SCHOOL OF
ECONOMICS AND
MANAGEMENT

How well does the DSM Directive balance copyright and AI development?

Filip Nilsson

DEPARTMENT OF BUSINESS LAW

Master's Thesis in European and International Trade Law

15 ECTS

HARN63

Spring 2024

Contents

Abstract

In this thesis, the interests behind AI development and copyright have been investigated. Furthermore, it has also examined how Articles 3 and 4 of the DSM Directive balance these interests.

During the thesis, it has become clear that AI has many possibilities and that the interest in continued development encompasses a wide scope. When an AI is developed, it may need to be trained with a dataset consisting of works in order to learn. These works have the potential to be protected by copyright. Furthermore, in the thesis the focus has been on understanding the copyright. Here, it has become clear that the works trained by AI can be covered by a protected work with exclusive right of reproduction. This is where copyright and AI development comes across. Just as there is a broad interest in AI development, there is also an interest in copyright from EU level down to individual level. In the last part, it was examined how well articles 3 and 4 of the DSM directive balance these interests. Regarding articles 3 and 4, it was clear that articles 3 and 4 open up the possibilities to use works but to a limited extent. Article 3 through specific users for a specific purpose and Article 4 with an "opt-out" mechanism.

During this essay, the conclusion could be drawn that there is a broad interest in both AI development and copyright. Furthermore, it could also be concluded that articles 3 and 4 of the DSM directive balance these interests, but perhaps not in the way that was intended.

Keywords: AI, Algorithm, Data, Copyright, Input, Output, Rightsholder

Abbreviations

AI	Artificial Intelligence
CJEU	Court of Justice of the European Union
EU	European Union

1 Introduction

1.1 Background

The development of technology has led to a significantly greater amount of data being produced that can lead to completely new levels of value creation, provided that it is compiled and used. The European Commission even highlights that this change is as big as the industrial revolution.[1]

Artificial intelligence is undergoing rapid development and something that can bring both benefits and consequences. This applies to both the economy and society in all business sectors and social activities. Artificial intelligence further covers a wide area and can be used for everything from supporting social and environmental aspects to providing important competitive advantages for companies and the European economy.[2]

Concerning the information above, there must be a balance between taking advantage of new technology and at the same time making it compatible with the fundamental values, fundamental rights and principles of the Union.[3]

With the latest types of artificial intelligence, the computer program is no longer just a tool. Computer programs can nowadays make many decisions themselves without human intervention in the machine-driven creativity.[4] Rapid technological developments continue to change the way works are created, produced, distributed and used. Through this development, new business models and actors are emerging.[5] With these technological developments, the Internet has become the most important marketplace for distribution of copyrighted content.[6]

This means that relevant legislation needs to be future-proofed with the aim of not limiting technological development. On the other hand, technological developments have contributed to the fact that there is some uncertainty regarding the legal situation of rightsholders and users regarding the use of works in the digital environment.[7]

In summary, the European Union has highlighted the importance of creating a data ecosystem that benefits both the society and the economy, including the individual

[1] Communication from the commission, 'Shaping the Europe's digital future', COM/2020/67 final [19.2.2020]

[2] Proposal for a regulation of the European and of the Council on laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain Union legislative acts, COM/2021/206 final [21.4.2021] p. 1 [cit Proposal AI Act].

[3] Ibid., p.1.

[4] Andres Guadamuz., 'WIPO MAGAZINE - Artificial intelligence and copyright', (WIPO 2017) < https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html> Accesed: 18.4.2024 (cit. Guadamuz (2017).

[5] Council Directive 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9EC and 2001/29/EC [2019] OJ L130/92, Recital 3 [cit DSM Directive]

[6] Proposal AI Act p.1.

[7] DSM Directive, recital 3.

citizen, business and public interests.[8] But this must also be accompanied by adequate protection of intellectual property rights. This is to ensure fair compensation for creators.[9]

That being said, artificial intelligence is rapidly evolving and involves various stakeholders to consider. Hence, it is important to balance copyright interests and at the same time enable technical development. According to this background, this thesis will take a closer look at this topic.

1.2 Purpose and research questions

The purpose of this thesis is to investigate which opposing interests that can stand against each other concerning copyright and artificial intelligence. Furthermore, this thesis also intends to examine how well the EU regulation has been adapted to balance these interests.

Hence, the research questions for this essay will be;

1. In what way are the interests in AI development and copyright in conflict with each other?

2. Do articles 3 and 4 of the DSM directive have the ability to balance these interests?

1.3 Delimitations

A delimitation for this thesis is that related and neighbouring rights not will be considered. This since it is not needed for the understanding regarding to the purpose and research question in the thesis. A further delimitation made in this thesis is to exclude Directive 96/9/EC on the legal protection of databases mentioned in article 3 and 4 in the DSM Directive and Directive 2009/24/EC on legal protection of computer programs mentioned in article 4 of the DSM Directive. The reason for this is also that it is not needed to the understanding of the research question and purpose of the thesis.

[8] European Commission, 'White paper on Artificial intelligence – A European approach to excellence and trust, COM/2020/65 final [19.2.2020] p. 2.

[9] Council Directive 2001/29/EC of May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10, recital 10. [cit Infosoc Directive].

1.4 Method and materials

In order to answer the thesis's purpose and research questions, an EU legal method in combination with the legal dogmatic method will be applied with a theological interpretation.

The legal dogmatic method is based on the text of law, case law, preparatory works and doctrine.[10] It can be said that laws and precedents from the court of precedents have a formal authority and even legislative acts are sometimes given a formal authority. Doctrine on the other side rather functions as persuasive to the arguments presented.[11]

The EU law consists of the binding sources of law, which includes the primary law and the charter of rights, binding secondary law, international agreements and general legal principles. It also applies in most cases to the case law from the CJEU. With that being said, the law adjudicator must act based on them. In contrast to the binding sources of law, non-binding secondary law is found. These include preparatory works, the opinions of the Advocate general, the EU legal doctrine and economic theories which are only considered to be indicative.[12] Against this background, the essay will primarily consider primary law and binding secondary law and then move on to doctrine.

Primary law is the superior source of law within Union law[13] and consists of treaties.[14] These treaties are created by a unanimous decision of the member states and need to be approved by every member state in accordance with Article 48 FEU.[15] According to Article 288 (2) TFEU, the treaties shall have a general application, be binding and directly applicable in each member state. In accordance with Article 51 TEU, protocols and annexes shall also form an integral part of these treaties. This means that these also belong to the primary law.[16] Other legal acts which have been decided in a corresponding order in unanimity within the EU and approved in accordance with the constitutional provisions of the respective member states also belong to primary law.[17]

Furthermore, the secondary law covers the legal acts that can be carried out with the support of the treaties.[18] It is these legal acts that can be carried out with the support of Article 288 TFEU. Article 288 (2) states that when the institutions exercise their powers, they shall adopt directives, decisions, recommendations and opinions. Even these are binding except from the recommendations and opinions. Secondary law also usually includes international agreements that the EU has concluded with

[10] Maria Nääv and Mauro Zamboni (ed), *Juridisk metodlära* (2th edn, Studentlitteratur, 2018) p.21. [cit. Nääv and Zamboni (2018)].

[11] Ibid., p.28.

[12] Jörgen Hettne and Ida Otken Eriksson (ed.), *EU – Rättslig metod* (2edn, Nordstedts Juridik, 2011) p.40 [cit Hettne and Otken Eriksson (2011)].

[13] Ibid., p.42.

[14] Ibid., p.41.

[15] Ibid., p.41.

[16] Ibid., p.41.

[17] Ibid., p.42.

[18] Ibid., p.41.

external countries or organizations. In addition, if the primary law is in conflict with the secondary law, the primary law takes precedence.[19]

In the thesis primary law is used to a limited extent. The EU charter will be used once, which belongs to primary law. However, the primary law is limited within the subject the thesis is about with regard to the purpose of the research. Against this background, binding secondary law will then be considered through the relevant provisions in the Infosoc- and DSM directives.

To interpret these provisions, case law from the CJEU will primarily be used in cases where there is practice. These also constitute binding secondary law. This will be followed by secondary law in the form of recitals and preparatory work for the directives. Moreover, the WIPO Copyright Treaty and the Berne Convention will also be a guide if needed. These also belong to secondary law since the WIPO Copyright treaty is an international agreement to which the EU is a party.[20] By WIPO Copyright Treaty Article 1 (4), the European Union is bound to consider Articles 1-21 of The Berne Convention. Doctrine will further be used in cases where more guidance is needed. An example of a source that is used here is the Swedish proposal Prop. 2021/22:278:278. This in order to get further guidance. The choice of this guidance comes from the fact that Sweden is a member state that must apply the provisions of the directives according to the directives purpose and goals according to article 288 TFEU. Through this order of legal sources, it also results in a combination between EU law and the legal dogmatic method.

