REPUBLIC OF TURKEY

ANKARA UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

DEPARTMENT OF INTELLECTUAL PROPERTY, TECHNOLOGY POLICIES AND INNOVATION MANAGEMENT

# FROM WORKS TO DATA – ARTIFICIAL INTELLIGENCE AND COPYRIGHT INFRINGEMENTS

Master's Thesis

BANU EYLÜL YALÇIN

Ankara, 2022

**REPUBLIC OF TURKEY**

**ANKARA UNIVERSITY**

**GRADUATE SCHOOL OF SOCIAL SCIENCES**

**DEPARTMENT OF INTELLECTUAL PROPERTY, TECHNOLOGY POLICIES AND INNOVATION MANAGEMENT**

MASTER OF LAWS (LL.M.) in INTELLECTUAL PROPERTY

# FROM WORKS TO DATA – ARTIFICIAL INTELLIGENCE AND COPYRIGHT INFRINGEMENTS

**Master's Thesis**

**Banu Eylül YALÇIN**

Assist. Prof. Dr. Zehra ÖZKAN ÜNER

Supervisor

**Ankara, 2022**

**DECLARATION**

**REPUBLIC OF TURKEY**

**ANKARA UNIVERSITY**

**To the Directorship of Graduate School of Social Sciences**

I hereby, declare that all information in my master's thesis titled "From Works to Data – Artificial Intelligence and Copyright Infringements, (Ankara, 2022)" which has been prepared under the supervision of Assist. Prof. Dr. Zehra ÖZKAN ÜNER has been gathered and submitted in compliance with academic rules and ethical conduct principles and as required by these rules and principles, I have fully indicated and cited all sources that are not original to this work. I also declare that, I have acted according to scientific research and ethical rules during the study process and if it is proven otherwise, I will accept all legal consequences.

Date: …/…/….

**Banu Eylül Yalçın**

TABLE OF CONTENTS

LIST OF ABBREVIATIONS ........ 4
INTRODUCTION ........ 6

CHAPTER ONE
ARTIFICIAL INTELLIGENCE AND COPYRIGHT

I. ARTIFICIAL INTELLIGENCE ........ 10
A. Definition of Artificial Intelligence and its Brief History ........ 10
1. Definition of Artificial Intelligence ........ 10
2. A Brief History of Artificial Intelligence ........ 11
B. Different Kinds of Artificial Intelligence ........ 13
C. Machine Learning ........ 15
1. Machine Learning Stages ........ 17
a. Corpus Compilation Stage ........ 18
b. Corpus Preprocessing Stage ........ 18
c. Computational and Statistical Analytical Processing ........ 20
d. Knowledge Discovery ........ 21
D. Legal Status and Authorship of Works Generated by Artificial Intelligence ........ 23
II. COPYRIGHT ........ 25
A. European Union ........ 27
B. United States of America ........ 33
C. Türkiye ........ 36

CHAPTER TWO
EXCEPTIONS AND LIMITATIONS

I. INTRODUCTION ........ 42
II. INTERNATIONAL INSTRUMENTS ........ 43
III. APPROACH OF THE EUROPEAN UNION ........ 46
A. Exception for Temporary Acts of Reproduction ........ 47
B. Text and Data Mining Exceptions ........ 50
1. Research Exception ........ 52
2. Article 4 of the Digital Single Market Directive ........ 53
IV. APPROACH OF THE UNITED STATES OF AMERICA ........ 57
A. Purpose and Character of the Use ........ 59
B. The Nature of the Copyrighted Work ........ 64
C. Amount and Substantiality of the Portion Copied ........ 65
D. Effect Upon the Potential Market or Value of the Copyrighted Work ........ 66

V. APPROACH OF TÜRKİYE ... 68
VI. EXTENDED COLLECTIVE LICENSING – POSSIBLE SOLUTION? ... 70
A. European Union ... 72
B. United States of America ... 74
C. Türkiye ... 75
D. Is Extended Collective Licensing a Possible Solution? ... 76

CONCLUSION ... 78
ABSTRACT ... 82
ÖZET ... 94
BIBLIOGRAPHY ... 82

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| **AI** | : | Artificial Intelligence |
| **API** | : | Application Programming Interface |
| **Berne Convention** | | Berne Convention for the Protection of Literary and Artistic Works |
| **CJEU** | : | Court of Justice of the European Union |
| **CMO** | : | Collective Rights Management |
| **DL** | : | Deep Learning |
| **DSM Directive** | : | Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market |
| **Draft Law** | : | Draft Law Regarding Making Amendments in the Law No. 5846 on Intellectual and Artistic Works |
| **ECL** | : | Extended Collective Licensing |
| **etc.** | : | Etcetera |
| **EU** | : | European Union |
| **HDL** | : | Hathitrust Digital Library |
| **i.e.** | : | *id est* (in other words) |
| **Ibid** | : | *Ibidem* (in the same place) |
| **Infopaq** | : | Infopaq International A/S v. Danske Dagblades Forening (2009) Case no. C-5/08. Court of Justice of the European Union, ECR I-06569. |

| | | |
|---|---|---|
| **Infosoc Directive** | : | Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society. |
| **IP** | : | Intellectual Property |
| **LIAW** | : | Law 5846 on Intellectual Property and Artistic Works |
| **ML** | : | Machine Learning |
| **para.** | : | Paragraph |
| **p.** | : | Page |
| **pp.** | : | Pages |
| **TDM** | : | Text and Data Mining |
| **TRIPS** | : | Agreement on Trade-Related Aspects of Intellectual Property Rights |
| **US** | : | United States of America |
| **WIPO** | : | World Intellectual Property Organization |
| **WCT** | : | World Intellectual Property Organization Copyright Treaty |
| **WTO** | : | World Trade Organization |

## INTRODUCTION

AI has been evolving for years; and it is now reaching a critical point in its growth and implementation. Advanced techniques such as deep learning and the advancement in available data and computer processing power allow the AI technology to be applied in many areas of our lives.

The research by PwC, a data analyst company, has revealed that "*AI could contribute up to USD 15.7 trillion to the global economy in 2030, more than the current output of China and India combined*"[1].

[1] Sizing the Price What's the real value of AI for your business and how can you capitalise? 2017. Available at https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html [Accessed January 25, 2022].

AI is everywhere. AI systems can be used to identify an x-ray as cancerous or not[2], recognize speech[3] as it is the case for Siri, Alexa etc., translate languages or predict traffic[4]. AI systems can even write novels[5], paint portraits or compose songs[6].

To be capable of facilitating these examples, a technology named "machine learning" is utilized. As the name itself suggests, with this technology machines "learn". But how? We, as humans, learn from our experiences, whereas machines learn from data. For instance, an AI project analyzed 100.000 news reports in 65 different languages in

---

[2] Erdaw, Y. & Tachbele, E., 2021. Machine learning model applied on chest X-ray images enables automatic detection of covid-19 cases with high accuracy. *International Journal of General Medicine*, Volume 14. Available at: https://www.dovepress.com/getfile.php?fileID=73078 [Accessed January 25, 2022].

[3] Gordon, C., 2021. A market to harness: Speech recognition artificial intelligence (AI) innovations on the rise. *Forbes*. Available at: https://www.forbes.com/sites/cindygordon/2021/12/23/a-market-to-harness-speech-recognition-artificial-intelligence-ai-innovations-on-the-rise/?sh=368012a8134d [Accessed January 25, 2022].

[4] Navarro-Espinoza, A. et al., 2022. Traffic flow prediction for smart traffic lights using machine learning algorithms. *Technologies*, 10(1). Available at: https://www.researchgate.net/publication/357743745_Traffic_Flow_Prediction_for_Smart_Traffic_Lights_Using_Machine_Learning_Algorithms [Accessed January 25, 2022].

[5] An AI written novel named "The Day A Computer Wrote a Novel" was created by researchers at the Kimagure Artificial Intelligence Writer Project and it passed the first round of a Japanese literary contest. See Jozuka, E., 2016. A Japanese AI almost won a literary prize. *VICE*. Available at: https://www.vice.com/en/article/wnxnjn/a-japanese-ai-almost-won-a-literary-prize [Accessed January 25, 2022].

[6] Ralston, W., 2005. Copyright in Computer-Composed Music: HAL Meets Handel. *Journal of Copyright Society* U.S.A, 52.

order to predict virus outbreak patterns[7]. For the Microsoft system to create literary works a training corpus involving 1.027 Chinese poems was used[8]. For the AI to become a "painter", 15.000 portraits are fed into the machine, which are taken from the online encyclopedia "WikiArt"[9]. AI has even been used to complete the Schubert's "Unfinished Symphony"; by analyzing the essential musical components of the first two movements, a new melody for the missing third and fourth movements were created[10].

With regards to intellectual property law, the more popular discussion on AI focused on whether the "output" generated by AI can be considered as "work" and, if the response is in the affirmative, who will be the "author" of these works. For the sake of the integrity of the subject and of the discussions surrounding copyright and AI, the legal status of AI-generated works and the concept of AI-authorship will be metioned briefly. However, the main focus of this study is on the "inputs" of AIs. As the examples show, for the training process of the AI, large sets of data are needed. This raises the legal

[7] Prosser, M., 2020. How AI Helped Predict the Coronavirus Outbreak Before it Happened. *Singularity Hub*. Available at: https://singularityhub.com/2020/02/05/how-ai-helped-predict-the-coronavirus-outbreak-before-it-happened/ [Accessed January 25, 2022].

[8] Gershgorn, D., 2018. A Microsoft chatbot composes poetry by looking at photographs. *Quartz*. Available at: https://qz.com/1354736/a-microsoft-chatbot-composes-poetry-by-looking-at-photographs/ [Accessed January 25, 2022].

[9] Anon, 2018. Is artificial intelligence set to become art's next medium?: Christie's. Available at: https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx [Accessed January 25, 2022]. See also Nugent, C., 2018. Can artificial intelligence create art? buyers think so. *Time.* Available at: https://time.com/5357221/obvious-artificial-intelligence-art/ [Accessed January 25, 2022].

[10] Huawei Presents 'unfinished symphony'. Available at: https://consumer.huawei.com/au/campaign/unfinishedsymphony/ [Accessed January 25, 2022].

question if these datasets involve unauthorized copies of copyrighted works, which is often the case if there is a “creative” AI-output.

Within the scope of this study, the definition of AI and its development will be discussed and the differences between weak and strong AI will be mentioned. In this context, the concepts of machine learning and deep learning will be described. Then, the stages of machine learning will be explained and potential copyright infringements during these stages will be discussed within the scope of the European Union, United States of America, and Türkiye regulations. Within this scope, possible exceptions will be reviewed and a comparison of exceptions among the European Union, United States of America, and Türkiye will be made. Lastly, it will discussed whether creating an extended collective licensing system could be a possible solution.

# CHAPTER ONE

# ARTIFICIAL INTELLIGENCE AND COPYRIGHT

## I. ARTIFICIAL INTELLIGENCE

### A. Definition of Artificial Intelligence and its Brief History

#### 1. Definition of Artificial Intelligence

AI has been a part of our lives for many years. Elevators can determine whether people are standing in the doorway and warn them to clear away or cars can warn people when the doors are not locked etc.[11]. Although AI is used widely on daily matters it still has no definition that is universally accepted. As a matter of fact, in November 2018 the AI Group of Experts created by the "Organisation for Economic Co-operation and Development" formed a subgroup with the task of providing a definition of AI system[12]. At the same time, the following definitions can be found in the literature:

- Technically, it constitutes a part of science whose purpose is to produce intelligent machines which can act like human beings[13].
- It is a computer system which has the ability to carry out tasks such as speech identification, visual sensation, decision-making and translation of languages which are acts usually necessitating human intellect[14].

---

[11] Mukherjee C., 2014. *Artificial Intelligence: What Everyone Needs to Know.* s.l. Lulu.com. p.4.

[12] Ibid.

[13] Moore, J. C., 2018. *Artificial Intelligence: What Everyone Needs to Know Today About Our Future,* CreateSpace Independent Publishing Platform.

[14] Nicholson, C.V., 2019. *Skymind Part 1: An Introduction to Artificial Intelligence.* [online] Skymind Inc. Available at: <https://skymind.ai/ebook/Skymind_Introduction_to_Artificial_Intelligence.pdf> [Accessed 26 November 2019], p.3.

- A system which is composed of three essential parts: sensors, operational system of logic and actuators[15].

In the broadest sense, AI can be defined as computer functionality designed through an interdisciplinary study by applying various principles and tools of computation, mathematics, logic, mechanics and biology[16]. AI, also referred to as synthesized intelligence, consists of algorithms for understanding, modeling and imitating human intelligence and cognitive processes[17]. AI has various sub-uses such as; natural language processing, robotics, machine learning, expert systems, speech and voice recognition, intelligence automation, and computer vision[18].

## 2. A Brief History of Artificial Intelligence

In 1950, Alan Turing authored an article in which he imagined the possibility of computers programmed for simulating intelligence[19]. In the paper named *Computing Machinery and Intelligence*[20], among other elements of AI, Turing examined and

---

[15] OECD., 2019, *Artificial Intelligence in Society.* [online] Paris: OECD Publishing. Available at: < https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en> [Accessed 26 November 2019], p.22.

[16] Frankish K., Ramsey W. M., The Cambridge Handbook of Artificial Intelligence, CUP, Cambridge-UK, 2014, p. 1.

[17] Ibid.

[18] For further information see Gollapudi S., 2019. Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs, Appress, Berkley-USA, pp. 7-29.

[19] Stone, P., 2016. *Artificial Intelligence and Life in 2030, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Stanford University, p. 50.

[20] Turing, A. M., *Computing Machinery and Intelligence*. 1950. Oxford University Press on behalf of the Mind Association. [online] Available at: <https://phil415.pbworks.com/f/TuringComputing.pdf> [Accessed 27 November 2019]

discussed how machines might learn automatically and how intelligence might be assessed[21]. He foresaw that it will be possible to mention of a thinking machine and he formulated an experiment named as "Turing Test" that would allow a computer to be evaluated as a conscious and thinking machine[22]. The ideals of Turing could not be put into action due to the deficiencies in the recourses of computing, however his ideas inspired AI inevitably[23].

Another important name in the AI history is John McCarthy. He was a computer scientist at Stanford University and he is the man behind the term "Artificial Intelligence" and is considered as the father of AI[24]. It is thought that the first steps of AI were formally taken at a conference organized by McCarthy in 1956[25]. He gathered the best researchers from different fields to discuss AI and although everyone agreed that AI was possible, due to not being able to agree on standard methods, the conference was not so successful as McCharty expected[26].

---

[21] Stone, p. 50.

[22] O'Regan, G., 2008. *A Brief History of Computing.* Springer Science & Business Media, p.53.

[23] Stone, p. 50.

[24]Stanford University, Contribution and Impact. [online] Available at:< http://jmc.stanford.edu/contributions/index.html> [Accessed 27 November 2019].

[25] Stone, p. 50.

[26] Harvard University Science in the News, 2017. *Special Edition on Artificial Intelligence: History of Artificial Intelligence*. [online] Available at: <http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> [Accessed 27 November 2019].

Since 1956, the interest in AI differed, first it flourished but then huge funds were lost in the area called "AI Winter"[27]. The imagined success of AI in the 1950s became a reality thanks to Internet which provided the necessary data to train AI systems[28].

Due to the accessibility of big data, cloud computing and enhanced storage capacity, the impact and extent of AI has been increased over the years[29]. The growth of AI does not only happen in computer science, it is also effective in fields such as health, medicine, biology and finance[30].

## B. Different Kinds of Artificial Intelligence

The distinction between AI techonoliges relies on the different algorithms of AI and their ability to perform certain acts. There are two main ways to classify AI. According to the first classification, AI is divided into two types as "strong AI" and "weak AI."

For a very long time, AI systems were not completely self-governing and therefore not actually "intelligent"[31]. Weak AI, contrary to strong AI, uses software which does not include the entire range of human cognitive skills[32]. Simply put, they perform acts which are destined by humans and do not have the ability to learn by themselves.

---

[27] Lawal, A.B., 2019. *Artificial Intelligence Fundamentals: The Foundations & History of Intelligent Machines.* AB Prominent Publisher, p.3.

[28] Stone, p. 51.

[29] OECD, p. 23.

[30] Ibid.

[31] Yanisky-Ravid, S., 2017. *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model*. Michigan State Law Review, Volume: 2, Issue 4, p. 674.

[32] Kumar, E., 2008. *Artificial Intelligence*. I. K. International Pvt Ltd, p.14.

A chess program can be given as weak AI software[33]; Siri, the smart assistant of Apple; applications that perform Internet searches, detect faces and perform translations; smart houses; autonomous cars are also examples of weak AI[34].

Through long-lasting efforts of researchers, AI has gained its current reputation of being intelligent and able to learn from its own experiences. Strong AI is able to conduct any intellectual task that a human being is able to do, meaning that it has a human-like intelligence[35]. Strong AI has the ability to reason, take decisions, plan, learn, communicate and to incorporate these abilities to reach a goal[36].

The second and a more detailed classification of AI has four types: reactive AI, limited memory AI, theory of mind AI, and self-aware AI[37]. "*Reactive AI*" specializes in only one area and it does not have the ability to interact socially like humans. However, it perceives the elements in the contextual environment on which it is built and takes action to react accordingly[38]. IBM's chess-playing AI named "Deep Blue" is one of the examples belonging to this category.

"*Limited Memory AI*" has a limited memory capacity to make appropriate decisions and applies current data using the limited amount of past experience it is programmed to[39]. Autonomous cars are representative of his category.

---

33 Ibid.

34 McPherson, S.S., 2018. *Artificial Intelligence: Building Smarter Machines*. Twenty-First Century Books, p. 8.

35 Ryan, M., 2014. *The Digital Mind: An Exploration of Artificial Intelligence*. Micheal Ryan, p. 6.

36 Goodman, M., 2015. *Future Crimes: Inside the Digital Underground and the Battle for Our Connected World*. London: Random House, p. 479.

37 Gollapudi, p. 4. Also See, Yilmaztekin, H.K., 2020. Yapay Zeka Eser Sahipligi, Galatasaray Üniversitesi Hukuk Fakültesi Dergisi, pp.1521-1523.

