



**Artificial intelligence
and copyright. From
Thaler's dilemma to
“the right to read is the
right to mine” doctrine**

José María Anguiano

November, 2023

**ARTIFICIAL INTELLIGENCE AND COPYRIGHT.
FROM THALER'S DILEMMA
TO "THE RIGHT TO READ IS THE RIGHT TO MINE" DOCTRINE**

José María Anguiano

NOVEMBER, 2023

**Artificial intelligence and copyright.
From Thaler’s dilemma to
“the right to read is the right to mine” doctrine**

© INSTITUTO DE DERECHO DE AUTOR, 2023

Author

José María Anguiano

Editorial direction and co-ordination

Marisa Castelo

Instituto de Derecho de Autor
Bárbara de Braganza, 7. 28004, Madrid
www.institutoautor.org

Cover design

Miguel Sueiro. Baliente

Layout

María García Asensio

Translation into English

Cathal Monaghan

TABLE OF CONTENTS

1. INTRODUCTION	1
2. WHAT DO WE MEAN BY THE TERM ARTIFICIAL INTELLIGENCE (AI)?	2
2.1. <i>DEEP LEARNING</i> AND NEURAL COMPUTATION	3
2.2. GENERAL AND SPECIFIC ARTIFICIAL INTELLIGENCE AND THE EVOLUTION TOWARDS GENERAL AI ..	4
2.3. THE FRANKENSTEIN COMPLEX	4
3. LEGAL QUESTIONS ON ARTIFICIAL INTELLIGENCE	6
4. ARTIFICIAL INTELLIGENCE AND COPYRIGHT	8
4.1. PROTECTION OF AI MODELS THEMSELVES	9
4.1.1. Protection of AI as computer software	9
4.1.2. <i>Sui generis</i> protection of AI	11
4.1.2.1. <i>Integrated circuits</i>	12
4.1.2.2. <i>Databases</i>	13
4.1.3. Delving further into the <i>sui generis</i> protection of AI	14
4.1.3.1. <i>What is being protected?</i>	15
4.1.3.2. <i>Originality of the implementation</i>	16
4.1.3.3. <i>The size of the investment made</i>	16
4.2. PROTECTING AI CREATIONS	19
4.2.1. Administrative and judicial decisions on the authorship of AI creations	20
4.2.1.1. <i>Decisions when human intervention is relevant</i>	21
4.2.1.1.1. <i>In China</i>	21
4.2.1.1.2. <i>In the European Union</i>	22
4.2.1.1.3. <i>In the U.S.</i>	23
4.2.1.2. <i>Decisions when human intervention is anecdotal The Artificial Inventor Project and the DABUS Case</i>	23
4.2.1.2.1. <i>What is DABUS?</i>	23
4.2.1.2.2. <i>What is the Artificial Inventor Project?</i>	25

4.2.1.2.3. <i>Decisions in the DABUS case</i>	26
4.2.1.2.3.1. In the United States	26
4.2.1.2.3.2. In the United Kingdom	28
4.2.1.2.3.3. In the European Union	29
4.2.1.2.3.4. In Germany	30
4.2.1.2.3.5. In Australia	31
4.2.1.2.3.6. In New Zealand	31
4.2.1.2.3.7. In other jurisdictions	32
4.2.1.2.3.8. The decision regarding <i>A Recent Entrance to Paradise</i> . .	32
4.2.1.2.4. <i>Conclusions regarding the DABUS case</i>	37
4.3. PROTECTION OF PROTECTED CREATIONS UTILISED TO “TRAIN” AI. COPYRIGHT AND TDM	40
4.3.1. What is text and data mining (TDM)?	41
4.3.2. What are the expected uses of TDM?	42
4.3.3. What is the controversy surrounding TDM?	43
4.3.4. Text and data mining in the USA. The fair use doctrine	44
4.3.4.1. <i>Analysis of court rulings to conclude on fair use, depending on the factors considered</i>	47
4.3.4.2. <i>Conclusions. “The right to read is the right to mine” doctrine</i>	48
4.3.5. The exception for TDM in the UK	48
4.3.6. TDM in the European Union From the fair use doctrine to <i>sui generis</i> law	49
5. CONCLUSIONS	54

INTRODUCTION

Many are reflecting on the consequences of AI's advance into the legal profession, whether human jurists can be replaced by AI¹ and whether in the not-too-distant future legal work will be taken over by these machines.

My approach is different. It is also important to reflect on the legal consequences of this technology's advance into our lives. Especially considering that existing legal systems are written by and for humans. The emergence of a new "participant" that is not covered by current regulations and whose actions have legal relevance in the environment seems likely to raise a number of legal challenges that will have to be addressed.

Copyright is a prime example of this. There is a global regulatory debate regarding the legal status of "artificial intelligence (AI) models".

There are three important issues at the heart of the matter. The first on how to protect AI itself; the second on the status of the creations generated by AI, whether they can be recognised as authors and rights holders of their creations; and the third on the protection of the rights of the owners of the works fed into AI: data, formulas, texts, images or sounds protected by copyright or patent.

¹ This term is utilised throughout this text to refer to artificial intelligence (AI). It is intended to be an umbrella term for the various forms in which this technology can be presented. In any case, I am thinking of software running on hardware whose variables are fed by any type of "data".

WHAT DO WE MEAN BY THE TERM ARTIFICIAL INTELLIGENCE (AI)?

To reflect upon this question with the necessary rigour, it is useful to start with the meaning we attribute to the term “artificial intelligence” (AI). I say this because there is no clear consensus. For the purposes of this reflection, we will understand that we are dealing with an AI system when it has “cognitive independence and independence of actions” —it learns and acts by itself— without the need for human intervention. To understand this, allow me to tell you the story of GO.

A number of years ago a division of the U.S. multinational Google —Google Deep Mind— announced that a machine it had created, called AlphaGo Zero, had beaten its predecessor (an earlier version of the machine —AlphaGo—) by a resounding 100 to 0. Both versions were AI designed to play a board game called “Go”.

It is a Chinese game, more complex than chess, played on a board containing 19 x 19 squares with black and white playing pieces called stones. One player uses the white stones and the other black. It is a game of strategy. The aim is to surround a larger area of the board with your stones than your opponent. A stone or group of stones of the same colour is captured and removed from the game if after one move it has no intersections; it is completely surrounded by stones of the colour played by the opponent. The player who finishes with the most territory wins the game.

Although the rules of Go are simple, the strategy is extremely complex and involves managing many variables, some of them contradictory. For example, placing stones close together helps to keep them connected, but placing them far apart gives you influence over a larger portion of the board, opening up the possibility of taking over a larger territory and thus winning the game. The strategic difficulty of the game lies in balancing the two alternatives. Player movements are both offensive and defensive and must combine short and long term strategies.

AlphaGo, the first machine designed to play the game, was trained based on human experience. It was fed with data from thousands of games and millions of moves made by human players. After training it was able to beat Lee Sedol, world Go champion, in March 2016.

The second version developed by DeepMind —AlphaGo Zero— executes new algorithms that allow it to learn to play without prior supervised training. It relies on a neural network based on “reinforcement learning”: the machine teaches itself by practising with itself until it reaches a capability far superior to previous versions. This is just the same as Dr B., a character in *The Chess Story*, a novella written in 1941 by Stefan Zweig. Nazis subject the protagonist to total isolation, and he is not even able to read. He only manages to escape the madness by stealing a book of past masters’ chess games. He memorises each game, analysing possible variations on each move and separates his psyche into two different players —the white player and the black player—. In this way, he can play time after time against himself and achieve a remarkable degree of proficiency in his performance. After his release, he easily defeats the world chess champion.

The foregoing means that AI have been created that are capable of learning from human experience (*machine learning*) and, above all, machines with cognitive independence, which acquire knowledge autonomously, without human involvement (*deep learning*). In short, it is the realisation that artificial intelligence is a reality. As a result, a number of countries have already begun to plan for and regulate the consequences of what has come to be known as the Fourth Industrial Revolution.

