by database copyright, but the same cannot be said of the open web. The risk from only using licensed data is that this constrains the kinds of data that can be used, introducing bias. For a large company, this may not be a major issue, as they can potentially afford to buy licences. But start-ups may not have this luxury, instead being limited to cheap, low-quality, or free data.

The alternative of course is that companies attempt to clear the individual rights to every single thing that is copied, but the transactions costs involved in this are so astronomically high that even the wealthiest internet giants cannot afford to do this (take the case of Google Books, which has been forced to restrict access to older books with uncertain copyright ownership, because the costs of identifying the owners of even books with a named authors and publishers are too great). Any amount of promoting licensing will still come up against the fundamental constraint that the transaction costs are too high, and that it benefits larger and more established AI-using companies over smaller and entrepreneurial ones.

4. Is there a need to provide additional protection for copyright or database owners whose works are used by AI systems? Please provide any evidence to justify this.

Answer: No. Database owners seem to be well aware of their rights, and can restrict access to their databases anyway, for example through paywalling. The law is currently in their favour, and would remain so even if an exception was introduced for text-and-data mining for all purposes. As for the public at large, there seems little reason that they should be remunerated just because some software looked at their copyrighted material. There is no loss of a market at all. Indeed, the majority of people upload photographs, text, and music freely for the very reason that they want it to be experienced, shared, and provoke responses. There seems no reason that we should restrict artificial minds from doing what human brains do when experiencing others' work.