Furthermore, the teleological interpretation method is used with an analysis in accordance with the legal dogmatic method.

The theological interpretation is the method of interpretation that the Union courts are best known for using. The theological interpretation is mainly applied when the wording or contextual context of a provision is unclear. This interpretation can be said to fulfil three different purposes within Union law. These are to promote the purpose of a provision, counteract unreasonable consequences of a literal interpretation and fill gaps that would otherwise exist in Union law.[21] The main reason for the theological interpretation is that one wants to avoid a result that is unreasonable with regard to Union law in general and in particular its general principles and regulations.[22]

Hence, for example, recitals in the Infosoc directive and the DSM directive have been considered in the thesis with its purpose. This with the goal of understanding the wording of the regulations in their context, which in the thesis reflects copyright in the context of artificial intelligence and vice versa. In other words, by using case law from the CJEU and interpreting the provisions based on their purpose with the help of recitals, among other things, the theological interpretation method is used.

[19] Ibid., p.43.
[20] European Commission - Shaping Europe's digital future 'The EU copyright legislation' (European Commission, 11.3.2024) https://digital-strategy.ec.europa.eu/en/policies/copyright-legislation Accesed: 22.4.2024.
[21] Hettne and Otken Eriksson (2011) p.168.
[22] Ibid., p.169.

In a legal dogmatic analysis, it is about to interpreting how a rule of law is perceived in a certain concrete context.[23] Within the legal dogmatic analysis there are two ways to build an argument. This is through de "lege lata" and "de lege ferenda" argumentation. De lege lata is based on trying to explain what the legal situation is, while de lege feranda rather suggests solutions to problems that are apparently unsolved so far.[24] This means that a legal dogmatic analysis will be used. This since the thesis examines how the interests between artificial intelligence and copyright are weighed against each other in the context of article 3 and 4 in the DSM Directive. Since the provisions thereof is interpreted in their context, it become a "de lege lata" interpretation.

1.5 Outline

The second chapter will briefly deal with what AI is, how an AI model works, from collecting works to generating output and who the stakeholders of continued AI development are. This concludes with a summary part. The purpose of this chapter is to create a basic understanding of AI and its possibilities in order to later also understand how copyright is affected. In the long run, this also provides a fruitful knowledge for later in the thesis to understand what is balanced together with copyright in article 3 and 4 of the DSM directive.

The third chapter deals with copyright. This chapter is structured with a background to copyright, what a copyright-protected work consists of, what exclusive rights the rights holder gets and what exceptions there are to these rights. This will be followed by a summary and analysis that connects chapters 2 and 3. This is to put AI development in relation to the interest of copyright. During this part, special focus will be placed on the exclusive right of reproduction since this is most relevant to understanding articles 3 and 4 of the DSM Directive. This even since these articles make an exception to the exclusive right of reproduction.

The fourth chapter takes aim at the question of how well articles 3 and 4 of the DSM Directive balance AI development with copyright. This chapter contains a background on the directive and the exception in general to then go deeper into article 3 and 4. This chapter concludes with a summary and analysis of whether these articles actually balance the interests of AI development and continued strong copyright protection.

The fifth and final chapter briefly summarizes the main findings in the thesis and answers the questions for the research.

[23] Nääv and Zamboni (2018). p.26.
[24] Ibid., p.36.

2 Understanding AI

2.1 Basics of AI

Humans have made creative works with the help of a computer since the 1970's and still do so today. To produce computer-generated work in the past, the computer was in the most cases dependent on a human creative input, where the machine has functioned as a tool or instrument. But with the new technology, the interaction between computers and the creative process is different.[25]

AI can be described as the ability of a machine to display human-like traits such as reasoning, learning, planning and creativity.[26]Another way to describe AI is through a collection of techniques that combine data, algorithms and computing power.[27] On the top of that, AI has the ability to enable technical systems to perceive their surroundings, manage what it perceives and solve problems in order to achieve a certain goal. This is because the computer collects information, which is already prepared or collected through its own sensors, processes the information and responds. In addition, AI systems are to some extent capable of adapting their behaviour by analysing the effects of previous actions, and of working independently.[28]

This paradigm shift can be seen in the light of the development of machine learning software. Machine learning is a subcategory of AI that produces autonomous systems that are capable of learning.[29] Machine learning is also an AI process that uses algorithms and statistical models to enable computers to make decisions without necessarily having to program it to perform the task.[30]

Furthermore, it can be good to simply describe the technical grounds for an AI model. Admittedly**,** these may differ for different areas of machine learning and for different algorithms within the same area of machine learning. This means that the technical steps and the legal consequences may vary.[31] In contrast, the different models from a structural perspective for generative AI have essentially common features.[32] In short, an AI system can be seen as a system that is based on 4 different

[25] Guadamuz (2017).

[26] European Parliament – 'What is artificial intelligence and how is it used' (European Parliament, 23.9.2020) > https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used > (cit. European Parliament, What is AI and how is it used, 2020) Acceded: 24.4.2024.

[27] European Commission, 'White paper on Artificial intelligence – A European approach to excellence and trust, COM/2020/65 final [19.2.2020] p.1. (cit AI White paper 2020).

[28] European Parliament, What is AI and how is it used (2020).

[29] Guadamuz (2017).

[30] WIPO – '*Technology Trends 2019: Artificial intelligence'* WIPO (2019), p.146.

[31] Thomas Margoni, 'Artificial Intelligence, Machine learning and EU copyright law: Who owns AI?' CREATe Working paper (2018), p.4-5 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3299523> Accessed: 21.4.2024 (cit Margoni (2018)).

[32] Jessica Fjeld and Mason Kortz, 'A Legal Anatomy of AI-generated Art: Part I', Harward journal of Law & Technology (2017) < https://jolt.law.harvard.edu/digest/a-legal-anatomy-of-ai-generated-art-part-i> (cit Fjeld and Kortz 2017) Accessed: 25.4.2024.

steps consisting of (1) Input, (2) Learning Algorithm, (3) Trained algorithm and (4) Output.[33]

In the (1) Input, the system is fed with existing works to be trained with[34] where this set of works is called "corpus".[35] This set of inputs can be anything from a small homogeneous type of corpus to a large data set of a huge amount of works.[36]. This depends on what you want to achieve. To get a more accurate result for the purpose, the set of data should be as close as possible to the corpus that the algorithm later should be applied to.[37] In that way, the output can be controlled in some way through the selection of the corpus.

In the next step there is (2) the learning algorithm. This algorithm consists of the machine learning system that operates on the inputs. The learning algorithm identifies the relevant features of the input and then stores that information in a data structure that then forms the trained algorithm.[38] However, it should be mentioned that although the learning algorithm is influenced by the input, it is still two separate elements.[39] The central point here is that the learning algorithm learns to analyze input to identify and compare pattern through this, the system thus produces different prediction rules for the trained algorithm.[40]

The third step consists of the (3) trained algorithm. The trained algorithm is the one that consists of the information that the learning algorithm has generated from its operations on the input together with the instructions to transform the information into a work. It could be good to have in mind that depending on the approach, this information may include recognizable fragments of the inputs or something more abstract such as a series of concepts or decision points.[41] Thus, the learning model contains a file where this file contains an abstraction of the data that the training algorithm has found useful to accurately distinguish between the data in the input and which data that will be retained. In addition, this file also contains the “memory” used by the AI to analyze new and unknown data and to adapt its behavior to this new reality. At this point, the original data set that was used as training material is no longer necessary for the AI system to work, it is only the trained algorithm that is needed.[42] While the input and the training algorithm can be developed independently of each other, a trained algorithm is something that is unique to the individual project. The trained algorithm ultimately consists of a data structure that is produced by the operation of the learning algorithm on the inputs which then results in a new

[33] Johan Axhamn, ’Copyright and Artificial intelligence – with a focus on the area of music’ in Festskrift til Jørgen Blomqvist, ExTuto publishing (2018), < https://canvas.education.lu.se/courses/25535/modules/items/988569 > p.43. (cit Axhamn (2018).

[34] Fjeld and Kortz (2017).

[35] Margoni (2018) p.4.

[36] Fjeld and Kortz (2017).

[37] Margoni (2018) p.4.

[38] Fjeld and Kortz (2018).

[39] Ibid.,.

