

Annex - Response form

After you have read the consultation document, please consider the questions below. There is no expectation or requirement that all questions are completed. You are welcome to only answer the questions that are relevant to you, your business or organisation.

A copy of this response form is available to download from GOV.uk.

There are two sections on this form:

A. Questions arising from this consultation

B. Information about you, your business or organisation

When you are ready to submit your response, please email this form and any other supporting documentation to Alcallforviews@ipo.gov.uk.

The closing date for responses is at 23:45 on 7 January 2022.

The options for computer generated works, text and data mining and patent inventorship are summarised in the following tables.

Computer generated works	
Option 0	Make no legal change
Option 1	Remove protection for computer-generated works
Option 2	Replace the current protection with a new right of reduced scope/duration

Text and Data Mining (TDM)	
Option 0	Make no legal change
Option 1	Improve licensing environment for the purposes of TDM
Option 2	Extend the existing TDM exception to cover commercial research and databases
Option 3	Adopt a TDM exception for any use, with a rights holder opt-out
Option 4	Adopt a TDM exception for any use, which does not allow rights holders to opt out

Patent Inventorship	
Option 0	Make no legal change
Option 1	"Inventor" expanded to include humans responsible for an AI system which devises inventions
Option 2	Allow patent applications to identify AI as inventor
Option 3	Protect AI-devised inventions through a new type of protection

Section A

Copyright – computer generated works (CGW)

In general, option 2: replacing the current protection with a new right of reduced scope/duration is relatively preferable whereas option 0 and option 1 are the least preferred choices. We will use the music industry as an example.

Incentives

We can see that artificial intelligence (AI) technology is reshaping many creative industries including the music industry. It can be seen that there are at least three possible market segments in the future music industry with AI technology involvement: self-entertainment, environment simulating and music creation.¹

First of all, it is no doubt that AI applications are positively engaged in professional music creation. In the current stage, many AI applications are used for music inspiration purpose and they can help both amateur and professional musicians create music. An Australia-based start-up Popgun (rebranded as Splash) announced their AI product Splash Pro to ‘explore various musical styles and controls, experiment with a number of AI-powered instruments and get inspired by a range of music genres.’² Musicians mention they use AI to write lyrics and melodies to the actual music and use them as a source of inspiration.³ We can even see many AI-generated music are released by popular artists. The use of AI applications such as PhonicMind, Magenta and LANDR achieve that AI has been substantially engaged in the whole music creation process from generating music for inspiration and generating new models, to mastering music tracks and achieving the best separation quality. While AI-produced music becomes a basis for musicians to write lyrics and melodies, it perhaps results in two situations: one is musicians only treat AI-produced music as a source for their inspiration of creating their own music; another is, to some extent, some AI-produced music would be a half-completed music work. Musicians then modify AI-produced music with their own interests and complete this work.

However, no matter which situation is, it can be seen that AI-produced music plays an important role in the pre-human-creation process and it maximized stimulates human creators’ inspiration and imagination in the following creation activity and then produce more music production. AI-produced music is more likely pre-trail or trail versions for the human musicians to develop their final work. From this perspective, they embrace an obvious footprint of musicians’ personal needs (such as required mood, style and rhythm etc.) and directly link and contribute with human musicians’ final work, and make musicians successfully create their final work. They are different from existing prior art (such as existing music works, melodies and music elements that are already been known or created). Therefore, providing relevant protection to the AI-produced work is necessary for stimulating human creation and production.

¹ Luo Li, ‘Artificial Intelligence: An Earthquake in Copyright Protection of the Digital Music’ in In Damian Bielicki (ed) *Regulating Artificial Intelligence in Industry* (Routledge 2021).

² Splash <<https://www.splashhq.com/tools>> accessed 6 January 2022.

³ Glitch Digital, ‘How Artificial Intelligence is Changing the Music Industry’ (*Influencive*, 5 August 2021) <<https://www.influencive.com/how-artificial-intelligence-is-changing-the-music-industry/>> accessed 12 December 2021.

Secondly, AI-simulated music with little human intervention would be highly likely to be used in the background music market. AI-generated music has the advantage of low cost and less procedure (compared with human-created works' license fee and licensing process). This would be competitive for not only traditional businesses like restaurants and department stores as well as other public places, but also would be very attractive to those small enterprises with a limited budget such as some start-up game development companies or longer-tail projects with a tight budget. For example, Melodrive is able to 'compose an infinite stream of original, emotionally variable music in real-time – the idea being that it adapts to what's happening within the game at a particular point in time'.⁴ This application could help many small companies reduce massive costs which would be extremely important for them. The potential advantages of AI-simulated music and its potential market would be the most initiatives for the companies to invest AI technology. In fact, providing protection of AI-simulated music is protecting and stimulating the investment.

