Q295-SR-C-2025

Summary Report

by Ralph NACK, Reporter General
Guillaume HENRY and Johanna FLYTHSTRÖM, Deputy Reporters General
Rafael ATAB, Klaudia BLACH-MORYSINSKA, Mamta RANI JHA, Yanfeng XIONG
Assistants to the Reporter General

2025 – Study Question
AI & Copyright

Introduction

The inextricable connection between artificial intelligence (AI) and intellectual property rights (IP rights) has increasingly been under the legal spotlight, as the rapid development of AI systems continues to challenge the limitations of existing jurisdictional rules and exposes the gaps in our legal systems.

As the integration of AI systems into daily life becomes more widespread, such systems are fed to all sorts of data for purposes of creation, provoking questions concerning the applicability and adequacy of existing Copyright regimes.

The AIPPI has previously published several works in this field. In 2019, at the London Congress, the AIPPI notably adopted the resolution on "Copyright in artificially generated works", which aims to determine if and under what conditions Copyright should be available for artificially generated works.

Since the adoption of this resolution, the AI landscape has shifted, with the launch of GENAI like ChatGPT. Large language models (LLM) have particularly challenged Copyright legislation, due to their reliance on web scraping and text-and-data mining (TDM) for training purposes, practices that typically involve the collection of data without obtaining authorization from the original creators.

In response to these developments, multiple legal actions have been initiated against AI developers, including OpenAI and Stability AI, by various Copyright holders alleging unauthorized use of protected works for training purposes. Concurrently, several legislative measures have been introduced to address and regulate the deployment of AI systems. For instance, the European union adopted the AI Act on June 13, 2024, while Japan amended its Copyright Act in 2018 to allow the use of copyrighted works without the permission of the Copyright holder in cases where enjoyment of the work is not the purpose sought by the exploiter.

These legal acts, as well as proposals summarized in this Report, are instrumental in deciding the future trajectory of AI. Indeed, the decision on whether copyrighted works can be used as training data could either slow down the development of AI systems or, conversely, potentially cause irreparable damage to creators and authors alike. A balance must therefore be struck between further AI development and fairly promoting human creation and rights.

This Study Question thus seeks to establish whether and under what conditions a Copyrighted Data can be used as training data to train an AI System. This Study Question also aims to determine the sanctions for using Copyrighted Data without Copyright holder consent.

The following definitions have been used in connection with this Study Question:

- In the context of this study, the term "**Copyright**" means the rights associated with copyright as set forth in the Berne Convention and all other copyright-type rights, e.g. "related rights", "neighbouring rights", "sui generis rights", etc.

- The term "**Economic Rights**" means the exclusive rights of Copyright, e.g. the right of reproduction, representation, adaptation, etc.

- The term "**Moral Rights**" means the rights to Copyright apart from Economic Rights, e.g. the right to object to distortion of the work, right to authorship, etc.

- The term "**Exception(s)**" means any legal or jurisprudential exception or limitation to the Economic Rights and/or Moral Rights. An exception can be subject or not to indemnification/compensation to Copyright holders.

- The term "**AI System**" means a machine-based system (AI, GENAI, etc.) that:

- Is designed to operate with varying levels of autonomy and may exhibit adaptiveness after deployment; and
- from the input (training data) it receives, generates output such as new contents, etc.

- The term "**Copyrighted Data**" means one or more data (music, images, videos, texts, etc.), protected under Copyright law.

- The term "**Use of Copyrighted Data to Train AI System**" means Copyrights Data that are:
  - scraped off the Internet (or by any other means); and
  - possibly structured and incorporated into a dataset; and
  - used to train AI System.

- The term "**Provider of an AI System**" means a natural or legal person that develops an AI System and places it on the market or puts the AI System into service.

The Reporter General has received Reports from the following Groups and Independent Members in alphabetical order: Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Central America and Caribbean, Chile, China, Croatia, Denmark, Egypt, Finland, France, Germany, Hungary, Chinese Taipei (Independent Member), India, Israel, Italy, Japan, Latvia, Luxembourg, Malaysia, Mexico, Netherlands, Paraguay, Poland, the Republic of Korea, the Russian Federation, Singapore, Spain, Sweden, Switzerland, Turkey, the United Kingdom (UK), Ukraine, and the United States of America (US).

39 Reports were received in total[1]. The Reporter General Team thanks the Groups and Independent Members for their helpful and informative Reports. All Reports may be found in AIPPI's library at www.aippi.org.

The Reports provide a comprehensive overview of national and regional laws, practices, and policies relating to artificial intelligence and copyright, set out in three parts:

- Part I – Current law and practice

[1] Reports received after 6 May 2025 are listed above but are taken into consideration by the statistics in Parts II and III. This includes the Israel, Switzerland, Ukraine and UK Reports.

- Part II – Policy considerations and proposals for improvements of your Group's current law

- Part III – Proposals for harmonisation.

This Summary Report does not summarise Part I of the Reports received. Part I of any Report is the definitive source for an accurate description of the current state of the law in the jurisdiction in question.

This Summary Report has been prepared on the basis of a detailed review of all Reports (including Part I) but focuses on Parts II and III, given AIPPI's objective of proposing improvements to, and promoting the harmonisation of, existing laws. As it is a summary, if any question arises as to the exact position of a particular Group in relation to Parts II or III, please refer to the relevant Report directly.

In this Summary Report:

- references to Reports of or responses by one or more "Groups" may include references to Independent Members;

- where percentages of responses are given, they are rounded to the nearest 5%; and

- in Part IV below, some conclusions have been drawn in order to provide guidance to the Study Committee for this Study Question.

I. Current law and practice

For the replies to Questions 1) – 12) set out in the Study Guidelines for this Study Question, reference is made to the full Reports. The Study Guidelines may be accessed in AIPPI's library at www.aippi.org.

II. Policy considerations and proposals for improvements of your Group's current law

Copyright regime

13) Could any of the following aspects of your Group's current law or practice

relating to the Use of Copyrighted Data to Train AI System be improved?

- 23 Groups (92%) indicate that their current legislation could be IMPROVED with respect to the Use of Copyrighted Data to Train AI System, with a majority agreeing that such use should require prior authorization from the Copyright holder and that certain exceptions should justify the bypass of this authorization. Most Groups were also in accordance with the belief that their current legislation could be improved with regard to the consequences encountered due to illicit use of Copyrighted Data to Train AI System.

- 2 Groups (8%) indicate that their current legislation requires no further improvements relating to the Use of Copyrighted Data to Train AI System.

a. Use of Copyrighted Data to Train AI System requires prior authorization of the Copyright holder?

- 29 Groups (83%) indicate that their current legislation could be improved regarding prior authorization of the Copyrighted Data to Train AI System, for various reasons:

  A majority of groups consider that there is legal ambiguity around whether the use of Copyrighted Data to train AI Systems constitutes an act of reproduction and therefore wish for clarity in this respect.

  The US Group mentions that since US Courts are given great flexibility in applying the fair use mechanism without statutory intervention, this can lead to inconsistencies in its application.

  The Chinese Group say that: *"in current practice, there are significant difficulties and high costs associated with obtaining such authorization. Therefore, it is necessary to continuously explore and expand methods of obtaining authorization in future practice, such as introducing collective management and collective negotiation as viable forms of authorization when conditions are ripe."*

  The Chilean Group indicates that: "*The current Chilean framework would be improved by explicitly clarifying when prior authorization is required or establishing specific conditions and licensing mechanisms, thereby*

reducing uncertainty and fostering innovation while safeguarding author's rights."

- 6 Groups (17%) consider that no prior authorization is necessary.

b. Exceptions authorising the Use of Copyright Training Data without Copyright holder consent.

- 27 Groups (79%) indicate that their current legislation could be IMPROVED with regard to the exceptions authorizing the use of Copyright Training Data without the consent of the Copyright holder for purposes of clarification and/or enlarging the exceptions at disposal of AI providers.

  - Current legal provisions could be clarified as to whether and in what circumstances use of Copyrighted Data as training data could be considered "fair dealing" (Canada Group) or "fair use" (US Group).

    Concerning the "opt-out" mechanism for right holders wishing to exclude their works from AI training, the Belgian Group comments: *"First, the Belgian group considers that the provisions of the Belgian Code of Economic Law (CEL) implementing the TDM exception for general purposes (Article 4 DSM Directive, opt-out clause) are blanket and general, meaning that an author can either fully allow or fully prohibit the use for any and all TDM purposes, whereas in practice, the author might agree only with some purposes (for example, development of AI model for drug discovery) and oppose to certain others (AI model for weapons). The Belgian group considers this an area for improvement. The need for clarification is even more urgent under Belgian law, given the question as to whether the right of destination can come into play. Second, (…) clarifications are needed at the very least on EU level, given the fragmentation of AI-related exceptions to Copyright at international level, and the fact that the EU legislator already acted towards harmonization."*

  - The French Group states that: "*the scientific research exception could be broadened to bring it into line with the definition in article 3 of Directive 2019/790 on text and data mining: the use of protected works by research organisations and cultural heritage institutions, to which they have lawful access, for the purpose of training an AI*

system. The French group is in favour of a more general exception under the following conditions: the use of protected works by any provider of a general-purpose AI system to which it has lawful access, for the purpose of training a general-purpose AI system provided that the use of the protected works has not been expressly reserved by their rights holders in an appropriate manner."