[40] Axhamn (2018) p.43

[41] Fjeld and Kortz (2018)

[42] Thomas Margoni and Martin Kretchmer, ’A Deeper Look into the EU Text and Datamining Exceptions: Harmonisation, Data Ownership and the Future of Tecknology, GRUR International, Volume 71, Issue 8 (2022) < https://academic.oup.com/grurint/article/71/8/685/6650009> p.688. (cit Margoni and Kretchmer (2022) Accessed: 4.10.2024

trained algorithm. The trained algorithm is later also the element that generates the output.[43]

The fourth step is (4) the output. The output consists of the works produced by running the trained algorithm. The generated output is usually derived from the works in the input and the learning algorithm that have given the trained algorithm a starting point. However, it should be mentioned that the output also can be created from random starting points or without restrictions.[44]

From a user perspective, with the AI models' users can simply write an instruction to the model, a so-called prompt, and then receive an output generated by the AI model. An example is that a user can ask the AI model to answer a certain question, write a certain type of text, create a painting, count, make videos and sounds and much more. To give an example, we can mention Chatt GPT which is an AI model which is used by millions of people where this so-called chatbot can answer questions, write stories, write web codes and process very complicated questions. Things we humans previously could spend a long time to research and understand, an AI model like Chat-GPT has the potential to produce a well-written alternative in a few seconds.[45] A narrower example is the AI model Next Rembrandt which generates paintings based on Rembrandt's work.[46]

2.2 Opportunities conferred by AI

As mentioned in the previous section, AI models can be used in many different ways and for many different purposes. What takes, for example, time for humans to learn and create something from, an AI model can provide an alternative in a few seconds.

With the new technology, automated computational analysis of information in digital form is enabled.[47] Through this, a significantly larger amount of data can be collected, where this can lead to completely new levels of value creation. However, it is important to point out that this value creation requires that the data is actually compiled and used. The European Commission even believes that this change is big as the industrial revolution.[48]

The rapid development of AI has the potential to generate benefits for both the economy and society in all business sectors and social activities. To give examples, AI can contribute to supporting social and environmental aspects and provide important competitive advantages for companies and the European economy.[49]

---

[43] Fjeld and Kortz (2018).

[44] Fjeld and Kortz (2018).

[45] Alex Hughes, 'BBC Science Focus – CHATT GPT:Everything you need to know about OpenAI's GPT4 tool', (BBC 2023) < https://www.sciencefocus.com/future-technology/gpt-3 > Accesed 25.4.2024.

[46] Axhamn (2018) p.43.

[47] DSM Directive, recital 8.

[48] Communication from the commission, 'Shaping the Europe's digital future', COM/2020/67 final [19.2.2020].

[49] Proposal AI Act, p.1.

According to the European Parliament, Europe's growth and wealth are closely linked to how data and connected technologies are used within the Union.[50] EU countries are already far ahead in the digital industry and business-to-business applications.[51] Furthermore, AI can also create new jobs while AI-driven industries develop and change. The European Parliament estimates that AI has an expected calculated increase in labor productivity of 11-37% linked to AI up to the year 2035 [52] For companies, AI can enable the development of new generations of products and services, including sectors where the EU already has strong positions. This in areas such as the environmental and circular economy, machines, healthcare and more.[53] From this perspective, AI is estimated to be able to help reduce global environmental emissions by 1.5 - 4% by the year 2030.[54] At the same time, AI can facilitate smoother and more optimized sales chains, improve machine maintenance, increase production and quality, improve customer service and save energy.[55] Based on these aspects, it is easy to understand why the European Union highlighted the importance of creating a data ecosystem that is beneficial to society and the economy, including the individual citizen, business and public interests.[56]

On the other side of the coin is an additional aspect, which is the negative consequences of underusing AI. According to the European Parliament, underuse of AI could even pose a threat to the EU. This is due to missed opportunities such as poor implementation of major programs, lost competitive advantages in relation to other regions, economic stagnation and fewer opportunities for residents. An underuse could be due to public and business mistrust of AI, poor infrastructure, lack of initiative and low investment, or, as machine learning relies on data, from a fragmented digital market.[57]Another potential reason for the underuse of AI could be copyright regulations. Although it is difficult to determine an exact impact on the creative economy, there is still a strong likelihood that it will have a dampening effect on investments in automated systems. If one cannot use the creations generated by AI, incentives to invest money in it are also affected. At the same time, AI can still be a good tool for being time efficient, which in itself can provide incentives for investment.[58] Finally, it is also interesting that the global scientific community generates over 1.5 million new scientific articles and approximately one new essay every 30 seconds. Data continues to show that approximately 90% of all published scientific articles are never cited and 50% are never read by anyone other than their authors, referees and journal editors. This shows an inefficient system where resources are spent on duplicating knowledge that already exists and goes undiscovered.[59]

[50] European Parliament – 'Artificial intelligence: threats and opportunities' (European Parliament, 20.6.20) < https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities > (cit. European Parliament, AI opportunities and threats (2020) Accesed:24.4.2024.
[51] Ibid.,
[52] Ibid.,
[53] Ibid.,
[54] Ibid.,
[55] Ibid.,
[56] AI Whitepaper 2020, s.2.
[57] European Parliament, AI opportunities and threats (2020).
[58] Guadamuz (2017).
[59] Margoni and Kretchmer (2022) p. 689.

2.3 Summary

In summary, an AI could be described as a machine's ability to exhibit human-like traits, such as reasoning, learning, creativity and planning. AI also has the ability, through technical systems, to perceive its surroundings, process what it perceives and solve problems in order to achieve a certain goal. On top of that, AI systems are also capable of adapting behaviour by analysing the effects of previous actions and working independently. This works by feeding the computer with a data set consisting of pre-existing works that it temporarily processes and then responds to. Such a process can consist of 4 different steps which include (1) Input, (2) Training algorithm (3) Trained algorithm and (4) the output. It is on this path that there are great opportunities with AI.

The first aspect is the opportunities that come with AI that include opportunities in all business sectors and societal activities. The possibilities in AI are great and include everything from contributing to science, new business models, giving the European Union competitive advantages and reaching environmental goals. This shows that the possibilities of AI cover a wide range of areas. An additional example is that AI is expected to generate economic growth for the EU and provide the internal labour market with new job opportunities, which also shows that the opportunities within AI both extend from the Union level all the way down to the individual level.

There are many possibilities and what takes time for a person can an AI do in seconds. Having said that, AI can make things more efficient and at the same time cover a larger amount of data and information. In this way, artificial intelligence can bring benefits to all business sectors and social activities.

The other aspect of AI is the danger of underusing and not keeping up with the development of AI. However, it could pose a threat to the EU where the Union loses competitive advantages in relation to other regions, missed opportunities, economic stagnation and fewer opportunities for the inhabitants. Just as there are many opportunities at different levels for actors and areas through AI, these opportunities can be lost by underusing. These opportunities can also become a threat if other regions take advantage of them while the EU does not.

In short, it can be said that there is a general interest in the continued development of AI. However, it is important to keep in mind that AI in many cases may have its basis in existing data that could be protected by copyright. Concerning this background, there is an interest in making it possible to use as much data as possible. Having said that, the next part should aim at the other side of the coin.

3 Understanding the EU Copyright framework

3.1 Basics of the EU Copyright regulation

The root of copyright can be found in Article 17 about the right to property in the EU Charter. Article 17 (1) of the EU Charter states that every person has the right to own, use, dispose of and bequeath his lawfully acquired possessions. Its further states that no one shall be deprived of their possessions. This applies as long as it is not in the public interest or prescribed by law and reasonable compensation is paid to the right holder for their loss. This use of the property shall be regulated by law to the extent that it is necessary for the public's interests. In accordance with Article 17 (2) in the EU, this also applies to intellectual property rights.

Before the 1980's when the harmonization of EU legal regulation of intellectual property rights began to take shape, the EU regulation was limited in the area.[60] Since then, the harmonization of intellectual property rights has developed and today there are 13 different directives and 2 regulations that harmonize intellectual property rights within the Union.[61]

The European Council emphasized at its meeting in Corfu in 1994 the need to create general and flexible legal frameworks at community level in order to promote the development of the information society in Europe. It was emphasized here that copyright is of great importance in this context because these rights protect and stimulate the development, marketing of new goods and services as well as the creation and the use of the creative content therein.[62] In order for the authors to be able to continue with their creative activities, it is necessary that they receive fair compensation for the use of their work. Adequate legal protection of intellectual property rights is therefore considered necessary to ensure such compensation.[63]

Hence, the harmonized legal framework for copyright comes through an increased legal certainty and through a high level of protection being created in the field of intellectual property law. This with the aim of encouraging significant investments in creative and innovative activities with the purpose of providing the union with growth and competitiveness.[64] A lack of harmonization within the Union can lead to significant differences in copyright protection and thereby restrictions on the free movement of goods and services that contain or are based on intellectual property

[60] Mireille va Echod, Pierre Bernt Hugenholtz, Stef van Gompel, Lucie Gubault and Natalie Helberger, 'Harmonizing European Copyright Law: The Challenges of better law making', International Law Series, Volume 19, Alphen aan den Rijn: Kluwen Law International 2009 (chapters 1 and 9), Amsterdam Law School Research Paper No.2012-07, Institute for information Law Research Paper No.2012-07, (2012) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2049935> p.2. Accessed: 4.20.2024.