Finally, many AI applications such as Jukedek and Amper Score could help users easily generate a piece of music in a minute with the user pre-choice of mood, style, patterns etc. From this perspective, users' intervention is minimized as the produced music highly relies on AI self-operation and generating function. Most users are not professional musicians and the purpose for them to use AI applications to generate music is just for fun (eg. generate music to be used as background music of their short video uploaded in a social media platform). The application of AI technologies massively encourages public users to participate in music entertainment and even towards music creation. Especially, the convenience of AI applications and easy operation would encourage more talents and amateurs to produce more quality music works.

Protection Scope and Term

Obviously, giving legal protection for AI-produced works has quite a positive role in both incentivising more productions and investment in AI technology. Nevertheless, it needs to be noted that AI-produced works should be given limited protection to distinguish them from human-created works in a copyright context. Before we consider the protection scope and protection term, we need to admit: on the one hand, machines do not need economic incentives to produce works but the companies investing AIs to produce works need legal recognition to guarantee their investment; on the other hand, the frequency of AI production is much higher than that of human creators. A long-term protection of AI-produced works is not necessary considering AI could produce massive works everyday and the fashion/popularity trends would be changed more frequently as well following the frequent production. Providing an appropriate protection term could help companies and individuals who invest and use AI to produce works get appropriate investment return (no matter money investment or

⁴ Stuart Dredge, 'Melodrive Debuts AI-music Generation System for Games' (*MusicAlly*, 23 November 2018) <<https://musically.com/2018/11/23/melodrive-ai-music-generation-games/>> accessed 3 January 2022.

other investment) but also make sure the public could freely use AI-produced works as soon as possible. The protection scope could refer to that of secondary works and a 5 to 10 years' protection term would be an appropriate choice.

Besides, it is worthy to identify the party that enjoys such protection. The Copyright, Design and Patent Act 1988 (CDPA) provides a unique section about computer-generated works, which was treated as the closest situation of AI-generated works. Section 9(3) identifies authors of computer-generated works as 'the person by whom the arrangements necessary for the creation of the work are undertaken.'⁵ It is worthy to notice that this section does not provide a precise meaning of the word "arrangements". We could understand this word as plans and preparation to make things happen (In the music industry, for example, it would mean planning and preparation to create music) and the person linked with such arrangements are varied which can be from users, program designers to software investors or instructors training and instructing the programmer.⁶ In the case of *Nova Productions Ltd v Mazooma Games Ltd & Ors*, the Court held the game's designer is the person by whom arrangements were undertaken rather than users who play the game when frame images generated during the playing.⁷

Nevertheless, it does not mean the above conclusion in the *Nova* case could be applied in all other industries. For example, in the situations of self-entertainment and the environment simulation, it is easier for us to understand who is the person to make necessary arrangements in an AI-produced work – that would be the programme designer since the users' intervention is minimized. However, while we look at the area of music creation in which most users are professional musicians, there is a sub-situation that human musicians may develop the final work based on AI-produced half-completed work. In this case, the final work embraces both human musicians' creativity and AI's effort. Obviously, the concept of computer-generated work is not appropriate to this situation since this concept only applies where there is no human author's effort. However, this would be a reality that legislators need to face now although this call for the consultation does not consider this aspect yet.

Copyright – text and data mining (TDM)

The following parts are from my report to WIPO Conversation on Intellectual Property and Frontier Technologies.

To understand whether we should consider a choice of extension of the existing TDM exception or the choice improving licencing, it is necessary to analyse the nature of how the data has been used in the text and data mining and machine learning process and what the protection scope of copyright law would be. In fact, text mining and data mining have common features and different features. Data mining means 'the

⁵ Copyright, Designs and Patents Act 1988, s 9(3).

⁶ Jani McCutcheon 'Curing the Authorless Void: Protecting Computer-Generated works following ICETV and Phone Directories' (2013)1 *Melbourne University Law Review* 53–56.