The Paraguayan Group adds that: *"Paraguay lacks modern exceptions such as TDM. Introducing narrowly defined exceptions for scientific research could be beneficial."*

In the same vein, the Australian Group indicates that: "*The lack of a text of data mining exception hampers the development of new, innovative technology in Australia, and a limited exception (at least where the use is not economically significant) seems desirable. However, the exception must operate in a way that adequately compensates the Copyright holder.*"

- 7 Groups (21%) consider that their current law does NOT need to be improved upon concerning exceptions, either due to sufficient legal provisions, or due to strict interpretations of what constitutes an act of reproduction:

  The Dutch Group says: *"Training of AI Systems is merely a new form of reproduction but in essence not different from other forms of reproduction. Accordingly, to the extent "Use" involves acts reserved to the copyright holder, the use should require the Copyright holder's consent. The current development of AI resembles in this respect the transition from paper to digital media, where reproduction on CD-ROM's or the Internet was still held to require consent from Copyright holder, despite the interests of publishers and others in being able to transfer to the digital area."*

c. <u>Consequences</u> of illicit Use of Copyrighted Data to Train AI System.

i. What can be qualified as infringing products, e.g. outputs and/or AI System itself? Please explain if needed.

- 26 Groups (74%) indicate that greater clarity is needed concerning whether the AI system can be found infringing if it utilizes Copyrighted Data in order to be trained or used, even in the cases where the output itself does not constitute a reproduction of a copyrighted work.

  The Swiss Group indicates that *"one question that remains whether an AI System that can reproduce Copyright-protected Data is, in itself, to be considered as infringing Copyrighted Data, aside from its training and one-off outputs of the AI System. For example, once trained, a large language model (LLM) may, given certain user prompts, come to reproduce (portions of) texts protected by Copyright. The question then arises as to whether this LLM can, in itself, be compared to a Database storing various (portions of) text protected by Copyright."*

- 9 Groups (26%) consider that their current law does not require further improvements with regard to the qualification of infringing products.

  The Taiwanese (Independent Member) Group summarized this through their response: *"the current law and policy regarding outlets and AI system are acceptable. The output may remain being subject to higher regulatory control, as it now, for the sake of the copyright owner's interest. The user of an AI system can determine what to prompt or omit when generating content. On the contrary, the AI system is neutral and should be jointly liable with illicit use of AIDGC."*

ii. The sanctions that should be available in case an AI System has been recognized to have been trained with infringed Copyrighted Data without authorisation? Please explain if needed.

- 19 Groups (56%) indicate that improvements could be made to their current legislation with regard to the sanctions that should be available in case an AI System has been recognized to have been trained with infringing Copyrighted Data without authorization.

Clarification is needed as to the proportionality of an injunction towards the provider of an AI System that has been trained only in part with infringed Copyrighted Data without consent, especially if the removal of such data is questionably feasible.

The Mexican Group suggests the introduction of explicit fines and penalties, allowing copyright holders to claim a proportional, reasonable share of profits generated by AI models using their works, implementing data set audits to ensure compliance with copyright laws.

The Finnish Group responds that: "*While current sanctions may well be sufficient, further clarity and specificity in applying the current law in AI System context would be need in form of case law especially in relation to establishing an effective basis for claiming compensation in connection with the fault under law: it has been publicly declared (in connection with the preparation of the now-suspended proposed AI liability directive, Article 4) that it can be challenging for claimants to establish a causal link between such non-compliance and the output produced by the AI System of the failure of the AI System to produce an output that gave rise to the relevant damage.*"

- 15 Groups (44%) believe their current legislation to be adequate in this regard, mostly because their current copyright sanctions prove to be sufficient.

  A majority of groups believe that the case-by-case application of the principle of proportionality when deciding sanctions is sufficient (see, for instance, the Italian, Dutch and Spanish reports).

14) Are there any other policy considerations and/or proposals for improvement to your Group's current law falling within the scope of this Study Question? If YES, please explain.

- 16 Groups (50%) recommend other policy considerations for improvement to their current legislation, including:

- The imposition of mandatory transparency obligations (see, for instance, the Paraguayan, Chinese, Luxembourg Reports).
- The establishment of fair compensation mechanisms for Copyright holders (ex. Blanket licensing) (see, for instance, the Mexico Report).
- The clarification of the Copyright status of AI-generated content and of liability in case of copyright infringement (see, for instance, the Austria and Mexico Reports).
- The development of a broadly recognized technical solution to make it easier to comply with the opt-out expressed by right holders and streamline the process of rights reservation (see the Belgian Group report).
- Defining the role of collection societies in case of compensation (see the Argentinian Group report).
- The Austrian Group suggests that there should be *"More precise provisions on liability for infringements, remuneration and information rights."*

- The Danish Group adds: "*There does not exist the litigation tools to meaningfully pursue infringement cases against the providers/developers of AI systems. This makes pursuing infringement de facto impossible.*"

- 17 Groups (50%) consider that no policy improvements were needed.

III. Proposals for harmonization

15) In your opinion, should Use of Copyrighted Data to Train AI System be harmonized?

- 34 Groups (100%) consider that harmonization is desirable, mostly due to the cross-border nature of AI Systems that therefore calls for a global, harmonized legal approach to Copyright law. Without such harmonization, forum shopping is likely to ensue among AI developers and harm the AI landscape by centralizing development in certain areas only and favoring AI developers in those countries.

  The position of the Groups can be summarized by the German Group opinion: *"AI systems can run on one or more servers that may be located in different jurisdictions. Usually, AI systems will be trained using copyrighted data from authors from different countries. The act of training as such may be already not limited to one jurisdiction. Based on the location of the server or servers and based on the country of origin of the copyrighted data automatically several jurisdictions and several legal systems are involved. In addition, the trained AI will be used in more than one country again falling under different national laws. To create legal security for all parties including copyright holders, AI providers and AI users, a harmonization of the law is desirable."*

- The UK Group is not in favor of harmonization with regard to the Use of Copyrighted Data to Train AI System.

I. Should the Use of Copyrighted Data to Train AI System be authorized by the Copyright holder?

16) Should the Use of Copyrighted Data to Train AI System fall within the scope of the Economic Rights monopoly, *i.e.* should it require the authorisation of the Copyright holder as a matter of principle, unless covered by an Exception?

No matter their response to this question, the majority of groups emphasize the need to strike the right balance between the interests of authors and those of AI system developers.

- The vast majority of Groups (95%) answer YES to this question.

- Within the groups that answer YES to this question, the majority believe that the Use of Copyrighted Data to Train AI System should fall within the scope of the Economic Rights monopoly, and that authorization of the Copyright holder should therefore be required, unless covered by an exception.

  For instance, the Australian Group indicates: *"In principle, a Provider or User of the AI System derives economic benefit from the creation output of the author of the Copyrighted Data. The Use of Copyrighted data should fall within the Economic Rights monopoly as a matter of principle, unless covered by an Exception."*

- Some groups (minority) that answer YES to this question, considering that authorization of the authors should be necessary in any circumstance, even if the training does not involve any reproduction.
  For instance, the Canadian Group states that: *"Authorization of the Copyright holder should likely be required in all circumstances, even if the "Use" does not involve reproduction, representation etc."*.

  In this regard, it should be noted that several Groups point out that it is unclear whether the training of an AI system systematically involves the reproduction of training data.
  For example, the German group states that: "*The question of whether the use of copyrighted data to train AI systems falls within the scope of the copyright holder's economic rights is fundamental, as it determines all subsequent legal issues, including the applicability of exceptions, remuneration obligations, transparency requirements, and liability considerations. Given its significance, the training process itself should be recognized as involving a relevant act of reproduction under copyright law. The exact technical mechanisms of AI training—particularly how information is processed within neural networks and whether/to what extent training data is stored —remain unclear* ».

- Only 2 Groups (5%) consider that authorization of the Copyright holder should NOT be required as a matter of principle.

  The Japanese Group considers that: *"In principle, the use of a copyrighted work for machine learning on an AI system should be legally permitted to the extent deemed necessary and should be illegal on an exceptional basis if the interests*

of the copyright owner would be unreasonably prejudiced, as in the current Copyright Act of Japan."

The Singapore Group hold that: *"it is likely to be unwieldy for the development of AI Systems if authorization needs to be actively sought."*

17) Should there be Exceptions that allow the Use of Copyrighted Data to Train AI System without Copyright holder consent:

a. General Exceptions to Copyright, *i.e.* which are not special to the Use of Copyrighted Data to Train AI System (*e.g.* fair use, intermediate storage/temporary reproduction, etc.)?

At first, the responses may seem divided, but it emerges from the reports that a large majority of the Groups consider that the General Exceptions to Copyright should apply (if the conditions are fulfilled).

- 18 Groups (55%) answer YES and agree that General Exceptions to Copyright Law should be applicable (if conditions of such exceptions are met).

  Most cited General Exceptions are "fair use" type exception and temporary reproduction exception.
  Regarding fair use, the conditions/criteria for AI developers to benefit from this exception are diverse.