[61] European Commission - Shaping Europe's digital future 'The EU copyright legislation' (European Commission, 11.3.2024) https://digital-strategy.ec.europa.eu/en/policies/copyright-legislation Accesed: 22.4.2024.

[62] Infosoc Directive, recital 2.

[63] Ibid., recital 10.

[64] Ibid., recital 4.

rights. This can further lead to fragmentation of the internal market and inconsistency in legislation as a result. The effect of such legal differences may become more extensive through the continued development of the information society, which has already led to a greater use of intellectual property rights across national borders.[65]

With that being said, the basis for the harmonization of copyright must be to have a high level of protection because these rights are of decisive importance for the continued intellectual creation. The protection of these rights contributes to preserving and developing creativity and benefits the authors, consumers, culture, business and the public.[66]

Through EU legal harmonisation, some central directives of this essay have been shaped. The most essential directive for the essay is the Infosoc directive (directive 2001/29/EC) on the harmonization of certain aspects of copyright and related rights in the information society. This directive aims at the legal protection of copyright in the internal market with particular emphasis on the information society for authors of their works in accordance with Article 1(1) and Article 2(2) respectively of the Infosoc Directive. It is these rights that the DSM directive on copyright and related rights in the digital single market and amending Directives 96/9 EC and 2001/29/EC makes exceptions and restrictions for text and data mining in accordance with articles 3 and 4 of DSM Directive.

3.2 What constitute a protected work

Copyright arises formless, which means that there are no formalities such as registration or deposit.[67] The infosoc directive states that the object of protection must be a work according to article 2(2) in the Infosoc directive. On the other hand, the directive does not define what a work consists of or what the requirements are for the work to obtain protection.

In Union law, the CJEU has stated the requirement that the work must achieve originality in order to obtain protection. Furthermore, the work must also be original in such a way that the work is the author's own intellectual creation and gives expression to it.[68] In addition, the work must also be the author's own intellectual creation.[69] For the work to be considered the author's own intellectual creation, the work must reflect his or her personality.[70] So there must be a so-called "personal touch" from the author that is expressed in the work. This is considered to be the

[65] Ibid., recital 6.

[66] Ibid., recital 9.

[67] Ulf Bernitz, *Marknadsföringsrätten: Svensk och europeisk marknadsföringsrätt* (2th edn, Nordstedts Juridik, 2020) p.71.

[68] Case C- 5/08 *Infopaq international V Danske Dagbladets forening* [2009] ECR I -6624, paras 37 and 39. (cit Case – 5/08 Infopaq).

[69] Case C-310/17 *Levola Hengelo BV V Smilde Foods BV* [2018] ECLI:EU:C:2018:899, para 36. (cit Case C-310/17 Levola Hangelo).

[70] Case C-145/10 *Eva-Maria Painer V Standard VerlagsGmbH and Others* [2011] ECR I-12594, para 88. (cit Case C-145/10 Painer).

case when the author has been able to make free and creative choices in connection with the creation of his work.[71] An insight from this is also that it is not about establishing any quantitative measure but rather qualitative if it has originality or not.

The CJEU has also stated that parts of works should not be judged differently. With this said, shall parts of a work also receive protection if these contribute to the originality of the work as a whole.[72] For example, it was considered that 11 coherent words could express an intellectual creation with originality.[73] Notably, words as such by themselves cannot be considered to constitute an intellectual creation with originality.[74] This can be understood to the background that it is through the choices, disposition and combination that the expression of an intellectual creation with originality is made.[75]

Further, the CJEU also refers to Article 2 (1) of the Berne Convention in relation to works. This since the EU is bound to this provision by Article 1(4) of the WCT, to which the EU is a treaty party.[76] From the Berne Convention article 2 (1) it appears that works must include every production in the literary, scientific and artistic field. This is independent of its form or expression. Furthermore, the CJEU also refers to Article 2 of the WCT[77] which provides that protection may cover expressions but not extend to ideas, procedures, methods or mathematical concepts as such.

To link back to the previous chapter, Margoni believes that many works used within the framework of an AI model have the potential to be protected by copyright. This is in the light of the fact that the concept of originality is an essential requirement for protection within copyright and that a small amount of originality can make a work worthy of protection.[78]

3.3 Protection that EU Copyright law confers

The Infosoc directive intends to provide authors with a high level of protection that enables them to receive fair compensation for the exploitation of their works.[79] To protect this interest, some exclusive rights have been awarded to the author. This by the exclusive rights in article 2-4 in the Infosoc directive. Exclusive rights mean that consent is required from the rights holder to use the work in a way that is covered by the exclusive rights. This is because the rightholder has the right to prohibit or allow such use. The exclusive rights include, in accordance with article 2-4 of the Infosoc directive, the right to reproduction, the right of communication to the public and the right of distribution.

[71] Ibid. Para 92.
[72] Case – C 5/08 Infopaq, para 38.
[73] Ibid., para 48.
[74] Ibid., para 46.
[75] Ibid., para 45.
[76] Case C-310/17 Levola Hangelo, para 39.
[77] Ibid.,
[78] Margoni (2018) p.4-5.
[79] Infosoc Directive, recitals 9-10.

Through article 2 (a) in the Infosoc directive about the exclusive right of reproduction, this exclusive right applies to the right holder to authorise or prohibit reproduction of their work directly or indirectly, temporarily or permanently by any means and in any form in whole or in part. The CJEU stated that this protection must also be given a broad interpretation.[80] Furthermore, there is a general principle laid down in the directive based on the fact that any reproduction of protected works requires the consent of the copyright holder.[81] One can thus see this exclusive right in the light of this principle since the right holder is the only one that can give this consent to reproduction.

There is no definition of the concepts of reproduction or partial reproduction in the Infosoc directive.[82] On the other hand, the CJEU states that these concepts must be interpreted through the provision's wording, context and the goals pursued in the directive and international law.[83] On the top of that, CJEU also states that the broad interpretation of what constitutes a reproduction can be seen in the light of the wording "directly or indirect ", “temporary or permanent”, “In any means” and in “any form” that is used.[84]

Reproducing parts of a work can constitute a partial reproduction.[85] With the wide interpretation, the protection can for example extend to parts of sentences in a text with the justification that these parts contribute to the originality of a work. This through an "element" as in itself constitutes an expression of the author's own intellectual creation.[86] The CJEU further said that it cannot be ruled out that eleven consecutive words from a protected work being reproduced can result in an extract constituting a partial reproduction in accordance with the Article 2 of the infosoc directive.[87] However, this shows a low requirement by the CJEU for what that can constitute a partial reproduction.

An additional exclusive right can be found in article 3 of the Infosoc directive which aims at the right of communication to the public. This stipulates the exclusive right for the author to permit or prohibit any communication to the public of their work by wire or wireless means. This means that the works are made available to the public in such a way that individuals can access these works from a place at a time of their choosing. In the same way, the exclusive right to make the work available to the public shall also apply in accordance with the second paragraph of the article. Furthermore, the CJEU points out that the wording "comminucation to the public" must be given a broad interpretation.[88] The communication in question must also be to a different public than the one intended by the original communication, in other words, a new public.[89] To be considered to communicate to the new public, it is

[80] Case – C5/08 Infopaq, para 43.
[81] Ibid., para 57.
[82] Ibid., para 31.
[83] Ibid., para.32.
[84] Ibid., para 42.
[85] Ibid., para 48.
[86] Ibid., para 47.
[87] Ibid., para 48.
[88] Case C-306/*05 SGAE V Rafael Hoteles [2006] ECR I-11543, para 36.*
[89] Ibid. para 40.

sufficient that the work is made available in such a way that the persons belonging to this public can have access to the work.[90]

Furthermore, in Article 4 of the Infosoc directive, there is an additional exclusive right through the exclusive right of distribution. This article states that an author has the exclusive right to permit or prohibit any distribution to the public, by sale or otherwise, of the original of their work or copies thereof. In addition, the exclusive right of distribution only extends to works that are part of a physical product.[91] Having said that, the CJEU held that this right does not apply to services and in particular not to online services where this right only applies to the distribution of works in physical form.[92]

To relate back to the previous chapter, it is necessary for a computer to make copies in order to "text and data mine" information in digital actions. This applies at least to temporary, indirect or transitory copies.[93] In this way, every digital copy, temporary or permanent, directly or indirectly, has the potential to infringe copyright.[94] Having said that, when AI models make their copies, these are probably reproductions where there is a risk that these are covered by the exclusive right of reproduction.