38 Gollapudi, p. 4.

39 Ibid., pp. 4-5

Third type of the first AI classification is "*theory of mind AI*", which is capable of understanding the thoughts and emotions that impact human behavior[40]. This kind of AI has the ability to learn actively and can engage socially. Today, its development is not yet completely finished.

Last type of the second classification category is "*self-aware AI*". As the name suggests, this type of AI is aware of its environment and existence. It is a upgraded version of "theory of mind AI" but different from that, it has intuition and consciousness. It can recognize its inner world, can make predictions about the feelings of those around them; can make abstractions and inferences. In terms of this type, a prototype has not yet been developed; but it is envisaged that in the future machines will be like this[41].

### C. Machine Learning

Being a branch of computer science, AI is also related to psychology, neuroscience, biology, mathematics, sociology and philosophy[42]. The term AI is usually used as an umbrealla term that includes a variety of approaches and algorithms.

Machine learning is one of the techniques of AI. Although AI and machine learning are two different concepts, in this study, AI and machine learning will be used interchangeably as the study only focuses on the machine learning technique of AI.

---

[40] Bozkurt Yüksel, A.E., 2018. Yapay Zekanın Buluşlarının Patentlenmesi, Uyuşmazlık Mahkemesi Dergisi, pp. 590- 591.

[41] Yilmaztekin, p. 1522

[42] Bozkurt Yüksel, p. 588

Machine learning is an area of study that aims at giving computers the capacity to learn and develop from experience without being expressly programmed[43]. This includes the ability of the system to obtain new information and skills as well as the ability to reconstruct the existing knowledge based on the new information acquired[44].

Current machine learning is a computational proceeding that begins with a large amount of data and attempts to deduce a rule or procedure that interprets the data[45]. However in the previous "*expert system*", programmers had to sit with experts to study the rules and measures used to make decisions and translate those rules into software code; which is a system contradictory to the process of learning from data[46].

Another technique under the AI is called deep learning, which is one of the machine learning techniques[47]. In other words, DL is a subsection of machine learning using multilayered (deep) architecture to detect the links between inputs and outputs[48]. Besides, it employs structures generally similar to human brain, comprising of a set of modules or neurons[49].The reduction of information presented in a complicated format

[43] Nagy, Z., 2018. *Artificial Intelligence and Machine Learning Fundamentals: Develop Real-World Applications Powered By The Latest AI Advances*. Birmingham: Packt Publishing Ltd, p. 2.

[44] Saloky,T., Šeminský, J., 2005. *Artificial Intelligence and Machine Learning*. [online] Available at:< http://uni-obuda.hu/conferences/SAMI2005/SALOKY.pdf> [Accessed 29 November 2019]

[45] Executive Office of the President, 2016. *Preparing For the Future of Artificial Intelligence.* [online] Available at: <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> [Accessed 29 November 2019].

[46] Ibid.

[47] WIPO, 2019. *WIPO Technology Trends 2019: Artificial Intelligence*. [online] Geneva: World Intellectual Property Organization, p.14.

[48] Wiley, J.F., 2016. *R Deep Learning Essentials*. Birmingham: Packt Publishing Ltd, p.2.

[49] Executive Office of the President, p. 9.

(such as pixels) to simple explainable labels (such as an item group) by layering the processing of information is included in the DL[50].

Google used DL in over 100 services such as Street View, Inbox Smart Reply and Voice Search[51]. Following these services, Google improved the quality of its translation programme through the DL process in 2016. Nearly in one night, language translation changed from a fragmented random mixture of phrases to flawless sentences[52]. Before the adapted DL methods, computers looked for word combinations that could be translated together, however DL searches for correlation within the whole sentence[53].

### 1. Machine Learning Stages

Due to the different kinds of ML techniques and practices there is no general illustration of how ML works. The preferred ML technique is reliant on the aim and the circumstances surrounding analysis on the data. However, it is possible to name four common stages: 1-corpus compilation stage, 2- corpus preprocessing stage, 3-computational and statistical analytical processing and 4-knowledge discovery stage[54]. This section will briefly explain the necessary stages in ML workflow. The focus will only be on the main aspects, primarily those relevant from a copyright point of view.

---

[50] Calo, R., Evtimov, I., Fernandes E., and Kohno, T., O'Hair, D., 2018. *Is Tricking a Robot Hacking?*, University of Washington Tech Policy Lab, Legal Studies Research Paper No: 2018-05, p.4.

[51] Sejnowski, T.J., 2018. *The Deep Learning Revolution*. Massachusetts: The MIT Press, p.7.

[52] Ibid.

[53] Ibid.

[54] Chiou, pp.402-404.

### a. Corpus Compilation Stage

The ML process usually requires making digital copies of the data to assemble training examples for the machine. To create the corpus, the works are digitalized, accumulated and then they are stored as data.

For instance, AI bots which are used in customer support services, home appliances, phones or watches communicate through voice. To do all these interactions the AI learns how to speak. But how do these computers learn the human language? Like humans storing information in the relevant part of the brain, AIs also store training information to learn. In terms of language learning, the dataset generally consists of linguistic resources such as books which are classified and collected in order to capture specific elements of the relevant language[55].

This stage of the ML process could cause the infringement of authors' reproduction right, if necessary authorizations are not obtained and any of the exclusions foreseen by the law are not applicable.

### b. Corpus Preprocessing Stage

After the first stage, once the works are digitized, they need to be converted into a machine readable format, because generally, the works constituting the training dataset are found in a number of formats, depending on how they were created[56]. For the AI to process them, they all have to be converted into a uniform format[57]. Through this stage

---

[55] Margoni, T., section II.

[56] Wlodarczak, P., 2020. *Machine Learning and its Applications*. CRC Press.

[57] Indurkhy, N., Zhang, T. and Weiss, S., 2010. *Fundamentals of Predictive Text Mining*. Springer, p.15.

of the process, the works are pre-processed to standardize them into a format that is readable by the machine (for instance, converting the PDF documents into XML)[58].

This stage of the process includes also adaptive use of the works since the works are converted to another form. Nevertheless, it is argued in the copyright law realm that this would not fall under the category of adaptation as they do not bear originality[59].

Once the works are copied (i.e. digitalized), they are usually gathered in a database called the "training dataset of the project". This database is preserved within a different location, resulting once more in the reproduction of works[60].

Lastly, depending on the training method, the training data may be subject to labelling. In the "supervised learning" approach, the labelling of the data is necessary whereas for the "unsupervised learning" the process relies on unlabeled data[61]. For instance, if the aim of the ML process is for the machine to recognize cats, a data set containing cat pictures is fed into the machine. However, in supervised learning approach, each picture is labelled with the information indicating whether there is a cat in the picture or not. The machine is then able to identify from an unlabeled picture whether there is a cat in it. This method might be used in a self-driving car to stop in order to avoid to run

[58] Geiger C., Frosio G., & Bulayenko O., 2018b. Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market' in Seuba X., & Geiger C., & Pénin J. (eds.), Intellectual Property And Digital Trade In The Age Of Artificial Intelligence And Big Data, Ceipi/Ictsd publication series on "Global Perspectives and Challenges for the Intellectual Property System", Issue No. 5, Geneva/ Strasbourg, pp. 97-111.

[59] Chiou, pp.402.

[60] ibid. See also Geiger et al. (2018b) p.98.

[61] For more information see: Shalev-Shwartz, S. and Ben-David, S., 2014. *Understanding Machine Learning From Theory to Algorithms*. New York: Cambridge University Press, pp.22-24.

over cats on the street.[62] Currently, this method of learning is the widely used one. However, during the process of labelling the works the programmers are inserting the labels, thus making changes on the training works.

### c. Computational and Statistical Analytical Processing

The ML process' fundamental training phase is called the computational and statistical analytical processing. Along this phase, the machine "reads" the works and applies the ML algorithm to detect and extract empirical findings, such as characteristics and patterns from the training data set[63]. Thus a dataset "mining" is carried out at this phase.

The aim of this stage is to allow the machine to identify technical and artistic components, as well as other creative aspects that are present in the training works[64]. The prediction of patterns or characteristics associated with a particular label within the training works may be another focus of the machine.

For instance, a machine, which is fed with a training data set consisting of different pop songs[65], can identify correlations between the songs and determine aspects

---

[62] Drexl, J., et al. 2019. Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective, Max Planck Institute for Innovation & Competition Research Paper No. 19-13 [online] Available at https://ssrn.com/abstract=3465577, accessed February 10, 2022.

[63] Chiou, p. 403

[64] Ibid.

[65] In the example of 'Flow Machines', the AI was trained with a total of 13,000 songs to generate a song called 'Daddy's Car'. See: Flow Machines. 2016. *AI makes pop music – Flow Machines*. [online] Available at: <https://www.flow-machines.com/history/events/ai-makes-pop-music/> [Accessed 6 January 2022].

such as theme, rhythm, melody, tempo etc., that may be the reason of a song's popularity and then utilize that finding to produce a song of its own[66].

This stage of the process requires works to be copied in the machine's memory to utilize it later for the processing and analyzing of the works[67]. This however, might not be the case "*if the software only 'crawls' through the text or the data, and processes them 'one by one', without copying the whole text but only one data or word or just a few of them at time.*"[68]. The act of copying may also be temporary and incidental if the machine does not need to store them after running through its system and destroying it at the end of the process[69].

#### d. Knowledge Discovery

In the final stage, based on its analysis activities carried out in the previous stage, the machine creates a set of rules. The machine will utilize this set of rules to make intelligent (automatic) decisions, which will be the 'machine output', in relation to "*new and unknown future input*"[70]. The machine output is a result of transforming and analyzing data and putting it in a way that presents information in a human readable

---

[66] Gervais, D., 2019. The Machine as Author. *Iowa Law Review*, [online] Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359524> [Accessed 3 January 2022].

[67] Triaille, J., d'Argenteuil, J. and de Francquen, A., 2014. Study on the legal framework of text and data mining (TDM). [online] Available at: <https://op.europa.eu/en/ publication-detail/-/publication/074ddf78-01e9-4a1d- 9895-65290705e2a5/language-en> [Accessed 4 February 2022], pp. 29 and 49. Schönberger, D., 2018. Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML). [online] p.16. Available at: <https://ssrn. com/abstract=3098315> [Accessed 4 February 2022].

[68] Triaille, p. 31.

[69] Schönberger, p. 16.

[70] Chiou, p. 404.

form[71]. The output may alternatively be called a "report". The output may not include any of the mined original works; it may be the case where only some information of the analyzed works are being retained[72]. It is therefore argued that the output would not be within the scope of copyright, because it only reports the findings of the analysis, which generally does not involve parts or extracts of the works that are mined[73].

The trained model *i.e.* the output, will be utilized especially for the "creative choices" that will result in the generation of art by the machine[74]. This model consisting of bundle of rules can be stored in a permanent place as the output of the training[75].

Consequently, during the different stages of the ML, works may be copied. These copies can be divided into two; as basic reproductions and adaptive uses of the training works. The adaptive uses may be classified as basic reproductions since they do not always enable the person who is in charge of the ML process to make "*free and creative choices*" [76]. The ML software has a limited amount of settings and other parameters that may be changed, but they are typically relatively standardized[77].

It is certain that a reproduction of works occurs during different stages of the ML process, either temporarily or permanently. If the input work is reproduced on the absence of author's permission, copyright infringement may occur. However, as the right of reproduction is not absolute, the act of reproduction can be considered lawful when subject to an exception or limitation.

---

[71] Triaille, p. 31.

[72] Ibid.

[73] Geiger et al., (2018b) p.99

[74] Ibid.

[75] Margoni, section II.

[76] Chiou, p. 404

[77] Margoni, section III.

### D. Legal Status and Authorship of Works Generated by Artificial Intelligence

AI systems have gone from simple calculations to writing novels, painting portraits, composing songs and other complicated creative activities. Since the copyright regulations were not formed with AI in mind, application of traditional copyright rules to AI-generated works cause uncertainty.

With the latest developments, AI technology has reached a point where it is no longer a tool creating works but in the creation process it actually takes individual decisions without a human intervention. Because machines thrive on creating works, the difference between human and machine generated works is blurred . The debate is whether a work created by the AI, is eligible for copyright protection as an original work.

While generating a work, humans can utilize an AI program as an instrument like as a pencil, paper or music instrument. In that case, it is undisputable that the created work meets the requirements of copyright protection because the work is obtained by the appropriate instructions given to the machine which are clearly planned and determined therefore result of human creativity. Here, AI technology simply functions as a tool to reflect the human creativity into an exterior article. A photograph taken by a computer generated object such as camera or cell phone can be compared to the generation of an art utilizing an AI technology . Both procedures are almost automatic therefore it could be argued that an AI, similar to camera, is simply an instrument used by an author to communicate his or her idea in a physical medium.

The second possible scenario is where an AI independently creates the work; however a human selects which of the created works are valuable[78]. For instance, an AI can compose music, however sometimes the algorithms may not be accurate and the AI

[78] Bohler, H.M., 2017. EU Copyright Protection of Works Created by Artificial Intelligence Systems, Master's Thesis, University of Bergen, p.19, Available at: <http://bora.uib.no/handle/1956/16479> [Accessed 5 December 2019]

may not be capable of assessing the appeal of each of the created songs. Therefore a human has to listen the songs and determine which one of them is worth the protection and distribution[79]. In that case, the issue is whether the involvement in the selection process makes a human the author of the work.

In cases where a work is completely generated by AI, the subjective condition recognized in most of the copyright legislations, which requires the work to reflect its author's characteristics, becomes an issue. For a work to fulfil that condition, it is common understanding[80] that the author of the work should be a human.

As the copright laws were legislated at times when the technology of AI was rare and the capacities of computers were still very limited, the scope of the authorship does not extent to non-humans. However, with the emergence of AI generated works, several discussions evolved on whom the ownership of rights should be given if the work created by AI is creative enough to acquire copyright status. Although there are different opinions in the doctrine, it is possible the categorize possible author groups for AI-generated works such as the programmer, user, AI itself, data owner, and AI project investor[81].

---

[79] Ibid.

[80] United Kingdom has a specific provision for computer-generated works. See Article 9 of the Copyright, Designs and Patents Act 1988, "In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken."

[81] For more information see Bridy, A., 2012. Coding Creativity: Copyright and the Artificially Intelligent Author. Stanford Technology Law Review, p.21 [online] Available at: <https://web.law.columbia.edu/sites/default/files/microsites/kernochan/09.materials-Bridy.pdf>; Denicola, R. C., 2016. Ex Machina: Copyright Protection for Computer-Generated Works. Rutgers University Law Review, Vol. 69:251; Yilmaztekin, H.K., 2020. Türk Fikri Haklar Hukuku Yapay Zeka Tarafından Meydana Getirilen Eserleri Korumak İçin Hazır mı?, Galatasaray Üniversitesi Hukuk Fakültesi Dergisi, 2020/2.; Guadamuz, 2016. The Monkey Selfie: Copyright Lessons for Originality in Photographs

## II. COPYRIGHT

Copyright legislation is a subset of IP law, which concerns the creations of the human mind. Intellectual and artistic works contribute greatly to the cultural creativity and development of a country. The progress and well-being of humanity relies on its ability to create new works. In order to encourage further creations, providing adequate legal protection to those creations is essential. By granting innovators and creators rights over their works, IP rights safeguard those who create.

IP has two branches: industrial property and copyright. There are various industrial properties such as patents, designs, and trademarks. Copyright, on the other hand, is concerned with literary and artistic creations, such as paintings, films, and music.

---

and Internet Jurisdiction. Internet Policy Review: Journal on Internet Regulation Volume 5, Issue 1. Available at: http://policyreview.info/articles/analysis/monkey-selfie-copyright-lessons-originality-photographs-and -internet-jurisdiction; Guadamuz, A., 2017. Artificial Intelligence and Copyright. WIPO Magazine. [online] Available at: <https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html>; Hedrick, S. F., 2019. I 'Think', Therefore I Create: Claiming Copyright in the Outputs of Algorithms. NYU Journal of Intellectual Property & Entertainment Law, Vol. 8, No. 2.; Hristov, K., 2017. Artificial Intelligence and Copyright Dilemma. IDEA: The IP Law Review, Volume 57, No. 3., p. 435, Available at: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2976428>; Osha, J. Et al., 2019. AIPPI Study Guidelines: Copyright in Artificially Generated Works. [online] Available at: https://aippi.org/wp-content/uploads/2019/08/Study-Guidelines_Copyright_Copyright-in-artificially-generated-works_23January2019.pdf>; Pearlman, R., 2018. Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law, Richmond Journal of Law & Technology, Vol. 24, Issue 2.; Samuelson, P., 1986. Allocating Ownership Rights in Computer-Generated Works. University of Pittsburgh Law Review. Vol. 47:1185, p. 1196 [online] Available at: < https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=2067&context=facpubs>; Schlackman, S., 2018. Who Holds the Copyright in AI Created Art. [online] Available at: <https://alj.artrepreneur.com/the-next-rembrandt-who-holds-the-copyright-in-computer-generated-art/>

Copyright covers only the form of expression of ideas, not the ideas themselves. These creations of mind are called "intellectual product", *i.e.* product of the mind, ideas and expression[82]. A painting, choreography of ballet or piano concerto has no "existence", they usually become concrete on an article such as a paper, canvas or marble[83]. However, intellectual product and the article on which this product is embodied are different. For instance, in a sculpture the intellectual product is neither the rock nor the marble but the thought and intellectual activity that shapes the marble[84]. Furthermore, it is accepted that intellectual product and the products on which they are embodied are subject to different areas of law. Depending on its characteristics, intellectual product is governed by either intellectual and artistic works law, patent and utility models or industrial design law, however the article on which it is embodied is governed by property law[85].

Copyright protects the creative works by reason of its selection, organization and means of expression such as shapes, melodies, colors, and words[86]. Copyright safeguards the owner of exclusive property rights from others who copy or otherwise appropriate the unique way the original work was presented[87].

The principle of territoriality is applied in copyright law. However, there are three main international treaties concerned with copyright: Berne Convention for the Protection

---

[82]Hirsch, E. E., 1948. Fikri ve Sınai Haklar. Ankara: Ankara Basımevi, p.1; Tekinalp, Ü., 2012. Fikri Mülkiyet Hukuku. 5th Ed. İstanbul: Vedat Kitapçılık, p. 5; Ateş, M., 2012. Fikri Hukukta Eser Sahipliği. Ankara: Adalet Yayınevi, p. 10.