  For instance, the Turkish Group considers that: *"General exceptions can be applied to AI-generated output that are classified as copyrighted works. Fair use principle or temporary reproduction exceptions, under certain conditions (e.g. non-commercial, educational or research) can be considered for Use of Copyrighted Data to Train AI System). Fair use criteria can be considered as (i) the Type of Copyrighted work (creative or factual) (ii) the amount of Copyrighted work used for the AI system training (snippet or substantial amount of the Copyrighted Data/work) and (iii) use could hurt the market for the original work. It's all about balancing those factors to see if it falls under an exception."*

  The Canadian Group summarizes this position: *"There should be a "fair use" (i.e. fair dealing) type exception; for example, for the purpose of*

research, private study and education. This exception should apply as an exception to use of Copyrighted Data for training AI System. Whether the use is fair should be based on multiple factors including 1. the purpose of the use; 2. the character of the use; 3. the amount of the use; 4. alternatives to the use; 5. the nature of the use; and/or 6. the effect of the use on the work".

In the same vein, US Group considers that general copyright exceptions, such as fair use, should apply to AI training and should *"requires the evaluation of 4 factors: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) The effect of the use upon the potential market for or value of the copyrighted work".*

The Chinese Group adds that: *"If the use of copyrighted data to train AI systems is not aimed at utilizing the original expression of copyrighted works and does not affect the normal use of copyrighted works or unreasonably harm the legitimate interests of relevant copyright holders, it may constitute fair use."*

The Paraguayan Group also considers that: "*A general exception like fair use or temporary reproduction may be appropriate if the use is transformative, non-commercial, and does not substitute or compete with the original work."*

- 15 Groups (45%) answer NO to this question.

- It emerges from the reports that the majority of these Groups are not opposed to the application of General Exceptions, but consider that they are not sufficient.

  For instance, the Belgian Group considers that: *"the technicalities of the reproduction and communication acts carried out on Copyrighted Data to Train an AI System are of such a specific nature, in comparison to the reproduction and communication acts generally covered by the General Exceptions to Copyright, that the Belgian Group is of the opinion that*

General Exceptions to Copyright to data are not fit for the purposes and a Special Exception should be provided for".

The Dutch Group also expresses that general exceptions are likely to meet difficulties in their application, as Copyrighted Data used to train AI may not meet all the requirements of this exception : *"While the temporary reproduction exception (…) is tangentially related to AI training, it will likely be challenging to argue that all the requirements of this exception are met when Copyrighted Data is Used to train an AI System. (…). Other general exceptions will likely meet comparable difficulties regarding applicability".*

The Spanish Group also considers that general exceptions are generally not sufficient to cover the extensive data processing required for training AI.

- Only very few Groups consider that as a matter of principle, there should be no general exceptions applicable to the Use of Copyrighted Data to Train AI System (see for instance the Luxembourg report).

b. Special Exceptions to the Use of Copyrighted Data to Train AI System (*e.g.* TDM Exception, etc.)?

- 29 Groups (90%) answer YES to this question and consider that Special Exceptions to the Use of Copyrighted Data to Train AI System should exist, often called Text and Data Mining (TDM) Exception.

- The majority of Groups that support a Special Exception for AI training systems indicate that the reason for this is that there is uncertainty as to whether general exceptions (such as temporary storage or processing, fair use, etc.) are sufficient.

For instance, the Chilian Group indicates that: *"General exceptions (such as broad fair use doctrines) are typically too ambiguous or inadequate to address the scale and specificity of AI training. Thus, it is advisable to establish clear, specialized exceptions tailored specifically for AI-related uses".*

To the same effect, the Bulgarian Group also upholds that: *"Fair use cases involving generative AI training on copyrighted works will be highly fact dependent. While some AI-related uses may qualify as fair use, unauthorized use of copyrighted material to train AI systems cannot be handwaved by a broad fair use exception that disregards the right of creators."*

Several Groups, such as the Brazilian and the Swedish Groups indicate that application of such Special Exception should comply with the three-step test provided by Article 9(2) of the Berne Convention.

Furthermore, several groups indicate that the TDM exceptions should cover only the train of AI System and should NOT apply to any acts after the training of the AI System, for instance any act of communication.

- Furthermore, many EU Groups (such as the Danish, Belgian, Polish, Dutch, French, Spanish, German, Italian and Hungarian Groups reports), but also Groups from other jurisdictions, such as the Turkish Group and the Chinese Group, are in favor of an opt-out system, giving the right to Copyright holders to refuse the use of their works for training of AI Systems. In such a system, as a matter of principle, every copyrighted work lawfully accessible on the Internet can be used to train AI System, except if the Copyright holder has expressly refused such use (opt-out). Several Groups indicate that the opt-out should be made in an appropriate manner such as machine readable form for contents made publicly available online and the deletion of reproductions should take place when they are no longer necessary for TDM.

  There is a debate within the groups that are in favor of a TDM Special Exception concerning the scope of such TDM exception.

  - The majority of the Groups that are in favor of a Special Exception consider that the Special Exception should be limited only to specific situations involving public interest such as scientific research, culture, education, information, non-profit activities, etc.

    For instance, the Mexican Group indicates that: *"Use for training purposes limited to special cases, including educational, and non-profit uses, legitimate research, hot news, and non-profit purposes, non-*

commercial AI development, systems security research, as well as the rights of persons with disabilities. Any exception must be narrowly tailored to avoid undermining the economic rights of authors and ensure a fair balance between technological development and the protection of creative effort".

The Australian group indicates that: *"A text and data mining exception should be limited only to circumstances that do not materially affect the economic interests of copyright owners. For example, use for scientific or non-commercial purposes could be permitted on a broader basis".*

With regard to the scientific research exception, the French Group responds that: *"The French group is in favor of an exception for scientific research purposes under the following conditions (…): The use of protected works by research organizations and cultural heritage institutions, to which they have lawful access, for the purpose of training an AI system. The French group is in favor of a more general exception under the following conditions (…): the use of works protected by any general-purpose AI system provider to which they have lawful access, with a view to training a general-purpose AI System; provided that the use of the protected works has not been expressly reserved by their rightsholders in an appropriate manner*".

- A small minority of Groups in favor of a TDM exceptions propose that it should apply generally to the Use of Copyrighted Data to Train AI System.

For instance, the Japanese Group states that: "*In principle, the use of a copyrighted work for machine learning on an AI System should be legally permitted to the extent deemed necessary and should be illegal on an exception basis if the interests of the copyright owner would be unreasonably prejudiced*".

In the same vein, the Croatian Group notably developed a model in response to this question: *"We propose a model based on open access to copyrighted data (i.e., without author's authorization) given that appropriate remuneration is provided to authors ("Data model")."*

- Only 3 Groups (10%), answer NO to this question, such as the US and the Canadian groups consider such exceptions to not be necessary, as existing General Exceptions are sufficient (see also the UK report).

18) Should the Exceptions provide for financial compensation for Copyright holders (*e.g.* a royalty paid to a collecting society, etc.)?

- The majority (23 Groups - 70%) uphold that the Exceptions should provide for financial compensation for the Copyright holders, to strike a fair balance between innovation and protection of creators' economic interests.

- The majority of the Groups point out that collecting societies should play a major role in the administration/negotiation of licenses with AI developers.

  The Egyptian Group resume this majority opinion: *"Exceptions for the use of Copyrighted Data to Train AI System should provide for financial compensation to copyright holders. (…) it is important to ensure that creators are fairly compensated when their works are used, even in contexts like AI training. Incorporating provisions for compensation, such as royalties or payments to collecting societies would balance the interest of copyright holders and the public, encouraging innovation while respecting the economic rights of creators. This approach aligns with international trends and helps create a sustainable ecosystem for both AI development and copyright protection, preventing the exploitation of copyrighted works without fair remuneration."*

- Some Groups explore specific options and forms of remuneration, such as financial contribution to cultural activities.

  For instance, in case the Copyright holder opts-out, the Dutch Group indicates that: *"after having opted out from the application of the TDM exception, right holders enter into licensing agreements to authorise TDM of works and other protected subject matter by generative AI model providers, some appropriate and proportionate remuneration should be provided to authors and performers when they have transferred or licensed their rights to such right holder. Since remuneration in such a case of training on massive numbers of works and performances might be rather minimal for authors and performers or difficult to determine, it would be useful to organise some other forms of appropriate compensation (such as a*

residual remuneration right or collective remuneration models or, beyond the copyright regime, other compensation mechanisms such as a financial contribution to cultural funds/activities or to the impacted creative sector). »

In the same vein, the Danish Group considers that an alternative model could be considered: "*it may instead be more appropriate and practically feasible to impose a fee on the distributors of AI systems. This small fee would then be administered either by collecting societies or, for example, the Ministry of Culture, who are then responsible for the money going to the creative fields whose works are most often used as training data".*

- Only 9 Groups (30%) believe that financial compensation is NOT desirable, for diverse reasons such as practical obstacles to implementing financial compensation, in order not to hinder the development of AI System or because these Groups are in favor of an opt-out system (see, for instance, the US, Japanese, French, German, Italian, Mexican, Swedish, and Brazilian reports).

  For instance, the Swedish Group considers that if the Copyright holder opts-out, no financial compensation is needed, because in such case, no exception applies and AI developer must contract with the Copyright holder before using copyrighted work to train AI System: "*the question of whether copyright holders should be financially compensated when the use is governed by a copyright exception is highly complex and multi-faceted. Although collective licensing has been proposed as a potentially pragmatic solution, the Swedish group underlines that such a model would introduce a range of practical and conceptual difficulties that must not be underestimated. One key challenge concerns the criteria for determining eligibility and the level of compensation. (…) The Swedish group also wishes to highlight the significant administrative challenge of distributing compensation fairly. (…) Moreover, the group expresses concern that introducing compensation obligations could substantially increase the cost of AI development regardless of sector. While large companies may be in a position to absorb such costs, smaller players could be effectively excluded from the market, thus distorting competition and hindering innovation. (…) Having regard to these various factors, the Swedish group leans towards the view that exceptions should not provide for mandatory financial compensation."* (see also the German report).

