I. Current law and practice

1) Does your current law / practice contain laws, rules, regulations or case law decisions specifically relating to the Use of Copyrighted Data to Train AI System?
Please answer YES or NO. Please explain.

No. Kind of.

Through the CDSM Directive of the EU, Denmark has implemented rules on data mining. These rules are implemented in the Danish Copyright Act, so they undoubtedly regulate, at least certain aspects of copyright and data mining. The Directive was implemented in 2019, and neither the Directive nor the Danish law nor its preparatory works make any specific mention of AI.

However, section 11 b (1) implements a definition of data mining that refers to »*any automated analytical method aimed at analysing text and data in digital form in order to generate information, including patterns, trends, and correlations.*«

Though such phrases as "generative AI" were not part of the common vocabulary in 2019, it is easy to see how these rules apply to data mining done to train an AI system. Previously it was referred to as a generative neural network.

Though Section 11(b) applies to data mining, including for AI, it is not clear whether generative AI models per se are permitted under this provision. Section 11(b) defines data mining as an automated analytical method used to generate »*information, including patterns, trends, and correlations*.« However, generative AI models do more than extract insights – they generate new content, such as text, images, or code, based on the mined data. It is uncertain whether data mining for such purposes falls within the scope of what is allowed under the data mining exception. It is uncertain whether such content creation falls within the scope of the provision, as the law does not explicitly address whether transforming analysed data into new outputs remains protected under the data mining exemption. This ambiguity leaves room for legal interpretation, particularly regarding whether courts will adopt a narrow or broad reading of what constitutes permissible data mining.

Does the Use of Copyrighted Data to Train AI System have to be authorised by the Copyright holder?

2) Does the Use of Copyrighted Data to Train AI System fall within the scope of the Economic Rights monopoly, i.e. does it require the authorisation of the Copyright holder unless covered by an Exception? Please answer YES or NO. Please explain, e.g. if the authorisation is necessary only if acts of reproduction, representation, etc. are carried out, or in any circumstances (even if no act of reproduction, representation, communication, etc. has been carried out).

YES.

Training of AI systems entails certain processes that would likely – but not necessarily - include reproductions of Copyrighted works, and reproduction falls within the Economic Right Monopoly of the copyright holder.

One must keep in mind that though data processing for AI systems does not entail creating copies that are recognisable to humans, they may still constitute infringing reproductions under copyright law. This is supported by the fact that computer programs are protected by copyright, cf. Section 1(3) of the Danish Copyright Act. The concept of a program includes source code and object code. Object code is also not understandable for humans, but it is nonetheless protected under copyright law.

The principle likely means that though the dataset will not be understandable to humans once it has been processed and encoded for it use in training AI systems, it will still be the same protected work. This is further substantiated by section 2(2) of the Danish Copyright Act, where reproduction is defined to also include »the act of transferring the work onto devices that can reproduce it«, i.e. digital storage of works.

The steps that may entail the reproduction of works include collection, storing, and processing of datasets. This requires (maybe temporary) copies of the data and, if that data is protected under copyright law such reproductions will fall within the exclusive rights of the copyright holder. Copying training data, even temporarily, could qualify as a reproduction under Danish law.

These steps usually also include a preprocessing step, where the data is normalised, tokenised, and/or transformed. Preprocessing steps such as normalising, tokenising, or transforming data may or may not constitute reproductions, depending on whether they preserve protectable elements of the original works. If preprocessing results in a modified version that retains expressive elements of the copyrighted content, it could qualify as a reproduction under Danish law.

Finally, encoding the data for use in AI models may entail actions that may constitute infringing reproductions. Though AI models do not store direct copies of the data, the encoded statistical representations may still retain elements of the training data. If courts determine that these representations preserve copyright-protected features of the original works, encoding could be considered a reproduction under Danish law. If a model can reconstruct or generate outputs that resemble its training data, courts may be convinced that it still contains a form of reproduction. If courts interpret encoding as a form of reproduction, AI models could be reproducing copyrighted content.

This would prima facie seem to mean that the use of copyrighted data to train AI systems requires permission from the copyright holder or the applicability of an exception.

3) Are there any Exceptions that authorise 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.)? Please answer YES or NO. Please explain, e.g. the conditions of application of each Exception separately.

NO.

It should be mentioned that the applicability of the below-mentioned exceptions is untested by the Danish courts and is thus – for the time being – theoretical.