3.4 Exceptions that EU Copyright framework confers

We have no seen that there are three exclusive rights to the protected works in article 2-4 in the Infosoc Directive. These exclusive rights also have exceptions in article 5, which we will have a closer look at below.

Article 5 (1) in the Infosoc directive provides an exception for temporary forms of reproduction. In order for this provision to be applicable, it is required that the reproduction is transient or incidental and constitutes an essential part of a technical process. It is further required that the reproduction has the purpose of enabling a transfer in a network between third parties via an intermediary, or a legal use of a work that has no independent economic significance. This article is an exception to the general principle that consent must be obtained from the right holder for any reproduction of a protected work.[95] As mentioned before, this consent is also required for parts of works that may constitute partial reproduction.[96]

The CJEU has described temporary acts of reproduction as a fleeting or subordinate use which make an integral and essential part of the technological process. The purpose shall moreover be exclusively to enable this transfer between third parties

[90] Ibid. para 43.

[91] Infosoc Directive, recitals 28-29.

[92] Case C-263/18 *Netherlands Uitgeversverbond and Goep Algemene Uitevers V Tom Kabinet Internet BV and others* [2019] ECLI:EU I-1111, paras 51-52.

[93] Margoni and Kretchmer (2022), p.689.

[94] Margoni (2018), p. 2.

[95] Case – 5/08 Infopaq, para 57.

[96] Ibid. para 48.

or the lawful use. Regarding that use, it is considered legal if it is done with the right holder's consent or is not prohibited by law.[97]

The use is considered "transient" if it exists for the time that is needed for the completion of a technical process. For example, this applies to the case where the process is automated in such a way that the form of reproduction is automatically deleted, without human intervention, as soon as it no longer fulfills any function for the implementation of the technical process. Furthermore, in case of temporary reproduction, the use should not exceed what is necessary for the technical process to work properly.[98] The use can, for example, be considered temporary in the event that the reproduction is deleted at the end of the process of an electronic search.[99]

The regulation also seems to include measures that make it possible for browsing and caching to take place.[100] The CJUE states that the exception covers acts that enable browsing as well as acts of caching to take place. This includes those that enable a transmission system to function efficiently. Even here they point out that such actions are created and deleted without human intervention.[101]

Additionally, the exception in the article is limited by Article 5(5) in the Infosoc directive. This part of the article constitutes what is called "the three-step test".[102] Article 5 (5) of the Infosoc directive states that this exception in the article shall only (1) be applied in special cases (2) that do not conflict with normal use of the work and (3) does not unreasonably infringe on the legitimate interests of the right holder. Since all these conditions according to the article must be met, these can be said to be 3 cumulative conditions. According to the CJEU, the exceptions and limitations must be interpreted restrictively concerning the wording of Article 5(5) and concerning the background that provisions constituting exceptions to a general principle laid down in the same directive must be interpreted restrictively.[103]

Article 5 in the Infosoc directive is the only mandatory limitation in the Directive. It applies, as known, to transient or incidental copies whose sole purpose is to facilitate the internet service provider's operations or certain legal uses.[104] It follows further exceptions to Article 5(2) - 5 (3) in the Infosoc directive with an exhaustive list of optional limitations for the Member States which otherwise apply to the rights of reproduction, communication to the public and distribution of the works.[105]It should also be mentioned that the conditions in Article 5 are also cumulative, which means that all the conditions must be fulfilled and if any of them are not fulfilled, the article does not apply.[106]

---

[97] Ibid. para 61.

[98] Ibid. para 60-61.

[99] Ibid. para 60.

[100] Infosocc Directive, recital 33.

[101] Case – C 5/08 Infopaq, para 63.

[102] Joao Pedro Quintais ' The New Copyright in the Digital Single Market: A Critical Look' European Intellectual Property Review 2020(1) (2019) < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3424770> p. 9 (cit Quintais (2019).

[103] Case C-435/12 *ACI Adam BV and Others V Stitching de Thuiskopie, Stiching Onderhandeligen Thuiskopie vargeoeding* [2014] ECLI:EU:C:2014:254, paras 22 and 24.

[104] Quintais (2019) p. 5.

[105] Ibid.,

[106] Case–C 5/08 Infopaq, para 55.

Finally, it is that the temporary copies in web pages that are made in the cache of computers or tablets are only possible because of this exception. Otherwise, surfing on internet could very likely constitute an infringement of copyright[107] In other words, it is the Article 5 (1) of the Infosoc directive that makes it possible to surf the Internet according to EU legislation.[108]

3.5 Summary and analysis

In this part, it has been examined what constitutes a protected work, what rights an author has to a protected work and what kind of exceptions there are.

A work that can receive protection is one that both constitutes a work and possesses originality and constitutes the author's own intellectual creation. Such a work is created by the author's choices, combination and disposition in it. Such a work is further protected by the exclusive right to reproduction, the exclusive right to communication to the public and the exclusive right to distribution. This in order to protect the author from exploitation of his work that interferes with his legitimate interests. In order to circumvent these exclusive rights, an approval by the rights holder is required or that the use goes under the exception in Article 5 in the Infosoc directive.

The right that primarily becomes relevant is the exclusive right of reproduction. This is because an AI system, such as Machine learning, needs to at least partially and temporarily reproduce works in the input and learning step. Having said that, an input of data is needed when creating such models in order to be able to get outputs later. In this way, the working method of an AI model becomes infringing on copyright interests right from the beginning. This is problematic since copyright fulfills an important function by giving authors incentives to create. That is why the author should receive fair compensation and that their work is not used in a way that infringes on their legitimate interests. This is what is achieved with the exclusive rights. From an AI development point of view, this is problematic because it limits the possibilities for using training data. It can be said that there are one/or several reproductions of works during the process of an AI model from input to output. This is because the input consists of existing works that are temporarily used by the training algorithm. In this way, one could argue for a kind of friction between the interest in developing AI and at the same time protecting copyright.

If we break this down on an individual level between a potential creator of an AI system and an author, the creator of the AI system wants to have access to works for his training of the AI model and on the other hand, the author does not want their work to be reproduced without their permission.

Just like the development of AI, there is an interest in maintaining copyright. In a similar way, this interest also extends from the individual level to the societal level. This applies to the creators' interest in being able to create works that receive

[107] Margoni (2018) s.2.
[108] Margoni (2018) s.17.

protection that later on can generate benefits for the general public. This can, for example, concern research, culture or innovations. Having said that, a high level of protection is needed and important for intellectual creativity and innovation that can benefit both the individual creator and the EU as a Union.

Concerning above there is kind of a breaking point between copyright and the development of AI. The AI models that can benefit many cannot be trained without data from the very beginning. Therefore, it is of interest for AI development to provide access to works that can constitute training data. On the other hand, if the author has no protection, the incentives to create works that can benefit many are reduced. Through this, the regulation gets a significant role in balancing these interests. It will be difficult to open up the use of existing works and at the same time not do it at the expense of the authors' rights.

What is clear through these two chapters is that both AI development and copyright are of interest on a societal level as well as an individual level. Hence, a kind of balance must be found. The next part will therefore take a closer look at how well articles 3 and 4 of the DSM directive balance these interests.