The Japanese Group provides a different perspective: *"The economic value of a copyrighted work is usually realised in the market when those who view or otherwise consume the work pay a price in order to enjoy the ideas or emotions expressed in the work and obtain the effect of satisfying their intellectual or spiritual needs. However, the use of a copyrighted work for machine learning on an AI system is an act not intended to enjoy the ideas or feelings expressed in the work, and it does not limit the opportunity to collect compensation from those who wish to enjoy the ideas or feelings expressed in the work. Therefore, financial compensation should not be required if the exceptions apply".*

19) Should the author be able to object to the Use of Copyrighted Data to Train AI System on the basis of his/her Moral Right (*e.g.* on the basis of the right to integrity, paternity, etc.)?

- 21 Groups (65%) answer YES to this question.

- A majority of Groups that answer YES to this question consider that the author should be able to object to the Use of Copyrighted Data to Train AI System on the basis of his or her Moral right, only to the extent that it causes injury to the work, such as through distortion or misrepresentation of the work.

  For instance, the Chilean Group indicates that: *"authors should retain the possibility to invoke their moral rights (integrity and authorship) to object to AI training if the resulting use or output negatively impacts the author's honor, reputation, or integrity of their works."*

  The Brazilian Group states that: *"A possible line of interpretation that the Judiciary might adopt is that denying authors the ability to object to using their works for AI training could erode their right to integrity. Since AI-generated outputs may incorporate elements of their original creations– potentially in ways that distort meaning, dilute artistic intent, or associate their work with undesirable contexts – such use could fundamentally alter how their work is perceived and attributed, impacting both their reputation and creative legacy."*

- A minority of Groups consider that the author should generally be able to object to the Use of Copyrighted Data to Train AI System on the basis of his/her Moral right (see for instance Spanish Group report).

- 11 Groups (35%) oppose the author's objection on the basis of his or her moral rights, for different reasons.

  The Canadian Group states that: "*The right to paternity and right to integrity of the work do not arise in the context of training of an AI system such that a moral right of the author would be engaged.*"

  The Central American and Caribbean Group responds: "*Since the moral right of an author is limited to the extent of its copyrighted work (e.g. right of attribution of the work, right to integrity of the work) the author should only be entitled to object the use of a specific "output" of the trained AI System which may amount to an infringement of their moral rights for specific copyright protected works.*"

20) Should the Provider of an AI System be obliged to make public the training data used to train the AI System (<u>transparency obligation</u>)? Please answer YES or NO. If YES, please explain, *e.g.* the degree of detail that should be required (all works, categories according to sources-websites, etc.) and whether the company should keep the list of training data or the training data itself for a certain period of time?

- A vast majority of the Groups (29 Groups - 85%) uphold that imposing transparency obligations is desirable, as it would ensure fairness and accountability from AI Providers and would facilitate enforcement and oversight.

  There is debate within the Groups that are in favor of a transparency obligation on the scope of such an obligation.

  The most commonly mentioned scopes of the transparency obligations are:

  - <u>All works</u> used, categorized according to sources (e.g. websites, databases, etc.) (see, for instance, the Spanish and Brazilian reports);

- List of training data used to train the AI System (without making public the works themselves) (see, for instance, Dutch, Canadian and Mexican reports);

- Meaningful summaries or categorization of works used (e.g. types of works, general sources) (see, for instance, Chilian, Polish and Chinese reports);

- URLs or website of works used (see, for instance, French and Polish reports);

- To provide sufficient details to make possible for the Copyright holders to actually enforce their rights (Danish, Belgian, Australian and Italian reports).

The German Group states that: "*Providers should be obligated to store a full record of the training data for a defined period and make it available to third parties for review. However, it is important to ensure that transparency requirements do not impose excessive bureaucratic burdens that could hinder AI development. Therefore, such obligations should be designed proportionately to protect right holders without stifling innovation. Disclosure would also facilitate financial compensation as it would be directly linked to the actual use of training data.*"

The Belgian Group proposes a two steps method: *"the transparency obligation as a principle should be applicable to all Providers of AI Systems anywhere in the world. Regarding the granularity of the data to be made public, the guiding principle should be to provide sufficient details to facilitate parties with legitimate interests to exercise their rights, and this includes right holders but also data subjects, consumers and other persons. In this regard, the focus should be on the sources of the databases, and on the description of the datasets (data size, modalities and characteristics) used for training the AI Model and System. This should apply to all sources of content used, from the pre-training to the fine-tuning stage. This data should be made public in the form of a summary on the content used for training and should be kept up to date. The way in which this data is presented in the summary should enable an average person without technical expertise nor legal*

knowledge to understand it. Subsequently, parties with a legitimate interest should be able to obtain more information to preserve and safeguard their rights, based on the data that the providers should still keep, that should include the data used for training and more insights into how the algorithm works".

- Only 5 Groups (15%) oppose the implementation of transparency obligations, the majority of whom believed that this would encroach upon privacy, confidentiality and trade secrets.

  The Luxembourgian Group summarizes this opinion: "*The provider of an AI System should not be obliged to make public the training data used to train the AI system, considering the trade secret protections and the sui generis protection of databases. (…) Trade secrets are vital for AI Systems providers due to the valuable and confidential nature of training data, which offers a competitive edge (…) Requiring AI Systems providers to disclose their training data could undermine this protection and might reduce the incentive to invest in creating high-quality databases.*"

  The US Group is on the same opinion: *"A general requirement of this nature would be impractical, likely to inhibit research and development, overly burdensome, and may implicate concerns over privacy, confidentiality, and trade secrets."* (see also Japanese, Swedish and Austrian report).

II. The consequences of unauthorised Use of Copyrighted Data to Train AI System

To answer questions 21 to 24, please consider that the Use of Copyrighted Data to Train AI System falls within the scope of the Economic Rights monopoly (no Exception can be invoked) but is made without the consent of the Copyright holder.

A. Infringing items

21) Should the contents created (output) by the AI System be qualified as Copyright infringement in the following cases?

a. The output contains characteristic elements of one or more Copyrighted Data to Train AI System?

- The vast majority of the groups (31 Groups - 90%) consider that an output that contains characteristic elements of one or more Copyrighted Data Used to Train AI System should be qualified as Copyright infringement.

  The US Group answers positively: *"if the output is substantially similar to a copyrighted work used in training and copies protectable elements of expression and no exception, such as fair use, applies. Copyright infringement occurs when a work copies protectable elements of another work in a manner that is substantially similar".*

- Only 3 Groups (10%) do NOT believe that output containing characteristics elements of the Copyrighted Data should qualify as copyright infringement (see Canadian, Swedish and Croatian reports).

b. The output is in <u>the same style</u> as one or more Copyrighted Data used to Train AI System?

- Only 8 Groups (25%) believe that output in the same style as Copyrighted Data Used to Train and AI System should be qualified as Copyright infringement.

  For instance, the Italian Group considers that: "*The lawfulness of copying another author's style is justified under traditional copyright law to ensure that people have adequate margins of artistic freedom and to allow them to express themselves also by joining artistic movements. With respect to the output of generative AI systems, these same requirements clearly do not apply. Therefore, we believe that the author's interest in obtaining adequate protection must prevail over the generation of synthetic content in direct competition with that which has been ingested by the machine in the training phase".*

  The Polish Group also states that: *"each author, person has a distinctive, different style or writing, painting, etc. this is their sort of "trade-mark", distinguishing the style from other artists. Each such style should be protected by legal standards."*

- 26 Groups (75%) do NOT think that output in the same style as Copyrighted Data Used to Train AI System justifies its qualification as copyright infringement.

  The majority considers that creating works in a similar style is generally not considered Copyright infringement, so the same should apply with respect to AI Systems (only expressions and not ideas are protected by Copyright).

  Nevertheless, the Belgian Group proposes to establish a presumption that the AI System has been trained with Copyrighted Data used to train AI System, if outputs are in the same style as this Copyrighted Data: *"an output generated in a style which merely constitutes the expression of a general aesthetic, should not be qualified as Copyright infringement. Nonetheless, the Belgian group notes that AI Models can be good at generating styles of authors, if the works of those authors are sufficiently present in the Copyrighted Data used to train AI System. This was the case, during the preparation of this report, with the "Studio Ghibli" style images. In this regard, if an AI System can produce output in a style characteristic to a certain author, this could be considered as a presumption or maybe evidence that the AI System has been trained with the author's Copyrighted Data".*

c. The output is in all cases infringing if it has been trained with one or more infringed Copyrighted Data.

- Only 7 Groups (20%) consider that the output should be deemed in all cases infringing, if AI System has been trained with one or more infringed Copyrighted Data.