The Danish Copyright has implemented Article 5(1) of the EU's Infosoc directive (Directive 2001/29/EC) which allows for Temporary acts of reproduction, »*which are transient or incidental, which are an integral and essential part of a technological process and the sole purpose of which is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance*«.

There is some speculation as to whether this exception should allow for using Copyrighted data to Train AI Systems. The present authors do not believe that this exception would apply in such instances. The reproductions would likely not be »*transient*« if they are stored by the supplier of the AI system for a period of time that is longer than what is strictly necessary.

Furthermore, Danish law has implemented the data mining rules found in Articles 3 and 4 of the CDSM Directive. These have been implemented in the Danish Copyright Act in Sections 11 b and 11 c.

Section 11 b (1) implements a definition of data mining that refers to refers to »*any automated analytical method aimed at analysing text and data in digital form in order to generate information, including patterns, trends, and correlations.*«

Though such phrases as "generative AI" were not part of the common vocabulary in 2019, it is easy to see how these rules apply to data mining done to train an AI system. Previously it was referred to as a generative neural network.

Though Section 11(b) applies to data mining, including for AI, it is not clear whether generative AI models per se are permitted under this provision. Section 11(b) defines data mining as an automated analytical method used to generate »*information, including patterns, trends, and correlations*.« However, generative AI models do more than extract insights – they generate new content, such as text, images, or code, based on the mined data. It is uncertain whether data mining for such purposes falls within the scope of what is allowed under the data mining exception. It is uncertain whether such content creation falls within the scope of the provision, as the law does not explicitly address whether transforming analysed data into new outputs remains protected under the data mining exemption. This ambiguity leaves room for legal interpretation, particularly regarding whether courts will adopt a narrow or broad reading of what constitutes permissible data mining.

b. Special Exceptions to the Use of Copyrighted Data to Train AI System (e.g. Text and Data Mining -TDM Exception, etc.)? Please answer YES or NO. Please explain e.g. the conditions of application of each Exception separately.

NO. Not specifically mentioning AI, but we do have rules targeting a more general issue of datamining.

There exist rules relating to data mining. These were implanted before AI models became what they are today, so the rules were not specifically aimed at AI Models. The data mining rules are, however, of particular significance specifically when assessing AI systems.

The Directive was implemented in 2019, and neither the Directive nor the Danish law nor its preparatory works make any specific mention of AI. However, section 11 b (1) implement a definition of data mining that refers to refers to »*any automated analytical method aimed at analysing text and data in digital form in order to generate information, including patterns, trends, and correlations.*«

Though such phases as "generative AI" were not part of the common vocabulary in 2019, it is easy to see how these rules apply to data mining done to train an AI system.

4) Do the Exceptions provide for financial compensation for Copyright holders (e.g. a royalty paid to a collecting society, etc.)? Please answer YES or NO. Please explain.

NO.

The mentioned exceptions do not provide for monetary compensation for copyright holder(s). Furthermore, the provisions in the Danish law regulating Collecting Societies (§ 50) do not refer to § 11B and C on datamining.

5) Can the author object to Use of Copyrighted Data to Train AI System on the basis of his/her Moral Rights (for example on the basis of the right to integrity, paternity, etc.).

YES (not on the basis of moral rights per se)

The laws on data mining provide the copyright holder the option/right to opt out of their works/data being data mined. The wording of the Section 11 b(2) of the Danish Copyright Act is the following:

> »A person who has legal access to a work may extract and make copies of the work for the purpose of text and data mining, provided that the right holder has not expressly reserved the use of the work in an appropriate manner, cf. however, section 11c(1)«

This possibility for the right holder to opt out is, however, only available when the data mining is not done for the purposes of »scientific research«. In such cases, the right holder may not object to their work being data mined.

6) Does the Provider of an AI System have to make public the training data used to train the AI System (transparency obligation)? Please answer YES or NO. If YES, please explain the degree of detail required (all works, categories according to sources - websites, etc.) and whether the developer / operator must keep the list of training data or the training data itself for a certain period of time?

NO (not yet).

According to the recently passed EU Regulation, (EU) 2024/1689 (the AI Act), providers of so-called general-purpose AI models will have an obligation to »*draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office*« was introduced, cf. Article 53(1)(d). According to Article 113, this provision and obligation enters into force on the 2nd of August 2025.