2.1. DEEP LEARNING AND NEURAL COMPUTATION

How does AlphaGo Zero manage to emulate Dr B and, without prior human supervision, acquire such skill in the game? It does so with mathematical algorithms and “neural computation models”.

Neural networks, also known as “connectionist systems”, are computational models based on a large number of artificial neurons —computers— connected together and forming a random number of layers. They resemble the functioning of neurons in the brain, and three distinct layers can be distinguished in both. While our neurons acquire information through the dendrites, there is a hidden processing layer (soma) and there is another layer for information output to other neurons (axon); in artificial neurons, the input layer, the hidden processing layer and the output layer for information output to other artificial neurons (computers) are clearly distinguishable. Generally, connections are made between neurons in different layers, but there may be intra-layer or lateral connections and also feedback connections in the opposite direction —towards the input—. Each of the artificial neurons has a “memory” capable of holding algorithms that process the information received. They also have a transfer function which, depending on the inputs and execution of the code in the memory, provides an output or an alteration of the code (algorithm) stored in the memory (code rewrite). As they themselves, depending on their own experience, can rewrite the code of their internal memory, their actions in the environment become unpredictable. No one knows in advance what rewrite the artificial

neurons will produce from the code stored in their memory, nor what weighting the algorithms will give to the incoming data and, therefore, no one knows what the actual performance of the machine will be.

It should also be foreseen that, while machines can evolve independently, and the first generations of AI will be linked to their human creators, this will not be the case for successive generations, where the possibility of disengagement increases the longer the lineage.

2.2. GENERAL AND SPECIFIC ARTIFICIAL INTELLIGENCE AND THE EVOLUTION TOWARDS GENERAL AI

The gradual development of artificial intelligence is for the time being directed towards the creation of special-purpose AI —devices capable of solving specific problems with higher-than-human capacity and speed— are being created. However, general artificial intelligence —capable of undertaking any task that the human mind can spontaneously perform— has not yet been achieved. Although scientists do not agree on “when”, there is consensus that there will come a time when AI surpasses human intellectual capacity. This uncertain date is known as “singularity day” —the moment when machines will surpass the intellectual capacity of humans—.

2.3. THE FRANKENSTEIN COMPLEX²

The dangers associated with robotics and artificial intelligence were already the subject of analysis when the phenomenon was only science fiction. In 1942 Isaac Asimov began to incorporate in his novels what he called “the Three Laws of Robotics”, which are a set of rules devised by the writer, which in his novels must be incorporated into the “code” of the robots and which serve to counteract the “Frankenstein complex” (that machines rise up against their creators). The Three Laws are:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Seventy-five years have passed and they remain relevant today. The world is beginning to question the need for laws that are up to date with the non-fictional reality we live in. In this context, on 31 May 2016, the European Parliament's Com-

² Expression coined by the Russian-born science writer and populariser Isaac Asimov. It describes man's fear of machines rebelling against their creators.

mittee on Legal Affairs published a report³ in which it made recommendations to the Commission to analyse the possibility of issuing civil law regulations to respond to the paradigms that robotics raised in this jurisdiction. The draft report lists the recitals that make the promulgation of this type of regulation advisable.

From the Committee's perspective, we were on the threshold of a new industrial revolution. Of an era of robots, bots, androids and other increasingly sophisticated forms of artificial intelligence that will affect all strata of society. On the one hand, they will bring efficiency and savings and prevent humans from being exposed to dangerous situations. However, it is also true that the development of robotics and artificial intelligence may lead to "AI" taking over much of the work now done by humans, raising questions about employability or the viability of tax and social security systems. There are also concerns about the physical safety of humans in the face of deliberate glitches in or attacks committed by these AIs or, as we have seen, the increasingly autonomous and therefore unpredictable nature of machines. There are also questions about the privacy of humans, not only about how their data may be utilised, but even about its ownership. The same is true for assistive robotics in relation to the effects that artificial intelligence might have on the dignity of the humans with whom it interacts. However, the European Parliament's main concern is the future; the moment when machines surpass human intellectual capacity (singularity day) which "could challenge mankind's ability to control its own creation and thus perhaps also the ability to be master of its own destiny and ensure the survival of the species".

For years now, many authoritative voices⁴, in the face of continuous advances in the discipline, have suggested stopping and reflecting on the future and artificial intelligence. They understand that it is inappropriate to embark on a frantic race towards ever more autonomous and sophisticated AI without first analysing the consequences and discussing an effective governance model.

Moreover, the European Parliament warned that other non-EU countries were considering or had initiated legislative processes and that it did not want Member States to lose regulatory control over these matters and thus be forced to adopt regulation from countries outside the EU.

³ Report A8-0005/2017 of the European Parliament, with recommendations to the Commission on robotics.

⁴ Stephen Hawking, Bill Gates, Elon Musk and many more.

LEGAL QUESTIONS ON ARTIFICIAL INTELLIGENCE

The approach to the concept of artificial intelligence leads to the conclusion that it has characteristics that are potentially significant from a legal perspective. Two issues are of particular interest in civil jurisdiction.

Firstly, AI is a “new source of knowledge”, not based on previous human experiences and biases, and it is predictable that it will not have the limitations of our “intracranial” intelligence⁵. This feature is of particular relevance to the analysis at hand. The absence of human intervention entails opportunities and obstacles. Opportunities in that opens the doors to a new source of knowledge, heretofore unknown and anticipating exponential advances in a multitude of fields. Obstacles as, in addition to the unknowns already mentioned when we referred to the “Frankenstein complex”, and from a legal perspective, AI clashes with current regulations which, without exception, places human beings at the centre of legal systems and also of the creative process, and grants them the exclusive rights that incentivise the creator. For this reason, AI cognitive independence and independence of action gives rise to legal controversies about the attribution of authorship and also about the ownership of their creations.

Secondly, that AI can be unpredictable for humans, which raises an interesting debate on the allocation of liability when the damage derives from AIs involvement. It seems that the classical schemes of attribution of responsibility could falter in the face of the continuous and exponential advance of artificial intelligence.

⁵ *Superintelligence*; Nick Bostrom. Darkside Books, 9/05/2018. One of the scientific reasons for the exponential advance of artificial intelligence. From a scientific perspective, the limited size of the brain, conditioned by the size of the skull, is a handicap for the growth of human intelligence. The reason is that the layering of neurons is limited by the size of the container of those layers (the brain). If intelligence is taken out of the skull, the limitation of the size of the skull disappears. It will always be possible to increase the number of layers and, consequently, the interrelation capacity of the neurons they are composed of.

Reflection on the legal implications of these two paradigms has something in common. Legal systems are created by and for human beings. Humans are the pivot on which all legal systems revolve; without exception. For this reason, the cognitive independence and independence of actions of AI generates obvious clashes in legal systems designed by and for humans. Jurists addressing these issues point to the possibility of solving this paradigm by granting legal personality to AI. They argue that we are already in an expansive process of conferring legal personality and that to grant AI legal personality would not be novel. They recall the assignment of legal personality to corporations, as well as the advances in animal rights promoted by animal rights organisations.

ARTIFICIAL INTELLIGENCE AND COPYRIGHT

Assuming the potential legal impact of the advance of artificial intelligence, it is appropriate to focus on the specific analysis of the legal questions raised by the advance of creative artificial intelligence.

Ever since human beings have been aware of the importance of science, talent, ingenuity and creativity in human progress, they have tried to promote them. To this end, a regulation was created that, on the one hand, encourages creative effort by granting the rights holder an exclusive exploitation right for a specific period of time and, on the other hand, guarantees free access to the creation with its incorporation into the “public domain” after the period of exclusive exploitation.

Since its inception, intellectual property regulations have been adapting to social changes. Technological changes have also taken place, including the digitisation of creations, their broadcasting, the possibility of downloading them onto memory devices, etc. This evolution has not been without teething difficulties, which have shown that technological changes affect the human rights holders of creations. Proof of this is the forced reconversion of the cultural sector to adapt their business models to the digitalisation of creations, for example, being forced in many cases to renounce the exploitation of their works through the sale of copies⁶.