  Furthermore, some groups nuance their position. For instance, the Italian Group considers that: "*in principle the output generated by an AI system that has been set up through unlawful training should be regarded as a form of unlawful exploitation of the ingested content. At the same time, it has also been observed that some caution should probably be taken to avoid that this line ends up unduly prejudicing those who have incurred in the*

training activity in unknowing infringements despite having taken appropriate and proportionate measures to avoid the offences and thus acted diligently. For example, it seems to us that different solutions are required for a system trained with millions of posts published on social networks or for a system trained (perhaps from a general model) on a dataset consisting of the entire production of a novel author".

- A large majority of Groups (27 Groups - 80%) consider that the output should not be considered to be infringing, for the sole reason that the AI System has been trained with one or more infringed Copyrighted Data.

  The German Group summarizes this idea: "*There should not be a strict liability for infringement solely based on the use of copyrighted training data. The infringement depends on the nature of the output, not just the unlawful nature of the training input. If the output does not reproduce protected elements of the original works, there is no reason to assume an infringement – even if the training process was unlawful. However, the training process itself may still be infringing (separate from the output), but this does not automatically render every output infringing.*"

  The US Group replies that: "*Copyright infringement should be determined by whether the output itself is substantially similar to a protected work, not merely by the legality of the training process. Potentially, the output could be enjoined if the system is found to be infringing."*

  The Mexican Group also upholds that: "*The fact that Copyrighted Data has been used for training purpose does not automatically result in an infringing output. To determine infringement, the underlying work must be recognizable in the output."*

d. In other cases?

- 10 Groups (30%) answer YES to this question.

Most Groups that answer YES to this question cite other cases in which output could be found infringing, includes:

- If the output is substantially similar to a copyrighted work (see, for instance, the German, US and Swedish groups reports);
- AI System imitates a fictional character or setting that is itself protected (see the German Group report);
    - In the case of deliberate or intentional alternation to disguise resemblance with the underlying work (plagiarism, paraphrasing) (see the Mexican Group report).

- 22 Groups (70%) do NOT consider that there are any other cases in which output should be found infringing.

22) Should the AI System itself be considered a Copyright infringement?

The Groups are quite divided on this question. Nevertheless, it emerges from the reports that the majority of the Groups consider that the AI System itself should be considered as Copyright infringement ONLY in certain specific circumstances, like, for example, when the AI System:

- is designed to create Copyright infringements; or
- systematically or massively uses copyrighted works as training data.

- 16 Groups (45%) answer YES to this question, considering that the AI System itself should be qualified as Copyright infringement. Diverse reasons are mentioned to justify such sanctions:

    - If the AI System has been trained with massive Copyrighted Data.

    For instance, the Italian Group states that: *"It depends on factors such as the methods of acquisition and processing of the data used for its training, its functionalities, and particularly its capacity to generate infringing output. An AI-system created by a massive, systematic and reckless infringement of third parties' rights would probably not be considered as lawful from the point of view of copyright law, at least because its very existence and actual configuration depends on a violation of copyrights. An AI-system that memorize substantial portions of protected works without the consent of rightholders would*

not qualify as well as lawful. The same could be said about an AI system capable of regurgitating input materials, particularly if such defect depends on negligence, or even intent, of the AI-developer."

- For instance, if the AI System is going to systematically reproduced copyrighted works, the Mexican Group mentions the following situations: *"If an AI system is designed in a way that systematically reproduces copyrighted content, it could be deemed an infringing tool. Some of the key factors would be:*
  - *o Whether the AI system directly generates the reproduction of copyrighted works without authorization.*
  - *o Whether it stores and retrieves copyrighted content in a way that substitutes for the original works.*
  - *o Whether its developers knowingly trained it on infringing material and failed to implement safeguards."*

  If the AI System has incorporated Copyrighted works in such a way that they have become inseparable from the AI System or its dataset, and that removing said works is no longer possible or extremely difficult."

- The Japanese Group upholds that: "*If an AI system generates outputs that are similar to a copyrighted work used for its machine learning with a high probability, the AI system itself may be regarded as being made through an act of infringement, and a request of destruction of the AI system could be available."*

- 18 Groups (55%) do NOT believe that the AI system itself should be considered infringing, because it does not infringe copyrighted works itself.

  For instance, the Chinese Group responds that: *"Generally during the data training process, the AI system learns statistical patterns through training data, and its parameters are abstract representations of the data, usually without directly storing the original data. Therefore, it is generally believed that even if the use of copyrighted data to train AI system without the authorization of the copyright holder constitutes infringement, it should not be directly concluded that the AI system itself constitutes copyright infringement just because the data training was unauthorized. This exception applies unless the AI system itself is primarily designed as a*

device or tool for creating infringing content, or there is an infringement of software copyright in relation to the AI itself."

The Central American and Caribbean Group adds that what matters is the presence of characteristics elements in the output: *"The output must contain characteristic elements of one or more Copyrighted Data in order to be considered a copyright infringement."*

B. Infringer

23) Who should be liable in case of Copyright infringement by the outputs?

There is a consensus among the Groups that in case of copyright infringement by the output, the following persons/entities should be held liable:

- The Provider of the AI System;
- The user who exploits commercially the AI System;
- The final user if she/he is of bad faith.

a. The Provider of an AI System?

- The vast majority of the Groups (30 Groups - 90%) consider that the Provider of the AI System, i.e., the person/entity that develops the AI System and places it on the market, should be held liable if outputs infringe copyright.

  The Belgian Group considers that the liability regime should be comprehensive, and adds that: *"the AI System and/or Model provider might be expected to put in place the appropriate technological measures that can be expected from a reasonably diligent provider in its situation in order to counter credibly and effectively Copyright infringements".*

  The Hungarian Group justifies such liability as follows: *"Here the same logic could be followed as that of the private copy remuneration (manufacturer and/or importer of equipment and media is liable for the compensation that is caused by private copying made by end user of such equipment and media). Those persons should be held liable who made it possible to produce*

and use "products" that are suitable to substitute the Copyrighted Data that was used to train the system."

- Only 4 groups (10%) consider that the Provider of an AI System should NOT be held responsible for the outputs produced by its system.

  Nevertheless, these groups often cite specific circumstances in which the Provider of AI System should be liable.

  For instance, the Japanese Group upholds that: *"The Provider of an AI System should not be considered as the doer of a copyright infringement under ordinary circumstances. However, such a provider could be the doer of a copyright infringement, if their AI system generates infringing items with high frequency, or if they develop and/or provide an AI system but do not take any measures to prevent the generation of items that are similar to an existing copyrighted work, in spite of their knowledge that their AI system generates such similar items with high probability."*

b. The user who exploits commercially the AI System?

- 30 groups (95%) answer YES to this question, agreeing that a user who exploits commercially the AI System should be found liable in case the outputs infringe Copyright.

  The Danish Group draws a distinction: *"It should depend on how the AI system is used. Instinctually, there is a difference between using AI systems to replace creative creation in creative professions and as a research tool in non-creative professions. The use of AI in creative fields to replace human creativity should make the user liable if the generated output constitutes an infringement. In other instances, it probably should not. The present authors find – based primarily on an instinctual sense of justice – that it, in most cases, is more appropriate to find users of AI systems liable for sharing the reproduction instead of simply for creating it."*

The Paraguayan Group responds that in cases of conscious use of infringed and copyrighted works, the users who exploit commercially the AI System should be held liable *"especially if they used infringing data knowingly or failed to prevent foreseeable infringement."*

- Only 2 groups (5%) consider that the user should NOT be held liable in this case (see the Egyptian and Polish reports).

c. The final user? Even if acting in good faith or unaware of the infringement?

Whether the answer to this question be YES or NO, the groups seem to agree that if the final user is in good faith, i.e., that he or she is not aware that the output infringes Copyright, he or she should therefore not be held liable.

On the contrary, if the final user uses an AI System in order to knowingly create infringing outputs, he/she should be liable.

- 24 groups (75%) answer YES to this question and believe that the final user should be held liable in case of Copyright infringement by the outputs.

- The French Group precises that the final user should be liable in case of Copyright infringement by the output: *"since good faith is ineffective in matters of infringement. This position is unanimously shared by the industrialists interviewed."*

- Some Groups nuance their answers by considering that a final user should be liable, but only in certain circumstances.

  For example, the Australian Group considers that the final user should be liable "*but not if the final user uses the system in good faith and is not aware of the Copyrighted Data. This is because infringement only occurs if the final user, by using the AI System to create the output, has 'copied' the Copyrighted Data. If the final user is acting in good faith and is not aware of the Copyrighted*

Data, then it cannot be said to have 'copied' the Copyrighted Data, and so would not have done an infringing act."

The Brazilian Group considers that: *"Depending on the type of use, the final user can be liable for Copyright infringement. If they act as an applicator – thus an AI agent – they will be liable. On the other hand, if the final user is a natural person using the AI System exclusively for particular, non-economic purposes, they will not be liable* (…)."

- 8 Groups (25%) consider that the final user should NOT be held liable in case of Copyright infringement by the outputs, except in some circumstances, for example, when the user is of bad faith.

  The Mexican Group states that: *"Unless the final user knowingly uses the AI System with the intention of infringing Copyrights (e.g. in a situation similar to that of the ULTRAMAN case decided in China, where the user knowingly prompted the generation of an image in the shape of this well-known character)."*

  The Italian Group similarly holds that the final user should not be liable: "*unless the end user has maliciously solicited the AI system to create content that infringes the rights of third parties. In principle, the position of the end user appears similar to that of the net user who, thanks to search engines, accesses content that may have been illegally published on the net. However, whereas the net surfer accesses content that already exists and has been made available by third parties, the user of a generative AI system solicits the system to create 'new' content. It should therefore be considered that when he knowingly solicits the machine to generate content that infringes the rights of third parties, the creation of such content is attributable to him and he can be held liable for it".*

d. Any other person?