Recital 107 of the AI Act provides a degree of guidance in determining what should be published. Recital 107 indicates that it is adequate that providers of general-purpose AI models draw up and

make publicly available a sufficiently detailed summary, and that this summary should be generally comprehensive in its scope instead of technically detailed to facilitate parties with legitimate interests, including copyright holders, to exercise and enforce their rights under Union law, for example by listing the main data collections or sets that went into training the model, such as large private or public databases or data archives, and by providing a narrative explanation about other data sources used.

The »*template provided by the AI Office* « has not yet been published. On 19th of December 2024, the second draft of the *Code of Practice* was published. The second draft does not provide much further guidance than the provision itself. Presently, it is expected that the final version of the Code of Practice will be published in May 2025.

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

To answer questions 7 to 10, 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.

Infringing items

7) Can the contents created (output) by an 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? Please answer YES or NO. Please explain if needed.

YES.

This is untried in practice, but is assumed that the same principles of infringement that apply in "regular" copyright infringement cases apply to AI copyright cases: (1) Identity between the original work and the copy; (2) imitation/copy of the original; (3) the imitation/copy concerns an original/characteristic and thus protected element of the original work; and (4) the copied element must be more than simply an abstract idea.

In the present scenarios, if these conditions are met, the AI-generated output will probably constitute an infringement of the original works on which it was based.

   b. The output is in the same style as one or more Copyrighted Data used to train AI System? Please answer YES or NO. Please explain if needed.

NO.

Probably not. Under Danish law, it is generally not considered to be copyright infringement to copy the style of someone else. A style will in most cases be too abstract to be protected under copyright.

If, however, the copied style can be adequately specified so that it concerns specific, individual elements that are in and of themselves original, copying these stylistic elements may constitute copyright infringement.

The above-mentioned principles are presumed to apply to copies generated by AI.

c. The output is in all cases infringing if it has been trained with one or more infringed Copyrighted Data. Please answer YES or NO. Please explain if needed.

NO.

It is a prerequisite for copyright infringement that original, characteristic elements of the original/copied work can be found in the copy (in casu: the AI output). If this condition is not met, the copy will not constitute copyright infringement.

There is currently no reason to assume that this principle should not also apply to cases where the potentially infringing product is AI-generated.

d. In other cases? Please answer YES or NO. Please explain if needed.

NO.

8) Can the AI System itself be considered to infringe Copyright? Please answer YES or NO. If YES, please explain, e.g. under what conditions.

PROBABLY.

The present authors have identified two distinct situations which we shall refer to as the "model-as-container" and the "model-as-algorithm + external data cache" (or the retrieval model). In the model-as-container situation, the AI system itself stores encoded representations of the data on which it was trained within its algorithm. In the model-as-algorithm situation, the AI system itself is just an algorithm that does not in itself contain any (potentially protected data). It instead has access to the (potentially processed and optimised) data through which the algorithm can "search".

The model-as-container may be considered an infringement, as it stores/encodes within the algorithm itself parts of the training data. Such storage – even if encoded – of copyright-protected works within the AI system would likely make the AI system per se illegal, if the stored data has retained its original and protected characteristics; see 2 above.

On the other hand, if the AI algorithm itself does not contain any of the training data per se – model-as-algorithm – but instead the algorithm has access to protected data, then the algorithm itself likely does not infringe upon the right related to the data. It then becomes a question of definition, is the AI system the algorithm and the data or are they separate? If they are separate, then the mere existence of the algorithm (which is definitionally the same as the AI system) does not infringe upon the right to the data. If the AI system is instead defined as being the algorithm and the date, then the model-as-algorithm + may also constitute an infringement, as the data be storage – and thus reproduction – of protected material.

In practice, the most common model is the model-as-container, but some models have the functionality to search the internet, which is comparable to the model-as-algorithm. In this manner, one may not meaningfully be able to distinguish between the two.

If the AI system is able to reproduce, for example, specific quotes from a copyright-protected book, this would heavily imply that this protected book is somehow "stored" (encoded) within the AI's algorithm (or at least the algorithm is such that it can withdraw specific passages of protected work if needed). This would likely be an illegal reproduction of the original work as the copyright holder of

the original work holds the exclusive right to reproduction »*in any way or in any form. A reproduction is defined as any direct or indirect, temporary or permanent and in whole or in part reproduction in any way and in any form. The production of copies is also considered to be the fact that the work is transferred to devices that can reproduce it.*«

Infringer

9) Who is liable in case of Copyright infringement? The Provider of an AI System? Please answer YES or NO. Please explain if needed.
    a. The user who exploits commercially the AI System? Please answer YES or NO. Please explain if needed.

If the system itself constitutes an infringement, the provider and/or developer of that system will likely be liable for that infringement; see 8 above.