As we set out above, three main copyright issues arise in the wake of the emergence of AI: (i) the protection of AI themselves, (ii) the authorship and ownership of AI creations, and (iii) the possible infringement of the rights of the owners of protected creations utilised to “train” AI.

⁶ Remember some years ago when “digital piracy” provoked a strong reaction from rights holders against P2P (*peer-to-peer*) downloading and access to protected works via “links”. The ultimate consequence of these practices has meant that exploitation models based on the sale of copies of many works have all but disappeared.

4.1. PROTECTION OF AI MODELS THEMSELVES

One of the questions that arises at the intersection between AI and copyright is about how AI itself is protected in their capacity, in principle, as a human creation. Indeed, the creation of artificial intelligence models requires human effort, talent and investment. If we want to encourage human initiative to create these AI, we will have to articulate the necessary protection mechanisms to stimulate the effort and investment in these types of creations, as well as to avoid unjust enrichment.

4.1.1. Protection of AI as computer software

One possibility is to turn to classical intellectual property rights and understand that an “AI model” is computer software. In fact, today, “artificial intelligence models” are protected as if they were one of these creations. We are not talking about something occurring in the registers in charge of granting exclusive rights. AI is, after all, —computer software— running on hardware —a computer— whose variables are fed with the appropriate data (always depending on the algorithm embedded in the code). In all AI regulations enacted to date, AI has, without exception, been considered to be a type of computer software. This is logical. It seems obvious that the AI “brain”, which drives them to achieve the result, is the installed computer software. The size of the machine on which the software runs, as well as the selection of the data to be fed into it, will be carried out according to the computational needs of the programmed code. Hardware and data therefore have an instrumental and programming-driven involvement. It therefore seems logical to understand that artificial intelligence, AI, is computer software. However, this does raise some questions.

The first is that copyright protects software as a literary work⁷. For a literary work to qualify for protection it must be original. To be original, the work must be the result of the “creative choices of its author”. This is where it is difficult to bring AI under the umbrella of copyright regulations. Firstly, because copyright does not protect concepts or ideas in general and therefore does not protect works whose design is intended to implement specific functionalities. It is therefore questionable whether copyright provides genuine protection for “AI models”. What they do is protect specific software that implements machine learning or deep learning in an original way (embodying the creative choices of its author). However, it is also possible to protect other computer software that develops the same idea in an equally original way. This is very well understood with “word processors”, which are also computer software. If the first registered word processor had closed off the possibility for others to develop the same idea —word processing— in other ways, the quantity and quality of supply would not be what we enjoy today.

⁷ A large part of the legal systems in force (USA, UK, Spain, among others) understand that computer software is a type of literary work, which is why up to now they have been considered as an author’s own work. An expression of human creativity.

The usual argument for the protection of AI as computer software is that existing regulations are flexible enough to deal with new challenges that may arise. In fact, it has been common for an old rule to be applied differently in response to novel circumstances that were not initially foreseen. One of the examples of the flexible application of the regulation is the protection of computer software.

The second question is authorship, which is closely related to originality. The original idea must be human.

To determine the originality of a trained AI model, we will at least have to be able to distinguish it from the traditional —entirely human-created— computer software. The need to distinguish them is because when it comes to traditional computer software, the human creates computer code for the purpose of obtaining a result previously anticipated or devised by that human and, consequently, the “conception” is human. However, in trained machine learning models, where the human author merely enables a computer system that “distills” the data provided to it and autonomously generates the model, the conception is not human⁸. While it is true that the human is the creator of the original code, it is also true that the code and the algorithm executed to obtain the “result” are rewritten autonomously. The autonomy of the machine is not only in obtaining the result, but also in the adjustments of the code and algorithm necessary to achieve it. The human not only loses control over the outcome, but also over the “algorithmic refinement” that achieves it. In short, what happens when an AI model is conceived that is capable of creating what the author of that AI model has not devised? I insist, this is an issue that affects both the code (the computer software executed) and the result obtained, which we will reflect on in depth later on.

This autonomy in the creation of the AI model is undoubtedly one of the most controversial issues. To explain this, it is useful to recall the ability of AI to rewrite its code and reformulate its algorithms. As this happens, there is an increasing human disengagement from the code, which could lead to a detachment from “conception”. It may be the case that the final result of the execution of the code was neither anticipated nor devised by a human. Recall the unpredictability of AI performance.

This brings us to another question raised by the protection of AI as computer software; about the dynamic condition of the protected creation (the software that governs the AI). This dynamism —the rewriting of code and algorithms— is also new in the field of protection. Think of a logo (trademark), a novel (copyright) or chemical formula (patent), which remain unchanged over time. This circumstance —fixed photograph— makes it possible to delimit more precisely what is being protected. When the scope of protection relates to dynamic —and evolving— software, the delimitation of the work to be protected becomes more complicated. One wonders from the outset whether protection is automatically extended to suc-

⁸ “Explainable Artificial Intelligence”, Wikipedia, https://en.wikipedia.org/wiki/Explainable_artificial_intelligence.

cessive versions. Do you therefore need to know the exact date of the infringement? Will the comparison that will determine plagiarism or not be of the initial version or of other later versions? Should the contours of “derivative work” be reconsidered?

4.1.2. *Sui generis* protection of AI

When the software that governs the AI model is not the result of the “creative choices of a human”, as the code is rewritten autonomously —without human involvement— there is a potential breach of the originality requirement. Ultimately, the human conception of software and therefore of AI is questioned. This could lead to a lack of protection for AI and, as a consequence, to a disincentive to creating it.

Before the digital revolution in which we are immersed, the USA considered the possibility of *sui generis* protection for these creations. In the end, however, U.S. Congress, in the Copyright Act of 1976, decided to consider computer software as literary works.

Those calling for a new form of protection for computer software warned of the economic impact of failing to consider them as copyrightable creations, as works that are costly to create but easy and cheap to copy. In short, they warned of the risk of unjust enrichment and, therefore, of the need to devise regulations to combat it.

From the promulgation of the Copyright Act until today, there have been increasing difficulties in applying the legal principle of copyright to classical works —novels, dramatic works, paintings, sound works, etc.— to cases in which there was no literal copy of the code of a computer software. However, with computer software it is common for the infringer to take advantage of the “core” of existing software, only modifying the user interface. In this way, the main lines of code are utilised to achieve the required functionality, but the new software is given a different look and feel. Whoever is executing the software, who only has access to the user interface, perceives important differences between the two pieces of software, yet the main lines of code to achieve the required functionality are the same. To address this issue, U.S. courts have created specific tests to conclude whether or not intellectual property rights have been infringed. Applying the fair use doctrine, they have granted protection to computer software rights holders against those who copied or rewrote portions of their software. They have understood that some of these codes were “derivative works” of pre-existing works (of the rights holders claiming protection) and that their undisturbed exploitation led to unjust enrichment. For many authors⁹, what the courts have essentially done is to create a “*sui generis* sub-category of law” to bring computer software under copyright regulations.

“*Sui generis*” is a Latin expression that translates to “of its own kind.” It refers to anything that is peculiar to itself. It is utilised to indicate something or someone unique or of its own kind or class. In copyright, the fundamental purpose of the *sui*

⁹ “*Sui generis* intellectual property protection for computer software”, by John C. Phillips, *Washington Law Review* (1992).

generis right is the protection of creations that, under the protection of intellectual property regulations, could be left unprotected. It differs from traditional classical rights in that the granting of this right protects the investment and effort required to bring about creation rather than protecting creativity or a particular form of expression. It grants the rights holder the prerogative to prohibit the extraction or reuse of protected works.

The *sui generis* right has frequently been utilised for the protection of integrated circuits, boat hull designs, fashion designs or collections of plant. In the European Union it has also been utilised for the protection of databases.

As our objective is the protection of "AI" (software, hardware and data), it seems appropriate to analyse their use for the protection of integrated circuits and databases.

4.1.2.1. *Integrated circuits*

The *sui generis* protection of integrated circuits covers images or designs representing a pattern (blueprint) for the construction of the different layers of a "semiconductor chip".

