Copyright – Option 2

Study Question – 2025 – Explanatory Note

AI & copyright

Artificial Intelligence (AI) has generated much interest recently, particularly in relation to the patentability of inventions generated by AI. AIPPI has studied the topic in Resolution Q276 (Inventiveness and sufficiency of disclosure in AI inventions, Online, 2021), Resolution Q269 (Copyright in artificially generated works, London, 2019) and Resolution Q272 (Inventorship of inventions made using Artificial Intelligence, Online, 2020). In particular Resolution Q269 examined whether there would be copyright in works generated using AI, and resolved that AI generated works should only be eligible for protection by copyright if there is human intervention in the creation of the work and provided that the other conditions for protection are met.

AI systems which are modelled on a neural network, similar to the brain of a human, learn when they are trained. However, because the training data sets need to be very large, unforeseen interactions between individual training sets will result in a certain amount of randomness in what the AI learns. Similarly, some AI systems deliberately 'add randomness' at the training stage so as to simulate creativity. Further, AI systems can also be trained on their own output data sets, in which the output data set becomes an input training data set. Thus the output of an AI can be dissimilar to what was used as the input to the AI, but nonetheless the output is indirectly a function of the totality of the data/images used for the input.

The purpose of this Study Question is to examine the enforcement of copyright in connection with AI systems trained using a very large number of copyright works (each a 'training data set'), such that the training of AI involves an infringement of copyright in each of the training sets but - once trained - the AI system can then 'independently' produce images or other data that does not contain a sufficient amount of material from any single given training set. The net result is that the output of an AI might not infringe copyright in the copyright works that were used to train the AI.

AI systems are portable. The training of an AI can take place in jurisdiction A (where there is infringement of copyright in the training data sets), but the AI can be moved to and used in jurisdiction B where the copyright infringement analysis could be different.

This Study Question will focus on whether the later use of a trained AI system should result in some form of liability, if it is assumed that the earlier training of AI system involved infringing acts in relation to copyright works. For the use of the trained AI system to lead to liability, does the training data set need to involve just one copyright work of the rightholder, or be composed entirely of copyright works of the same rightholder, or include some proportion of copyright works of the same rightholder?