If an AI system generates infringing output it seems likely that the provider of the system will be liable (at least in part) for that infringing output.

    b. The final user? Even if acting in good faith or unaware of the infringement? Please answer YES or NO. Please explain if needed.

The final user cannot be held liable for infringement related to the training of the system. The final user may be liable if they make their own additional "custom dataset" to fine-tune an existing model. Such upload of data is a reproduction of the data if that data is protected under copyright law.

In terms of infringing output, the end user can likely be liable, but there can likely also be instances where output is technically infringing but falls below a de minimis threshold. This will likely depend on which prompts the end user gave the AI system, whether the end-user was actively trying to reproduce a known work, and whether there exists a certain diligence obligation on the part of the user to examine whether the output is similar to existing works that the user can reasonably be expected to know.

    c. Any other person? Please answer YES or NO. Please explain if needed.

Contributory negligence is recognised under Danish law. An online platform or other intermediaries which continue to host the AI system may thus be found to be liable, notwithstanding being informed about the infringement.

Sanctions

10) What sanctions can be imposed if it is found that Copyright has been infringed in order to train the AI System (because Copyrighted Data has been used with no authorisation)?
    a. Injunction, destruction, etc. of Copyrighted Data used to Train AI System still present in the data set or in the AI System? Please answer YES or NO. Please explain if needed.

Under Danish Law, one can get an injunction against an infringing party and one can get a court decision/order requiring that infringing party to destroy infringing replicas. In theory, these options exist for AI systems also. However, it is difficult to envisage how injunctions or orders of destruction can be handled practically against AI systems. If the infringing AI system falls under the model-as-container model (see 8 above) one could not meaningfully separate one or more specific protected works from the algorithm.

As an analogy, one could think of baking a cake. The cake is made of flour, cinnamon, eggs, sugar, and milk. These are the training data. The ingredients are mixed and baked. This is the preprocessing and encoding processes of the AI system. This results in a baked cake. The baked cake is the AI system. Once baked, the individual ingredients cannot be meaningfully extracted and distinguished. If, for example, the eggs were copyright-protected, it would not be practically possible to destroy them or enforce an injunction against only the eggs.

b. Injunction, destruction, etc. of the AI System itself deemed to be infringing, because it has been trained with infringed Copyrighted Data? Please answer YES or NO. Please explain.

YES in theory. In practice difficult to envisage.

Following the cake analogy, this would mean that because cinnamon is present in the cake to such an extent that it is recognisable in the baked cake, the entire cake could be ordered destroyed. This would likely be deemed to be too far-reaching and disproportionate.

In theory, if Cinnamon, eggs, sugar, and milk all sued together, it could be argued that it would now be proportionate to destroy the cake, as an adequately large proportion of its ingredient (training data) were infringing.

In practice, however, this would probably not be possible, as AI systems are not trained on 5 ingredients, they are trained on millions of pieces of data. Some of this data is protected and some is in the public domain. Some of the training data has identifiable owners, some of it has not.

c. Injunction, destruction, recall from commercial channels, etc. of outputs deemed to be infringing? Please answer YES or NO. Please explain if needed.

YES. This more closely resembles the “classical” example of copyright infringement, where someone has created something that infringes upon an existing work.

d. The award of damages, including punitive damages? Please answer YES or NO. Please explain if needed.?

YES. Damages and compensation. Often awarded as a discretionary lump sum by the courts, as it is very difficult to document the full loss in practice. Punitive damages are not available under Danish

law. It is only possible to obtain damages for actual losses, including loss of market opportunity or reasonable compensation.

e. Confiscation of all or part of the profits generated by the operation of the AI System? Please answer YES or NO. Please explain if needed.

YES in theory – however the specific provisions in the Danish Copyright Act do not allow to actual confiscation of the profit.

f. Any other sanctions? Please answer YES or NO. Please explain if needed.

Publication of the decision may be ordered.

In which other situations can outputs be qualified as Copyright infringement?

11) Please explain, if and under what conditions outputs generated by an AI System are qualified as infringement of a Copyrighted work, e.g. because the output contains characteristic elements of the Copyrighted work, in the following cases:
    a. In case the Use of the Copyrighted Data (work) to Train the AI System is covered by an Exception. Please answer YES or NO. Please explain if needed.

Probably.