4 Does the DSM Directive balance copyright and AI?

4.1 The DSM Directive in general

The DSM Directive deals with copyright in the digital single market and includes 32 articles and 86 recitals. The DSM directive takes aim at copyright with particular concern to digital and cross-border use of protected content. In addition, the directive also contains provisions on exceptions and limitations of copyright in the digital cross-border environment as well as measures to facilitate licensing and ensure a functioning internal market for the exploitation of works and other materials.[109] Moreover, the DSM directive is intended to contribute to the harmonization of European legislation with the purpose to provide good conditions for a well-functioning internal market, stimulate innovation, creativity, investments and the production of new content in the digital environment as well.[110]

The technological development has resulted in new ways in which works and other protected works are created, produced, distributed and used. Therefore, light is shed on the importance that relevant legislation must be future-proofed in order to not to limit the technological development.[111] At the same time, through the development, new technology has created new areas of use in innovation and research, as well as new actors and business models.[112] This has tried to be handled in the directive by adapting and supplementing to the existing union copyright framework. This has also been done with consideration of a continued high level of protection.[113] The purpose of the exceptions and limitations in the directive is to achieve a fair use of rights and interests between rights holders and users. These exceptions and limitations are intended to be applicable only in special cases that do not conflict with the normal use of the work and do not unreasonably infringe on the legitimate interests of the right holder.[114]

However, it must be emphasized that the protection of technical measures established in the Infosoc directive is still of utmost importance to ensure protection and the effective exercise of rights granted to authors and other right holders.[115] As a counterweight to this, the digital environment has also led to an intensified cross-border use of works where new opportunities for consumers to access copyright-protected content have arisen. The proposal points out that the EU's regulation

[109] DSM Directive, article 1 and recital 3.

[110] Ibid., recital 2.

[111] Ibid. recital 3.

[112] DSM Directive, Recital 5 and Proposal for a regulation of the European and of the Council on copyright on the Digital Single Market, COM/2016/593 final [14.9.2016] p.1 (cit. DSM Proposal).

[113] DSM Directive, recital 3.

[114] Ibid., recital 6.

[115] Ibid., recital 7.

around copyright is still good but needs to be adapted to the new reality where exceptions and limitations in copyright are harmonized at union level.[116]

4.2 The exceptions in article 3 and 4 of the DSM Directive

In some cases, text and data mining may include works protected by copyright. This particularly applies to reproduction of works, which for example occurs when data is normalized within the framework of text and data mining. In those cases where there is no exception or restriction that can be applied, permission is required from the right holder to carry out such an act.[117]

In Article 3 and 4 respectively of the DSM directive, an exception is made in copyright for text and data mining. What is excluded by the exception is the exclusive right of reproduction in accordance with article 2(1) in the Infosoc directive.[118]

In the DSM directive, text and data mining is defined as an automated analysis technique used to analyze text and data in digital form to then generate information, including, but not limited to patterns, trends and correlations according to Article 2(2) in the DSM Directive. In other words, this could be seen as a computer-assisted analysis of information where text and data mining process large amounts of information with the aim of gaining new knowledge and discovering new trends.[119] Margoni and Kretchmer state that this definition is broad and appropriately identifies the potential of a tool that can analyze autonomously or semi-autonomously large amounts of data. One of the most used approaches at the moment to perform such a task today is machine learning which is a subclass of AI.[120] The definition found in the directive therefore captures most activities where technology is used to analyze information and extract data. Against this background, it can be argued that the definition found in the DSM directive is future-proof because it covers and regulates the most areas within AI/machine learning that are now known or developed in the future that rely on data analysis.[121] In addition, it should not be forgotten that there are types of text and data mining that can be based on simple facts or data that are not copyright protected. Sometimes there are also cases where the text and data mining does not involve reproduction or where the reproduction in question is covered by the exception for temporary reproduction as stipulated in article 5 (1) in the Infosoc directive. In the recitals of the DSM directive it is highlighted that Article 5 (1) should continue to apply to text and data mining which does not include copying that goes outside the scope of the exception. [122]

[116] DSM Proposal, p.2.
[117] DSM Directive. Recital 8.
[118] See Chapter 3.4.
[119] DSM Directive, recital 8.
[120] Morgoni and Kretchmer (2022) p. 687.
[121] Ibid.,
[122] DSM Directive, recital 9.

With that being said, the definition of text and data mining to Article 3-4 of the DSM Directive captures a wide range of automated data technologies used to analyze information in digital form.[123] This includes most modern and data-driven forms of AI such as traditional machine learning and more advanced forms of deep learning and neural network functions.[124]

4.2.1 Specific for article 3

Article 3 of the DSM Directive aims at text and data mining with the purpose of scientific research. In the first paragraph, it is stated that this exception covers reproductions and extractions that research organizations carry out for research purposes. This in order to carry out text and data mining of works to which they have legal access to in accordance with Article 3 (1). These copies of works must also be stored in a way that guarantees an appropriate level of security and shall be retained for research purposes, including for the verification of the research results in according to Article 3(2) of the directive. Furthermore, in accordance with Article 3(3) in the directive it must be added that the right holder shall have the right to take measures to ensure the security and integrity of networks and databases where the works are found. However, these measures shall not extend beyond what is necessary to achieve this goal. In article 3(4) the fourth and last paragraph of the article, it appears that Member States shall encourage right holders, research organizations and cultural heritage institutions to establish commonly agreed best practices regarding the application of the obligation and the measures referred to in paragraphs 2 and 3.

As stated above, this provision is for the benefit of research organizations and cultural heritage institutions. Article 2 of the DSM directive defines these terms. The term "Research organization" is defined in article 2(2) as a university or other research institute or any other entity whose primary purpose is to conduct scientific research. This in a non-profit way or where the profit is reinvested in its scientific research or to carry out a mission of public interest recognized by a member state. Furthermore, in article 2 (3) a cultural heritage institution is defined as a publicly accessible library or museum, an archive or a film or an audio heritage institution.

In the preamble to the directive, it is stated that research organizations cover a wide range of entities whose primary task is to conduct scientific research or to do that in combination with providing educational services. This must also, as mentioned, be done on a non-profit basis or be in the public interest recognized by a Member State. The term research organization also extends beyond universities and other institutions of higher education where hospitals and research institutes that conduct research also should be included for example.[125] The other type of user covered by the article is cultural heritage institutions. In the directive's preambles, it was considered that these should be regarded to include both public libraries and museums. This applies regardless of the type of work they hold in their permanent collections, such as archives and institutions for film and sound archives. The term also includes national libraries and national archives. This also applies to their

[123] Margoni and Kretchmer (2022) p.688.
[124] Margoni and Kretchmer (2022) p.689.
[125] DSM Directive, recital 12.

archives, publicly accessible libraries, educational institutions, research organizations and public broadcasting companies.[126]

The second limitation in the article is that research organizations and cultural heritage institutions must carry out the exception with a purpose of scientific research. What is meant by the term scientific purpose is not explicitly defined in the directive.[127] On the other hand, it should be if the aim is to conduct research and also do this without profit or within a mission of public interest recognized by the Member State.[128] The important thing is that the main purpose is to conduct scientific research or to conduct educational activities that also include scientific research.[129] However, it is stated in Recital 12 of the directive that natural science research and humanistic research at least should be covered.[130] Having said this, it can be about other types of research as well.

A further limitation in this exception is that the text and data extraction must be done on works to which research and cultural heritage institutions have legal access. The term legal use should be understood to include content to which they have access to through either an open access policy or through agreements with rights holders. This may, for example, concern subscriptions or other legal means. Furthermore, it is established that legal access also includes content that is freely available online.[131]

Furthermore, a responsibility is stipulated in the article for research- and cultural heritage institutions, and a right for rights holders, regarding the storage and use of the copies. The liability is based on the fact that the exceptions are limited to producing copies and therefore do not intend to make them available to the public. Having said this, the copies shall only be kept for such purposes as are necessary for text and data extraction purposes, which also includes verification of the research result.[132] These copies must be stored in a way that also guarantees an appropriate level of security. The right that the right holder receives through this article is to take measures in cases where there is a risk that the security and integrity of the systems or databases may be compromised.[133]This right to take measures extends as long as they are proportionate and applies even if it affects the user's ability to produce copies.[134]

To simplify this a bit, it can further be said that there is a double limitation in article 3 in the DSM directive which stipulates that the article is only applicable to research- and cultural heritage institutions and that the article is only applicable to scientific research within text and data mining. Therefore, this regulation is not applicable to commercial companies or universities that act with a purpose other than research.[135] In addition to this, contractual agreements that limit the exception in article 3 are not

[126] Ibid., recital 13.
[127] Swedish government Proposal 2022/22:278 ‘Upphovsrätten på den digitala inre marknaden’ [7.7.2022] p.45 (cit. Prop.2021/22:278).
[128] DSM Directive, recital 12.
[129] Prop. 2021/22:278 p.43
[130] DSM Directive, recital 12.
[131] Ibid., recital 15.
[132] DSM Directive article 3 and Prop. 2021/22:278 s. 45.
[133] Prop. 2021/22:278 s.48
[134] DSM Directive, recital 16.
[135] Margoni and Kretchmner (2022) p.694.

possible according to article 7 of the DSM Directive. On the other hand, The goal of Article 3 is to introduce a mandatory exception under the Union's copyright legislation that exempts reproduction acts and extracts performed by research organizations and cultural heritage institutions for the purpose of performing text and data mining in scientific research.[136] It should also be said, while this article opens up the use of certain types of text and data mining, any potential harm to the right holder would be minimal according to the preamble in the directive. This with regard to the nature and scope of the provision.[137]

4.2.2 Specific for article 4

Article 4 of the DSM Directive aims at text and data mining in a more general perspective compared to Article 3 of the Directive. Through this article, the exception shall be made for reproduction and extracts of legally available works for text and data mining purposes according to article 4(1) in the directive. In accordance with article 4(2) should these reproductions and excerpts further be retained as long as necessary for text and data mining purposes. This exception is conditioned by the fact that the use of works referred to above has not been expressly reserved by their rightsholder. This must have been done in an appropriate way through, for example, machine-readable methods in the case of content made publicly available online according to article 4(3). In the article 4(4) it is stated that this article shall not be affected by article 3 of this directive. What distinguishes this article from article 3 is that there is no user limitation or purpose limitation in this provision. At the same time, it is limited by the possibility for rights holders to contract out their works from the provision.