[83] Tekinalp, p. 6

[84] Ibid.

[85] Ibid., p. 7

[86]WIPO, 2016. Understanding Copyright and Related Rights. Available at https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf

[87] Ibid.

of Literary and Artistic Works ("Berne Convention")[88], WIPO Copyright Treaty ("WCT")[89], and Agreement on Trade-Related Aspects of Intellectual Property Rights ("TRIPs Agreement")[90].

Two type of rights are protected by copyright, namely economic and moral rights. Economic rights enable right holders to obtain renumeration from others' use of their work. Moral rights, on the other hand is designed to protect the special emotional and/or intellectual link between the creator and its work.

Most copyright laws grant exclusive rights to right holders over their work. Thus, the right holder of a work is entitled to determine how to use the work and can disable others from using it without authorization. Generally, right holders of a work can permit or forbid certain acts such as reproduction, distribution, translation and adaptation of the work. In this paper, only the right of reproduction under the European Union, United States of America, and Türkiye will be reviewed.

### A. European Union

The EU copyright law is constituted of eleven directives and two regulations[91]. With the copyright *acquis,* the EU's aim is to eliminate specific differences in the national legislation of the Member States and thus ensure an effective enforcement. However, at

---

88 Berne Convention for the Protection of Literary and Artistic Works as amended on 28 September 1979, https://wipolex.wipo.int/en/text/283693.

89 WIPO Copyright Treaty (WCT) 1996. https://wipolex.wipo.int/en/text/295157.

90 Agreement on Trade-Related Aspects of Intellectual Property Rights (Marrakesh, Morocco, 15 April 1994), Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, THE LEGAL TEXTS: THE RESULTS OF THE URUGUAY ROUND OF MULTILATERAL TRADE NEGOTIATIONS 321 (1999), 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994) (TRIPS).

91 European Commission, 2022. The EU Copyright Legislation, available at <https://digital-strategy.ec.europa.eu/en/policies/copyright-legislation> [Accessed August 2, 2022].

this point a unified copyright law at the EU level is not present. Nonetheless, the case law of the CJEU, whose role is to ensure that EU law is understood and implemented the same in every Member State, has helped shaping the harmonization of national copyright laws.

Article 2 of the Infosoc Directive, introduced a harmonized right of reproduction. According to the article, the right of reproduction is an "*exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part.*"[92] .

The concept of "reproduction in part" was dealt with in CJEU's *Infopaq* decision[93]. The same decision also led to the de facto harmonization of "originality" requirement for works[94]. The case concerned extracts of eleven words from selected newspapers that Infopaq International had prepared to sell to its customers. The extracts were obtained as a result of scanning the original text of the newspapers with the data capture technique. The question standing before the CJEU was whether the reproductions of extracts of eleven words would be considered as reproduction "in part" within the scope of Article 2 of the Infosoc Directive?

Firstly, the Court noted that either the concept of "reproduction" or "reproduction in part" are defined in the Infosoc Direcitive[95]. Thus, the Court decided that the concept of "reproduction in part" should have a uniform meaning within the EU.

---

[92] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167.

[93] Infopaq International A/S v. Danske Dagblades Forening (2009) Case no. C-5/08. Court of Justice of the European Union, ECR I-06569 ("Infopaq").

[94] Rosati E., 2019. Copyright and the Court of Justice of the European Union. Oxford University Press, p. 87.

[95] Infopaq, para. 31.

In the decision, the Court started by noting that the right of reproduction is intended to cover "work"[96]. The concept of work has its roots from the Berne Convention and accordingly, "*protection of certain subject-matters as artistic or literary works presupposes that they are intellectual creations*"[97]. Although the Berne Convention does not have a definition, it implies that subject matter is protected if it is sufficiently *original* in the sense that it is an intellectual creation[98].

The threshold of originality set in the *Infopaq* case was further explained by the CJEU in the *BSA*[99], *FAPL*[100], *Painer*[101] and *Football Dataco*[102] cases. In conclusion, for a work to be considered original: i) it must be the author's own intellectual creation[103]; ii) the author must reflect its personality by making creative and free choices while creating the work[104]; iii) there shouldn't be one way of expressing an idea or the expression should not be determined by a purpose or dictated by narrow rules, there should be an opportunity for creative choices[105].

Consequently, the CJEU clarified in *Infopaq* that "reproduction in part" should not be understood in a quantitative manner. Parts of the work thus enjoy protection if the

---

96 Infopaq, para. 33.

97 Infopaq, para. 34.

98 Rosati, p. 87.

99 CJEU, Case C-393/009, Bezpečnostní softwarová asociace v Ministerstvo kultury, E.C.R. I-13971, (2010).

100 CJEU, Cases C-403/08 and C-429/08, Football Association Premier League Ltd and Karen Murphy, E.C.R. I-09083, FAPL (2011).

101 CJEU, Case C-45/10, Eva-Maria Painer v Standard Verlags GmbH and others, E.C.R. I-12533, (2011).

102 CJEU, Case C-604/10, Football Dataco v Yahoo! UK Ltd and others, 2 C.M.L.R. 24 (2012).

103 Infopaq, paras. 34-48.

104 Football Dataco, paras. 38-39; BSA, para. 50; Painer, paras. 89 and 92.

105 FAPL, para. 98; Football Dataco, para 39.

relevant parts include elements which are the expression of the author's own intellectual creation[106].

The CJEU, while interpreting the reproduction right's scope, adopted a wide approach with the *Infopaq* decision[107]. Thus, whether temporarily or permanently, directly or indirectly, any digital copy of a work used during the training of AI, even in part- that "*reproduces the creative expression*" of the original work, has the probability to cause infringement, no matter how transient, short, or non-commercial it is[108].

The concept of "reproduction in part" has been further explained by the CJEU in its *Pelham*[109] decision. The case concerned the electronic copying ("sampling") two seconds of a rhythm sequence from the song "Metall auf Metall" and using it in an endless loop in another song called "Nur Mir". The question before CJEU this time was whether taking a sample of two seconds could be considered as "reproduction in part" and whether prior authorization from the phonogram producer is needed.

Phonogram producers are granted the exclusive right of reproduction in Article 2(c) of the Infosoc Direcive[110]. In the decision, the Court firstly pointed out that the Infosoc Directive does not have a definition for the concept of "*reproduction…in whole or in part*". Nevertheless, the meaning and scope of those words should be determined keeping in mind their usual meaning in everyday language[111].

---

[106] Infopaq, para. 39.

[107] Infopaq, paras. 39 and 43.

[108] Chiou, T., 2019. *Copyright lessons on Machine Learning: what impact on algorithmic art?* 10 JIPITEC 398 para 1.

[109] CJEU, Case C-476/17, Pelham GmbH, Moses Pelham, Martin Haas v Ralf Hütter, Florian Schneider-Esleben, (2019).

[110] Article 2(c) of the Infosoc Directive: "*Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part: (…) (c) for phonogram producers, of their phonograms*".

[111] Ibid., para. 28.slight a

The Court further notes that the reproduction by a user of a sound sample, even if very short, of a phonogram must, in principle, be regarded as a reproduction 'in part' of that phonogram[112]. However, the Court led to the conclusion that, "*where a user, in exercising the* ***freedom of the arts****, takes a sound sample from a phonogram in order to use it, in a modified form* ***unrecognisable*** *to the ear, in a new work, it must be held that such use does not constitute 'reproduction'*"[113] (emphasis added).

From the Court's conclusion it might be inferred that the exclusive rights of phonogram producers are limited to copies that are recognizable in the new work. Court's finding is jusitified in the way that a balance is struck between the right of reproduction and other fundamental rights[114]. Hence, sampling is permitted to "afford the opportunity to take part in the public exchange of cultural, political and social information and ideas of all kinds". Consequently, in the Pelham decision, the CJEU employs the concept of a fair balance to exclude specific uses from the reach of right of reproduction[115]. It should be mentioned that the Pelham decision set forth that, unlike reproduction of works in part; for phonograms, it is not necessary to evaluate whether the copied part is original or not[116].

---

[112] Ibid., para. 29.

[113] Ibid., para 31.

[114] Jütte, B., Quintains, J., 2021. The Pelham Chronicles: sampling, copyright and fundamental rights. Journal of Intellectual Property Law & Practice, Vol. 16, No. 3, p. 216.

[115] Ibid., p. 217.

[116] See Opinion of Advocate General Szpunar in Case C-476/17 Pelham GmbH, Moses Pelham, Martin Haas v Ralf Hütter, Florian Schneider-Esleben, para 30: *"(...) in the case of a work, it is possible to distinguish the elements which may not be protected, such as words, sounds, colours etc., from the subject-matter which may be protected in the form of the original arrangement of those elements, such a distinction is not, however, possible in the case of a phonogram. A phonogram is not made up of small particles that are not protectable: it is protected as an indivisible whole. Moreover,* ***in the case of a phonogram, there***

Adaptation generally refers to a modification of a work into a different form that is not equivalent to reproduction, for instance turning a book into a movie. The Berne Convention defines the right of adaptation in its Article 12. However, the Infosoc Directive does not refer to that right. In its decision of Allposters International BV *v.* Stichting Pictoright, the CJEU indicates: "*With regard to the adaptation right, it is true that Article 12 of the Berne Convention confers on authors of literary or artistic works an exclusive right of authorising adaptations, arrangements and other alterations of their works and that there is no equivalent provision in Directive 2001/292"*[117].

Although there is a lack of harmonization within the EU regarding the right of adaptation, the right is claimed to apply as a more general expression of right of reproduction[118]. It is argued that some adaptations might be also characterized as reproductions and could be considered in the reach of right of reproduction, to the extent that the alterations made in the work bring out further reproductions of previous works, without adding any creativity[119].

***is no requirement for originality, because a phonogram, unlike a work, is protected, not by virtue of its creativeness, but rather on account of the financial and organisational investment"*** (emphasis added).

[117] CJEU, C-419/13, Allposters International BV v. Stichting Pictoright (2015), para. 26.

[118] Geiger, C., Frosio, G. and Bulayenko, O., 2018a. Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?. *IIC - International Review of Intellectual Property and Competition Law*, 49(7), p.819.

[119] Chiou, pp.401 and Margoni, T. 2018. Artificial Intelligence, machine learning and EU copyright law: who owns AI? CreateWorkingPaper. [online] Available at: https://ssrn.com/abstract=3299523. (Accessed 2 February 2022).

### B. United States of America

The Constitution of the United States empowers Congress, the legislature branch of the federal government, to pass legislation establishing a copyright system in the country:

"*The Congress shall have Power ... To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries ...*"[120]

The Congress passed the first federal copyright legislation in 1790, since then there have been multiple revisions, general ones being in 1831, 1870, 1909 and 1976. Currently, the Copyright Act of 1976 is in force. The Act has often been subject to judicial interpretation by the federal courts, including US District Courts, the US Courts of Appeal, and the US Supreme Court. With the 1976 Act, the dual federal/state scheme was replaced with a uniform federal structure[121].

The Act lists the categories of protected works in Section 102 in a non-exhaustive way[122]. The Judiciary Committee report on the legislation notes that the "*list sets out the general area of copyrightable subject matter, but with sufficient flexibility to free the courts from rigid or outmoded concepts of the scope of particular categories*."[123]

---

[120] Article I, Section 8 of the U.S. Constitution.

[121] Reid A., Port K.L.; Tu S. 2021. Fundamentals of United States Intellectual Property Law : Copyright, Patent, and Trademark, Kluwer Law International, p. 11.

[122] See 17 U.S. Code § 102: *"(...) Works of authorship include the following categories: literary works, musical works, dramatic works, pantomimes and choreographic works, pictorial, graphic, and sculptural works; motion picture and other audovisiual works; sound recordings: and architectual works*."

[123] 17 U.S. Code § 102 - Subject matter of copyright. House Report no. 94–1476.

According to the 1976 Act, copyright protection is available for "*original works of authorship fixed in any tangible medium of expression*"[124]. Thus, any work not fixed in a physical medium is outside the protection scope.

The idea itself is not protected, rather it is the "expression" of an idea that is protected by the Act. Section 102(b) provides: "*In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.*"[125]

Besides fixation, originality is another requirement for copyright protection. The US Supreme Court explained the concept of originality in its *Feist*[126] decision, with regards to a telephone listing. The Court noted that original "*means only that the work was independently created by the author (as apposed to copied from other works), and that it possesses at least some minimal degree of creativity*"[127]. Yet, the creativity that is seeked in works is extremely low; in view of the Court, even a slight amount is sufficient. The House Report to the 1976 Act noted; "*This standard does not include requirements of novelty, ingenuity, or esthetic merit, and there is no intention to enlarge the standard of copyright protection to require them*."[128]

With the *Feist* case, the doctrine called "sweat of the brow" was rejected. Under this doctrine, where a sufficient amount of effort was given into the creation, the work was considered original. The court explained its stance by stating that the aim of the

---

[124] Ibid.

[125] Ibid.

[126] Feist Publications Inc v Rural Telephone Service Co 499 U.S. 340 (1991).

[127] Ibid.

[128] House of Representatives, 1976, Report No. 94-1476, at 51.

copyright was not to remunerate those who collect information but rather encourage creative expression.

Derivative works, which are based upon existing works need to contribute "*a distinguishable variation from the older works [which] variation must be substantial and not merely trivial*".[129]

Section 106 of the Copyright Act of 1976 lists the economic rights of right holders. One of these rights is the right to reproduce the copyrighted work. The right is infringed when "copies" or "phonorecords" are made by someone not allowed by the right holder. The concepts "copies" and "phonorecords" are defined comprehensively in the Act: "*Copies are material objects, other than phonorecords, in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device"*[130]. Phonorecords are material objects in which sounds are fixed.

The right of reproduction is considered to be infringed when the unauthorized copy includes "*substantial*" parts of the original work[131]. The 2nd Circuit noted that "*if the secondary work sufficiently transforms the expression of the original work such that the two works cease to* ***be substantially similar****, then the secondary work is not a derivative work and, for that matter, does not infringe the copyright of the original work*" (emphasis added).[132]

---

[129] Norma Ribbon & Trimming, Inc. v. Little, 51 F.3d 45, 47 (5th Cir. 1995); Entertainment Research Group., Inc. v. Genesis Creative Grp., Inc., 122 F.3d 1211 (9th Cir. 1997).

[130] 17 U.S. Code § 101.

[131] Balganesh S., Manta I.,. and Wilkinson-Ryan T., 2014. Judging Similarity. 1*00 Iowa Law Review 267, U of Penn Law Schoo*l, Public Law Research Paper No. 14-15, Hofstra Univ. Legal Studies Research Paper No. 2014-09, Available at SSRN: https://ssrn.com/abstract=2409811, [Accessed May 22, 2022].

[132] Castle Rock Entertainment, Inc. v. Carol Publishing Group., Inc., 150 F.3d 132, 143 n.9 (2d Cir. 1998).

Different circuit courts embraced somewhat different methods to evaluate the substantial similarity between works. However, majority of the courts are on the same opinion that the comparison should be made in a comprehensive manner, considering the entirety of the work[133]. "*Total concept and feel*"[134] of the work is taken into account rather than concentrating on the individual components of the work. The two leading creators of the substantial similarity standard, the Second and Ninth Circuits, emphasize that the test should have a quantitative and qualitative aspect[135]. The quantitative aspect concerns determining how much of the work is copied. On the other side, the qualitative aspect is about determining if the allegedly infringer party took particularly important and significant portions of the work.

### C. Türkiye

Turkish copyright law is embodied in the Law 5846 on Intellecutal Property and Artistic Works (LIAW)[136]. The subject of the rights recognized under the LIAW is intellectual and artistic works.

According to the law, "work" is any intellectual or artistic product bearing the characteristic of its author, which is deemed a scientific and literary or musical work or

---

[133] Ibid.

[134] See Atari, Inc. v. N. Am. Philips Consumer Elecs. Corp., 672 F. 2d 607, 614 (7th Cir. 1982).

[135] Balganesh, Manta and Wilkinson-Ryan, p. 7

[136] Law No. 5846 of 5.12.1951. (amended by: Law No. 2936 of 1.11.1983; Law No. 4110 of 7.6.1995; Law No. 4630 of 21.2.2001; Law No. 5101 of 3.3.2004; Law No. 5217 of 14.7.2004; Law No. 5571 of 28.12.2006; Law No. 5728 of 23.01.2008 Law No. 6279 of 29.02.2012 Law No. 6552 of 10.09.2014) Transleted by Nilsson G. O. and Şehirali F. H, Available at https://en.unesco.org/sites/default/files/turkey_lawintellectualartisticwork_entof.pdf . (Altough the translation of the Law is based on its previous version, in this paper the translation of unaltered provisions are used.)

work of fine arts or cinematographic work. As per this Article, in order to regard intellectual and artistic work as "work" within the context of LIAW and grant them protection, there are various conditions. The regulation of these conditions can be justified as it would be pointless and challenging to consider every intellectual product as "work" and grant protection[137]. For instance business letters are also an intellectual product however they can be created by anyone therefore they do not qualify as work[138]. Besides, the protection should be given to intellectual properties which contribute to the culture of the society and have a creative component[139].

There are two conditions to regard an intellectual property as work; one being subjective and the other one being objective[140]. The subjective condition means that the "work" should be an outcome of the creative effort by the author and bear the characteristic of the author[141]. The author should reflect his creative spirit and personality to the work[142]. The characteristic of the author shows itself in the expression or the style used in the work[143] .

It is also accepted by the Article 9/2 of TRIPS and Article 2 of the WCT that copyright protection covers the expressions[144]. The relevant Articles have the same

---

[137] Erel, Ş.N., 2009. *Türk Fikir ve Sanat Hukuku.* 3rd Ed. Ankara: Yetkin Basımevi, p. 51.

[138] Ibid

[139] Ibid; Hirsch, E., p.130.

[140] Hirsch, p. 130

[141] Erel, p.52; Turkish Court of Cassation, 11th Civil Chamber, Case No: 2006/934, Decision No: 2007/4555, 13.03.2007.

[142] Suluk, C., Orhan, A., 2005. Uygulamalı Fikri Mülkiyet Hukuku Cilt II: Genel Esaslar Fikir ve Sanat Eserleri, 1st Ed. İstanbul: Arıkan Yayınları, p. 144.

[143] Tekinalp, Ü., p. 106.