It is not obvious that simply because an exception allows for using protected data as training data, that this would allow for generating output that is infringing. Section 11 b and 11 c of the Danish Copyright Act does not seem to allow for this; see 1 and 3 above.

b. In case the Copyrighted work has NOT been used to train the AI System. Please answer YES or NO. Please explain if needed.

NO.

If an AI system has not been trained on a specific copyrighted work and subsequently generates an output that is highly similar to that work, such output will generally not constitute copyright infringement, provided the similarity arises from independent creation rather than copying. Under Danish copyright law, as under EU law more broadly, copyright infringement requires that the allegedly infringing work involves a reproduction – whether direct or indirect – of the protected work. If the original work was not included in the AI system's training dataset, nor accessed through dynamic search capabilities (e.g. retrieval-augmented models), then it is unlikely that the AI system had the opportunity to copy the original, and infringement would be difficult to establish.

Furthermore, coincidental similarity does not constitute infringement if access to the original work cannot be demonstrated, and the output was generated independently. This reflects the traditional copyright principle that independent creation is not infringing, even if the result resembles an existing protected work.

One can debate the originality – and thus protection – of the original work, if an AI system can generate something that is very similar to the original work without "being aware" of it. In Danish legal theory in particular, reference is often made to the criterion of double creation as a test of originality – If two people can/would come up with the same work independently of each other, it is unlikely that the work is original. This principle is not known in EU law, and Danish law does allow for independent creation.

Originality under Danish law and EU jurisprudence does not require uniqueness or novelty, but rather that the work reflects the author's own intellectual creation. Thus, the possibility that an AI may reproduce something similar by chance does not strictly retroactively affect the protectability of the original work. Especially since the AI system may have been trained on works derived from the original work that was not itself part of the training data.

c. In case the Use of the Copyright Data to Train AI System has been authorised, is the content generated (output) always licit, even if some for instance outputs contain characteristic elements of the Copyrighted work? Please answer YES or NO. Please explain if needed.

While "always" may be too strong, it is reasonable to argue that if a copyright holder enters into a license agreement with an AI developer or provider that permits the use of the rightsholder's works as training data, then the rightsholder can be expected to anticipate that the AI system may, as a matter of course, generate outputs that reflect some similarity to the training material.

This expectation arises from the nature of generative AI systems, which internalize patterns, styles, and structural elements of the training data. However, the precise scope of permissible similarity will depend on the terms of the agreement, and whether it expressly or implicitly covers output generation — especially where outputs are close to verbatim or constitute derivative works.

12) Who is liable in case outputs infringes Copyright?
    a. The Provider of an AI System? Please answer YES or NO. Please explain if needed.

Potentially YES.

As the developer/provider has programmed and trained the AI system, it seems reasonable that they would bear the risks that this includes.

b. The user who exploits commercially the AI System? Please answer YES or NO. Please explain if needed.

Potentially YES.

It depends on what is meant by the question. If an author or visual artist (i.e., someone who has as their primary profession to generate works of art; a creative profession) uses generative AI to "write" their books or "paint" their paintings, and the artist then shares these, it seems likely that the artist may be liable.

On the other hand, if someone who does not have generating protected works as their primary profession (a non-creative profession) uses generative AI, it is less obvious that this would be infringing. If, for example, an attorney, uses generative AI to write a writ that is handed into the courts or a contract that is signed by the parties to it this would (if the writ and contract are deemed to be infringing) prima facie be infringement, but the present authors are not convinced that a court would rule for infringement.

Similarly, if generative AI is used in a non-creative profession as a search engine or for information gathering, the output may, in theory, constitute an infringing recreation even if it is not shared, but the present authors find it hard to believe that the user will be held liable.

This distinction between creative and non-creative industries should, in theory, not make a difference in the assessment of whether an infringing recreation has occurred, but in practice, it may.

The present authors find it more likely that users of generative AI systems will be found liable for infringing the exclusive right to communicating reproductions of works to the public than that the users will be found liable for infringing the exclusive right to reproduce the work if these are not shared. Unfortunately, this assessment is not based on a rigid legal standard or reference but more so on a general insight into the tendencies of the Danish courts.

c. The final user? Even if acting in good faith or unaware of the infringement? Please answer YES or NO. Please explain if needed.

Depends on how it is used.

If the AI system is simply used in everyday life as a search engine, this would probably fall within the de minimis exception and thus not be an infringement.

If, however, the private end user generates something infringing and shares it with other people, for example, on social media, it would probably constitute infringement. Infringement does not have to be active.