A requirement that the article sets is that these reproductions only are allowed to be kept as long as it is necessary for the purpose of text and data mining. Having said that, the storage of the copies is thus permitted as long as it is required within the context of text and data mining.[138] There does further not seem to be any requirement that these copies shall be stored in a certain way.[139]

Just as in Article 3, the exception in Article 4 applies as long as the user has obtained access to the works by legal means or when the works have been made available online.[140] Legal use is a use that is not reserved by the right holder or that is not restricted by applicable legislation.[141] Having said that, article 4 is further limited by the condition that the right holder not have reserved the right to make reproduction and extractions for text and datamining purposes in an appropriate way.[142] Regarding such works that have been made available online, it should be considered applicable

[136] Ibid., p.686.
[137] DSM Directive, recital 17.
[138] DSM Directive, recital 18.
[139] Prop. 2021/22:278 p.48.
[140] DSM Directive , recital 18.
[141] Margoni and Kretchmer (2022) p.692.
[142] DSM Directive, recital 18.

to preserve those rights only through machine-readable means, including metadata and terms of a website or service. Other use according to the recitals should not be affected by reservation of rights and conditions for text and data mining.[143] In other cases, it may be appropriate to reserve the rights in other ways. Agreements or unilateral declarations can be examples of such reservations. Rights holders should thus be able to act to ensure their reservation in this regard.[144] What such a reservation should look like is not defined in the legal text, but in the Swedish proposal it was said that how such a reservation should look should be judged in the individual case. However, it should be pointed out that the decisive factor should be that the reservation has been designed in such a way that both in terms of form and content it is possible for the user to receive it and understand its meaning.[145]

In the last paragraph of the article, it appears that the exception shall not affect the mandatory exception in article 3 prescribed in this directive. Furthermore, this provision shall not either affect the existing exception for temporary forms of reproduction provided in 5(1) of the Infosoc directive.[146]

In summary, unlike Article 3, Article 4 is not limited to certain users and is therefore potentially available to everyone. On the other hand, this article has an extension compared to article 3 in the form that right holders can "Opt-out" their work from the provision by a reservation. This "opt out" mechanism possibly makes the provision less effective. This since the number of available works may become narrow if many right holders choose to do so. In other words, Article 4 is actually available to all types of users for all types of purpose but can thus be "opt-out" by the right holders. This is in contrast to Article 3, which is mandatory.[147]

According to Geiger, this provision could have opened up the stricter regulation in Article 3 if it were not for the fact that the provision contained an "opt-out" mechanism that enables the holder to restrict the use. Geiger believes that by conditioning this exception on goodwill, one also risks that the effectiveness of the provision depends on whether the right holder chooses to "opt-out" their work or not.[148]

4.3 Summary and analysis

In this chapter we took a closer look at the purpose of the DSM Directive and the exceptions that are stipulated in Article 3 and 4 of the DSM Directive. In this final

[143] Ibid.,

[144] Ibid.,

[145] Prop. 2021/22:278 p.49.

[146] DSM Directive, Article 7(1) and recital 7.

[147] Margoni and Kretchmer (2022) p.686.

[148] Cristophe Geiger, 'The missing Goal-Scorers in the Artificial intelligence Team: Of Big Data. The Fundamental right to Research and the failed Text and Data Mining Limitations in the DSM Directive', Apphen aan den Rijn, Kluwer Law International. (2021) > https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829768< p.9.

part, it also has become time to analyse how well these two articles balance the interest of copyright protection and AI opportunities.

The exceptions in these articles open up the possibility of text and data mining, which is an integral part of most AI systems built around machine learning. In Article 3, this applies to research organizations and cultural heritage institutions. These must also carry out the text and data mining for a research purpose. On the top of that, this exemption is also mandatory and cannot be overridden by agreement. Article 4, on the other hand, covers text and data mining without targeting a specific group of users or a specific purpose. However, this article can be put out of action by the right holder "opting-out" his work. This in contrast to article 3.

It has become clear from the previous chapter that there is an interest in maintaining strong copyright protection and at the same time taking advantage of the possibilities of AI. Here the question is whether the DSM Directive balances these interests well.

As we have seen in the recitals of the directive, among other things, there is a seemingly high ambition to promote and maintain a strong copyright protection without limiting technological development. In the directive, action has been taken on this through the directive's articles 3 and 4 with the exception of text and data mining. The definition that has been set of “text and data mining” is broad and can therefore capture AI models. At the same time, this has been balanced with limitations such as specific users and purposes in Article 3 and the possibility for rights holders to "opt-out" their work in Article 4. This to the advantage of maintaining strong protection for copyright.

For article 3 of the DSM Directive, text and data mining must be carried out by research organizations or cultural heritage institutions with a scientific purpose. The term Research organizations covers a wide range of organizations that conduct research without profit or conduct research based on a public interest test recognized by a Member State. Cultural heritage institutions include public libraries or museums. In order to be covered by the term scientific purpose, the activity must be conducted without profit or be covered by public interest. This article is also mandatory which means that it cannot be taken out of play by rights holders. At the same time, there is a small opening for rights holders to take measures if the use takes place in a way that unreasonably interferes with their legitimate interests. Thereof, this article makes it possible to conduct research using AI systems. On the other hand, the article is limited to a certain type of user and to a certain purpose. For example, the article does not cover for-profit research organizations that also could contribute to science and innovation. This means that opportunities with AI in research can be lost.

Article 4 of the DSM Directive takes a more general approach by not being limited to a certain type of users and purposes. On the other hand, the exception here is conditional on the right holder not having explicitly reserved this type of use of the work. This further gives the rights holder the opportunity to choose whether their work will be applicable for Article 4 or not. This is the most decisive difference to Article 3 of the Directive. Having said that, Article 4 has the potential to promote the conditions for AI systems to develop, but this opening can also be closed by

rights holders. In practice, this could lead to there being few works for an AI model to be trained on and that we thereby miss out on important innovation.

In summary, it can be said that the reason and purpose of the directive promises a lot about not limiting the development of technology and at the same time maintaining exclusive rights for the rightholder. At the same time, a regulation has been made which in practice may not fulfil this goal.

With the limitations in Article 3 to users and purposes of use and the possibility for right holders to "opt-out" their works in Article 4, there is a risk that these regulations will not take advantage of the possibilities of AI. In conclusion, the interest in AI development and the interest in copyright have been balanced by adding these regulations. On the other hand, these articles may not be sufficient because they are limited to users and scope of use in Article 3. Article 4, which would open up the possibilities for technological development more, is instead dependent on rights holders.

5 Summary and conclusion

In this thesis, the interests behind AI development and copyright have been investigated. Furthermore, it has also examined how Articles 3 and 4 in the DSM Directive balance these interests. The research questions for the thesis to be answered here is;

1. In what way are the interests in AI development and copyright in conflict with each other?

2. Do articles 3 and 4 of the DSM directive have the ability to balance these interests?

It has become clear during the essay that AI has many possibilities and the interest in continued development covers a wide range. When an AI is developed, it may need to be trained with a data set consisting of works to learn, and these works have the potential to be copyrighted. For this reason, the focus of the essay has further been on understanding the copyright. Here, it has become clear that the works trained by AI can be covered by a protected work with exclusive right to reproduction. Just as there is a broad interest in AI development, there is also an interest in the copyright from a union level to an individual level. These interests are something that needs to be balanced, and that is why Articles 3 and 4 of the DSM Directive have been added to the EU copyright framework. These are exceptions to the exclusive right of reproduction in article 2 of the Infosoc Directive.