[144]*Agreement on Trade-Related Aspects of Intellectual Property Rights*. 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994) [online]

wording which reads as: "*Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.*"[145].

Expression or style is not used only in stories, articles or novels; but a film, painting, dance chorography or a sculpture also has an "expression"[146]. Since knowledge is a personal concept meaning that it differs within persons, in scientific works the characteristics show itself in the usage of knowledge and scientific method[147].

According to Hirsch, if a work can be created by anyone, there is no public interest in protecting such works, as they will have no characteristic[148]. Anyone who has a camera can take pictures or who knows how to write can write down his thoughts or emotions and therefore works that cannot be created by everyone and that has a character should be protected only[149]. With this approach, Hirsch does not refer to masterpieces, however he makes a distinction between works that are generated by everyone due to their duties and that do not reflect any effort, comment or thought and works that are an outcome of an intellectual process and reflect the talent of the author[150].

For instance, a court minutes that conveys a fight from the testimony of a witness is very different than a well-known author's description of a fight with his unique sentences and technique[151]. Every lawyer can write petitions, experts can prepare reports,

---

Available at: <https://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm> [Accessed 1 December 2019]; *World Intellectual Property Copyright Treaty*. 1996. [online] Available at: < https://wipolex.wipo.int/en/text/295157> [Accessed 1 December 2019]

[145] Ibid.

[146] Tekinalp, p. 106.

[147] Ibid.

[148] Hirsch, p. 131

[149] Ibid.

[150] Tekinalp, p. 10

[151] Güneş, İ. 2015. Uygulamada Fikir ve Sanat Eserleri Hukuku. 2nd Ed. Ankara: Seçkin Yayınları, p. 59.

court minutes can be organized by judges, yet it is very exceptional that these works qualify as a "work" within the scope of Law No. 5846[152].

The view of Hirsch has been criticized by some of the academicians[153]. In view of Tekinalp, while the expression determines the characteristic of the author, it does not have to be original[154]. There is no doubt that works that are original are qualified as "work" within the Law No. 5846, however if high standards such as "originality" or "being intrinsic to the author" are set to protect the work, the scope of the protected works will be very narrowed[155]. The stance of Court of Cassation is similar to Tekinalp, the Court stated in its decision that keeping the "originality" threshold high will make it very difficult to create new works and therefore the broad interpretation of the term "originality" is beneficial both for the authors and the public[156]. Yet, the Court of Cassation also notes that the characteristic of the author requires "not being ordinary" and "a certain level of creativeness"[157].

In doctrine, it is argued that the work can only bear characteristics if the author is a real person, thus the work in the sense of Law No. 5846 can only be generated by real persons[158].

---

152 Ibid.

153 Arslanlı H., 1954. Fikri Hukuk Dersleri II. Fikir ve Sanat Eserleri. İstanbul, p. 6-7; Akın, B., 2006. Fikir Hukuku Dersleri. İstanbul: Arıkan Yayıncılık.

154 Tekinalp, p. 107

155 Ibid.

156 Turkish Court of Cassation, 11th Civil Chamber, Case No: 2016/3508, Decision No: 2017/5950.

157 Turkish Court of Cassation, 11th Civil Chamber, Case No: 2015/12923, Decision No: 2017/2724, 09.05.2017.

158 Tekinalp, p. 113; Kılıçoğlu, p. 202; Öztan, p. 82-87; Ateş, p. 13, 32; Bozbel, p.19

The objective condition on the other hand, requires the work to exist physically and be included in one of the categories such as; scientific and literary works; musical works; artistic works; or cinematographic works.

According to Article 20 of the LIAW, "*the right to exploit, in any manner or form, a work that has not yet been made public belongs exclusively to the author. The exclusive right granted to the author to exploit a work that has been made public consists of the rights stipulated as economic rights by the Law*". There are five economic rights granted to the author, which are the rights of adaptation, reproduction, publication, representation, and broadcasting and public transmission. For the purpose of this paper, only the right of reproduction will be explained below.

The right of reproduction is regulated by Article 22 of the LIAW. Accordingly, "*the right to reproduce the original or copies of a work in any form or by any method, in whole or in part, directly or indirectly, temporarily or permanently belongs exclusively to the author".* As per the second paragraph of the Article 22, "*the making of a second copy of the original of the work or the recording of the work* ***on all types of devices now known or to be developed in the future*** *enabling the transmission or repetition of signs, sounds and images, all kinds of sound and music recordings (...) are deemed reproduction*". (emphasis added).

The article's wording stating that the "recording of the work on all types of devices now known or to be developed in the future" is a reproduction, shows that the law-maker did not want to limit the scope of reproduction right with today's technologies.

The work does not have to show exactly the same physical characteristics as the original for it to be accepted as a copy; it is sufficient that the characteristics in the work belonging to the author of the work is transferred to the copy. Reproduction can also be partial. In that case, it is essential that the characteristic of the author can be understood from the copied part[159].

Whether the reproduction is permanent or temporary does not make a difference according to the LIAW. Hence, with this law the right holders of works enjoy a wide-ranging protection

Finally, the third paragrapgh of Article 22 provides that "*the right of reproduction also covers the acts of loading, displaying, running, transmitting and storing a computer program to the extent that such acts require the temporary reproduction of the computer program*". Caching, which is the process of automatically creating a copy of the file in temporary memory for accessing any content on the computer, or viewing web pages via a browser (browsing) is also within the scope of reproduction rights in Türkiye[160].

---

[159] Suluk C., Karasu C., Nal R., 2019. Fikri Mülkiyet Hukuku, Ankara, Seçkin Yayınevi, p. 92.

[160] Güneş, p. 147.

## CHAPTER TWO

## EXCEPTIONS AND LIMITATIONS

### I. INTRODUCTION

A big number of data is required to train AI. Sometimes, the data employed to teach and train AI may consist of works, which enjoy copyright protection. For an AI to write a book or compose a song for instance, it is trained by a large set of data consisting of maybe hundreds or thousands of books or songs.

In 2016, a portrait named "The Next Rembrandt" has been revealed to the world[161]. Being a 3D-printed painting, the Next Rembrandt was made exclusively from data obtained from 346 known paintings by Rembrandt[162]. The aim of the project was to find out whether an algorithm could be generated, using conventional techniques of data analysis, that could create a physical work of art that imitates the appearance of a real Rembrandt painting[163]. A software system was devised to learn the style of the famous painter which could understand him based on his use of geometry, combination and painting tools [164]. The most distinctive geometric forms used by Rembrandt are specified and categorized through a facial recognition algorithm then the learned principles are used to duplicate the character and create new facial aspects for the painting[165]. The Project is

---

[161] Schlackman, S., 2018. *Who Holds the Copyright in AI Created Art.* [online] Available at: https://alj.artrepreneur.com/the-next-rembrandt-who-holds-the-copyright-in-computer-generated-art/ [Accessed 3 December 2019]

[162] Thompson, J. W., *The Next Rembrandt.* [online] Available at: https://www.jwt.com/en/work/thenextrembrandt [Accessed 3 December 2019]

[163] Schlackman, p.

[164] The Next Rembrandt: Generating the Features. [online] Available at: https://www.nextrembrandt.com/ [Accessed 3 December 2019].

[165] Ibid.

an example of how the AI is trained to further produce an artwork. However in this case, copyright was not an issue since the copyright term for the paintings of Rembrandt had expired. Yet, the issue of copyright, especially the right of reproduction may come up if the AI is trained with works without obtaining a prior authorization from the author.

As discussed above, during the ML process, use of copyrighted materials as data to train the AI, may result in the infringement of the reproduction right and there is no specific exception or limitation regarding this issue. However, there are some legal notions, which may be relevant. Within the limits of this study, firstly the regulations in the international area, then the legal regime of exceptions and limitaitons within the EU, US and Türkiye will be examined.

## II. INTERNATIONAL INSTRUMENTS

Copyright laws enable specific limitations and foresee certain exceptions to economic rights to balance the right holders' interests and users of the works.

On international level, the Berne Convention laid down a three-step test in the area of exceptions and limitations. It is a clause that provides three cumulative conditions: "*It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate* *interests of the author."*[166] The three-step test is embodied also in the TRIPS Agreement[167] and WIPO Copyright Treaty[168].

In 2000, the dispute-settlement panel of the World Trade Organization (WTO) interpreted the three conditions because the European Communities asked for consultations with the United States in relation to Section 110(5)[169] of the US Copyright Act[170]. According to the WTO Panel, the first step of the test, in other words, the requirement of a "certain special case", indicates that an exception or limitation in national legislation must be clearly defined[171]. This, however, does not mean that every possible situation to which the exception could apply, must be identified. The term "special" means "*having an individual or limited application or purpose*"[172]. The Report further notes that "*an exception or limitation should be narrow in quantitative as well as qualitative sense*"[173].

---

[166] Article 9(2) of the Berne Convention.

[167] Article 13 of the TRIPS Agreement: "Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder."

[168] Article 10 of the WIPO Copyright Treaty: "Contracting Parties may, in their national legislation, provide for limitations of or exceptions to the rights granted to authors of literary and artistic works under this Treaty in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author.

[169] Section 110(5) of the US Copyright Act permitted, under certain conditions, the playing of radio and television music in public places (bars, shops, restaurants, etc.) without the payment of a royalty fee.

[170] See Panel Report of 15 June 2000, United States-Article 110 (5) of the US Copyright Act, WT/DS160/R. Available at https://www.wto.org/english/tratop_e/dispu_e/cases_e/ds160_e.htm

[171] Ibid, p. 33.

[172] Ibid.

[173] Ibid.

The second step requires the prohibition of a conflict with a "normal exploitation" of the work. The Panel first identifies the meaning of "exploitation" and notes that the term means "making use of" or "utilising for one's own ends"[174] To explain what "normal" exploitation means, the Panel states that it is "something less than full use of an exclusive right"[175]. Thus, the step of normal exploitation was decided to include consideration of the forms of exploitation that "*currently generate significant or tangible revenue*" for the right holder as well as those which, are likely to be of "*considerable economic or practical importance*" in the future[176].

The final step of the test concerns the unreasonable prejudice to the legitimate interests of the right holder. For this step, the Panel determined that "*the prejudice to the legitimate interests of right holders reaches an unreasonable level if an exception or limitation causes or has the potential to cause an unreasonable loss of income to the copyright owner*"[177].

The above explained three-step test allows countries to introduce national exceptions for the exclusive rights that are not specifically provided in international copyright treaties. Below, the exceptions of reproduction right adopted by the EU, US, and Türkiye will be clarified and it will be discussed whether the existing exceptions allow the use of works as data during the machine learning stages of AI.

---

[174] Ibid, p. 44.

[175] Ibid.

[176] Ibid, p. 48; Geiger C., Gervais D., Senftleben M., 2015. Understanding the "Three-Step Test". In D. J. Gervais (Ed.), International Intellectual Property – A Handbook of Contemporary Research, Edward Elgar, p, 175.

[177] Ibid., p. 59.

## III. APPROACH OF THE EUROPEAN UNION

The EU acknowledged the fact that rapid technological developments change the creation, production and exploitation of works. Moreover, it is stressed by the Union that the data is an essential part of the economy and to improve the innovative environment, big data needs to be utilized[178]. Hence, to grasp the full potentiality of the data, the EU made some legislative reforms and introduced some exceptions in that regard.

In fact, the European Commission report highlights the importance of data as: "*Digital innovation, driven by the combination of Big Data, Cloud Computing, Mobile technologies and Social media, is one the most powerful drivers of change and the best opportunity for Europe to move back to a growth path. Within this technology cluster, Big Data considerably improves decision-making capabilities and, ultimately organizational performances. Leveraging datadriven innovation, enabled by Big Data technologies, is therefore a powerful value generator and constitutes a real benefit for Europe's economy as a whole.*"[179]

Article 5(5) of the Infosoc Directive shows that the Berne Convention's three-step test is embraced within the EU. The Article provides: "The exceptions and limitations provided for in paragraphs 1, 2, 3 and 4 shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder."

The Infosoc Directive foresees a special exception for temporary acts of reproduction. However, to grasp the full potentiality of the data, another reform was the entry of force of the Digital Single Market Directive in 2019, which includes exceptions that are significant to review for ML purposes.

---

[178]Shaping Europe's digital future. 2022. *Big data*. [online] Available at: <https://digital-strategy.ec.europa.eu/en/policies/big-data> [Accessed 7 February 2022].

[179] European Commission, 2017. The European Data Market Study: Final Report, p. 5.

### A. Exception for Temporary Acts of Reproduction

Temporary acts of reproduction are regulated with the Article 5(1) of the Infosoc Directive. It is the only mandatory exception foreseen by the Directive, which means that the provision is implemented in all Member States[180]. The article reads as the following: "*Temporary acts of reproduction referred to in Article 2, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose 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, shall be exempted from the reproduction right provided for in Article 2*."

As stated in the Recital 33 of the Directive, the exception is designed to make browsing and acts of caching[181] possible to allow specific kind of copies constituting a part of a technological process[182]. Thus, the exception is not specifically tailored for ML,

[180] Triaille, p. 43.

[181] According to the Cambridge dictionary, cache refers to an area or type of computer memory in which information that is often in use can be stored temporarily.

[182] See Recital 33 of Infosoc Directive: "*the exclusive right of reproduction should be subject to an exception to allow certain acts of temporary reproduction, which are transient or incidental reproductions, forming an integral and essential part of a technological process and carried out for the sole purpose of enabling either efficient transmission in a network between third parties by an intermediary, or a lawful use of a work or other subject-matter to be made. The acts of reproduction concerned should have no separate economic value on their own. To the extent that they meet these conditions, this exception should include acts which enable browsing as well as acts of caching to take place, including those which enable transmission systems to function efficiently, provided that the intermediary does not modify the information and does not interfere with the lawful use of technology, widely recognized and used by industry, to obtain data on the use of the information. A use should be considered lawful where it is authorized by the rightholder or not restricted by law.*"

but it may be applicable to the ML activities as long as it meets all the conditions of the Article.

The CJEU made a determination on the *Infopaq* case regarding which acts are assessed as 'transient' under the scope of the Article 5(1). According to the Court, "*the duration should be limited to what is necessary for the proper completion of the technological process in question, it being understood that the process must be automated so that it deletes that act automatically, without human intervention"*[183]. The Court does not explain the necessary length of the copy for it to be characterized as 'transient'.

Another temporary act of reproduction referred to in the Article 5(1) is the 'incidental' ones. The reproduction duration is longer than a 'transient' copy in case of 'incidental' copies[184]. An incidental act of reproduction may take place when the copy is incidental regarding the principal act of exploiting the work, on the condition that the reproduction is not permanent[185].

As discussed above, reproduction might be transient during the analytical processing stage and the exception can be applied to the activities carried out at that stage

---

[183] Infopaq, paragraph 64. See also CJEU, C-360/13, Public Relations Consultants Association (5 June 2014), ECLI:EU:C:2014:1195, paras. 43 and 50 (noting that an act of reproduction is ''incidental'' ''if it neither exists independently of, nor has a purpose independent of, the technological process of which it forms part'')

[184] For the difference between transient and incidental copies see United Kingdom Supreme Court, Public Relations Consultants Association Limited v The Newspaper Licensing Agency Limited and Others, [2013] UKSC 18. The Court identified incidental copy as: "*If, as I consider, the copies made in the internet cache or on screen are "transient", it is strictly speaking unnecessary to consider whether they are also "incidental". But I think it clear that they are. The software puts a web-page on screen and into the cache for the purpose of enabling a lawful use of the copyright material, i.e. viewing it. The creation of the copies is wholly incidental to the technological process involved*"

[185] Triaille, p. 43

of the ML. However, this might not be the case for other stages of the ML process. For instance, during the stages of corpus compilation stage and preprocessing of the training material the reproductions are generally not transient. The work used during the ML process may need to be stored for a specific time to be analyzed or transformed.

A further condition of the exception is that the copy, either transient or incidental, must be made because it is an "*integral and essential*"[186] stage in a technical process. In terms of the reproductions taking place during the ML stages, it is clear that this condition is met as the reproductions are an integral and essential part of a technological process, namely ML.

Lastly, the temporary acts of reproduction cannot have an independent economic importance for the exception to be applied. To make an assessment about this condition, it is necessary to determine if an economic benefit is derived directly from the temporary reproduction[187]. It is possible for some stages of the ML to have an independent economic importance, depending on the choice of the person, who made the copies. If the person sells or licenses the extraction results[188] or the database (created during the compilation stage for instance)[189], then the act of reproductions would have an independent economic importance.

To conclude, the exception foreseen in Article 5(1) of the Infosoc Directive may not be applicable to the acts of reproductions carried out during the ML stages. The exception requires the conditions to be met cumulatively and it is interpreted strictly.

[186] Ibid., paragraph 54.

[187] Opinion of Advocate General Trestnjak delivered on 12 February 2009, Case C-5/08, Infopaq, par. 127.

[188] Triaille et al. p. 47.

[189] Chiou, p. 406.

Therefore, it is generally accepted[190] that this mandatory exception might only apply to limited reproductions of ML process.

### B. Text and Data Mining Exceptions

"*The development of AI leads to a growing relevance of Text and Data Mining regime and of its possible weaknesses*"[191].

To benefit from the flow of new information and to create knowledge and innovation, text and data mining (TDM) is a crucial instrument. TDM can be defined as: "*The automated processing of digital materials, which may include texts, data, sounds, images or other elements, or a combination of these, in order to uncover new knowledge or insights."*[192]

At this point, an explanation regarding the relationship between data mining and machine learning can be necessary. The literature lacks a clear differentiation between TDM and ML. Data mining, according to some sources, is concerned with extracting data patterns and identifying links between data, whilst ML is concerned on making

[190] See Triaille et al. pp. 50; Geiger et al (2018a) pp. 822, indicating that the sole mandatory exception has a very limited scope that would hardly cover most TDM activities and offers insufficient legal certainty. Margoni, pp. 19*:"[g]iven the cumulative, strict and partially unclear conditions that qualify article 5(1), a very careful case-by-case assessment should be performed before deciding to rely on this exception given the unavoidable degree of risk involved*."

[191] Holder C. (ed.), Iglesias, M. (ed.), Triaille J.-P., Van Gysegnem J.-M., 2019. Legal and regulatory implications of Artificial Intelligence. The case of autonomous vehicles, m-health and data mining, Publication Office, [online] Available at: https://op.europa.eu/en/publication-detail/-/publication/f962b17b-5c04-11e9-9c52-01aa75ed71a1/language-en/format-PDF (Accessed 7 February 2022).