It originates in the USA. In accordance with the "Semiconductor Chip Protection Act" of 1984, the design of the circuit resulting from the joining of the different layers is eligible for protection if it has been applied for through the U.S. Copyright Office. If so, the applicant benefits for a period of ten years from the exclusive right to reproduce the integrated circuit, as well as to import and distribute the chips incorporated therein or to entrust third parties to implement the development necessary to achieve this.

Exclusivity prohibits third parties only from copying, but also from autonomous implementation. It also authorises the use of "reverse engineering"¹⁰ techniques for instructional purposes, as well as their subsequent use in new original works.

One of the characteristics of semiconductor protection is that the originality requirement is demanding. The integrated circuit design must not only be an "independent creation", but it must also not be a "common design" or "familiar" to the semiconductor industry. Design variations which incorporate functional modifications, but which, taken as a whole, are not original, are also not acceptable¹¹.

In any case, the practical implementation of protecting an AI as an "integrated circuit" is complicated because, while variations in an integrated circuit are easily

¹⁰ Reverse engineering is a computer activity that, starting from a compiled code, obtains the source code. It has often been utilised to maintain and modify executable code when source code was not available. The application of algorithms and lines of code to reverse the compilation process of computer code. The advantage of accessing the source code is to access the logic that governs the software and to be able to modify or copy it at will.

¹¹ 17 U. S. C. § 902(b)(2).

distinguishable, the same is not true for machine learning models that may not even be distinguishable to the human authors of these creations. The opacity in obtaining them makes it exceedingly difficult for the human seeking protection to prove originality.

4.1.2.2. *Databases*

In contrast to the legal protection of integrated circuits, that of databases does not take into consideration the requirement of originality, but rather the investment effort and the value of what has been created. To understand this, it is necessary to recall the reasons that prompted this atypical means of protecting databases. The regulation is set out in the European Union Directive on the legal protection of databases of 1996.¹² At the dawn of the “information society”, it was the case that the “makers” of the databases were unprotected against illegal appropriation of their content. Back then, databases were a type of collection. The EU directive set out: “Collections of other people’s works or other elements or data which, due to the selection or arrangement of their contents, constitute intellectual creations, are also intellectual creations pursuant to this law, without prejudice, where appropriate, to the copyright that the original authors may have over the incorporated works.”

The problem was that the makers of databases often did not meet the mandatory requirement of originality—the criteria for the selection or arrangement of materials were often not original—so that they could not cope with the appropriation (extraction) and subsequent incorporation into other databases (re-utilisation) of those that were also available on the market.

Consequently, the reasons for the atypical protection of databases coincide with the reasons for exempting AIs from such protection. In both cases, atypical forms of protection are sought in the face of prior non-compliance with the originality requirement. This is so in databases because the creation is not original. The selection or arrangement of the “elements” that make up the database is often not novel and, in the case of AI, because the requirement for originality is fulfilled by a non-human third party: an AI model. Since the resulting algorithms are the product of “distillation” of the data and rewriting of the algorithms by the machine itself, it can be difficult to understand that the resulting algorithms are of human creation.

The *sui generis* right of database owners covers systematically arranged collections of data that allow individualised electronic access to each of the elements (data). The maker or owner of the database must demonstrate a substantial investment in the procurement, verification or presentation of the collection items, assessing the substantiality of the investment from both a quantitative and qualitative point of view. The protection entitles the owner to prohibit and prevent the ex-

¹² Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996.

traction or re-utilisation of all or a substantial part of the contents incorporated in the database for a period of 15 years without the prior registration of the database required to obtain the protection¹³.

Critics of *sui generis* protection of databases agree with proponents of "open source", arguing that this type of protection unnecessarily limits access to and use of information and recall that the granting of *sui generis* rights has not proved useful for the protection of the database market in the European Union. Furthermore, they warn that the granting of the right tends to be perpetual when the database owner makes successive substantial investments in the database, which might otherwise be necessary for its normal exploitation.

4.1.3. Delving further into the *sui generis* protection of AI

From what has been said so far, it can be inferred that it is difficult to protect AI with the classic models of intellectual and industrial property. Patent protection of AI models has the disadvantage that it only protects novel, non-obvious models, and that this type of protection is not homogeneous in all jurisdictions, with the obvious disadvantages that this entails.

Intellectual property protection has, as we have said, two fundamental drawbacks. The first is that it does not seek to protect the underlying idea, but the original expression of that idea, which is a handicap for the protection of AI models, since the same idea can be implemented with different algorithms.

The second drawback is that the machine rewrites its code autonomously, which makes it difficult to attribute authorship of the code that is finally executed, to the human authors of the initial code.

On the other hand, intelligent computing models are essential to achieve "artificial intelligence". They are utilised constantly. Not only to obtain the desired end result; they are also in charge of "distilling" the data to strip it of unwanted biases.¹⁴ Moreover, these AI could achieve exponential advances in knowledge acquisition.

The usefulness of these computational models has been noted, however it is also obvious that the procedure is costly. Locating and accessing the data, conceptual modelling of the system, programming it and, after implementation, verifying and presenting the results are potentially time-consuming tasks. They are also onerous. Moreover, a high cost of production coupled with a low cost of copying is fertile ground for those who pursue unjust enrichment, and sufficient cause to devalue AI and also its creations.

¹³ Directive 96/9/EC, Article 7.4.41 Articles 10.1. and 7.1.

¹⁴ Karen Hao, "This is how AI bias really happens—and why it's so hard to fix", MIT Technology Review (4 Feb. 2019), available at [https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand\[1\]why-its-so-hard-to-fix/](https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand[1]why-its-so-hard-to-fix/).

Starting from the premise that the unlawful appropriation of an AI model is possible and, moreover, at a low cost¹⁵, it must be asked in any case whether protection by means of a *sui generis* right could in any way affect the effective implementation of the safeguards that current intellectual property regulations provide for them.

To conclude on the incentive or disincentive of the new right, multiple factors will have to be considered. Among them, undoubtedly, the existence of a significant disproportion between the investment made by the original creator of the model and that of the plagiariser. The fact is that the consideration of this factor is the germ of protection by means of a *sui generis* right, as there is a comparison factor that could be objective. For example, the plagiarist's inability to prove certain thresholds of investment in infrastructure or personnel with adequate technical expertise would justify the application of the *sui generis* right by establishing an objective benchmark that could be indispensable both for the shaping of new rights and for the enforcement of existing ones.

In order to determine whether the AI model rights holder is worthy of protection, we analyse three circumstances: (i) what is protected, (ii) the originality of the implementation and (iii) the substantial nature of the investment made.

4.1.3.1. *What is being protected?*

In principle, as mentioned above, protection would apply to any computer code or mathematical expression (algorithm) that may be utilised. Consequently, this would be similar to how computer software is treated. Although by applying a *sui generis* protection we can circumvent the drawback of the absence of human creativity, we will in any case have to take into consideration the drawbacks already noted about considering AI as computer software.

Firstly, that only the specific expression that develops the underlying idea is protected, but that the underlying idea remains unprotected and can be freely utilised¹⁶. This inescapable premise forces us to question whether the protection of a specific type of mathematical expression is the appropriate means of protection for machine learning systems and ultimately for artificial intelligence or whether, on the contrary, it would be necessary to protect the idea and therefore the underlying capacity of the AI model.

Secondly, remember that it is also unclear what algorithm or computer code is being protected. Since the AI model itself, depending on the data it is fed with and the results it obtains after the execution of the code and algorithms, is able to rewrite its code or change its algorithm, a crucial question arises as to whether the protection is on the initial code and algorithm or also on the result after the rewriting or reformulation.

¹⁵ Not that low. A model with a similar purpose also requires an investment of time, labour and the infrastructure to support the computational and data storage effort required to implement the model.

¹⁶ John C. Phillips, "Sui Generis Intellectual Property Rights for Computer Software", *George Washington Law Review* (1992), available at https://cyber.harvard.edu/property/protection/resources/phillips_unedited.html.