The present authors find it more likely that users of generative AI systems will be found liable for infringing the exclusive right to communicating reproductions of works to the public than that the users will be found liable for infringing the exclusive right to reproduce the work if these are not shared. Unfortunately, this assessment is not based on a rigid legal standard or reference but more so on a general insight into the tendencies of the Danish courts.

d. Any other person? Please answer YES or NO. Please explain if needed.

Contributory negligence is recognised under Danish law. An online platform or other intermediaries, which continue to host the AI system may thus be found to be liable, notwithstanding being informed about the infringement.

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

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? If YES, please explain.

YES

Generally, the primary “issue” with copyright and AI currently is the lack of legal certainty. This could be solved either through law or it will be naturally solved as more and more jurisprudence accrues.

a. Use of Copyrighted Data to Train AI System require prior authorisation of the Copyright holder? Please answer YES or NO. Please explain if needed.

YES

The present authors are convinced that the use of protected works as AI training data prima facie falls within the exclusive rights of the right holder of the concerned works, as it will entail recreations of the works; see 2 above.

Legal certainty would still be nice, though.

The question that needs answering is, thus, whether these recreations may be legal pursuant to an exception, specifically Sections 11 b and 11 c of the Danish Copyright Act (Articles 3 and 4 of CDSM Directive).

b. Exceptions authorising the Use of Copyrighted Training Data without Copyright holder consent. Please answer YES or NO. Please explain if needed.

YES

It should – ASAP – be made clear whether sections 11 b and 11 c of the Danish Copyright Act (Articles 3 and 4 of CDSM Directive) apply to data mining with the view of training AI models. If so, the training of the models based on protected data should be legal as the lesser is included in the greater, i.e. if data mining is allowed to train AI systems, then simply training AI systems should also be allowed.

Alternatively, a separate statute could be added that states that using protected data to train AI systems is lawful or that it is not.

If the lawmakers decide that it is not lawful, such clarification should be implemented in a separate provision, preferably 11 d or as a subsection to the existing 11 b and 11 c.

c. Consequences 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 answer YES or NO. Please explain if needed

YES

It should – ASAP – be made clear whether the use of a work in training data constitutes an illegal reproduction. As stated above, the present authors believe that the training data – which is encoded into the algorithm itself – can constitute unlawful reproductions, but one can easily imagine that once the data has been encoded, it should no longer be considered an unlawful reproduction. In that case, legal clarity in whether the temporary reproductions of the data (prior to it being encoded) would be lawful, potentially according to Section 11 a on temporary and technically necessary reproductions.

ii. The sanctions that should be available in case an AI System has been recognised to have been trained with infringed Copyrighted Data without authorisation? Please answer YES or NO. Please explain if needed.

YES.

It should – ASAP – be made clear how the lawmakers expect the right holders to react in such cases: As outlined in section 10 above, there are no real practical options for right holders to stop infringing use as training data.

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? Please answer YES or NO. If YES, please explain.

Currently, there does not, in Denmark, exist the litigation tools to meaningfully pursue infringement cases against the providers/developers of AI systems. This makes pursuing infringement de facto impossible.

It could be considered whether a third leg of the exclusive right conferred by copyright should be implemented: Copyright would confer onto the holder the exclusive right to reproduction, the exclusive right to communicate the work to the public, and the exclusive right to use the work as AI training data. Then, the option of making mandatory licenses for this use could be established and be administered by collecting societies.

Denmark has implemented similar mandatory licenses administered by collecting societies with some success. This would allow for the innovation of AI systems by creating the legal certainty that the use of works as training data would not constitute infringement, and it would mean that the right holders of the works would be compensated.

There is, however, the significant issue that such a system would require all the relevant right holders to be identified. Even with the soon-to-be-implemented obligation in Article 53(1)(d) to draw up an overview of the training data, identifying the relevant right holders may be impossible. Also, this would, to a certain extent, require the evaluation of protection for all the training data, as this system would not require licenses for data that is not copyright protected and is part of the public domain: A lot of training material is very likely not protected by copyright. This may very well be too large of an administrative burden to bear.

It should also be noted that the primary legal battle ahead may not even be related to copyright but may very well be related to personality and image rights and data protection law. For example, the use of patient and health data as training data.

III. Proposals for harmonisation

Please consult with relevant in-house / industry members of your Group in responding to Part III.

15) In your opinion, should Use of Copyrighted Data to Train AI System be harmonised? Please answer YES or NO. For what reasons?

YES.