[192] Tiriaille et al. p. 17

predictions[193]. The goal of data mining is to evaluate data and uncover patterns that may be used to explain a phenomena[194].

Machine learning, on the other hand, seeks to make predictions by constructing models that can anticipate some future outcome[195]. According to *Wlodarczak,* ML may be considered as a "subarea" of data mining in which the rules are learned automatically[196].

As explained above, ML phases entail a computational and statistical analysis of the works, which are fed into the machine for training purposes. In the sense that it involves the mechanical processing of digital materials, which may involve writings, data, audio, pictures, or other components, or a mix thereof, to discover novel information or understanding, the analytical processing of training works is a sort of data mining[197].

Different from "expert" TDM systems, which usually follow rules engineers have programmed for them, ML systems are trying to learn and create their own rules to differentiate for instance cat pictures from dog pictures[198].

It is certain that for the machine to learn processing large sets of data is required and for that reason TDM is an important part of ML, thus AI projects. The essential role of the TDM in the field of AI has been expressed by 23 organizations consisting among others, universities, companies and researchers, in an open letter addressed to the European Commission. The letter emphasized the role of TDM with the following words: "*TDM is a building block for both machine and deep learning: without the ability of*

---

[193] Wlodarczak, p. 5

[194] Ibid.

[195] Ibid.

[196] Ibid.

[197] Chiou, p. 407.

[198] Lemley, M. and Casey, B., 2020. Fair Learning. *Texas Law Review*, pp.119-120.

*computers to analyse very large amounts of data, employ cognitive technologies to allow the learning of patterns, AI is not possible.*"[199]

Due to the importance of TDM within ML processes, the rules regulating TDM activities have an essential impact on ML. The Directive on Copyright in the Digital Single Market[200] (DSM Directive) introduced two exceptions in relation to TDM.

### 1. Research Exception

One of the exceptions allowing reproductions and extractions of works during the TDM workflow is regulated under the Article 3 of the DSM Directive. The first paragraph of the Article reads as:

*"Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.".*

For the exception to be applicable, the TDM project should be characterized as scientific research. If the project has a commercial aim the exception will not cover the act of TDM of works. Moreover, the exception only covers the TDM activities of research

[199] European Alliance for Research Excellence. 2018. *Open letter: Maximising the benefits of artificial intelligence through future-proof rules on Text and Data Mining - European Alliance for Research Excellence.* [online] Available at: <https://eare.eu/open-letter-maximising-the-benefits-of-artificial-intelligence-through-future-proof-rules-on-text-and-data-mining/> [Accessed 9 February 2022].

[200] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, 92–112.

organizations and cultural heritage institutions. Thus, the exception has a limited group of beneficiaries and a limited scope.

The Article only allows TDM activities on works, which the beneficiaries have lawful access to. Recital 14 sheds some light on what is considered as 'lawful access' in terms of the Directive. Accordingly, lawful access can be perceived as access to content, based on an open access policy or a contract. The Recital gives the subscriptions as an example of lawful access and adds that it also includes access to content that is "freely available online". However, the criterion of 'lawful access' does not indicate that right holders are entitled to prohibit TDM activities contractually (e.g. through the terms of agreement). Any contractual clause that is in violation of Article 3 is unenforceable[201].

**2. Article 4 of the Digital Single Market Directive**

Article 4 of the DSM Directive provides another exception on the right of reproduction:

"*1) Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.*

*(2) Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.*

*(3) The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.*

[201] See Article 3 of the DSM Directive.

*(4) This Article shall not affect the application of Article 3 of this Directive*."

Article 4 does not have limited beneficiaries and can be applied to everyone, as opposed to Article 3. Moreover, the Article allows TDM to be used for any purpose irrespective of whether it is economic or not. However, the Article limits its use to lawfully accessible works for which the author has not reserved the right to make reproduction. Thus, with this Article right holders are given the right to "opt-out"[202]. Put differently, data mining is allowed by default, unless this type of utilization of the works has not been explicitly reserved in an "appropriate manner"[203].

The Article does not provide an explanation as to how the rights can be reserved in an "appropriate manner" by their holders. The Recital 18 of the DSM refers to this issue in the following way: "*In the case of content that has been made publicly available online, it should only be considered appropriate to reserve those rights by the **use of machine-readable means**, including **metadata** and **terms and conditions of a website** or a service. Other uses should not be affected by the reservation of rights for the purposes of text and data mining. In other cases, it can be appropriate to reserve the rights by other means, such as **contractual agreements** or a **unilateral declaration**. Rightholders should be able to apply measures to ensure that their reservations in this regard are respected*." [emphasis added].

---

[202] Margoni, T. and Kretschmer, M., 2021. A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology. *Ducato, Rossana and Strowel, Alain M., Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out (February 1, 2021). European Intellectual Property Review*, [online] p.25. Available at: <https://ssrn.com/abstract=3829858> [Accessed 6 March 2022].

[203] Gatti, L., 2021. The European solution for text and data mining: a focus on the opt-out system provided by article 4 of the DSM Directive. *European Intellectual Property Review*, p.3.

There are certain tools to reserve the rights appropriately in the digital realm. One of them is adding a 'robot.txt' file, also known as 'robots exclusion standard', to the online content[204]. The robot.txt file provides instructions to inform web robots about which parts of the website should not be processed or scanned[205].

Although not directly related to the opt-out system of the TDM exception, in 2010, the German Federal Supreme Court issued a decision regarding Google's use of images taken from an artist' website[206]. Upon searching the artists' name, Google's search engine displayed thumbnail pictures of her work, obtained from her own webpage. As it is widely known, the search engine provides a feature that allows users to search for pictures that have been posted on the Internet by others. The Court determined that Google was not infringing any exclusive rights by showing these pictures as the artist did not employ the basic technological tool allowing her to prevent Google from indexing her webpage. The

---

[204] Hugenholtz, P., 2019. *The New Copyright Directive: Text and Data Mining (Articles 3 and 4) - Kluwer Copyright Blog*. [online] Kluwer Copyright Blog. Available at: <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/> [Accessed 20 February 2022].

[205] For more information about the robots exclusion standard see Google Developers. n.d. *How Google interprets the robots.txt specification | Documentation | Google Developers*. [online] Available at: <https://developers.google.com/search/docs/advanced/robots/robots_txt> [Accessed 21 February 2022].https://developers.google.com/search/docs/advanced/robots/robots_txt

[206] German Federal Supreme Court (Bundesgerichtshof), 29 April 2010, I ZR 69/08 ('Vorschaubilder'). For comment on the case, see Clark, B., 2010. Google Image Search does not infringe copyright, says Bundesgerichtshof. *Journal of Intellectual Property Law & Practice*, 5(8), p. 553. and Clark, B., 2022. *BGH: Google's image search is no copyright infringement*. [online] The IPKat. Available at: <https://ipkitten.blogspot.com/2010/04/bgh-googles-image-search-is-no.html> [Accessed 10 February 2022].

decision could set an example for right holders to deploy necessary tools, who wish to reserve their use of works for TDM[207].

In 2018, a research[208] revealed that various online mediums are restricting the TDM activities in their terms of service, which shows a general trend towards prohibiting TDM. In case of publishers, the newspaper publishers are of the opinion that licensing should be done individually for each case, even for non-commercial researches in order to avoid widespread exploitation[209]. Various publishers such as Oxford University Press and Springer Nature joined the Copyright Clearance Center's "Right Find RightFind™ XML for Mining"[210]. The platform offers a single source of full-text XML content on request from publications the users have subscribed to[211].

The exception might create legal uncertainty for people who are engaged in AI training[212]. The researcher or the person behind the training, should assess for each work,

---

207 Gatti, p. 4.

208 Ducato, R. and Strowel, A., 2018. Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility. *SSRN Electronic Journal*, p. 22.

209 Rosati, E., 2019. Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. *Asia Pacific Law Review*, 27(2), p.202.

210 For more details see https://www.rightsdirect.com/rightfind-xml-mining-faq/

211 Rosati, p. 203.

212 See, Rosati , pp. 214 and 217; Chiou, p. 409; Škiljić, A., 2021. When Art Meets Technology or Vice Versa: Key Challenges at the Crossroads of AI-Generated Artworks and Copyright Law. *IIC - International Review of Intellectual Property and Competition Law*, 52(10), p.7; Kretschmer, M. and Margoni, T., 2018. *The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs – CREATe*. [online] Create.ac.uk. Available at: <https://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directive-digital-single-market-not-what-eu-copyright-needs/> [Accessed 12 February 2022].

if the right holder has opted-out. Since the AI training requires generally thousands of works, this exception is not considered as an incentive for the use of AI technologies.

In light of all the above, the utility of the exception of Article 4 for the TDM activities is open to discussion, especially because of the opt-out mechanism foreseen by it. It is even argued that rightholders' reservation right might destroy the exception's whole purpose[213].

## IV. APPROACH OF THE UNITED STATES OF AMERICA

The approach of the US towards the limitations and exceptions is different from the EU. As seen above, the EU system adopted a specific list of exceptions whereas in the US a case-by-case approach is employed[214]. The courts examine four elements to decide if the use of a material protected by copyright without the right holder's authorization can be "fair" and not result in copyright infringement.

A doctrine that was developed for balancing the exclusive rights of the right holders and the public interest, fair use is named as the "*most troublesome in the whole law of copyright*"[215]. The doctrine seeks to determine if permitting the usage rather than prohibiting it, would be a better way to achieve the copyright's aim of "promoting the Progress of Science and useful Arts" [216].

Fair use is regulated in the Copyright Act of 1976, section 107 with the heading "Limitations on exclusive rights: Fair Use":

[213] Rosati, p. 215.

[214] Sections 108-122 of the Copyright Act include specific limitations to the exclusive rights. Absence of any specific exemptions, the Courts can still decide against infringement due to the doctrine of fair use.

[215] Dellar v. Samuel Goldwyn, Inc., 104 F.2d 661, 662 (2d Cir. 1939).

[216] Arica Inst., Inc. v. Palmer, 970 F.2d 1067, 1077 (2d Cir. 1992).

*"[…]the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include-*

*(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;*

*(2) the nature of the copyrighted work;*

*(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and*

*(4) the effect of the use upon the potential market for or value of the copyrighted work. [...]"*

The intention of the Congress by drafting the section, was to enable courts "*to adapt the doctrine to particular situations on a case-by-case basis*"[217]. The Congress noted in its Report that in a time where technological changes are rapid, instead of freezing the doctrine in the act, the bill certifies the aim and overall framework of the fair use doctrine[218].

The four factors of the fair use doctrine will be examined below, however, it should be noted that these factors are not to be "treated in isolation, one from another. All are to be explored, and the results weighed together, in light of the purposes of copright."[219]

[217] House of Representatives Report No. 94-1476, 1976, p.66 (1976), as reprinted in 1976 U.S.C.C.A.N. 5659, 5679. Available at https://www.copyright.gov/history/law/clrev_94-1476.pdf [Accessed August 7, 2022]

[218] Ibid.

[219] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 (1994); accord Castle Rock Entrtainment, Inc. v. Carol Publishing Group., Inc., 150 F.3d 132 (2d Cir. 1998).

### A. Purpose and Character of the Use

The first factor is considered as essential in deciding cases where the fair use defense is argued[220]. Whether the use is commercial or not is decisive in terms of the first factor. However, according to the US Supreme Court, the "central purpose" of the first factor is to establish whether the new work merely replaces the original one or it "*adds something new*", which has a further "*purpose or different character*".[221] In other words, the Courts examine if the new work is "transformative".

Transformation encompasses "*altering the original with new expression, meaning or message*"[222]. The Supreme Court decided that "*the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use.*"[223]

The Court's preference to use the conjunction "or" in the decision implies two possibilities: purpose or character[224]. The transformative character, in other words 'transformative use' means that the work is utilized to generate a new work, as it is the

---

[220] Beebe, B., 2020. An Empirical Study of U.S. Copyright Fair Use Opinions Updated, 1978-2019. *10NYU Journal of Intellectual Property & Entertainment Law 1*, p.23. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3758229 [Accessed February 26, 2022].

[221] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, paragraph 579 (citations omitted): "The central purpose of this investigation is to see, [..], whether the new work merely "supersede[s] the objects" of the original creation, or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is "transformative."

[222] Ibid.

[223] Ibid.

[224] Lee, E., 2010. Technological Fair Use. *Southern California Law Review*, 83(797), pp.837. Available at: https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=1356&context=fac_schol [Accessed February 27, 2022].

case with parodies for instance. Hence, the character of the work changes. Contrarily, 'transformative purpose' means that the work is copied entirely but repurposed. In sum, a new work is created through transformative use, whereas in cases where transformative purpose is concerned, an identical copy of work is taken and it is applied to a new purpose[225].

The US Courts did not expressly decide on a case where the ML process involved using copyrighted material, without the right holder's authorization. However, several circuit court decisions can be guiding to evaluate whether creating and modifying digital copies of works to train AIs is 'transformative'.

In the *Authors Guild v. HathiTrust*[226] case, an organization named HathiTrust was formed to operate a digital library called "HathiTrust Digital Library", (HDL). Members of the organization consisting of universities, non-profit organizations and colleges, made their collections of physical books available for scanning. Thus, HDL included over ten million digital reproduction of works. The HDL allowed the public to look for specific words within the digital copies[227]. The search result did not show any text from the work, hence the user neither could see the page, which included the searched word, nor any other part of the work. Solely the page number on which the search word is located inside the work and the term's frequency of appearance were displayed in the search results[228].

---

[225] Valente, M.-A., 2016. Transformativeness in the Age of Mass Digitization. *St. John's Law Review*, 90(233), p.238.

[226] Authors Guild, Inc., et al., v. HathiTrust, et al., 755 F.3d 87 (2d Cir.2014).

[227] See ibid., two more uses were permitted by the HDL: "*Second, the HDL allows member libraries to provide patrons with certified print disabilities access to the full text of copyrighted works. [..] Third, by preserving the copyrighted books in digital form, the HDL permits members to create a replacement copy of the work, if the member already owned an original copy, the member's original copy is lost, destroyed, or stolen, and replacement copy is unobtainable at a "fair" price elsewhere*."

[228] Ibid.

The legal issue that had to be solved by the Court was whether the HDL's use of the works represented copyright infringement or whether its actions were protected under the fair use doctrine. The Court acknowledged that to provide a full-text search of works, HDL had to create digital copies of the book in their entirety. However, in Court's opinion, creating a full-text searchable database was "*quintessentially transformative use*"[229]. The Court decided that HathiTrust's acts were within the scope of the fair use doctrine because "*the result of a word search is different in purpose, character, expression, meaning and message, from the page (and the book) from which it is drawn.*"[230].

Another significant decision in the field is *Authors Guild v. Google, Inc*[231]. Google scanned nearly 20 million books involving copyrighted ones, in collaboration with large libraries. Based on these scans, the company created a corpus of machine-readable texts, which is used to power its service called Google Books.

Being a search engine accessible to public, Google Books offers its users various services. It allows the users to receive a full list of books in the database involving the search word entered by the user. Google points out that the search function enables a researcher to find books, among millions, that use the word picked by the researcher, as well as those that do not. According to Google, this information would otherwise not be available throughout the course of a lifetime searching[232]. The search engine also provides information about the trend of word and phrase usages throughout the time via TDM[233].

---

[229] Ibid.

[230] Ibid.

[231] Authors Guild, Inc., et al., v. Google, Inc., 804 F.3d 202 (2d Cir.2015).

[232] Ibid.

[233] For the Court's example see ibid: "*tracking the frequency of references to the United States as a single entity ('the United States is') versus references to the United States in the plural ('the United States are') and tracking how that usage has changed over time*."

Different from the *HathiTrust* case, Google Books search program permits searchers to read small parts from the book searched.

Considering the circumstances of the case, the Court decided that both the search and the snippet function of GoogleBooks were transformative. The Court held that the "*purpose of Google's copying of the original copyrighted books is to make available significant information about those books*"[234].

Moreover, Google's function to present the frequency of words within the scanned copies, through TDM techniques is transformative in the opinion of the Court. According to the Court, the snippets added to the Google Books' transformative purpose as it revealed some parts of the books containing the searched word and made available just enough context for the searcher to determine whether or not she needs to obtain the work[235].

One of the Plaintiffs argument was that Google was a profit-motivated company and it had a commercial motivation, which should overrule the applicability of fair use. The Court pointed out that "*there is no reason for presuming that a commercial use is not a fair use*"[236]. Thus, it rejected the Plaintiffs claim due to Google Books' "*highly transformative*" purpose.

In 2021, the Supreme Court concluded the dispute between *Google LLC and Oracle America, Inc.*[237] after 10 years of litigation. The decision carries a specific importance, as it is a Supreme Court decision regarding fair use after over twenty-five

---

[234] Ibid.

[235] Ibid.

[236] The Court also noted: "Our court has since repeatedly rejected the contention that commercial motivation should outweigh a convincing transformative purpose and absence of significant substitutive competition with the original."

[237] Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021).

years. The Court decided that Google's unauthorized copying of parts of Oracle's Java application programming interface ("API") in its Android programming platform was fair use. In the context of the first factor, the Court examined Google's development of the Android platform, in which it "reimplemented" the API, as well as all of the additional products and programming capabilities that Android spawned[238].

The Court noted that Google's purpose for copying was to develop a new software for Android phones, which the Court found "*highly creative and innovative*" and in line with copyright's inventive aims. Google's copying of the API was identified by the Court as "*only insofaras needed to include tasks that would be useful in smartphone programs*".

Moreover, the Court defined Google's acts as "reimplementation" in the following words: "*Google did as 'reimplementation', defined as the 'building of a system . . . that repurposes the same words and syntaxes' of an existing system—in this case so that programmers who had learned an existing system could put their basic skills to use in a new one"*[239]. Thus, the Court found Google's act of copying as transformative.

The decision also touched upon the role of "bad faith" in determination of fair use. The Judges in this regard noted that "*[c]opyright is not a privilege reserved for the well-behaved.*"[240].

[238] Anon, 2021. Google LLC v. Oracle America, Inc.. *Harvard Law Review*. Available at: https://harvardlawreview.org/2021/11/google-llc-v-oracle-america-inc/#:~:text=141%20S.-,Ct.,functions%20from%20scratch.%E2%80%9D%20Id. [Accessed March 4, 2022].