⁷ *Nova Productions Ltd v Mazooma Games Ltd & Ors* (CA). Reference: [2007] EWCA Civ 21.

computational process of discovering and extracting knowledge from structured data'.⁸ Text mining is 'is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights'.⁹ As the most common data type, text data is divided into three categories: structured data, unstructured data and semi-structured data.¹⁰ Most data in the world stays in an unstructured format, which covers text from social media, product reviews, video and audio files for example.¹¹ Therefore, text mining research would be very valuable because it could transform these unstructured documents into a structured format for analysis purposes.¹² No matter data mining or text mining, 'it works by copying large quantities of material, extracting the data, and recombining it to identify patterns.'¹³

It can be seen that data is normally used in several ways during the text and data mining and the machine learning and training process.

Information Extraction

As text and data mining technique are to explore those undermined relationships in those structured and unstructured data, one of the important functions for this "mining" technique is information extraction (or called data extraction) so as to find out relevant patterns, trends and correlations. For the purpose of achieving data extraction, researchers need to copy and input a large number of data which would involve massive copyrightable materials. Therefore, there are two issues that need to be analysed and clarified: firstly, how to identify the actions of data copying and inputting made by researchers from a copyright perspective; secondly, what is the nature of the action of "data extraction" in the copyright context.

A universal copyright principle recognised in the world is that copyright law only protects expressions of ideas rather than the idea itself. This means copyright law protects those original expressions that underline the works but not apply to facts, ideas, procedures, methods, etc. In this case, if any data relates to non-original expressions, it can be freely used for text and data mining as well as machine learning and training without copyright infringement issues.

If the researchers lawfully access and read the data, it would be no risk of copyright infringement. However, if they copy a whole or substantial part of copyrightable works for text and data mining/machine learning and training, it might be a risk of involving infringement. The good thing is most copyright laws at the national level provide limitations and exceptions allowing the public use. That is to say, even if the data

⁸ Cambridge Libraries, 'Text & Data Mining: What is TDM?' (*Cambridge LibGuides*, 28 June 2019) <<https://libguides.cam.ac.uk/tdm/definitions>> accessed 6 January 2022.

⁹ IBM Cloud Education, 'Text Mining' (*IBM Cloud*, 16 November 2020) <<https://www.ibm.com/cloud/learn/text-mining>> accessed 6 January 2022.

¹⁰ Ibid.

¹¹ Ibid.

¹² Ibid.

¹³ UK National Archives, 'Text Mining and Data Analytics in Call Evidence Responses' (2014) <<https://webarchive.nationalarchives.gov.uk/ukgwa/20140603093549/http://www.ipso.gov.uk/ipreview-doc-t.pdf>> accessed 6 January 2022.

involves copyrighted works, the data can be used as long as it falls into exceptional situations. In most national copyright laws, those non-commercial research and private study are normally covered by copyright exceptions. Although some countries such as the Copyright, Design and Patent Act (CDPA) in the UK copyright law and EU's DSM Directive expressly allow the text and data mining as an exception,¹⁴ this exception is strictly limited. For example, the CDPA only allows making a copy of a copyrighted work for text and data analysis with the purpose of non-commercial research whereas the DSM Directive requires text and data mining for the purpose of scientific research only. In practice, both commercial and non-commercial entities have been engaged in machine learning and AI training sectors. It is difficult to point out certain text and data mining research or machine training would be purely scientific and/or non-commercial, considering there is a possibility of shifting from a non-commercial starting point to a commercial achievement. Therefore, such a conservative approach to text and data mining as well as machine training data perhaps would be an issue for the existing copyright system.

Another issue is about the nature of “data extraction”. Data extraction is the first step for text and data mining as well as machine learning and training. Data extraction refers to ‘the process of collecting or retrieving disparate types of data from a variety of sources, many of which may be poorly organized or completely unstructured’.¹⁵ While data extraction requires collecting, checking and temporary copying data for retrieving and extracting purposes, its application seems to be different from a normal collecting and copying purpose. Data extraction is not aimed at reproducing “expressions” of copyrightable materials but to extract informational value through these copyrightable materials. In other words, for example, approaching the data extraction method to literary works would only allow a text extract with annotation of labels such as sentiment tags, named entities and addresses, it would not refer to a replication of the literary works themselves. From this point of view, data extraction does not fall into copying “expressions” of the copyrightable materials but more like to dig and access ideas and facts rooted in these materials. Besides, the purpose of data extraction is not pretending to compete with authors of copyrighted literary works in the market. For instance, the text extraction method is used for machine natural language training so as to achieve automatic translation rather than reproducing copies of literary works via the extraction method.

Therefore, treating data extraction as a pre-potential-infringement nature is not appropriate and counting text and data mining as a copyright exception would be weird as well.