- A slim majority (21 groups - 60%) consider that other persons may be found liable in case of infringing output, with a majority citing third-party platforms.

For instance, the Mexican Group notably cites: "*Hosting platforms (e.g., AI-generated content sites) and search engines if they fail to take down infringing content after having gained knowledge of the infringing. Third-party AI trainers who commercialize training datasets might also be liable if they knowingly use copyrighted works without authorization".*

- 13 Groups (40%) do NOT believe that other persons may be held responsible in this case.

C. Sanctions

24) What sanctions should be imposed if it is found that Copyright has been infringed in order to train the AI System, because Copyrighted Data has been used?

A majority of groups underline that, in regard to sanctions, all sanctions should be available and should be imposed on a case-by-case application, all whilst being proportionate.

a. Injunction, destruction, etc. of Copyrighted Data used to train AI System still present in the data set or in the AI System?

- All groups (34 Groups - 100%) answer YES to this question.

- For instance, the Mexican Group is in favour of the destruction of such Data: *"Removal of copyrighted works from AI training data sets. This could include deleting infringing training data and requiring AI providers (and their users) to stop using certain datasets, or the infringing portions thereof".*

- However, many groups underline (i) the technical difficulty, in some circumstances, of putting in place such sanction and (ii) that such sanction should be imposed only if it is proportionate (see for instance Singapore, Chinese, Belgium, Danish, US, French and UK reports).

If the destruction of infringing Copyrighted Data is not technically possible or is not proportionate, then several groups uphold that alternative measures should be considered, such as financial compensation.

For instance, the Chinese Group states that: *"the implementation of injunctions, destruction, and other measures should be proportionate to the extent of the infringement consequences, and the difficulty and feasibility of technical implementation should be considered. In specific cases, damages or other remedies may be considered as alternatives to ceasing infringement".*

The Swedish Group also considers that: *"the availability of sanctions should reflect both the realities of current technology and the need to ensure a balanced and predictable legal framework for both rightsholders and AI developers. While the group reiterates the importance of enforcing copyright and ensuring that rightsholders can seek redress when their works have been used unlawfully, it underlines that any sanctions must adhere to the principle of proportionality.*
The Swedish Group adds that: *"Where infringing copyrighted data is still identifiably present in a dataset, the group considers that an order requiring the removal of that specific content may be justified, provided it is technically and practically possible to do so without disrupting the integrity of the system or dataset as a whole. However, the group notes that once training has occurred, it is often no longer possible to isolate or extract individual pieces of training data or to trace their influence within the trained model. In such cases, the destruction of the training dataset or the AI system itself would not only be disproportionate but may also be ineffective from an enforcement perspective.*
The Swedish group further observes that, in many instances, the inclusion of infringing material may represent only a small fraction of an otherwise lawful and diverse dataset. In this context, targeted remedies requiring the deletion of specific content from datasets that remain in use for future training or fine-tuning may offer a more appropriate balance between enforcement and innovation. Where removal is not possible or meaningful (e.g. after

model training has been completed), the group considers that other sanctions, such as monetary compensation, or the implementation of compliance measures for future conduct, may be more proportionate and effective."

Similarly, the US Group responds that: *"Remedies should be thoughtfully and carefully crafted to protect and compensate copyright owners for infringement, while also balancing the need for technological development and innovation. If the infringed inputs can be extracted and removed from the model, and/or filtered or excluded from outputs to prevent unauthorized use, this is preferred. However, if this is not possible, an injunction against future training use, and financial compensation for past unauthorized use under the prevailing statutory framework, would be appropriate".*

The French Group also points out that the destruction of infringing Copyrighted Data can be imposed only *"if the data is still used and/or identifiable in the system".*

The Danish Groups expresses a similar opinion: *"In model-as-container systems, where training data is statistically encoded into the model's parameters, it is practically impossible to isolate or remove individual works post-training. While courts should retain the option in theory, it is highly unlikely to be applied in practice, given the technical and proportionality constraints. That said, future technological developments, such as machine unlearning techniques, may eventually allow for the targeted removal of specific training data from AI models. If and when such tools become viable, injunctive relief or targeted "data removal" orders may become more practical and proportionate enforcement mechanisms".*

- On the contrary, the Hungarian Group considers that in case infringing data cannot be separated, the sanction should apply to the entire system.

b. Injunction, destruction, etc. of the AI System itself deemed to be infringing, because it has been trained with infringed Copyrighted Data?

The groups are very divided on this key issue.
Nevertheless, the majority of groups believe that the sanction of destroying/prohibiting the AI system itself should only be imposed in exceptional circumstances, in cases of bad faith, such as if the AI System was designed to be trained with infringing Copyrighted Data or to create copyright infringement, or is primarily used for systematically or massively uses copyrighted works as training data.

- Half of the groups (17 Groups - 50%) answer YES to this question, considering that the injunction/destruction of the AI System itself trained with infringing Copyrighted Data should be available.

  But many Groups that answer yes to this question consider that the destruction/injunction of the AI System itself should not be automatic and should be subject to the proportionality principle.

  Most cited criteria to determine if such sanction is proportionate or not are:

  - The circumstances of the train of the AI System, e.g. good or bad faith, awareness of the infringement, sufficient diligence taken to avoid infringement, AI developed to infringe copyright, etc.

  - The gravity of the infringement, e.g. amount/proportion/value of infringing data used.

  - The possibility to redesign the AI System.

  For instance, the Belgian Group considers that injunction or destruction of the AI System itself should be imposed, but "*upon the condition that this is subject to a proportionality test. In this group's opinion, a proportionality test should be included to ensure that the sanction is not disproportionate. This should be assessed on a case-by-case basis, allowing the court to consider*

all relevant circumstances (e.g., the amount of Copyrighted Data used, the purpose of the AI System, the scope of infringement, the value of the Copyrighted Data infringed, etc.)".

The Italian Group considers that such sanction should be imposed *"if such measures appear proportionate in the case at hand. In particular, care should be taken to distinguish those cases in which the training has been carried out in a sufficiently diligent manner from the point of view of respecting the rights of third parties to the content used as input from those in which, conversely, the training has been carried out systematically in disregard of those rights of others. If the violation of input rights appears to be occasional and unconscious, the injunction or destruction of the resulting AI system does not appear proportionate to the gravity of the conduct and should therefore be denied. If, on the other hand, the infringement appears conscious and systematic, the protection of input rights, as well as the protection of legitimate competitors, seems to us to justify the application of these remedies".*

- The Turkish report indicates that: *"AI Systems should be sanctioned as last resort in case there are no other proportionate remedies to cease infringement. A more proportionate remedy would be injunction, destruction, etc. of Copyrighted Data itself."* (see also the Chilean report).

- The Canadian Group responds: *"If the AI system can be redesigned to remove or otherwise avoid reproduction of the Copyrighted Data, then the System itself ought not to be enjoined, destroyed, etc. However, if the AI System cannot be re-trained or must rely on the prior infringement, then an injunction or destruction of the AI system is an appropriate sanction."* (see also UK and Dutch reports).

- Half of the Groups (17 Groups - 50%) answer NO to this question, and believe that the destruction/prohibition of the AI System itself should not be an available sanction, mostly because it would be a disproportionate sanction.

- For instance, the Mexican Group indicates that: *"An injunction ordering the AI System to cease operating would be disproportionate, could affect millions of users, or cause an adverse impact on society"* (see also French, Australian, Chinese, Singapore, Luxembourg reports).

  The Swedish Group emphasizes *"the importance of proportionality and legal certainty in determining appropriate remedies. Sanctions such as damages—particularly where there is a finding of negligence or intent—may provide a more effective and balanced means of addressing harm, especially in cases where only a portion of the training data is found to be infringing".*

- However, many Groups that answer NO to this question, indicate that in exceptional circumstances, e.g. if the AI System has been designed for the primary purpose of infringement, the prohibition/destruction of the AI System itself should be available.

  For instance, the Chinese Group considers that: *"For general-purpose large models, direct destruction may result in disproportionate consequences, unless the AI system is specifically developed for infringement or constitutes a tool primarily used for producing infringing copies"* (see also the US, Korean, Egyptian, Paraguay and German reports).

c. Injunction, destruction, recall from commercial channels, etc. of outputs found to be infringing?

- 32 Groups (95%) answer YES to this question.

  The Japanese Group considers that: *"If an AI-generated content constitutes an infringement of the copyright, a request for an injunction against the use of the said content or the destruction thereof should be allowed (a request for recall of content from commercial channels should also be available in those countries where it is permitted by their copyright law)."*

  The Belgian Group proposes to distinguish the two following scenarios: *"The output itself is infringing, then a proportionality test*

should be performed. The recall from commercial channels may sometimes be impractical or impossible or undesirable. - The output is infringing but Copyright protection could be claimed on said output. If the output is solely created by an AI System, it is unlikely to be protected by Copyright. However, if a natural person is deemed to have Copyrights on the output, the destruction of the output might infringe their Moral Rights."