[239] Ibid.

[240] Ibid. For more information regarding the Supreme Court's decision see Myers, G., 2022. Muddy Waters: Fair Use Implications of Google v. Oracle America, Inc. *Northwestern Journal of Technology and Intellectual Property*, 19(2). Available at: https://scholarship.law.missouri.edu/cgi/viewcontent.cgi?article=2014&context=facpubs [Accessed April 9, 2022].

Using the Supreme Court's logic, a programmer's act of scanning copyrighted works to build a corpus of training data for the AI may be considered as "transformative" thus protected in the context of the fair use doctrine[241]. Because the objective of an author in creating works is to communicate through artistic expression, make a living, and/or practice their skill. The AI programmer, on the other hand, turns works into training data so that an AI program may improve its algorithm[242]. Because the goal of digitizing copyrighted works to train an AI program differs from the objective of an authors' creation of works, digitization to train an AI program is likely to be qualified as transformative[243].

However, right holders may claim that unlike in the case of search engines, certain ML systems do not provide customers with a creative "end product" that is different in nature or purpose than the training work[244]. Instead, it is argued that ML is a user of copyrighted content, producing a commercial technology rather than novel creativity, and these for-profit users should pay for what they acquire[245].

**B. The Nature of the Copyrighted Work**

The second factor concerns if the work is factual or creative in nature and if it has been published earlier. If the work is mainly informative or factual, deciding in favor of a fair use is more likely.

In the case of *HathiTrust,* in terms of the second factor, the Court only noted that "*this factor 'may be of limited usefulness where, as here, the creative work…is being used*

[241] Gillotte, J., 2020. Copyright Infringement in AI-Generated Artworks. *UC Davis Law Review*, 53(5). Available at: abstract=3657423 [Accessed March 11, 2022].

[242] Ibid.

[243] Ibid.

[244] Lemley, M. and Casey, B., pp. 142.

[245] Ibid.

*for a transformative purpose*"[246]. Similarly, in the case of *Google Books,* the Court decided that the second factor's role in determining the case is very little[247]. The Court stated that if the copied books were fiction that would not change their decision, as Google's use was transformative[248].

### C. Amount and Substantiality of the Portion Copied

The third factor regards the amount and substantiality of the part used with respect to the copyrighted work. When little amounts are reproduced deciding for a fair use is more probable unless the little amount is the "heart" of the original work.

In the *Google Books* case, the program copied the books completely. However, the Court stated that copying entirely without changing anything has been considered as fair under two circumstances: firstly, the copying should be *"reasonably appropriate to achieve the copier's transformative purpose*" and it should be done in a way that *"it does not offer a competing substitute for the original*"[249]. The Court concluded that Google's reproducing of books in their entirety was not only appropriate for the transformative purpose but it was "*literally necessary*"[250]. The Court also took into account the fact that the authorized digital copies of books were not revealed to the public entirely but were only used to present certain information about the books[251]. For the second circumstance, the Court decided that the snippet view of the books did not reveal much that would amount to a competing substitute of the books.

---

[246] Authors Guild, Inc., et al., v. HathiTrust, et al., 755 F.3d 87 (2d Cir.2014).

[247] Authors Guild, Inc., et al., v. Google, Inc., 804 F.3d 202 (2d Cir.2015).

[248] Ibid.

[249] Ibid.

[250] Ibid.

[251] Ibid.

In its recent decision of *Oracle v. Google*, while acknowledging that Google's copying was large in nature, the Supreme Court focused on the purpose of copying and indicated that "*Google copied those lines not because of their creativity, their beauty, or even (in a sense) because of their purpose*."[252] According to the Court, where the quantity of the copying was connected to a "valid and transformative purpose", the "substantiality" factor will usually lean towards upholding of a fair use[253].

In the context of ML, copying the works entirely to set up a training corpus could be assessed  as fair use, as the aim of utilizing the works would be different than the author's initial aim in creating the work[254].

**D. Effect Upon the Potential Market or Value of the Copyrighted Work**

The last factor concerns if the unauthorized copy constitutes a competitive replacement for the original work.

In the famous cornerstone decision of *Campbell v. Acuff-Rose Music, Inc.,* the Supreme Court found that the as the degree transformative use increases, the likelihood of the use to hinder the market for the original work decreases[255]. However, the Court in Google Books case stated if taken place in a way that leads in broad disclosure of sufficiently important sections of the original work, copying may impair the value of the original work, despite the transformative purpose[256]. In the case, although the Court recognized that Google Books' snippet view function could result in some loss of sales, it concluded that it is not enough to make the copy an "*effectively competing substitute that*

---

[252] Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021).

[253] Ibid.

[254] Gilotte, p. 2685.

[255] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569.

[256] Authors Guild, Inc., et al., v. Google, Inc., 804 F.3d 202 (2d Cir.2015).

*would tilt the weighty factor in favor of the rights holder in the original*"[257]. According to the Court, "*there must be a meaningful or significant effect upon the potential market for or value of the copyrighted work*"[258]. In the *Oracle v Google* decision, the Supreme Court pointed out that a balance should be made between public interest and right holders' losses.

ML has the potential to change the balance of fourth factor because it may substitute not only for the original works on which it trains but also for the authors of those works[259]. Here, a distinction should be made. Firstly, the training data can be used for non-generative AI algorithms and process; in that case AI is capable of performing functions that do not result in expressive activity[260]. Secondly, the training data can be used to create further expressive works[261].

Some uses of ML could deprive authors of markets they benefit from, if the AI uses the training data to generate expressive works. Newspaper writers or journalists might be in danger, for instance, as there is a company, which creates articles based on an algorithm. The founder estimates that "more than 90 percent" of news reporting will be generate by computers by 2027[262]. Another example may be Jukedeck, a company that

[257] Ibid.

[258] Ibid.

[259] Sobel, B., 2017. Artificial Intelligence's Fair Use Crisis. *Columbia Journal of Law & the Arts, Forthcoming*, p.32. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3032076.

[260] Comments of the Authors Guild, Inc. submitted by Mary Rasenberger In the Matter of Impact of Artificial Intelligence Technologies on Copyright, pp.8. Available at: https://www.authorsguild.org/wp-content/uploads/2020/01/Authors-GuIld-Responses-to-USPTO-AI-NOI1.pdf [Accessed March 18, 2022].

[261] Ibid.

[262] Levy, S., 2012. Can an algorithm write a better news story than a human reporter? Wired. Available at: https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/ [Accessed March 23, 2022].

composes music via AI that is trained through ML. The company offers royalty-free uses of the songs therefore; it is possible that it can harm the recording market of human artists.[263] However, there is no consensus that AI generated works will be a substitute. It is also argued that AI-generated art is a new category of art; hence it is not very likely that it will overlap with the original work's market[264]. Besides, if the works are used by the ML to carry out a non-expressive function, it is argued that that the output would not compete with the original works used to train the AI[265].

To conclude, although the act of copying during the ML stages can pass the first three factors of the fair use, it is debatable whether it would pass the fourth factor. However considering that the fair use doctrine does not have clear-cut lines and it is decided on a case-by-case basis, it is difficult to reach a definite conclusion.

## V. APPROACH OF TÜRKİYE

In Turkish law, the LIAW makes it possible to limit the exclusive rights only in cases specifically specified in the law. Indeed, the LIAW regulates the relevant issue under the title of "Limitations" between articles 30-47. For instance, in Article 30, the reproduction of works are allowed for court proceedings or forensic purposes; article 35 regulates freedom of quotation and article 38 allows personal use of works without pursuing profit.

Each of these limitations, some are individual and some are related to public actions, are clearly defined in the law and if a use falls within the scope of the mentioned articles, it is considered an exemption from copyright infringement.

---

[263] Sobel, p. 30.

[264] Gilotte, p. 2689

[265] Comments of the Authors Guild, Inc., p. 9

According to Turkish Court of Cassation, in terms of copyright infringments, either the action must be within the scope of the specified limitations or the right holder must have given explicit written consent to the relevant use. However, the Court of Cassation started to refer to CJEU jurisprudence in the dissenting votes, by stating that it is necessary to evaluate if the use is fair or not based on the circumstances of the case[266].

To conclude, the LIAW does not have a "fair use" clause or specific exceptions such as the ones in the EU for the developing technology. The adequacy of the limitations specified in the LIAW in terms of meeting the needs of new technological developments is often the subject of debate. One of these debates concern the temporary reproductions due to technical necessities.

Although it is argued that temporary reproductions due to technical necessity should not be considered as copyright infringement, there is no special regulation in the LIAW that would permit it[267].

Due to Turkey's alignment process with the EU acquis and the need to address new technologies, the Ministry of Culture and Tourism published a draft in 2017 called "Draft Law Regarding Making Amendments in the Law on Intellectual and Artistic Works" ("Draft Law")[268]. In Draft Law, a new exception for temporary reproductions is envisaged. According to the draft exception; "*temporary or incidental reproductions, which are an integral and essential part of a technological transaction process, whose*

---

[266]Turkish Court of Cassation, 19th Criminal Chamber, Case No: 2015/19627, Decision No: 2018/7669; Turkish Court of Cassation, 19th Criminal Chamber, Case No: 2016/12740, Decision No: 2018/7674.

[267] Karaarslan, P. 2022. Fikir ve Sanat Eserlerinin Teknik Gereklilik Nedeniyle Geçici Çoğaltılması, TFM, C:8, S:1, p. 29-49. Available at : https://dergipark.org.tr/en/download/article-file/2125864 [Accessed July 30, 2022].

[268] Küçükali, C., 2017. 5846 Sayılı Fikir ve Sanat Eserleri Kanun Tasarısı Üzerine Bir Değerlendirme, Terazi Hukuk Dergisi, 12, no. 135.

*sole purpose is to ensure the lawful use of a work or to transmit it to third parties in a network, and which do not have an independent economic value, are allowed*"[269]. The exception is drafted in parallel with the Article 5(1) of Infosoc Direvtive. Browsing, caching and streaming data from legal platforms, can be given as examples of reproductions that meet all the conditions in the draft article[270].

It is not certain when the Draft Law will be accepted and integrated into the LIAW. Thus, it is safe to say that the current copyright law in Türkiye does not have any exceptions and/or limitations that would allow the copying happening during the machine learning stages.

## VI. EXTENDED COLLECTIVE LICENSING – POSSIBLE SOLUTION?

Considering the ambiguities of whether the possibly copyright-infringer acts of ML would be covered under the exceptions in the EU or the fair use doctrine in case of the US, another legal approach that balance the rights between right holders and the interests of the users in the digital world should be adopted.

AI companies need access to large volumes of data but the quantity itself is not enough. "*A common misunderstanding is about the quantity of data needed for machine learning when in reality the quality of data is really the key*."[271]. The high quality of the data is usually found in the form of copyrighted content[272]. The owners of these

---

[269] Karaarslan, p. 40.

[270] Ibid., p. 47.

[271] WIPO Conversation on Intellectul Property (IP) and Artificial Intelligence (AI), 2019. Available at: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_3_ge_20/wipo_ip_ai_3_ge_20_inf_5.pdf [Accessed March 24, 2022].

[272] Copyright Clearance Center, 2020. Written Comments of Copyright Clearance Center, Inc. (submitted 13 February 2019), in the content of the WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI) Second Session (Draft Issues Paper on Intellectual Property and Artificial Intelligence,

copyrighted contents want to be compensated for the use of their work. One way to create balance between these two interests could be the adoption of a "collective licensing".

During the course of ML when mass copying takes place, it is very impractical for AI companies to find each of the copyright holder and obtain permission on a work-by-work basis. Extended collective licensing (ECL) has been considered as a conceivable way to provide licenses for mass use[273]. In fact, it is argued that that ECL was designed to "*resolve the contradiction between copyright exclusivity and mass use*", and the problem with the mass use of works is "*how to grant licenses when the identification of individual use is not possible or requires an unreasonable amount of effort*"[274].

ECL was introduced to the legal sphere by the Nordic countries in the 1960s[275]. It stands in between of voluntary licensing and compulsory licensing[276]. In this system, a collective rights management organization (CMO) is authorized by the government to negotiate licenses for a certain class of works, (such as books or magazines) or a certain class of uses (such as reproduction for scientific aims)[277].

Jan. 14,2020). Available at: https://www.copyright.gov/disted/comments/init026.pdf [Accessed March 24, 2022].

[273] Comments of the Authors Guild, Inc., p. 10

[274] Huuskonen, M. 2006. Copyright, Mass Use and Exclusivity: On the Industry Initiated Limitations to Copyright Exclusivity, Especially Regarding Sound Recording and Broadcasting. Helsinki: Helsingfors universitet, p. 201.

[275] Stokkmo, O. 2019. The Role of Collective Licensing, International Publishers Association, p. 9.

[276] Office of the register of copyrights , 2011. Legal Issues in Mass Digitization: A Preliminary Analysis and Discussion Document. United States Copyright Office., pp. 35 Available at: https://www.copyright.gov/docs/massdigitization/USCOMassDigitization_October2011.pdf [Accessed March 24, 2022].

[277] Ibid.

The license, which is given by the CMO on behalf of its members, is automatically extended by legislation to all of the right holders irrespective of whether they are part of the CMO or not – unless they opt-out[278]. Different from the compulsory licensing regime, the royalty rates and terms of use is determined by a private entity instead of a government.

There are two safeguard measures for outsiders, namely the right to individual remuneration and the opt-out possibility.

By extending collective agreements to right-holders of works of the identical nature, it enables users to use all works in the identical field without raising the transaction cost to unacceptable levels.

Below, the legal situation in the EU, US, and Türkiye will be examined to evaluate whether it would be possible to create a ECL system that would allow the AI-developers to train the AI with the above-explained ML stages through the use of copyrighted material.

### A. European Union

ECL is explicitly recognized in Articles 8 and 12 of the DSM Directive. Article 12 is a non-mandatory provision that is applicable to all types of works, whereas Article 8 is a mandatory provision on the use of works that are out-of-commerce[279]. Despite being optional, Article 12 harmonizes Member States' regulations on ECLs and gives them a

[278] European Broadcasting Union, Extended Collective Licensing A Valuable Catalyst for the Creative Content Economy in Europe. , p.5. Available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Extended-Collective-Licencing_EN.pdf.

[279] European Copyright Society, 2020. Comment of the ECS on Arts. 8 and 12 DSM-D, 11JIPITEC 101 p. 101.

limited freedom. Hence, the rules on ECL in each Member State must be in accordance with the safeguards provided by Article 12[280].

For a CMO to be entitled to enter into an agreement with extended effect, Article 12(1) of the DSM Directive necessitates a number of conditions that need to be met cumulatively. The CMOs must be "***sufficiently representative***" of the types of works and of the rights, which are the subject of the licence[281]. A CMO that is sufficiently representative, shall conclude an agreement only within the "***well-defined areas of use***", in which "*obtaining authorisation from rightholders on an individual basis is typically onerous and impractical to a degree that makes the required licensing transaction, namely one involving a licence that covers all rightholders concerned, unlikely to occur due to the nature of the use or of the types of works or other subject matter concerned.*"[282]

Article 12(3) lists the safeguards that Member States must ensure in terms of the protection of rightholders who did not mandate the CMO offering the license[283]. These safeguards should be applied in a non-discriminatory manner, irrespective of rightholders' nation and or residency. Equal treatment of all right holders, especially in terms of access to information on the license agreement and the sharing of remuneration should be guaranteed[284].

Another safeguard provided in the Article 12, is the right to opt-out. Accordingly, "*rightholders who have not authorised the organisation granting the licence may at any time* ***easily*** *and* ***effectively*** *exclude their works*" [285] (emphasis added). The rightholders

[280] Ibid.

[281] Recital 48 of the DSM Directive.

[282] Recital 47 of the DSM Directive.

[283] Recital 48 of the DSM Directive.

[284] Ibid.

[285] Article 12(3)(c) of the DSM Directive.

have the right to opt-out before the conclusion of a licence and during the term of the licence[286]. Lastly, to let the rightholders know about the capacity of the CMO to license works, about the place of the license and their right to opt-out, necessary publicity measures should be taken[287]. The Recital 48 notes that the publicity measures should not involve "*imposing a disproportionate administrative burden on users, collective management organisations or rightholders, and* ***without the need to inform each rightholder individually****.*"[288] (emphasis added).

The DSM Directive noted that "*given the increasing importance of the ability to offer flexible licensing schemes in the digital age*" , Member States should be entitled to create ECL systems in compliance with their national traditions. Thus, the establishment of the ECL system in the EU is not aimed at creating cross-border licenses[289]. It is clearly mentioned in the Recitals that the effect of an ECL mechanism is restricted within the jurisdiction of the concerned Member State[290].

To conclude, the Member States can create their own ECL regime provided that they adhere to the minimum safeguards provided by the DSM Directive.

### B. United States of America

There is no existing ECL system for the use of works in the US. The Copyright Office proposed ECL for mass digitization in its report published in 2015[291]. The

[286] Recital 48 of the DSM Directive.

[287] Rosati, E., 2021. Copyrigth in the Digital Single Market. Oxford University Press, p. 230.

[288] Recital 48 of the DSM Directive.

[289] Rosati, p. 226. See also Recital 46 of the DSM Directive : "Such mechanisms should only have effect in the territory of the Member State concerned, unless otherwise provided for in Union law".

[290] Recital 46 of the DSM Directive : "Such mechanisms should only have effect in the territory of the Member State concerned, unless otherwise provided for in Union law".

[291] United States Copyright Office, p. 82.

Copyright Office, after receiving stakeholder views, proposed adopting a pilot program to establish a limited ECL scheme for the mass digitization of three types of works[292].

The Copyright Office asked for public's comments on its proposal and the comments were predominantly negative[293]. The most outstanding objection was that the fair use doctrine already allowing most of what the propsed ECL scheme would bring[294]. The libraries commented:

"*Despite the Office's assurances about inclusion of a fair use savings clause, we are nonetheless concerned that the proposed ECL system would cast a shadow over potential fair use assertions for mass digitization, steering organizations like ours away from using that important right and into more conservative and more costly licensing practices*"[295]

Consequently, the Copyright Office decided that there is no sufficient stakeholder agreement, hence implementation of an ECL scheme would be difficult.