Regarding the first research question, An AI model needs an input of data in order to later learn to create outputs. The right that primarily becomes relevant is the exclusive right to reproduction. This is because an AI system, such as Machine learning, at least needs to do partial and temporary reproduction of works in the input and learning step. In this way, the working method of an AI model is already in its basic construction violating copyright interests.

Further in the essay, it has become clear that both the interest in AI development and the interest in copyright covers a wide spectrum. The AI models that can benefit many cannot in its most basic structure be trained without data. Therefore, it is of interest for AI development to provide access to works that can constitute training data. On the other hand, if the author has no protection, the incentives to create works that can benefit many are reduced. So, there is both an interest in opening up the use of works but there is also an interest in protecting these works with copyright. Therefore, this applies both to the relationship between the author and the creator of an AI system and even to the relationship of AI development and copyright in the Union.

It is difficult to open up the possibilities of using works without it being at the expense of rights holders. It is also difficult to not open up the use of works without it being at the expense of AI development. This is a problem that shows the breaking

point between copyright and the development of AI. Hence, this needs to be balanced and the thesis has also examined how well articles 3 and 4 of the DSM directive balance these interests.

Regarding the second research question, the recitals and purpose of the DSM directive promise a lot. On the other hand, articles 3 and 4 of the directive may not fully respond to this.

As we have seen in the thesis, the recitals of the directive show a high ambition to maintain a strong copyright without limiting technological development. In the directive, action has been taken on this through the directive's article 3 and 4, with the exception for text and data mining. We have also seen that the definition they have set of text and data mining is broad and can therefore capture AI models. At the same time, this has been balanced with limitations such as specific users and purposes in Article 3 and the possibility for rights holders to "opt-out" their work in Article 4. The latter is to the advantage of maintaining strong copyright protection.

Article 3 is limited by the fact that the user is a research organization or cultural heritage institution. In addition, this article is also limited by a research purpose which makes it narrower. At the same time, it is also mandatory, which means that rights holders cannot make this article non-applicable by agreement. Article 4, on the other hand, has a more general and broader target group of users and has no requirement for a specific purpose either. However, there is the possibility for rights holders to "opt-out" their works from the provision. This means that this article will not be so broad in case many rights holders choose to “opt-out” their works. Having said that, this article becomes dependent on how rights holders act and potentially the range of works can become narrow. With the limitations in Article 3 to users and purposes of use and the possibility for rights holders to “opt-out” their works in Article 4, it is possible that these articles do not take advantage of the possibilities of AI.

Last but not least, the interest in AI development and copyright has been balanced by adding these regulations in Article 3 and 4 of the DSM Directive. On the other hand, these articles may not be sufficient because they are limited to users and purposes in Article 3 and dependent on rights holders in Article 4.

With these considerations, the conclusions for the research questions can be drawn. For the first research question, the interest in AI development and copyright covers a wide spectrum from the individual to the general public. Here, AI development and its developers want a more open access to works and the rights holders want a continued strong protection. Regarding the second research question, it can be concluded that Articles 3 and 4 balance these interests but in a way that may not be sufficient. The use of works is opened up but limited. However, this may result in losing benefits generated by AI in the future.

References

Official documents

European Commission

Consolidated Version Of The Treaty On The Function Of the European Union, OJ C 326, 26.10.2012

Consolidated Version Of The Version Of The Treaty On European Union, OJ C 326, 26.10.2012

Charter Of Fundamental Rights Of The European Union, OJ C 364, 18.12.2000

Council Directive 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9EC and 2001/29/EC [2019] OJ L130/92

Council Directive 2001/29/EC of May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10

Proposal for a regulation of the European Council on laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain Union legislative acts, COM(2021) 206 Final

European commission, Communication from the commission, Shaping the Europe's Digital future, COM(2020) 67 Final

European Commission, 'White paper on Artificiall intelligence – A Approach to excellence and trust, COM(2020) 65 Final

International Treaties

The Berne Convention For the Protection of Literary and Artistic Works, (Paris Act, 1971 and Amended in 1979)

WIPO Copyright treaty (WCT), (adopted in Geneva 1996)

International official documents

Swedish Proposal for copyright in the digital single market, Prop.2021/22:278

Literature

Johan Axhamn, *"Copyright and Artificial Intelligence – with a focus on the area of music"* in Festskrift til Jørgen Blomqvist (ExTuto publishing, 2018)

Ulf Bernitz, "Marknadsföringsrätten: Svensk och europeisk marknadsföringsrätt (2 th edn, Nordstedts Juridik, 2020)

Christophe Geiger, *"The missing Goal – Scorers in the Artificial intelligence Team: Of Big Data, The Fundamental right to research and the failed Text and Datamining Limitations in the CDSM Directive"* (Apphen aan den Rijn, Kluwer Law International, 2021) Available at: https://ssrn.com/abstract=3829768

Jörgen Hettne and Ida Otken Eriksson I (ed.), "EU – Rättslig metod" (2th edn, Nordstedts Juridik, 2011)

Mireille va Echod, Pierre Bernt Hugenholtz, Stef van Gompel, Lucie Gubault and Natalie Helberger, *"Harmonizing European Copyright Law: The Challenges of better law making'"*, (International Law Series, Volume 19, Alphen aan den Rijn: Kluwen Law International 2009 (chapters 1 and 9), Amsterdam Law School Research Paper No.2012-07, Institute for information Law Research Paper No.2012-07, 2012) Available at: https://ssrn.com/abstract=2049935

Thomas Margoni, *"Artificiall Intelligence, Machine learning and EU copyright law: Who owns AI?"*, (CREATe Working paper, 2018) Available at: https://ssrn.com/abstract=3299523

Thomas Margoni and Martin Kretchmer, "A *deeper look into the EU Text and Datamining Exceptions: Harmonization, Data Ownership and the future of technology*" (GRUR International, Vol 71, Issue 8, 2022). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886695

Maria Nääv and Mauro Zamboni(ed), "Juridisk metodlära (2th edn., Studentlitteratur", 2018)

Joao Pedro Quintais, *"The New Copyright in the Digital Single Market: A Critical Look* (Intellectual Property Review (2020(1), 2019) Available at: https://ssrn.com/abstract=3424770

WIPO, "*WIPO - Technology trends 2019: Artificial Intelligence*" (Worlds Intellectual property organisation, 2019)

Online sources

Alex Hughes, " BBC Science Focus – CHATT GPT: Everything you need to know about OpenAI's GPT4 tool, (BBC,2023) Available at https://www.sciencefocus.com/future-technology/gpt-3 Accessed: 27.4.2024

Andres Guadamuz, "WIPO MAGAZINE – Artificial intelligence and copyright (WIPO, 2017) Available at: https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html Acceded: 18.4.2024

European Parliament, “Artificial intelligence: Threats and opportunities” (European Parliament (2020) Available at: https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities Accessed: 24.4.2024

European Commission, “Shaping Europe’s Digital future – The EU copyright legislation’ (European Commission, 2024) Available at: https://digital-strategy.ec.europa.eu/en/policies/copyright-legislation Acceded: 22.4.2024

Europan Parliament, “What is artificial intelligence and how is it used” (European Parliament, 2020) Available at: https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used Accessed: 24.4.2024

Jessica Fjeld and Mason Kortz, A legal anatomy of AI- generated Art:Part 1”, (Harward journal of Law & Technology, 2017) Available at: https://jolt.law.harvard.edu/digest/a-legal-anatomy-of-ai-generated-art-part-i Accessed: 25.4.2024

Case law

Court of Justice of the European Union

Judgement of 19 December 2019, *Netherlands Uitgeversverbond and Goep Algemene Uitevers V Tom Kabinet Internet BV and others*, Case-263/18, EU:C:2019:1111

Judgement of 13 November 2018, *Levola Hangelo V Smilde Foods BV*, C-310/17, EU:C:2018:899

Judgement of 10 April 2014, *ACI Adam BV and others V Stitching de Thuiskopie, and Stitching Onderhandeligen Thuiskopie vergoedig*, C-435/12, EC:2014:254

Judgement of 1 December 2011, *Eva – Maria Painer V Standard VerlagsGmBH and others,* C-145/10, EC:C:2011:789

Judgement of 16 July 2009, *Infopaq A/S V Danske Dagbladets Forening*. C-5/08, EU:C:2009:465

Judgement of 7 December 2006, *SGAE vs Rafael Hoteles SA*, C-306/05, EU:C2006:764