Information Storage

Text and data mining and machine learning require data extraction, transformation and loading process, which is normally called ETL. During this process, those high-

¹⁴ Copyright, Design and Patent Act 1988, s 9A; DSM Directive, art 2(2), art 3–4 and 7.

¹⁵ Talend, ‘What is Data Extraction? Definitions and Examples’ (*Talend*)

<<https://www.talend.com/resources/data-extraction-defined/>> accessed 3 January 2022.

quality, extracted and refined data are then delivered to a data warehouse for storage and analysis. Therefore, text and data mining and machine learning often involve both temporary and permanent copies of copyrightable materials. Normally, data would be permanently stored for the data set preparation. Data set refers to ‘a file that contains one or more records’¹⁶ which are ‘the basic unit of information used by a program running on z/OS’.¹⁷ In other words, a data set is a collection of data. Researchers may also temporarily store data for the analysis of the data set. Therefore, both temporary storage and permanent storage would lead to the copying of data (it may cover both raw data and copyrighted materials such as literary, music and artistic contents). Copying raw data would not have any copyright issue but copying those copyrighted materials would be treated as infringing exclusive rights that copyright owners enjoy if such copying is unauthorised.

Nevertheless, it is worthy to consider whether such action of “storage” for text and data mining in the purpose of machine learning and training should be determined as the action of “reproduction”. The machine-learning-purposed data storage much likes to store memories (the data) into machine/AI’s neural network. While machines use those stored data for training, self-analysis and adjustment, it simply uses their remembered (stored) information (or copyrighted materials). In this case, the memories (data) can be concrete information; can be an abstract form (creative elements of copyrighted materials). While humans remembering a novel or painting would not be treated as infringing copyright, AI/machines remembering (storing) this novel or painting is determined as an infringement. In this case, it is necessary to review the action of “storage” in a machine-learning-purpose and distinguishing from the traditional way of digital storage in computers, as well as consider a possible expansion of the interpretation to the word “storage” from a digital copyright perspective.

Section B: Respondent information

A: Please give your name [REDACTED]

B: Are you responding as an individual, business or on behalf of an organisation?

- 1) Business – please provide the name of your business
- 2) Organisation – please provide the name of the organisation
- 3) Individual – please provide your name: [REDACTED]

C: If you are a responding on behalf of an organisation, please give a summary of who you represent.

D: If you are an individual, are you?

- 1) General public
- 2) An academic

¹⁶ IBM, ‘Zosbasics’ (IBM) <<https://www.ibm.com/docs/en/zos-basic-skills?topic=more-what-is-data-set>> accessed 3 January 2022.

¹⁷ Ibid.

- 3) A law professional
- 4) A professional in another sector – please specify
- 5) Other – please specify

E: If you are responding on behalf of an organisation, are you?

- 1) An academic institution
- 2) An industry body
- 3) A licensing body
- 4) A rights holder organisation
- 5) Any other type of organisation - please specify

F: If you are responding on behalf of a business or organisation, in which sector(s) do you operate? (choose all that apply)

- 1) Agriculture, forestry and fishing
- 2) Mining and quarrying
- 3) Manufacturing – Pharmaceutical products
- 4) Manufacturing – Computer, electronic and optical products
- 5) Manufacturing – Electrical equipment
- 6) Manufacturing – Transport equipment
- 7) Other manufacturing
- 8) Construction
- 9) Wholesale and retail trade; repair of motor vehicles and motorcycles
- 10) Transportation and storage
- 11) Information and communication – Publishing, audio-visual and broadcasting
- 12) Information and communication – Telecommunication
- 13) Information and communication – IT and another Information Services
- 14) Financial and insurance activities
- 15) Real estate activities
- 16) Scientific and technical activities
- 17) Legal activities
- 18) Administrative and support service activities
- 19) Public administration and defence
- 20) Education
- 21) Human health and social work activities
- 22) Arts, entertainment and recreation
- 23) Other activities – please specify

G: How many people work for your business or organisation across the UK as a whole? Please estimate if you are unsure.

- 1) Fewer than 10 people
- 2) 10–49
- 3) 50–249
- 4) 250–999
- 5) 1,000 or more

H: The Intellectual Property Office may wish to contact you to discuss your response. Would you be happy to be contacted to discuss your response? [Yes](#)

I: If you are happy to be contacted by the Intellectual Property Office, please provide a contact email address.

J: Would you like an acknowledgement of receipt of your response? [Yes/No](#)