- 2 Groups (5%) (Brazilian and Polish groups) answer NO to this question.

d. Award of damages, including punitive damages?

- All Groups (32 Groups - 100%) consider that damages, including punitive damages, should be an available sanction to compensate for material and/or moral prejudice caused by the Copyright infringement and the bypassing of Copyright holders' consent.

  Some Groups agree to the award of damages but NOT of punitive damages (see for instance German report), or only in specific circumstances. For instance, the Canadian Group states that: "*Punitive damages ought to be able in cases with particularly aggravating factors, such as malice, fraud, etc.*"

e. Confiscation of all or part of the profits generated by the operation of the AI System?

- The majority of the Groups (26 Groups - 75%) believe that confiscation of the profits generated should take place.

- The Finnish Group supports the confiscation of profits as damages: "*confiscation of assets to the state is not desirable. However, profits generated by the operation of the AI System should be available as damages (loss of profit) to the Copyright holder. In copyright offence cases, it should also be possible to confiscate financial benefit gained by the offence".*

- But many Groups consider that the confiscation of profits should be proportionate to the quantity of infringing data used:

  - The French Group suggests, for instance: *"the confiscation of the profit should be proportional to the quantity of infringing data used in relation with the total training date used."*

  - The German Group agrees that confiscation should take place: "*but only the part of the profits which was generated as a result of the infringement can be confiscated*."

  - The Italian Group expresses concerns over the feasibility of such a sanction: *"If by "confiscation" is meant the restitution of the infringer's profits, the answer may be positive, although with the caveat that it will probably not be easy in practice to reconstruct which part of the infringer's profits depends on the infringement of the rights of an individual input."*

- 8 Groups (24%) oppose the confiscation of profits.

  For instance, the Mexican Group asserts that: "*Damages should be awarded by authorities and courts, but confiscation of profits is an extreme measure. It is important to maintain a balance between Copyright protection, innovation, and the public service that AI Systems provide to communities*."

f. Any other sanctions?

- 23 Groups (74%) suggest various other traditional sanctions in case of IP infringement, such as publication (see for instance Swiss, Hungarian, Chilean and Italian reports), profit sharing, award of legal fees (Dutch and Canadian reports), fine on non-compliance with court's decision (Dutch and Chilean reports), information injunctions on the origin and use of training data (Swedish Group), etc.

  - For instance, the Italian Group suggests: "*Publication, either by traditional means (e.g. in newspapers) or online, of the judgement finding the infringement can help achieve an*

adequate level of deterrence as it may motivate other rights holders to act (and correspondingly discourage infringement)" (see also Chilean report).

- The Chinese Group proposes as a sanction: *"in specific cases, mandatory modifications to the AI algorithm (such as adjusting parameters or weights)".*

- The French Group adds: *"In the event that the AI model has been trained illegally and that it is not possible to withdraw the training works or destroy them, provision should be made for a profit-sharing key and a percentage on future profits, along the lines of what can be put in place for composite works. An injunction could also include an obligation to ignore protected works and to neutralize infringing outputs of protected works used to drive the system".*

- The US Group advocates harsher sanctions under judicial supervision: *"Courts may impose additional sanctions in extreme cases, particularly in instances of repeated infringement or willful disregard of copyright law"*.

  In the same vein, the Brazilian and German Groups consider that criminal sanctions should also be available.

- 8 Groups (26%) do NOT consider that other sanctions are necessary.

D. In which other situations should outputs be qualified as Copyright infringement?

25) Please explain, in the following cases, if and under what conditions outputs generated by an AI system should be qualified as infringement of a Copyrighted work, e.g. because the output contains characteristic elements of the Copyrighted work, or is in the same style as the Copyrighted Work, etc.:

a. In case the use of the Copyrighted work to train the AI System is covered by an Exception.

- 22 Groups (65%) consider that outputs generated by an AI System and containing characteristic elements of a copyrighted work used to train the AI should be qualified as infringement, even if the training was covered by an Exception (see, for instance, the Japanese, US, German, Canadian, French, Dutch, Australian, Spanish, Finnish, UK and Chinese reports).

  The majority opinion is summarized by the German Group: *"if the output contains recognizable characteristic elements of the Copyrighted work which reflect that author's own intellectual creation, thus if the AI-generated content would, if it were created by a human being, qualify as a copyright infringement. There is no reason to privilege a content for the mere reason that it has been generated by an AI system. Hence, the output should be assessed in accordance to the already applicable principles."*

  The French Group also considers that outputs should be qualified as Copyright infringement "*if outputs reproduce in whole or in part the characteristic elements of one or more prior works, provided that the element or elements reproduced are original and reflect the personality of the author of the prior work. In the case of the 'same style', the French group believes that the action to be taken should be a tort action for unfair competition and/or parasitism".*

  For example, the Danish Group justifies it as follows: "*Simply because an exception allowing for the training of AI Systems should exists, this should not give free reign to use the AI system to create infringing output.*"

  The UK Group adds: *"An Exception that permits one or more acts at the point of Training should not mean that further acts of infringement at the point of deployment should be covered by the same Exception automatically – those are separate acts of infringement that must fall within their own Exception."*

- 12 Groups (35%) consider that outputs should NOT be qualified as infringement in case the use of Copyrighted work to train AI System is covered by an Exception (see, for instance, Turkish, Indian, Italian, Mexican, Polish, Korean reports).

For instance, the Italian Group responds that: *"Exceptions and limitations apply 'in certain special cases'. If, therefore, an exception refers to the use of protected content for training purposes, there seems to be no reason to extend its application (if it makes sense) to the output generated by the system once trained. This is without prejudice to the possibility that the exception, by reason of its content and purpose, may apply to the output as such. Take, for example, the exception for research purposes: this may justify both the use of a protected content as input and the use of an output generated by the trained system."*

b. In case the Copyrighted work has NOT been used to train the AI System.

- 16 Groups (45%) consider that outputs generated by an AI System should be qualified as infringement of Copyright, even if this copyrighted work has not been used in the training process (see, for instance, Chinese, Canadian, Spanish, Swedish, Australian, Italian, French, Dutch, Mexican, Finnish and Polish reports).

- Nevertheless, two Groups bring up the chance encounter defense.
  For example, the French Group responds: "*a chance encounter could be invoked as a defence, for example if the prior work is not accessible.*"

  The Italian Group also cites the chance encounter, but debates its applicability: *"The situation described in the question recalls the traditional problem of chance encounters, i.e. the autonomous creation by different authors of overlapping works. However, it does not seem to us that there is full analogy between chance encounters due to the hand of human authors and respectively caused by the operation of generative AI systems. If in fact the chance encounter is an exceptional case when the creation takes place by the hand of man and in an effectively autonomous manner (i.e. without having access to the prior work of others), the same cannot be*

said when the creation is the result of the operation of an artificial intelligence system, which is oriented by its nature not to autonomously create something new but to recreate something that approximates what has been ingested in the training or a summation of elements found in the different inputs."

- The Chinese Group points out the difficulty of proving that the AI System has not been trained with (or without) copyrighted works: *"if the output meets the conditions of "access & substantial similarity", the court can determine that the content constitutes copyright infringement. To rebut the "access" element, the provider of an AI system needs to prove that copyrighted works were not used for training. However, due to the algorithmic black box, it is relatively difficult to provide evidence in practice."*

- 18 Groups (55%) consider that outputs generated by an AI System should be NOT be qualified as infringement of a copyrighted work in case this work has not been used in the training process (see, for instance, the US, German, Brazilian, Japanese, Indian, Austrian, Korean, Croatian, Danish and Turkish reports).

  The Brazilian Group explains that: *"Since (…) the material was not used as a database for training the AI, the system cannot have knowledge of the specific elements of the copyrighted work to infringe upon them. At most, the output may contain similar elements that happen to converge by coincidence, which cannot be classified as infringement. Practically speaking, it would be impossible to demonstrate a causal link leading to an infringement."*

  The Latvian Group adds that output *"not used commercially and if not infringes moral rights"* cannot be qualified as infringing if the copyrighted work has not been used for training the AI System.

c. In case the Use of the Copyrighted Data to Train AI System has been authorized, is the content generated (output) always licit, even if for instance some outputs contain characteristic elements of the Copyrighted work?

Regardless of the answer YES or NO, the majority of Groups seem to consider that in this case, the infringing qualification (or not) of the output should depend on the scope of the authorization. It should be taken into consideration if the authorization is limited to the training of the AI System, or if it is extended to the outputs generated.

- Only 7 Groups (20%) answer YES, considering that outputs should be qualified as infringing, even if the use of the Copyrighted Data in the training process was authorized by the Copyright holder.

  The Australian Group nuances its answer by agreeing to this statement, "*Unless the consent in fact extended to outputs.*"

  The Bulgarian Group also responds that "*It depends to the authorization given. If consent has been given no infringement should be considered.*"

  The Mexican Group responds that: *"even in the case of authorized training data, infringement may still occur if an output contains characteristic elements of the Copyrighted work."*

- 25 Groups (80%) answer NO, considering that outputs cannot be infringing when the consent of the Copyright holder to use his/her works has been given for training the AI System.

- But a majority of Groups that answer NO to this question agree that this case merits a nuanced approach and depends on the scope of the copyright holder's authorization.