4.1.3.2. *Originality of the implementation*

The problems with considering AI as software bring us to one of the most controversial issues: the lack of originality. To solve this problem, we can opt for a two-pronged approach: (i) we can apply the European model of database protection, which dispenses with originality as a requirement for protection; (ii) we can choose to require originality similar to that of integrated circuit designs, where, remember, to meet the requirement of originality it is not enough for the circuit design to be an "independent creation". It must also be a design not common or familiar to the semiconductor industry. The problem, as mentioned above, is that it would be difficult to meet the requirement of originality in an AI model because not even the authors of the code would be able to distinguish the modifications in the code.

Consequently, the only plausible approach is to grant protection to AI without valuing the creative input of humans, but rather the investment made to achieve the result. In short, it would be protected as a "business work"; as if it were computer software owned by a business.

4.1.3.3. *The size of the investment made*

The last factor to analyse for the protection of AI is the accreditation of having made a substantial investment in the training or implementation of a trained AI model. For example, protection would be granted to developers of trained AI models who were able to demonstrate independence in the creation of the model by providing documentation of the investment or work required for its training and implementation (for example, the provision of data or the hardware required to process it). As European Community regulations allow for an assessment of the materiality of the investment, both from a qualitative and quantitative point of view, it could, for example, be a qualitatively substantial investment to incorporate new algorithms that optimise learning into the model.

However, if we focus on the protection of AI by *sui generis* right, the right granted to database owners (the right to prohibit and prevent the extraction or re-utilisation of all or a substantial part of the protected work) applies to the three levels of protection of AI models (input, code and algorithms and output) and will necessarily be accompanied by other rights inherent to the one initially granted. For example, the prohibition of extraction should be accompanied by the right to combat the distribution of any media containing what has been unlawfully extracted.

Questions also arise about the documentation to be provided in the application to register AI. For two main reasons: (i) to enable competitors of the applicant to determine what is subject to exclusive exploitation and thus to determine what cannot be "extracted or re-utilised" and (ii) to determine what information is to be made available to the general public so that, after the end of the period of exclusive exploitation, they can have access to it.

The minimum information that should be provided in the application would be the model itself, including the code and algorithms that are part of the model. It is understood that the provision of this documentation would be sufficient to fulfil the dual objective of allowing replication of the model when the exclusive exploitation expires and informing potential competitors of what cannot be copied (extracted or re-utilised).

Consideration should also be given to whether the application for registration should be accompanied by some of the data utilised to “feed/train” the model. Forcing the applicant to hand over all the data utilised may seem excessive, unrealistic and potentially dangerous for the privacy of the persons to whom the data relate. On the other hand, the omission of the provision of information as a requirement for registration could mean that the provision of the model (code) and the algorithms incorporated in it would not be sufficient for the model to be replicated after the expiry of the exclusivity period. The solution may be to oblige the applicant for protection to describe the categories of data utilised or likely to be utilised in the training or execution of the model.

Another important issue to take into consideration in the creation of a protection model for AI based on the application of a *sui generis* right has to do with the liability of the infringer of the exclusive rights of the rights holder. If the model is built on the premise of the existence of strict liability, the rights holder is granted a right of an enhanced nature. Mere copying is sufficient for a finding of infringement even where the infringer has made an “independent development”. If the model is based on the intentionality of the infringer, it is possible to exclude “innocent re-utilisation” (where the infringer has obtained the model independently when they were also unaware of the protection of the pre-existing model). The logical way to resolve this issue could again be in the information provided by the rights holder at the time of registration. To the extent that such information is sufficient for the potential infringer to be deemed to have been informed of the existence of an exclusive right, the exploitation of the replicated model will have been preceded by negligence on the part of the infringer in not having consulted the register to learn of the existence of the exclusive right. It is also decisive that the rights holder can prove a personalised communication to the infringer informing them of the exclusive right and urging them to discontinue the unlawful exploitation.

Another potentially controversial issue in the application of *sui generis* law relates to the amount of material extracted or re-utilised. If it is necessary to conclude an infringement of the rights holder’s *sui generis* right. In order to solve this question, it may again be useful to refer to the European Community regulation on the legal protection of databases, which provides that there is an infringement when all or a “substantial” part of the database is extracted or re-utilised and, above all, when it provides that substantiality can be determined both qualitatively and quantitatively¹⁷.

¹⁷ Samuel E. Trosow, “Sui Generis Database Legislation: A Critical Analysis”, 7 Yale J. L. & Tech. 534 562 (2004).

In any case, it will be up to legislators and courts to adapt the *sui generis* right to the concrete situations that arise when disputes over the infringement of the right arise.

Another possible advantage for the *sui generis* protection of AI models is that, as has been shown, American courts have solved the vicissitudes in the protection of software by developing what is referred to in case law as a "*sui generis* soft law". Homogeneity in protection models is vital for their effectiveness in the global world we live in. Remember that U.S. courts have solved the "expensive to implement, cheap to copy" dilemma by enforcing proof of substantial investment by the rights holder or the absence of such proof by the potential infringer. And all this, remember, in application of the "fair use" doctrine.

The fair use doctrine may have relevance in different contexts of artificial intelligence. On the one hand, it is undeniably applicable in determining when the rights holders of an AI can use copyrighted data or works such as any literary, musical, etc., for the "training" of the machine or for any other purpose. I invite further reflection on this issue in the last section.

It can also be useful for the resolution of various disputes over the use and re-utilisation of computer code or algorithms. As mentioned above, U.S. court rulings have invoked the fair use doctrine to declare the infringement of the rights of the owner in cases in which, without there being a literal copy of the software code, a "derivative software" had been created based on a pre-existing one, on the understanding that there may be an unlawful action, as well as unjust enrichment.

In relation to output—the result—the fair use doctrine could have special relevance in view of the impossibility of covering the mandatory requirement of originality of traditional protections. In the absence of originality due to the lack of human authorship, one may choose to exempt its creators from protection, justifying it on the grounds of the substantial investments made by the "developer of the model" and the unjust enrichment of those who take unjustified advantage of them. Of particular relevance could be the "fair use" doctrine in determining the substantiality of the investment, as well as the qualitative or quantitative condition of the extraction or reuse.

Another important issue to consider is the duration of the registration. This is not a trivial issue. It should be borne in mind that a long registration is a disincentive for those who could innovate on the basis of the protected creation. If the registration is short, the return on investment to the developer of the model may be compromised. One possible way to solve this dilemma would be to grant an extension of the registration in cases where continued use of the protected creation can be proven. This would be a transposition of U.S. trademark regulations, which requires a declaration of use or excusable non-use every ten years from the registration date¹⁸.

¹⁸ "Definitions for Maintaining a Trademark Registration, USPTO: Trademarks" (15 Feb. 2020 12:01AM), <https://www.uspto.gov/trademarks-maintaining-trademark-registration/forms-file/definitions-maintaining-trademark>.

Another possibility is to revert to European Community regulations on the legal protection of databases, which grant rights holders successive extensions of protection provided they can prove substantial new investments in updating the database. However, as mentioned above, critics of *sui generis* database protection argue that this favours the granting of “quasi-perpetual” exclusive rights. In any case, an initial duration of 5 to 10 years is envisaged, with possible extensions depending on accreditation of use (USA) or substantial new investments (European Union).

Irrespective of the doubts already raised, it will also be necessary to consider how protection will work in the legal domain. How the assertions supporting the litigants’ claims will be substantiated and also whether the infringer can be compelled to cease or compensate the legitimate rights holder. The problem lies in the difficulty of proving infringement of exclusive rights in the field of AI, due to the opaque nature of this technology. Indeed, the characteristics of this technology make it a “black box”¹⁹ in terms of being understood by humans. The humans who design AI do not know the logic behind the output, nor are they able to reconstruct or explain the exact functioning of the “model”. This blurs the contours of what exactly is protected and therefore what cannot be extracted or re-utilised. The only plausible check is to compare the result of the claimant’s “black box” with that of the defendant, which does not seem to be a panacea either, since “black boxes” are also dynamic—they are rewritten according to the input and output data—.