More harmonisation "expands" the protection of the copyright holder as their legal position becomes more clear to them. Also, AI systems and models are not so kind as to care for borders, and the genie is out of the lamp, and it is very difficult to imagine existing AI models being rolled back from the markets on which they are already present. There is no reason to live under the illusion that this is even possible – instead, we must "simply" determine their legal status and how to best move forward.

If YES, please respond to the following questions without regard to your Group's current law or practice.

Even if NO, please address the following questions to the extent your Group considers your Group's current law or practice could be improved.

Should the Use of Copyrighted Data to Train AI System be authorised 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? Please answer YES or NO. Please explain, e.g. if the authorisation should be necessary only if acts of reproduction, representation, etc. are carried out, or in any circumstances.

To the extent that training requires reproductions of the work - YES.

One could argue that the use of Copyrighted Data to train systems should be a new, separate (third) leg of the exclusive right conferred by copyright, as it does not necessarily strictly fall within "reproduction" and/or "communication to the public", but third parties would gain from being allowed to commercially exploit the work, and the right holders should share in the gain derived from their works.

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.)? Please answer YES or NO. Please explain, e.g. the conditions of application of each Exception separately.

Probably. It would stifle innovation if we unduly restricted data mining and processing. Also, the genie is out of the lamp: AI systems are now a fact, and their revolutionary impact and influence cannot and therefore should not be stopped – there is no reason to live under such illusions. If Denmark (or the EU more generally) does not produce a legal status that can serve as a fertile foundation to foster innovation, Denmark (and the EU more generally) will be left behind. This would be untenable.

b. Special Exceptions to the Use of Copyrighted Data to Train AI System (e.g. TDM Exception, etc.)? Please answer YES or NO. Please explain e.g. the conditions of application of each Exception separately.

Almost all of copyright law in Denmark is based on EU law. Thus, almost all the applicable rules are derived from the EU, and for this reason, it would be good if their interpretation was harmonised. This goes for the interpretation of (1) whether the use of data as training material is a reproduction; (2) whether an AI algorithm where protected data has been encoded (model-as-container) constitutes a reproduction; (3) and whether the data mining exceptions apply to data mining with the goal of training an AI system.

If lawmakers do not get involved, the above will have to be interpreted by the CJEU, which could take five to ten years.

18) Should the Exceptions provide for financial compensation for Copyright holders (e.g. a royalty paid to a collecting society, etc.)? Please answer YES or NO. Please explain.

YES.

This could be connected to data mining exclusivity “proposed” above; see 14 above.

It seems reasonable that if there is any (financial) gain to be made from the use of protected work as training data, then the right holders of that data should share in that gain.

There would, however, likely be significant practical problems with implementing such a system; see 14 above.

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.)? Please answer YES or NO. Please explain if needed.

Yes. At least if the use is commercial.

20) Should the Provider of an AI System be obliged to make public the training data used to train the AI System (transparency obligation)? 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?

YES.

Both as a matter of principle and because if we wish to somehow reward the right holders of the data, this would be a practical prerequisite.

In order to be able to properly police the use of copyrighted material and because of the difficulties of proving possible infringement of copyrighted material, the supplier of the AI system should provided information about what rights they have exploited to such an extent that it becomes possible for the copyright holder to actually enforce their rights.

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.

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? Please answer YES or NO. Please explain if needed.

Yes.

Classical assessment of copyright infringement.

    b. The output is in the same style as one or more Copyrighted Data used to Train AI System? Please answer YES or NO. Please explain if needed.

“Style” is probably too abstract to be copyright protected.

Depending on the specifics of the case, it may be an infringement of different laws.

    c. The output is in all cases infringing if it has been trained with one or more infringed Copyrighted Data. Please answer YES or NO. Please explain if needed.

NO. It should be more closely based on the “classical” infringement assessment:

- Identity between the original work and the deviate
- Copying of the original work (i.e. independent creation is not an infringing).
- Copying of original/protected/characteristic elements of the original work
- Copying of more than abstract ideas

Proving the above criteria may be different in practice in instances where the derivative work is AI-generated, but the principles of assessment of infringement should be the same.

d. In other cases? Please answer YES or NO. Please explain if needed.

None that the present authors can think of.

22) Should the AI System itself be considered a Copyright infringement? Please answer YES or NO. If YES, please explain, e.g. under which conditions.

YES.

Model-as-container models should be considered prima facie infringing (if their algorithm contains encoded reproductions that meet the criteria of infringement). An exception should, however,apply to not stifle innovation.

Infringer

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

a. The Provider of an AI System? Please answer YES or NO. Please explain if needed.