### C. Türkiye

There is no regulation for an ECL organization in Türkiye. Collecting societies are the only authorized CMOs in Turkey. There is no definition of CMOs in the LIAW but the aim of the establishment of CMOs is provided in Article 42 of the LIAW as "*protect*

---

[292] Three types of works were: literary works, pictorial or graphic works published in literary works, and photographs.

[293] Reed, C.S., 2019. Next Great Copyright Act: U.S. Copyright Policy for the 21st Century, Edward Elgar Publishing Ltd, p. 65.

[294] Ibid.

[295] Ibid citing Letter from Karyn A. Temple Claggett, Acting Register of Copyrights, to Chairman Bob Goodlatte and Ranking Member John Conyers of the H. Comm. on the Judiciary 4–5 (2017).

*the mutual interests of their members, the management and pursuit of the rights granted by this Law and the collection and distribution of fees to the rightholders*."

### D. Is Extended Collective Licensing a Possible Solution?

If an ECL system is created, the AI developers can contact a single CMO to obtain license for the works needed to train the machine. Hence, the right holders would be compensated for the use of their works. Outside right-holders have the same right to remuneration as members and the CMO has the duty to actively search for those who have a right to remuneration under an ECL agreement.

Besides, licensing for copying purposes is a very old concept and it is becoming more widespread in relation with AI. For instance, Copyright Clearance Center, which is a pioneer in collective licensing, has made available collective licensing of published material for TDM for the purposes of research[296]. However, would it be a solution in practice for the copyright crises that AI developers are likely to face or already facing?

The operation of ECL systems in practice has been subject to skepticism. Its propitiousness, legitimacy and efficiency has been under question[297]. For the ECL system to operate effectively, certain safeguards for the rightholders must be provided and followed-up by an authority. Every step of the licensing such as the step of fees collection and redistribution, terms of the license, must be transparent. A "culture of collective management" must exist[298]. Considering the potential monopoly that the CMOs will have,

[296] Copyright Clearance Center, pp. 5. See also. Springer partners with Copyright Clearance Center to power text and Data Mining Solution. Copyright Clearance Center. 2021. Available at: https://www.copyright.com/springer-partners-with-copyright-clearance-center-to-power-text-and-data-mining-solution/ [Accessed March 24, 2022].

[297] IFLA, 2018. Background Paper on Extended Collective Licensing, p. 5.

[298] Ibid, p. 6.

it is especially important that the CMOs where the ECL system is operating, are sufficiently representative, transparent and have a good governance. Another possible barrier for creating a well-functioning ECL system is the expense and bureaucracy of the system, as both human and organizational resources are needed.

As discussed above, the EU has a clear provision allowing Member States to introduce an ECL system provided that the minimum safeguards listed in the relevant article are met. However, the scope of the ECL is limited to each Member State's own jurisdiction and there is no mutual recognition of these licensing agreements. The creation of an ECL system has been discussed in the US but did not receive positive feedback from the stakeholders. In Turkish legal system, it is not recognized by the law and its establishment has not been discussed. Moreover, it can be said thatthe culture of collective management of rights is not fully established in Türkiye.

To conclude, although creation of an ECL to cover mass-copying during the ML stages, could be a solution; in practice the well-functioning of such a licensing system depends on multiple conditions.

## CONCLUSION

The developments in artificial intelligence and the increasing importance of data caused many scholars and governments to rethink traditional copyright rules. The history shows that law and technology shaped each other continuously. This relationship has an influence on the development of technology as law regulates how that technology is used, who gets access to technology, at what costs and under which circumstances[299]. In the case of AI, which brought endless possible applications, the complexity of the relationship increased even more.

AI and its applications are a great asset to humans; it improves the efficiency, accuracy and effectiveness of human efforts. However, for the AI to flourish and perform its activities it needs data. Data is the "*new oil*"[300] and the "*lifeblood of economic development*"[301]. Through the developments in digitalization, copyrighted works can be part of this "data" as "training materials". Machine learning enables AI technologies to learn from these data, which may consist of thousands of unauthorized copies of copyrighted works. During the ML stages, the right of reproduction may be infringed.

Considering the limited scopes of the articles in the EU, the implementation of the TDM exceptions and the exception for temporary reproductions to the ML process would not be very straightforward. Article 3 of the DSM has a limited group of beneficiaries (research organizations and cultural heritage institutions) and a limited scope; whereas the Article 4 of the DSM has a broader scope but under this article, right

[299] Margoni, T. and Kretschmer, M., p. 33.

[300] The Economist, 2017. The world's most valuable resource is no longer oil, but Data. Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [Accessed March 26, 2022].

[301] European Commission, 2020. A European strategy for data. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066 [Accessed March 26, 2022].

holders are entitled to reserve the right to make reproductions, in other words they can "veto" the use of their works in ML process.

For the exception foreseen in the Article 5 of the Infosoc to be applicable, the acts of reproduction should be temporary, transient or incidental, which may or may not be the case in the ML process (depending on the employed technique). Considering all, it is not possible to say if the training process of AI would fall one of the exception categories, hence an individual analysis for each case would be required. However, this legal uncertainty and the risk of facing copyright infringement suits would hinder the researchers' incentive to work further in this field. Consequently, the rules in the field of copyright are an "*obstacle rather than enabler*" in the EU[302].

In the United States, the current fair use doctrine could be applied to the acts of copying during the ML stages. This might be the case when the training works are used for non-generative AI algorithms and process, which does not result in an expressive activity. In other words, works are merely used as "data" and none of the artistic expression is copied.

However if the training data is used to create further expressive works then the transformative character of the work may not be enough to pass the first factor of the fair use doctrine. Similarly, if the work is not considered transformative it is very likely for the AI-generated work to be a substitute of the original works on which it was trained. Nevertheless, from the legal doctrine it can be concluded that the fair use allows specific uses of works if the use transforms the initial work and does not constitute a threat to the original work's market. Therefore, businesses in the US are in a more advantaged situation than the EU-based businesses. The list of exceptions in the EU which have been enumerated, have less flexibility in adapting to advancements in the technological field.

[302] Rosati, p. 217

While it is possible to say that the open-ended fair use doctrine has a broader scope than the list of EU exceptions, it also bears to some extent legal uncertainty. The House Report on the 1976 Act referred to this problem as "*Although the courts have considered and ruled upon the fair use doctrine over and over again, no real definition of the concept has ever emerged. Indeed, since the doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts*."[303] As Judge Leval once stated, "*Writers, historians, publishers, and their legal advisers can only guess and pray as to how courts will resolve copyright disputes*."[304]

Lastly, Türkiye does not have any exception or limitation that would allow the copying of the copyrighted material during the ML stages.

On one hand, there are AI researchers who need to be incentivized to continue in investing and developing these technologies and on the other hand, there are right holders' copyrights. In order to resolve this dilemma, a balance needs to be established between the parties.

Considering the unpredictability of the fair use doctrine in the US and the narrow scope of the copyright exceptions in the EU, those who want to use the copyrighted works as training data without being in danger of facing a lawsuit need to obtain authorization from the right holders. However, it is very impractical for AI companies to find each of the copyright holder and obtain permission on a work-by-work basis. For this reason, extended collective licensing may be a conceivable way to provide licenses for mass use. Nonetheless, its well-functioning depends on multiple conditions and its effectiveness is

---

[303] House of Representatives Report No. 94-1476, at 65 (1976), as reprinted in 1976 U.S.C.C.A.N. 5659, 5679. Available at https://www.copyright.gov/history/law/clrev_94-1476.pdf [Accessed August 7, 2022]

[304]Leval P., 1990. Toward a Fair Use Standard, *Harvard Law Review* 103, Available at https://www.jstor.org/stable/1341457 [Accessed August 7, 2022].

widely disputed. Therefore it is safe to conclude that a new regulation governing this area is needed.

## BIBLIOGRAPHY

### BOOKS, JOURNALS AND ARTICLES

Akın, B., 2006. Fikir Hukuku Dersleri. İstanbul: Arıkan Yayıncılık.

Arslanlı H., 1954. Fikri Hukuk Dersleri II. Fikir ve Sanat Eserleri. İstanbul.

Anon, 2021. Google LLC v. Oracle America, Inc.. *Harvard Law Review*. Available at: https://harvardlawreview.org/2021/11/google-llc-v-oracle-america-inc/#:~:text=141%20S.-,Ct.,functions%20from%20scratch.%E2%80%9D%20Id. [Accessed March 4, 2022].

Ateş, M., 2012. Fikri Hukukta Eser Sahipliği. Ankara: Adalet Yayınevi.

Balganesh S., Manta I.,. and Wilkinson-Ryan T., 2014. Judging Similarity. *100 Iowa Law Review 267*, U of Penn Law School, Public Law Research Paper No. 14-15, Hofstra Univ. Legal Studies Research Paper No. 2014-09, Available at SSRN: https://ssrn.com/abstract=2409811, [Accessed May 22, 2022].

Beebe, B., 2020. An Empirical Study of U.S. Copyright Fair Use Opinions Updated, 1978 2019. *10 NYU Journal of Intellectual Property & Entertainment Law 1*. Available at:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3758229 [Accessed February 26, 2022].

Bozkurt Yüksel, A.E., 2018. Yapay Zekanın Buluşlarının Patentlenmesi, Uyuşmazlık Mahkemesi Dergisi.

Bohler, H.M., 2017. EU Copyright Protection of Works Created by Artificial Intelligence Systems, Master's Thesis, University of Bergen, Available at: <http://bora.uib.no/handle/1956/16479> [Accessed 5 December 2019]

Bridy, A., 2012. Coding Creativity: Copyright and the Artificially Intelligent Author. Stanford Technology Law Review, Available at: <https://web.law.columbia.edu/sites/default/files/microsites/kernochan/09.materials-Bridy.pdf>.

Calo, R., Evtimov, I., Fernandes E., and Kohno, T., O'Hair, D., 2018. *Is Tricking a Robot Hacking?*, University of Washington Tech Policy Lab, Legal Studies Research Paper No: 2018-05

Chiou, T., 2019. *Copyright lessons on Machine Learning: what impact on algorithmic art?* 10 JIPITEC 398.

Clark, B., 2010. Google Image Search does not infringe copyright, says Bundesgerichtshof. *Journal of Intellectual Property Law & Practice*, 5(8).

Denicola, R. C., 2016. Ex Machina: Copyright Protection for Computer-Generated Works. Rutgers University Law Review, Vol. 69:251.

Drexl, J., et al. 2019. Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective, Max Planck Institute for Innovation & Competition Research Paper No. 19-13 [online] Available at https://ssrn.com/abstract=3465577, accessed February 10, 2022.

Ducato, R. and Strowel, A., 2018. Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility. *SSRN Electronic Journal*.

Erdaw, Y. & Tachbele, E., 2021. Machine learning model applied on chest X-ray images enables automatic detection of covid-19 cases with high accuracy. *International Journal of General Medicine*, Volume 14. Available at: https://www.dovepress.com/getfile.php?fileID=73078 [Accessed January 25, 2022].

Erel, Ş.N., 2009. Türk Fikir ve Sanat Hukuku. 3rd Ed. Ankara: Yetkin Basımevi.

European Broadcasting Union, Extended Collective Licensing A Valuable Catalyst for the Creative Content Economy in Europe. Available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Extended-Collective-Licencing_EN.pdf.

European Commission, 2017. The European Data Market Study: Final Report.

European Commission, 2020. A European strategy for data. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066 [Accessed March 26, 2022].

European Commission, 2022. The EU Copyright Legislation, available at <https://digital-strategy.ec.europa.eu/en/policies/copyright-legislation> [Accessed August 2, 2022].

European Copyright Society, 2020. Comment of the ECS on Arts. 8 and 12 DSM-D, 11JIPITEC 101 p. 101.

Executive Office of the President, 2016. *Preparing For the Future of Artificial Intelligence.* [online] Available at <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> [Accessed 29 November 2019].

Flow Machines. 2016. *AI makes pop music – Flow Machines*. [online] Available at: <https://www.flow-machines.com/history/events/ai-makes-pop-music/> [Accessed 6 January 2022].

Frankish K., Ramsey W. M., The Cambridge Handbook of Artificial Intelligence, CUP, Cambridge-UK, 2014.

Gatti, L., 2021. The European solution for text and data mining: a focus on the opt-out system provided by article 4 of the DSM Directive. *European Intellectual Property Review*.

Geiger C., Frosio G., & Bulayenko O., 2018b. Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market' in Seuba X., & Geiger C., & Pénin J. (eds.), Intellectual Property And Digital Trade In The Age Of Artificial

Intelligence And Big Data, Ceipi/Ictsd publication series on "Global Perspectives and Challenges for the Intellectual Property System", Issue No. 5, Geneva/ Strasbourg.

Geiger, C., Frosio, G. and Bulayenko, O., 2018a. Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?. *IIC - International Review of Intellectual Property and Competition Law*, 49(7).

Geiger C., Gervais D., Senftleben M., 2015. Understanding the "Three-Step Test". In D. J. Gervais (Ed.), International Intellectual Property – A Handbook of Contemporary Research, Edward Elgar.

Gervais, D., 2019. The Machine as Author. *Iowa Law Review*, [online] Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359524> [Accessed 3 January 2022].

Gillotte, J., 2020. Copyright Infringement in AI-Generated Artworks. *UC Davis Law Review*, 53(5). Available at: abstract=3657423 [Accessed March 11, 2022].

Gollapudi S., 2019. Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs, Appress, Berkley-USA.

Goodman, M., 2015. *Future Crimes: Inside the Digital Underground and the Battle for Our Connected World*. London: Random House.

Google Developers. n.d. *How Google interprets the robots.txt specification | Documentation | Google Developers*. [online] Available at: <https://developers.google.com/search/docs/advanced/robots/robots_txt> [Accessed 21 February 2022].

Gordon, C., 2021. A market to harness: Speech recognition artificial intelligence (AI) innovations on the rise. *Forbes*. Available at: https://www.forbes.com/sites/cindygordon/2021/12/23/a-market-to-harness-speech-recognition-artificial-intelligence-ai-innovations-on-the-rise/?sh=368012a8134d [Accessed January 25, 2022].

Guadamuz, 2016. The Monkey Selfie: Copyright Lessons for Originality in Photographs and Internet Jurisdiction. Internet Policy Review: Journal on Internet Regulation Volume 5, Issue 1. Available at: <http://policyreview.info/articles/analysis/monkey-selfie-copyright-lessons-originality-photographs-and -internet-jurisdiction>.

Guadamuz, A., 2017. Artificial Intelligence and Copyright. WIPO Magazine. [online] Available at: <https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html>.

Güneş, İ. 2015. Uygulamada Fikir ve Sanat Eserleri Hukuku. 2nd Ed. Ankara: Seçkin Yayınları.

Harvard University Science in the News, 2017. *Special Edition on Artificial Intelligence: History of Artificial Intelligence*. [online] Available at: < http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> [Accessed 27 November 2019]

Hedrick, S. F., 2019. I 'Think', Therefore I Create: Claiming Copyright in the Outputs of Algorithms. NYU Journal of Intellectual Property & Entertainment Law, Vol. 8, No. 2.

Hirsch, E. E., 1948. Fikri ve Sınai Haklar. Ankara: Ankara Basımevi.

Holder C. (ed.), Iglesias, M. (ed.), Triaille J.-P., Van Gysegnem J.-M., 2019. Legal and regulatory implications of Artificial Intelligence. The case of autonomous vehicles, m-health and data mining, Publication Office, [online] Available at: https://op.europa.eu/en/publication-detail/-/publication/f962b17b-5c04-11e9-9c52-01aa75ed71a1/language-en/format-PDF (Accessed 7 February 2022)

House of Representatives Report No. 94-1476, 1976, p.66 (1976), as reprinted in 1976 U.S.C.C.A.N. 5659, 5679. Available at https://www.copyright.gov/history/law/clrev_94-1476.pdf [Accessed August 7, 2022].

Hristov, K., 2017. Artificial Intelligence and Copyright Dilemma. IDEA: The IP Law Review, Volume 57, No. 3., p. 435, Available at: < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2976428>

Hugenholtz, P., 2019. *The New Copyright Directive: Text and Data Mining (Articles 3 and 4) - Kluwer Copyright Blog*. [online] Kluwer Copyright Blog. Available at: <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/> [Accessed 20 February 2022].

Huuskonen, M. 2006. Copyright, Mass Use and Exclusivity: On the Industry Initiated Limitations to Copyright Exclusivity, Especially Regarding Sound Recording and Broadcasting. Helsinki: Helsingfors universitet.

IFLA, 2018. Background Paper on Extended Collective Licensing.

Indurkhy, N., Zhang, T. and Weiss, S., 2010. *Fundamentals of Predictive Text Mining*. Springer.

Jütte, B., Quintains, J., 2021. The Pelham Chronicles: sampling, copyright and fundamental rights. Journal of Intellectual Property Law & Practice, Vol. 16, No. 3.

Karaarslan, P. 2022. Fikir ve Sanat Eserlerinin Teknik Gereklilik Nedeniyle Geçici Çoğaltılması, TFM, C:8, S:1, p. 29-49. Available at: https://dergipark.org.tr/en/download/article-file/2125864 [Accessed July 30, 2022].

Kretschmer, M. and Margoni, T., 2018. *The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs – CREATe*. [online] Create.ac.uk. Available at: <https://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directive-digital-single-market-not-what-eu-copyright-needs/> [Accessed 12 February 2022].

Kumar, E., 2008. *Artificial Intelligence*. I. K. International Pvt Ltd.

Küçükali, C., 2017. 5846 Sayılı Fikir ve Sanat Eserleri Kanun Tasarısı Üzerine Bir Değerlendirme, Terazi Hukuk Dergisi, 12, no. 135.

Leval P., 1990. Toward a Fair Use Standard, *Harvard Law Review* 103, Available at https://www.jstor.org/stable/1341457 [Accessed August 7, 2022].

Lawal, A.B., 2019. *Artificial Intelligence Fundamentals: The Foundations & History of Intelligent Machines.* AB Prominent Publisher.

Lee, E., 2010. Technological Fair Use. *Southern California Law Review*, 83(797) Available at: https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=1356&context=fac_schol [Accessed February 27, 2022].

Lemley, M. and Casey, B., 2020. Fair Learning. Texas Law Review.