4.2. PROTECTING AI CREATIONS. CAN AI BE RECOGNISED AS AUTHORS AND HOLDERS OF INTELLECTUAL AND INDUSTRIAL PROPERTY RIGHTS FOR THEIR CREATIONS?

Having analysed the protection of AI, it is worth reflecting on the protection that can be afforded to the creations generated by it. In other words, how does one protect the result of the execution of the computer software (AI), when the originality of the creation does not belong to the party seeking protection, but to an “AI model”? Imagine, for example, an AI that autonomously writes literary works. Controversy arises from the outset over the attribution of authorship. Does the authorship of this literary work belong to the human who programmed the software or to the machine?

While many believe that the humanist view of authorship *de facto* excludes the possibility of AI acquiring such rights, others believe that this view is outdated in our times, and call for a dynamic and expansive interpretation of the concept.

In 2011, an interesting debate arose in the legal domain. David Slater, a British photographer, intentionally left a camera on a tripod in an Indonesian jungle where a community of macaques lived. When he retrieved the camera and developed the film, he found an abundance of selfies of one of them, whom he nicknamed Naruto, news of which quickly spread.

¹⁹ The term “black box” is utilised in computer jargon to refer to a computer process with opaque or “obscure” logic. It may be that computational randomness prevents us from knowing the logic behind it or that the programmer wishes to prevent access to it.

The legal controversy arose in 2015 when People for the Ethical Treatment of Animals (PETA) claimed that Naruto was entitled to benefit from the income derived from the images and sued Slater, his company Wildlife Personalities, and the publisher Blurb, which had published a book of the images.

A court in California (USA) ruled in April 2018 that the macaque Naruto could not benefit from the income generated by these images, as animals, lacking legal personality, cannot benefit from the protection afforded by copyright regulations²⁰.

The increasing use of artificial intelligence systems for creative purposes hits the above dilemma even harder. The emergence of cognitively autonomous AI, capable of acting autonomously—without human intervention—raises a multitude of questions that must be answered if we are to continue to encourage creative endeavour. The legal community is engaged in a gripping debate.

One issue has remained unchanged over time, regarding the granting of rights to authors of creations derived from ingenuity or creativity: the assignment of these rights to humans. No one, until now, had considered that creations and inventions could have an authorship other than human. Cases like Naruto's were easy to elucidate. A human conceived the photographic work. It was the photographer David Slater who had the idea and took the risk of leaving his camera at the disposal of the macaques. The whole process had been conceived by a human who had provided the necessary means to execute it and who had foreseen the result: that the macaques would play with the abandoned camera and the selfies would be obtained.

With that in mind, in the first steps towards autonomy, it was logical to understand that machines were nothing more than instruments at the service of humans that merely facilitated the expression of our creativity. However, the frenetic technological development invites us to rethink once again the intellectual and industrial property dilemmas that this AI raises. We are no longer talking about simple machines that are limited to the execution of functionalities devised, designed and mechanically or digitally executed by human initiative, but AI that have cognitive autonomy (they learn by themselves) and as a consequence have the capacity to generate improved lineage. Moreover, as has been said, the result of its execution is unpredictable for human kind, ultimately representing a new "source of creativity".

4.2.1. **Administrative and judicial decisions on the authorship of AI creations**

The emergence of these autonomous and creative AI changes the rules of the game. I say change and will not change because this issue is already controversial, the immediate consequence of which is that, in some jurisdictions, creations derived from the execution of an autonomous and creative system end up in the public domain.

²⁰ Judgment of the Court of Appeals for the 9th Circuit on case no. 16-15469 dated 13 April 2018.

Indeed, the latest report published by the U.S. Copyright Office states that “the office will not register works devised by a machine or as a result of a mere mechanical process that operates randomly or automatically without input or intervention by a human author”.

Moreover, given the exponential growth of AI, the more sophisticated machines becomes, the less human intervention it will require and, consequently, the greater the number of creations that end up in the public domain, causing a logical disincentive to creativity and, at the same time, depriving teachers and researchers of access to the information needed to replicate protected creations.

The logic that prevents protection in favour of AI is that it is impossible to attribute any intellectual or industrial property rights because such attribution is exclusively reserved for humans and because, in addition, machines would not have the legal personality to take legal action of any kind against potential infringers by invoking the regulations that grant them authorship. These reflections are not only theoretical; they are already reflected in the courts. It is useful to start by analysing the decisions made to date by administrative and judicial bodies in different jurisdictions when controversy has arisen on this issue.

In order to analyse them, it is useful to distinguish according to the degree of autonomy of AI. There are decisions that refer to cases in which their intervention can be considered instrumental and others on creations in which human intervention is anecdotal.

We will also distinguish between those decisions that refer to authorship in the creation of works and those that refer to the AI as inventor.

4.2.1.1. *Decisions when human intervention is relevant*

With regard to decisions that refer to applications for registration of creations where the human contribution is substantial, I examine the judicial or administrative decisions that have opened the way to outlining a solution to the problem raised in China, the USA and the EU, as I understand that these are the countries where the AI race is most accelerated and, furthermore, in the case of the European Union, because they deal with regulations and judicial or administrative decisions that directly affect us.

4.2.1.1.1. *In China*

In December 2019, the People’s Court of Shenzhen Nanshan District ruled on a news article written by AI called “Dreamwriter” owned by the Chinese company Tencent.

Dreamwriter is a —human— assisted artificial intelligence system developed by Tencent. It was created to write texts (literary creations) autonomously.

On 20 August 2018, the company’s website published a financial article written by Dreamwriter. A Shanghai tech company quickly reproduced it on its website without Tencent’s prior consent, prompting Tencent to sue the Shanghai company for infringement of its intellectual property rights.

Since Article 2 of China's copyright regulations state that literary, artistic and scientific creations shall be protected provided that they are original and can be reproduced in a tangible form, the court first concludes that the literary work whose rights are in dispute is original and can be reproduced in a tangible form, and therefore finds that the necessary requirements are met for Tencent's claim to be admissible.

To conclude on the originality of the controversial article, the court applies a two-step originality rule:

The first asks whether literary creation (article) has a sufficient degree of creativity. In support of its affirmative answer, the court argues that the content of the article reflects the reflection, analysis and conclusions based on the relevant published information on the securities market and the data available at the article publication date. The court also found that the article had a proper structure and its reasoning was based on logic, and that it was therefore considered to be a copyrighted work.

At the second step of the test, the Court concluded that it was an article that reflected the knowledge and personalised discernment of the Tencent team members who had overseen the construction of Dreamwriter. The Court argues that the programming team had succeeded in getting the machine to generate the article in four stages: (i) inputting data, (ii) tagging the data and writing the code that executes the algorithm, (iii) verifying the data obtained after the execution of the code, and (iv) determining the placement of each of these data in a logical and properly formatted argument structure. They understand that these are all intellectual activities of the programming team that are unquestionably reflected in the article.

The court contrasts all these activities with the execution of the computer code that spends two minutes generating the article and concludes that it would be unfair to regard the existence of the article as solely the product of the two-minute execution of that computer process. Accordingly, it considers that Tencent has standing to challenge the non-consensual reproduction and upholds its claims.

4.2.1.1.2. *In the European Union*

To analyse the situation in the EU, it is worth looking at a decision taken by the European Patent Office (EPO) in 2020. A patent application in favour of the company Äquivalenter Aortendruck for the use of a neural network to determine with high precision information about the heart rhythm of patients is rejected.

It is worth recalling the explanations we gave about neural networks when we began this reflection. Thousands of artificial neurons (algorithms executed on computers) exchanging information and changing the "importance" that the data has in the final equation; with unpredictable results for the human intellect and without the resulting algorithm and thus the combination of data being clearly traced back to conclude in a logical process for the human mind. We already referred to this drawback when we reflected on the problems of protecting AI as computer software.