Yes.

However, it would probably be advisable to implement standard of liability akin to the safe harbour liability scheme known from intermediary liability: The AI producer/provider should be able to avoid liability if they can prove that they have done what cane be reasonably expected from a diligent provider/producer to code the algorithm to avoid infringement.

Also, if the system is prompted to produce infringing material, the prompter/user should be liable and not the producer/provider. This would be an assessment that hinges on the individual facts of a given case.

b. The user who exploits commercially the AI System? Please answer YES or NO. Please explain if needed.

Yes.

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; see. 12) b above.

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.

c. The final user? Even if acting in good faith or unaware of the infringement? Please answer YES or NO. Please explain if needed.

Very rarely.

The use of AI systems as an alternative to search engines should not make he end-user liable for reproductions.

If AI-generated content is shared – for example on social media – this may be copyright infringing.

d. Any other person? Please answer YES or NO. Please explain if needed.

It should probably be possible in certain instances to hold intermediaries liable.

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 with no authorisation?
    a. Injunction, destruction, etc. of Copyrighted Data used to train AI System still present in the dataset or in the AI System? Please answer YES or NO. Please explain if needed.

YES.

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; se cake analogy in 10 above.

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 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.

b. Injunction, destruction, etc. of the AI System itself deemed to be infringing, because it has been trained with infringed Copyrighted Data? Please answer YES or NO. Please explain.

YES.

In theory, if it were proven that a given AI model had been trained exclusively on copyright-protected works without authorisation, and all relevant rights holders brought a suit, the courts could be in a

position to order the destruction of the model or issue an injunction prohibiting its use. This would follow from established principles in Danish and EU copyright law, where remedies such as injunctions and destruction of infringing materials are available in cases of clear and substantial infringement.

However, in practice, such outcomes are highly unlikely. The logistical and legal hurdles to identifying all relevant rightsholders, proving infringement at scale, and balancing the proportionality of such remedies make this virtually impossible. Courts would also need to weigh the significant public interest in the continued functioning of deployed AI models, particularly when they have widespread societal or economic impact.

Accordingly, while the legal possibility should exist – and already does in theory – the present authors consider it extremely improbable that a court would, in practice, require the destruction of a widely used AI model – particularly one whose outputs are not shown to cause significant market harm or reproduce protected works in a directly recognisable manner.

c. Injunction, destruction, recall from commercial channels, etc. of outputs found to be infringing? Please answer YES or NO. Please explain if needed.

Yes.

This is very close to the "classical" case of copyright infringement.

d. Award of damages, including punitive damages? Please answer YES or NO. Please explain if needed.

Yes. But not punitive damages

e. Confiscation of all or part of the profits generated by the operation of the AI System? Please answer YES or NO. Please explain if needed.

Yes.

f. Any other sanctions? Please answer YES or NO. Please explain if needed.

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. Please answer YES or NO. Please explain if needed.

No.

Simply because an exception allowing for training of AI systems exists, this should not give free reign to use the AI system to create infringing output.

b. In case the Copyrighted work has NOT been used to train the AI System. Please answer YES or NO. Please explain if needed.

No.

In most instances, this should not be copyright infringement. Copyright infringement requires the copying of an existing, protected work. If the AI is not aware of a given work, it probably cannot be said to copy the work.

However, there should be instances where a work that was not part of the training data can nonetheless be infringed upon – it is probably possible to prompt an AI system in such a way that it will create a "copy" of a work that it was not trained on. For example, one could probably prompt an AI system to make a song that includes the chord progression E♭m, G♭, A♭m, E♭m, B♭m, Am, A♭m, C♭, G♭, E, B, and D♭, and if that song turns out to sound like Jimi Hendrix's Little Wing, it will probably be an infringement of that song, even if it was not part of the training data.

c. In case the Use of the Copyright Data to Train AI System has been authorized, is the content generated (output) always licit, even if some for instance outputs contain characteristic elements of the Copyrighted work? Please answer YES or NO. Please explain if needed.

No.

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

a. The Provider of an AI System? Please answer YES or NO. Please explain if needed.

It should be possible; see 23) a above.

b. The user who exploits commercially the AI System? Please answer YES or NO. Please explain if needed.

Should be possible; see 23) b above.

c. The final user? Even if acting in good faith or unaware of the infringement? Please answer YES or NO. Please explain if needed.

Should be possible; see 23) c above

d. Any other person? Please answer YES or NO. Please explain if needed.

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.

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