  For example, the Belgian Group responds that: *"such a situation deserves a nuanced approach. The question as to whether the output generated by an AI system qualifies as a Copyright*

infringement in such case depends on the scope of the authorisation given in the first place. The authorisation given (upstream) for reproductions occurring in the training of an AI System should not necessarily legitimate any further reproduction or communication to the public (downstream).The authorization for training might pertain to the creation of the AI model, not the subsequent exploitation of its outputs, especially if those outputs infringe on existing Copyrights".

The Luxembourgian Group states that: "*even if the use of Copyright Data to Train an AI System has been authorized, the content generated by the AI System should not always be deemed licit. If the output contains substantial parts or characteristic elements of the Copyright Data, it may still be considered an infringement. The answer will depend on the scope of authorization/objection given by the Copyright holder, as the authorization for training does not extend to the reproduction of protected elements in the outputs and therefore harmonization and uniformization of the authorization/objection process is crucial in order to guarantee legal certainty.*"

- Some Groups (minority) were more decisive in their response.

26) Who should be liable in case of Copyright infringement by the outputs?

a. The Provider of an AI System?

- 29 Groups (90%) consider that the Provider of an AI System should be liable in case the output infringes copyright.

The US Group responds: "*the Provider of an AI system may be liable, directly or secondarily, depending on the facts of the case. A Provider of an AI System used by a third-party user to create an infringing work should be liable under prevailing theories of secondary liability, such as contributing infringement and vicarious infringement, absent an applicable exception or defense, such as fair use.*"

- 4 Groups (10%) (Australian, Swedish, Japanese and Russian Groups) consider that the Provider should NOT be held liable in this case.

  The Swedish Group emphasizes the role of the user: "*the Swedish group takes the position that, as a general rule, the Provider of an AI System should not be held liable for copyright infringement committed by users of the system. Analogously, the manufacturer of a computer, a network card, or an electronic storage device is not liable for how those devices or components are used by the user.*"

b. The user who exploits commercially the AI System?

- 31 Groups (all groups except one) maintain that a user who commercially exploits an AI System should be held liable when the outputs infringe Copyright.

- Only one Group considers that the user who commercially exploits the AI System cannot be held liable in this case (see Polish report).

c. The final user? Even if acting in good faith or unaware of the infringement?

- 24 Groups (75%) consider that the final user should be responsible for infringing output, if acting in bad faith.

  The Chinese Group points to the final user being conscious of the infringing nature of the output: "*Generally, the final user does not bear the liability for infringement of the output. However, if the user deliberately induces the AI system to output infringing content, or if the user is aware that the output is infringing but still takes actions such as publicly disseminating it to expand the impact of the infringement, etc., the user should bear the corresponding liability for infringement.*"

- 9 Groups (30%) do NOT think the final user should be held liable.

For instance, the Argentinian Group considers that the final user should not be held responsible unless he or she was given notice of the infringement.

In the same vein, the Latvian Group adds *"except if he/she knew about infringement or should knew or infringes moral rights of the author."*

d. Any other person?

- 20 Groups (65%) answer YES, with a majority of Groups considering that those persons directly involved in the distribution of the infringing output should be held liable.

  For example, the Finnish Group answers that: "*Any other party that causes a Copyright infringement, such as a distributor distributing an infringing output generated by the AI System, should also be liable.*"

  The Luxembourgian Group responds that: "*Any other person involved in the distribution or commercialization of the infringing outputs should also be held liable for Copyright infringement.*"

- 12 Groups (40%) do NOT think any other person should be held liable for infringing output.

27) Please comment on any additional issues concerning any aspect of the Use of Copyrighted Data to Train AI System you consider relevant to this Study Question.

24 Groups provide a response to this question.

The Japanese Group suggests: "*In addition to the harmonization of legal provisions related to exceptions and outputs mentioned above, the following issues may be considered: (…) (2) International standardization of data licensing. Develop an international licensing framework (e.g., a standard contract clause) that would allow AI developers and right holders to clarify the terms of use. (…) (5) Establishment of revenue-sharing model. Discuss a mechanism for distribution of revenue from AI-generated contents among the*

right holders (e.g., distribution through collecting societies). (6) Formulation of guidelines under the initiative od international organizations. Facilitate consensus-building internationally through a set of guidelines for harmonization on national laws, formulated by an international organization such as the WIPO. (7) Enhancement of copyright protection by the use of technological measures. Build a mechanism for preventing an AI system from using a copyrighted work for learning without permission by making it obligatory to include copyright information in metadata. (8) Balance between R&D and regulations. Draw a clear line between the use of works for research purposes (for which the use in AI learning is generally permitted) and the use for commercial purposes. (9) Promotion of use of works in the public domain. Increase the amount of lawful resources for AI learning by providing support to the creation of datasets that contain those works whose copyright protection period has expired. (10) Protection of cultural diversity. Encourage the use of diverse works to ensure that AI learning data sources are not concentrated in certain regions and cultures. Consideration should be given to the harmonization of benefit sharing, especially when traditional knowledge (TK) and traditional cultural expressions (TCEs), discussed at WIPO's IGC50, are used as learning data."

The Mexican Group responds that: "*Technological solutions that allow for the scalable identification and marking of contents, automated collection of royalties, authorization/prohibition to use, and traceability of works during their entire lifecycle can be an important complement to legal and policy developments. Universal, fair and reasonable blanket licenses might prevent conflicts and introduce certainty to the market. Automated opt-out mechanisms embedded in the works themselves will empower rights holders to freely determine how their works are to be used, or whether they should be used in relation to AI System at all.*"

28) Please indicate which industry sector views provided by in-house counsel are included in your Group's answers to Part III.

29 Groups provide responses to this question, with the following industry sectors being mentioned:

- Publishing sector
- Household appliance manufacturing sector
- Audiovisual sector
- Food and beverages sector

- Legal AI industry
- Film and music industry
- Newspaper publisher
- Energy industrialist
- Cosmetic industrialist
- Luxury industrialist
- IT consulting, engineering and digital transformation services
- Computer software, AI and data licensing sectors
- Manufacturing industry
- Television broadcasting industry

IV. Conclusions

- A clear majority of the responding Groups consider that harmonization regarding the use of Copyrighted Data for training AI is desirable due to the cross-border nature of AI systems and the risks of forum shopping by AI developers.

I. Authorization to use copyrighted works to train AI systems

- The majority of Groups consider that prior authorization from the copyright holder should be, as a matter of principle, necessary before the copyrighted work may be used to train an AI system, unless covered by an exception.

- The majority of Groups agree that general exceptions to copyright, such as the "fair use" and "temporary reproduction" exceptions, should apply to the training of AI Systems with copyrighted works, if the conditions to benefit from such an exception are met.

  Furthermore, the majority of Groups propose to introduce a special exception (TDM) to allow the use of copyrighted works to train AI Systems without consent from the copyright holder. Such a TDM exception should be limited to situations of public interest, such as for purposes of scientific research, education, information or not-for-profit.

  The slight majority of Groups are in favor of an opt-out system, giving the right to copyright holders to refuse the use of their works for training of AI System.

The majority of Groups also consider that exceptions should be provided for financial compensation for the copyright holder whose works have been used to train AI and should comply with the three-step test provided by Article 9(2) of the Berne Convention.

- The majority of Groups believe that there should be transparency obligations from the side of the AI developers providing sufficient details on works used to train the AI system, to make it possible for the Copyright holders to enforce their rights.

II. Consequences of the unauthorized use of copyrighted works to train AI System

a. Infringing items

- There is a strong consensus that an output that contains characteristic/substantial elements of a copyrighted work used to train AI, should be considered copyright infringement.

- In case use of a copyrighted work to train the AI System is covered by an exception or authorized by the Copyright holder, in order to determine if the outputs that reproduce characteristic/substantial elements of the copyrighted work are infringing or not, one should take into consideration the scope of the exception or authorization.
  If it is limited to the training of the AI System, the output should be qualified as infringement.
  If it covers also outputs created by the AI System, the output should not be qualified as infringement.

- An output that is in the same style as a copyrighted work used to train AI System, should not constitute copyright infringement.

- The majority of the groups consider that the AI System itself should be considered as copyright infringement in certain specific circumstances, like, for example, when the AI system:
  - is designed to create Copyright infringements; or
  - systematically or massively uses copyrighted works to be trained.

b. Infringer

There is a consensus among the Groups that in case of copyright infringement by the output, should be held liable:

- The provider of an AI System;
- The user who exploits commercially the AI System;
- The final user only if she/he is of bad faith, i.e. is aware that the output infringes copyright.

c. Sanctions

- Subject to the proportionality principle, there is a consensus among the groups that all traditional sanctions should be available (injunction, destruction, recall from commercial channels, etc.), such as:

1. Destruction or injunction of outputs found to be infringing.

2. Destruction or injunction of the AI System itself, but only under specific conditions.
   These conditions should include:

   - the circumstances in which the AI System was trained (e.g., good or bath faith, awareness of the infringement, sufficient diligence taken to avoid infringement, AI developed to infringe copyright);

   - the gravity of the infringement (e.g. amount/proportion/value of the infringing data used); and

   - the potential redesigning of the AI System, should determine what type of sanction is proportionate.

3. Destruction of copyrighted work that has been unlawfully used to train an AI System and the prohibition of its further use for training purposes. In case such destruction is not possible, alternative measures like financial compensation should be imposed.

4. Provision of damages and confiscation of all or part of the profits generated by the AI System.