In view of this issue, the EPO, in application of Article 83 of Directive 219/790/EC, considers that the patent application does not comply with the established requirements, because the applicant does not attach information to the application that is sufficient, clear and complete for someone with sufficient knowledge to be able to put it into practice. Ultimately, given the “black box” nature of neural computation, it is impossible for the patent applicant to comply with the requirement to clearly illustrate how the result has been obtained and, consequently, inventions resulting from the execution of this AI cannot be patented for non-compliance with this bedrock requirement. Ultimately, the clause that underpins the granting of the right to exclusive exploitation for long periods of time has considerations; not only that the creations or inventions pass into the public domain, but also that humanity can benefit from the creation by giving researchers and teachers access to the logic behind it.

4.2.1.1.3. *In the U.S.*

The aforementioned Naruto case perfectly illustrates the position of the U.S. courts. They understand that a human has “conceived” the work. The Naruto selfies would not exist without the ingenuity of the photographer David Slater, who predicted the famed macaque’s performance.

4.2.1.2. ***Decisions when human intervention is anecdotal The Artificial Inventor Project and the DABUS Case***

In order to analyse the decisions of administrative and judicial bodies when the application for protection refers to creations in which there is no significant human participation, it is worth looking at the Artificial Inventor Project and the DABUS case.

4.2.1.2.1. *What is DABUS?*

DABUS stands for Device for the Autonomous Bootstrapping of Unified Sentience. It is an artificial intelligence system created by Dr Stephen Thaler²¹, whose main mission is to emulate certain brain functions for creative purposes.

To understand this, one has to go back to Dr B, the character in Stefan Zweig’s novel who had the ability to play chess against himself by splitting his brain into the white player and the black player. In the case of DABUS there are two mental processes involved. You are observing something and suddenly your mind generates a progression of related thoughts that describe a new and useful application of it. It would be a relational or logical finding, typical of human rumination. Or let us imagine that your mind makes a representation that is morphologically or otherwise novel—the notion—and that the progression of related thoughts suggests to you a

²¹ Dr Stephen Thaler’s biography is available on his website <https://imagination-engines.com/founder.html>.

utility or potential value for that "notion". Let us also imagine that the machine is fed, among other things, with data obtained from the distillation by neural networks of the policy of patent offices and case law of courts on what is considered original or novel.

Let us stop imagining. These are some of the things that the machine devised by Thaler achieves by replacing biological neurons with artificial ones. This is a new AI paradigm that is utilised to autonomously combine simple concepts by transforming them into more complex ones and having the ability to anticipate the consequences —uses— of these complex concepts (ideas). In short, Mr Thaler has created what perfectly describes the name of the company of which he is CEO: a creative machine.

Creative machines are not new. For decades now, neural networks have been playing off one another, in cooperative or adversarial fashion, to create new ideas or action plans. These "Creativity Machines®" (Thaler, 1997 2008 2013), required at least two neural networks: an idea generator and a critic. The two permanently connected into a neural architecture capable of adjusting any parameter (for example, learning rate) through feedback connections to such a generative network to "steer" its artificial ideation in the desired direction.

However, DABUS (Thaler, 2019) is a different kind of creative machine. Part of a swarm of disconnected neural networks. Each has interrelated memories —for example, of a linguistic, visual or auditory nature—. These independent networks are constantly connecting and disconnecting due to the carefully controlled chaos that is introduced "within" and "between" them. Then, through cumulative cycles of learning and unlearning, parts of these initially independent networks are interconnected into structures representing complex concepts. These chains of complex concepts are connected to other chains capable of representing the anticipated consequences of any given concept. Thereafter, such ephemeral structures fade, as others take their place, in a manner reminiscent of what we humans consider "stream of consciousness"²².

Thus, the big difference between the initial creative machines and DABUS is that in the latter, ideas are not represented by on-off patterns, but by ephemeral structures formed by chains of networks that quickly materialise and dematerialise. If by chance one of these geometrically represented ideas incorporates one or more desirable outcomes, these forms are selectively reinforced, while geometries representing undesirable notions are weakened. Finally, the desired complex ideas are incorporated into a "memory repository", allowing DABUS to be queried about them.

²² "Vast Topological Learning and Sentient AGI", *Journal of Artificial Intelligence and Consciousness*, Vol. 8, No. 1 (2021) 1-30.

4.2.1.2.2. *What is the Artificial Inventor Project?*

It is a project of an eminently legal nature. In August 2019, a group of jurists from various countries, led by Professor Ryan Abbott²³, convinced of the dysfunctions of the current regulations to meet the challenges of protecting the creations that lie ahead—mainly in relation to smart machine inventions—decided to file applications for two specific DABUS inventions in the patent offices of up to 150 countries. The outcome of their work in October 2022 indicates that they were right. The application of the regulations in force could end up producing effects contrary to those usually pursued by these laws: to encourage research, effort, work, talent, ingenuity, creative capacity... I deeply appreciate for what is probably one of the most successful legal research projects in history.

And what is a group of jurists from various countries doing filing patent applications on inventions of an artificial intelligence system, naming the system itself as the inventor? Because they are convinced that doing so is sufficiently disruptive to consider regulatory changes.

They assume that in all intellectual and industrial property regulations authorship is reserved for human beings (natural persons). But they also note that all these regulations provide for cases in which the rights to exploit works or inventions are vested in third parties other than the creators or inventors. The most common case is the automatic assignment, unless otherwise agreed, of exploitation rights in favour of the employer when the creation is the result of an employment relationship. But it is not the only one; there is also an assignment of rights when the creation is “work made for hire”. In these cases and unless otherwise agreed, the contractual transfer of the exploitation rights is presumed. In this way, the price compensates the author or inventor for his effort: they renounce the economic rights derived from the exploitation, but retain the moral rights that guarantee them recognition as the author of the work or invention.

This group of jurists believes that the protection of AI-generated works will encourage innovation. While it will not directly motivate an AI—which neither feels nor suffers and therefore can hardly perceive the economic or reputational incentive—it will do so with the humans who develop, own and use AI. They claim that protecting AI-generated works will promote the development of inventive or creative AI, thus achieving greater societal development.

Allowing a person to be listed as the author or inventor of an AI-generated creation would not be unfair to the AI, which as stated above has no economic or reputational interest, but allowing people to take credit for work they have not done would devalue the human creation. It would put the work of someone who simply asks an AI to solve a problem on an equal footing with the work of those who are legitimately creating something new. In short, would you opt for the candidate whose CV acknowledges that their final degree project was done by Chat GPT?

²³ Ryan Abbott’s biography available on Wikipedia.

They do not defend the holding of rights or patents by AI. They are aware that they have no economic or moral rights, so their proposals do not involve granting rights to machines. They understand, however, that attributing authorship to the machine would protect the moral rights of traditional human inventors and the integrity of the protection system. As noted above, it is often the case that the author of a creation is not its rights holder. They believe that the rights holder should be the owner of the AI.

Moreover, they anticipate that creative AI may soon become an important part of inventive and artistic creation and, ultimately, development. When this happens, it will be important to have clear rules on the protection of these creations.

4.2.1.2.3. *Decisions in the DABUS case*

The existence of an intelligent and creative system (DABUS), together with the legal uncertainty about the reaction of the offices in charge of examining applications for protection, is the breeding ground for launching this experiment: applying in different jurisdictions for protection for the creations of an AI model, naming the model itself as the author.

These applications concerned both inventions (patents) and works (images). Ultimately, the legal debate that arises concerns both intellectual and industrial creations.

Three creations were selected: two patent applications (for a plastic food container based on fractal geometry, and a flashing light (or "neural flame") for attracting enhanced attention), as well as the protection of a computer-generated image, "A recent entrance to paradise". We first analyse the DABUS patent applications below, naming the machine itself the application in various jurisdictions. Below, we analyse the application for protection of the image entitled "A recent entrance to paradise".

4.2.1.2.3.1. In the United States

On 22 April 2020, the United States Patent and Trademark Office (USPTO) decided that, in accordance with U.S. patent regulations, only natural persons can apply for protection and rejected an invention application filed by Dr Thaler, naming DABUS as inventor.

The application stated that the first name of the sole inventor was "DABUS" and the surname "Invention generated by artificial intelligence". Stephen Thaler —creator and owner of the DABUS creative machine— was also named as the applicant. In the patent application, Dr Thaler told the USPTO that DABUS conceived the invention using trained neural networks without human intervention.