Margoni, T. 2018. Artificial Intelligence, machine learning and EU copyright law: who owns AI? CreateWorkingPaper. [online] Available at: https://ssrn.com/abstract=3299523. (Accessed 2 February 2022)

Margoni, T. and Kretschmer, M., 2021. A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology. *Ducato, Rossana and Strowel, Alain M., Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out (February 1, 2021). European Intellectual Property Review*, [online] p.25. Available at: <https://ssrn.com/abstract=3829858> [Accessed 6 March 2022].

McPherson, S.S., 2018. *Artificial Intelligence: Building Smarter Machines*. Twenty-First Century Books.

Moore, J. C., 2018. Artificial Intelligence: What Everyone Needs to Know Today About Our Future, CreateSpace Independent Publishing Platform.

Myers, G., 2022. Muddy Waters: Fair Use Implications of Google v. Oracle America, Inc. *Northwestern Journal of Technology and Intellectual Property*, 19(2). Available at: https://scholarship.law.missouri.edu/cgi/viewcontent.cgi?article=2014&context=facpubs [Accessed April 9, 2022].

Nagy, Z., 2018. *Artificial Intelligence and Machine Learning Fundamentals: Develop Real-World Applications Powered By The Latest AI Advances*. Birmingham: Packt Publishing Ltd.

Navarro-Espinoza, A. et al., 2022. Traffic flow prediction for smart traffic lights using machine learning algorithms. *Technologies*, 10(1). Available at: https://www.researchgate.net/publication/357743745_Traffic_Flow_Prediction_for_Smart_Traffic_Lights_Using_Machine_Learning_Algorithms [Accessed January 25, 2022].

Nicholson, C.V., 2019. Skymind Part 1: An Introduction to Artificial Intelligence. [online] Skymind Inc. Available at: <https://skymind.ai/ebook/Skymind_Introduction_to_Artificial_Intelligence.pdf> [Accessed 26 November 2019].

O’Regan, G., 2008. A Brief History of Computing. Springer Science & Business Media.

OECD., 2019, Artificial Intelligence in Society. [online] Paris: OECD Publishing. Available at: <https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en> [Accessed 26 November 2019].

Osha, J. Et al., 2019. AIPPI Study Guidelines: Copyright in Artificially Generated Works. [online] Available at: https://aippi.org/wp-content/uploads/2019/08/Study-Guidelines_Copyright_Copyright-in-artificially-generated-works_23January2019.pdf>

Office of the register of copyrights, 2011. Legal Issues in Mass Digitization: A Preliminary Analysis and Discussion Document. United States Copyright Office. Available at: <https://www.copyright.gov/docs/massdigitization/USCOMassDigitization_October2011.pdf> [Accessed March 24, 2022].

Panel Report of 15 June 2000, United States-Article 110 (5) of the US Copyright Act, WT/DS160/R. Available at: <https://www.wto.org/english/tratop_e/dispu_e/cases_e/ds160_e.htm> [Accessed August 2, 2022].

Pearlman, R., 2018. Recognizing Artificial Intelligence (AI) as Authors and Inventors Under U.S. Intellectual Property Law, Richmond Journal of Law & Technology, Vol. 24, Issue 2.

Prosser, M., 2020. How AI Helped Predict the Coronavirus Outbreak Before it Happened. *Singularity Hub*. Available at: https://singularityhub.com/2020/02/05/how-ai-helped-predict-the-coronavirus-outbreak-before-it-happened/ [Accessed January 25, 2022].

Ralston, W., 2005. Copyright in Computer-Composed Music: HAL Meets Handel. *Journal of Copyright Society* U.S.A, 52.

Reed, C.S., 2019. Next Great Copyright Act: U.S. Copyright Policy for the 21st Century, Edward Elgar Publishing Ltd.

Reid A., Port K.L.; Tu S. 2021. Fundamentals of United States Intellectual Property Law: Copyright, Patent, and Trademark, Kluwer Law International.

Rosati, E., 2019. Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. *Asia Pacific Law Review*, 27(2).

Rosati E., 2019. Copyright and the Court of Justice of the European Union. Oxford University Press.

Rosati, E., 2021. Copyrigth in the Digital Single Market. Oxford University Press.

Ryan, M., 2014. *The Digital Mind: An Exploration of Artificial Intelligence*. Micheal Ryan.

Saloky,T., Šeminský, J., 2005. *Artificial Intelligence and Machine Learning*. [online] Available at:< http://uni-obuda.hu/conferences/SAMI2005/SALOKY.pdf> [Accessed 29 November 2019].

Samuelson, P., 1986. Allocating Ownership Rights in Computer-Generated Works. University of Pittsburgh Law Review. Vol. 47:1185, p. 1196 [online] Available at:<https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=2067&context=facpubs>> [Accessed 29 November 2019].

Schlackman, S., 2018. *Who Holds the Copyright in AI Created Art*. [online] Available at: https://alj.artrepreneur.com/the-next-rembrandt-who-holds-the-copyright-in-computer-generated-art/ [Accessed 3 December 2019].

Schönberger, D., 2018. Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML). [online]. Available at: <https://ssrn. com/abstract=3098315> [Accessed 4 February 2022].

Sejnowski, T.J., 2018. *The Deep Learning Revolution*. Massachusetts: The MIT Press.

Shalev-Shwartz, S. and Ben-David, S., 2014. *Understanding Machine Learning From Theory to Algorithms*. New York: Cambridge University Press.

Shaping Europe's digital future. 2022. Big data. [online] Available at: <https://digital-strategy.ec.europa.eu/en/policies/big-data> [Accessed 7 February 2022].

Sizing the Price What's the real value of AI for your business and how can you capitalise? 2017. Available at <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html> [Accessed January 25, 2022].

Škiljić, A., 2021. When Art Meets Technology or Vice Versa: Key Challenges at the Crossroads of AI-Generated Artworks and Copyright Law. *IIC - International Review of Intellectual Property and Competition Law*, 52(10).

Sobel, B., 2017. Artificial Intelligence's Fair Use Crisis. *Columbia Journal of Law & the Arts, Forthcoming*, p.32. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3032076> [Accessed January 25, 2022].

Springer partners with Copyright Clearance Center to power text and Data Mining Solution. Copyright Clearance Center. 2021. Available at: https://www.copyright.com/springer-partners-with-copyright-clearance-center-to-power-text-and-data-mining-solution/ [Accessed March 24, 2022].

Stokkmo, O. 2019. The Role of Collective Licensing, International Publishers Association.

Suluk, C., Orhan, A., 2005. Uygulamalı Fikri Mülkiyet Hukuku Cilt II: Genel Esaslar Fikir ve Sanat Eserleri, 1st Ed. İstanbul: Arıkan Yayınları.

Suluk C., Karasu C., Nal R., 2019. Fikri Mülkiyet Hukuku, Ankara, Seçkin Yayınevi.

Stanford University, Contribution and Impact. [online] Available at:< http://jmc.stanford.edu/contributions/index.html> [Accessed 27 November 2019].

Stone, P., 2016. Artificial Intelligence and Life in 2030, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel. Stanford University.

Tekinalp, Ü., 2012. Fikri Mülkiyet Hukuku. 5th Ed. İstanbul: Vedat Kitapçılık.

Thompson, J. W., *The Next Rembrandt.* [online] Available at: https://www.jwt.com/en/work/thenextrembrandt [Accessed 3 December 2019].

Triaille, J., d'Argenteuil, J. and de Francquen, A., 2014. Study on the legal framework of text and data mining (TDM). [online] Available at: <https://op.europa.eu/en/ publication-detail/-/publication/074ddf78-01e9-4a1d- 9895-65290705e2a5/language-en> [Accessed 4 February 2022].

Turing, A. M., Computing Machinery and Intelligence. 1950. Oxford University Press on behalf of the Mind Association. [online] Available at: <https://phil415.pbworks.com/f/TuringComputing.pdf> [Accessed 27 November 2019].

United States Copyright Office, 2015.Orphan Works and Mass Digitization. Available at: <https://copyright.gov/orphan/reports/orphan-works2015.pdf> [Accessed March 24, 2022].

Valente, M.-A., 2016. Transformativeness in the Age of Mass Digitization. *St. John's Law Review*, 90(233).

Wiley, J.F., 2016. *R Deep Learning Essentials*. Birmingham: Packt Publishing Ltd.

WIPO, 2016. Understanding Copyright and Related Rights. Available at https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.

WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), 2019. Available at: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_3_ge_20/wipo_ip_ai_3_ge_20_inf_5.pdf [Accessed March 24, 2022].

WIPO, 2019. *WIPO Technology Trends 2019: Artificial Intelligence*. [online] Geneva: World Intellectual Property Organization.

Wlodarczak, P., 2020. *Machine Learning and its Applications*. CRC Press.

Yanisky-Ravid, S., 2017. *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era: The Human-like Authors Are Already Here: A New Model*. Michigan State Law Review, Volume: 2, Issue 4.

Yilmaztekin, H.K., 2020. Yapay Zeka Eser Sahipligi, Galatasaray Üniversitesi Hukuk Fakültesi Dergisi.

Yilmaztekin, H.K., 2020. Türk Fikri Haklar Hukuku Yapay Zeka Tarafından Meydana Getirilen Eserleri Korumak İçin Hazır mı?, Galatasaray Üniversitesi Hukuk Fakültesi Dergisi, 2020/2.

## **DECISIONS**

Arica Inst., Inc. v. Palmer, 970 F.2d 1067, 1077 (2d Cir. 1992).

Authors Guild, Inc., et al., v. Google, Inc., 804 F.3d 202 (2d Cir.2015).

Authors Guild, Inc., et al., v. HathiTrust, et al., 755 F.3d 87 (2d Cir.2014).

Atari, Inc. v. N. Am. Philips Consumer Elecs. Corp., 672 F. 2d 607, 614 (7th Cir. 1982).

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (United States Supreme Court, 1994).

Castle Rock Entertainment, Inc. v. Carol Publishing Group., Inc., 150 F.3d 132, 143 n.9 (2d Cir. 1998).

Dellar v. Samuel Goldwyn, Inc., 104 F.2d 661, 662 (2d Cir. 1939).

Entertainment Research Group., Inc. v. Genesis Creative Grp., Inc., 122 F.3d 1211 (9th Cir. 1997).

Feist Publications Inc v Rural Telephone Service Co 499 U.S. 340 (United States Supreme Court, 1991).

Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (United States Supreme Court, 2021).

Norma Ribbon & Trimming, Inc. v. Little, 51 F.3d 45, 47 (5th Cir. 1995).

CJEU, C-360/13, Public Relations Consultants Association (5 June 2014), ECLI:EU:C:2014:1195, paras. 43 and 50 .

CJEU, Case C-393/009, Bezpečnostní softwarová asociace v Ministerstvo kultury, E.C.R. I-13971, (2010).

CJEU, Cases C-403/08 and C-429/08, Football Association Premier League Ltd and Karen Murphy, E.C.R. I-09083, FAPL (2011).

CJEU, Case C-45/10, Eva-Maria Painer v Standard Verlags GmbH and others, E.C.R. I-12533, (2011).

CJEU, Case C-604/10, Football Dataco v Yahoo! UK Ltd and others, 2 C.M.L.R. 24 (2012).

CJEU, Case C-476/17, Pelham GmbH, Moses Pelham, Martin Haas v Ralf Hütter, Florian Schneider-Esleben, (2019).

CJEU, C-419/13, Allposters International BV v. Stichting Pictoright (2015).

CJEU, C-5/08, Infopaq International A/S v. Danske Dagblades Forening (2009).

German Federal Supreme Court (Bundesgerichtshof), 29 April 2010, I ZR 69/08 (‘Vorschaubilder’).

Opinion of Advocate General Trestnjak delivered on 12 February 2009, Case C-5/08, Infopaq.

Turkish Court of Cassation, 11th Civil Chamber, Case No: 2006/934, Decision No: 2007/4555, 13.03.2007.

Turkish Court of Cassation, 11th Civil Chamber, Case No: 2016/3508, Decision No: 2017/5950.

Turkish Court of Cassation, 11th Civil Chamber, Case No: 2015/12923, Decision No: 2017/2724, 09.05.2017.

Turkish Court of Cassation, 19th Criminal Chamber, Case No: 2015/19627, Decision No: 2018/7669.

Turkish Court of Cassation, 19th Criminal Chamber, Case No: 2016/12740, Decision No: 2018/7674.

United Kingdom Supreme Court, Public Relations Consultants Association Limited v The Newspaper Licensing Agency Limited and Others, [2013] UKSC 18.

## CODES AND REGULATIONS

Agreement on Trade-Related Aspects of Intellectual Property. 15.04.1994. **https://www.wto.org/english/docs_e/legal_e/27-trips.pdf**

Berne Convention for the Protection of Literary and Artistic Works as amended on 28 September 1979, https://wipolex.wipo.int/en/text/283693.

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, 92–112.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167.

Law No. 5846 of 5.12.1951. (amended by: Law No. 2936 of 1.11.1983; Law No. 4110 of 7.6.1995; Law No. 4630 of 21.2.2001; Law No. 5101 of 3.3.2004; Law No. 5217 of 14.7.2004; Law No. 5571 of 28.12.2006; Law No. 5728 of 23.01.2008 Law No. 6279 of 29.02.2012 Law No. 6552 of 10.09.2014) Transleted by Nilsson G. O. and Şehirali F. H available at <https://en.unesco.org/sites/default/files/turkey_lawintellectualartisticwork_entof.pdf>.

World Intellectual Property Copyright Treaty. 1996. [online] Available at: < https://wipolex.wipo.int/en/text/295157> [Accessed 1 December 2019].

## WEBSITES

Anon, 2018. Is artificial intelligence set to become art's next medium?: Christie's. Available at: https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx [Accessed January 25, 2022].

BGH: Google's image search is no copyright infringement. [online] The IPKat. Available at: <https://ipkitten.blogspot.com/2010/04/bgh-googles-image-search-is-no.html> [Accessed 10 February 2022].

Comments of the Authors Guild, Inc. submitted by Mary Rasenberger In the Matter of Impact of Artificial Intelligence Technologies on Copyright. Available at: https://www.authorsguild.org/wp-content/uploads/2020/01/Authors-GuIld-Responses-to-USPTO-AI-NOI1.pdf [Accessed March 18, 2022].

Copyright Clearance Center, 2020. Written Comments of Copyright Clearance Center, Inc. (submitted 13 February 2019), in the content of the WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI) Second Session (Draft Issues Paper on Intellectual Property and Artificial Intelligence, Jan. 14,2020). Available at: https://www.copyright.gov/disted/comments/init026.pdf [Accessed March 24, 2022].

European Alliance for Research Excellence. 2018. Open letter: Maximising the benefits of artificial intelligence through future-proof rules on Text and Data Mining - European Alliance for Research Excellence. [online] Available at: <https://eare.eu/open-letter-maximising-the-benefits-of-artificial-intelligence-through-future-proof-rules-on-text-and-data-mining/> [Accessed 9 February 2022].

Gershgorn, D., 2018. A Microsoft chatbot composes poetry by looking at photographs. Quartz. Available at: https://qz.com/1354736/a-microsoft-chatbot-composes-poetry-by-looking-at-photographs/ [Accessed January 25, 2022].

Huawei Presents 'unfinished symphony'. Available at: https://consumer.huawei.com/au/campaign/unfinishedsymphony/ [Accessed January 25, 2022].

Jozuka, E., 2016. A Japanese AI almost won a literary prize. VICE. Available at: https://www.vice.com/en/article/wnxnjn/a-japanese-ai-almost-won-a-literary-prize [Accessed January 25, 2022].

Levy, S., 2012. Can an algorithm write a better news story than a human reporter? Wired. Available at: https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/ [Accessed March 23, 2022].

Nugent, C., 2018. Can artificial intelligence create art? buyers think so. Time. Available at: https://time.com/5357221/obvious-artificial-intelligence-art/ [Accessed January 25, 2022].

The Economist, 2017. The world's most valuable resource is no longer oil, but Data. Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [Accessed March 26, 2022].

The Next Rembrandt: Generating the Features. [online] Available at: https://www.nextrembrandt.com/ [Accessed 3 December 2019].

ABSTRACT

Through long-lasting efforts of researchers, Artificial Intelligence (AI) has gained its current position being intelligent and able to learn from its own experiences. A key component of AI is Machine Learning (ML), which gives the capacity to learn and develop from experience without being expressly programmed. However this “learning”, process is reliant on huge numbers of data. Sometimes the data employed to teach and train AI may consist of unauthorized copies of copyrighted works.

This study analyses the potential copyright infringements that might occur during the ML stages within the legal scope of European Union (EU), United States of America (US) and Türkiye. Possible copyright exceptions are reviewed within the same jurisdictions and a comparison are made between them. Lastly, it is discussed if the concept of extended collective licensing could be a possible solution to create balance between the interests of rightholders and AI developers.

***Key words:*** *Artificial Intelligence, Machine Learning, Copyright, Text and Data Mining, Fair Use, Extended Collective Licensing*

# ÖZET

Araştırmacıların uzun süredir devam eden çabaları sonucunda Yapay Zeka (AI), akıllı ve kendi deneyimlerinden öğrenebilen mevcut konumunu gelmiştir. Yapay zekanın önemli bir bileşeni, açıkça programlanmadan, kendisine deneyimlerden öğrenme ve geliştirme kapasitesi veren Makine Öğrenimidir (ML). Ancak bu "öğrenme" süreci çok sayıda veriye bağlıdır. Bazen yapay zekayı eğitmek için kullanılan veriler, telif hakkına tabi eserlerin yetkisiz kopyalarından oluşabilir

Bu çalışma, Avrupa Birliği (AB), Amerika Birleşik Devletleri ve Türkiye Düzenlemeleri kapsamında ML aşamalarında meydana gelebilecek olası telif hakkı ihlallerini incelemektedir. Telif hakkı istisna ve sınırlamaları da aynı ülkeler bakımından incelenmiş ve birbirleriyle kıyaslanmıştır. Son olarak genişletilmiş toplu lisans konseptinin hak sahipleri ve AI geliştiricilerinin menfaatleri arasında denge kuracak olası bir çözüm olup olmadığı tartışılmıştır.

***Anahtar Kelimeler:*** *Yapay Zeka, Makine Öğrenimi, Telif Hakları, Metin ve Veri Madenciliği, Adil Kullanım, Genişletilmiş Toplu Lisans*