The USPTO sent the applicant a remedy notice indicating that the application did not identify the inventor by his legal name. Thaler reacted by claiming that DABUS created the claimed invention independently and autonomously and that it had not

received specific scientific training on the subject matter of the invention. Therefore, Thaler's view was that DABUS should be recognised in the patent application as the inventor.

In its reply, the USPTO noted that Title 35 of the U.S. Code repeatedly refers to inventors as natural persons. For example, 35 USC § 101 provides that "Whoever invents or discovers...", where the term "whoever" suggests a natural person. Also 35 USC § 115 which uses terms such as "himself", "herself", "individual" and "person". Ultimately, the USPTO understands that reinterpreting the term "inventor" to include machines "would contravene the plain reading of the patent regulations, which always refer to persons and individuals".

UPSCO also supports the conclusion that, in accordance with case law, an inventor must be a natural person. Thus, in *Max-Planck-Gesellschaft zur Förderung der Wissenschaften* it is concluded that a state cannot be an inventor because inventors must be individuals who conceive an invention and conception is a "formation in the mind of the inventor" and "a mental act". The USPTO understands that "conception", the touchstone of invention, must be by a "natural person". Furthermore, the office recalls that the *Manual of Patent Examining Procedure* defines "conception" as "the mental part of the inventive act" and "the formation in the mind of the inventor", thus reinforcing the view that conception must be carried out by a natural person.

The USPTO also addresses other arguments of the petitioner. For example, Mr Thaler argued that the U.S. patent office had previously granted other patents to DABUS and, in doing so, "implicitly legalised the process by which DABUS arrives at an invention". The USPTO rejected the argument, stating that the "grant of a patent by virtue of 35 USC § 151 for an invention involving a machine does not mean that patent regulations provides for that machine to be listed as an inventor in another patent application, any more than a patent for a camera makes the camera copyrightable".

Thaler appealed the decision taken by UPSCO [*sic*: USPTO] and, on 2 September 2021, the Court for the Eastern District of Virginia upheld the decision of the United States Patent and Trademark Office (USPTO) that an inventor, defined in 35 USC § 100 (f) as "the individual or, if it is a joint invention, the individuals who invented or discovered the subject matter of the invention must be a natural person". The court also relied on the Dictionary Act, as well as the use of personal pronouns before the term "individual" in the U.S. Civil Code, to conclude that Congress intended that term to have a "typical meaning".

Although the court acknowledged that "there may come a time when artificial intelligence reaches such a level of sophistication that it can satisfy the accepted meanings of the invention", it does not believe that the time has yet come; which is a contradiction in terms, since the USPTO never disputed a fundamental factual issue for this purpose; that DABUS, not a natural person, was the inventor of the subject matter of the patent claimed.

In any case, the court considered that it was up to Congress to decide what legal treatment to give to inventive AI.

This decision has been appealed to the Court of Appeals for the Federal Circuit; the oral hearing took place on 6 June 2022 and in April 2023 it dismissed Dr Thaler's appeal.

4.2.1.2.3.2. In the United Kingdom

Dr Thaler applied to the UK patent office for the grant of two patents, No. GB1816909.4 and No. GB1818161.0. In a letter dated 4 December 2019 (decision no. O/741/19), H Jones, the director responsible for deciding on the admissibility of the application submitted, considers that DABUS is not a person within the meaning of sections 7 and 13 of the English Patents Act and cannot therefore be considered as an inventor. Jones asserts that, even if he is wrong in his assessment, the applicant (Dr Thaler) is not entitled to apply for the patent because he is the owner of DABUS, as he has not satisfactorily explained the derivation of the right to the machine to the applicant.

The patent office's decision was subsequently upheld by the High Court of England and Wales. The judge writing for the court —Judge Marcus Smith— held that the Patents Act 1977 provides that a person —natural or legal— filing a patent application must have legal personality and that, therefore, a patent can only be granted to a "person" with legal personality. He found that since the inventor is by default the person entitled to the patent, it follows that the law in force requires the "inventor" to have legal personality. He also argued that, taking into consideration that patent rights are property rights, a machine is not legally entitled to hold and assign patent rights, as it lacks the legal personality necessary to assign property rights or even to hold them by creation.

Dr Thaler appealed the High Court's decision²⁴ and on 21 September 2021 a 2:1 majority of the Court of Appeal of England and Wales upheld the decision of the High Court of England and Wales, ruling on two fundamental issues raised in the appeal. Firstly, that only a natural person can be considered an inventor. Secondly, that there can be no derivative acquisition of any patent rights because the machine is not capable of assigning them either, as it has no legal personality.

For two of the three judges, Dr Thaler's status as the owner of DABUS —the rights holder of the computer software— is insufficient to establish derivation of ownership of the inventions claimed —those arising from the execution of the code which he is the rights holder of—.

They argue that there is no legal provision in the UK that grants such a right —that a new intangible, generated by tangible property, is owned by the owner of the tangible property—.

²⁴ *Stephen Thaler v Comptroller General of Patents Trademarks and Designs* [2021]. EWCA Civ 1374.

Colin Birss, the dissenting judge, agreed with the other judges that the wording of the UK Patents Act requires an “inventor” to be a natural person. However, he sees no impediment to patentability in the circumstance that the creator of an invention is a machine. He argues that the UK Patents Act does not require an investigation into whether a claimed inventor is actually the inventor or whether the non-inventor applicant has the necessary legal or contractual entitlement to benefit from the patent. For Mr Birss, the requirement to name the inventor in a UK patent application is met when the applicant names who they actually believe to be the inventor, accompanied by a statement of the legal circumstance from which the right entitling them to apply for the patent is derived.

4.2.1.2.3.3. In the European Union

On 17 October 2018 and 7 November 2018, Dr Thaler filed two patent applications with the European Patent Office (EPO). The first, EP 18 275 163 relating to “Food packaging” and the second, EP 18 275 174 relating to “Devices and methods for attracting increased attention”. No application named an inventor in the grant application.

On 28 January 2020 after an oral hearing with the applicant in November 2019 the EPO determined the refusal of the patent applications to the DABUS inventions.

As both stated that he was the inventor of a machine called “DABUS” and that his status as a non-inventor applicant was due to his ownership of the inventing machine, the EPO analysed the applicable European Union regulations and concluded that the inventor stated in a patent application must be a natural person. The office also noted that a number of courts in Member States had issued decisions to this effect.

It also recalls that the granting of the patent has legal consequences, such as benefiting from certain rights, and that in order to be able to exercise them, the inventor must have legal personality, which artificial intelligence systems do not enjoy. In short, the EPO understands that naming a machine does not fulfil the legal requirements for applying for a patent.

Dr Thaler appeals against the decision rejecting his applications and on 21 December 2021 the EPO Board of Appeal dismisses his appeal. The decision, which was not published until 5 July 2022, concluded that an “inventor” within the meaning of Article 81 of the European Patent Convention (EPC) must be a person with legal capacity.

The Board also rejected another application that did not identify an inventor, but stated that a natural person was entitled to apply for a patent as the owner and creator of an artificial intelligence system (DABUS). On this issue, the Board agrees with the decision under appeal that a machine has no legal capacity to assign any rights and therefore Dr Thaler could not be an assignee of the rights in DABUS’ creations within the meaning of Article 81 EPC.

However, the Board also stated that it was “not aware of any case law that prevents the user or owner of a device involved in an inventive step from designating himself as an inventor under European patent law”, and could always report in the application that an AI generated the invention. While this approach to patent filing may solve the problem of unpatentability, it is incompatible with jurisdictions such as the United States and the United Kingdom, where the inventor must have “conceived” or “devised” the invention and not just be someone who owns a computer. It is also unclear how this approach to patenting would work when the user and the owner of a device are different individuals or groups of individuals.

The Board concluded that the designation of DABUS as inventor did not comply with Article 81 of the EPC because the designated inventor has to be a person with legal capacity. After citing the *Oxford English Dictionary* definition of “inventor”, which refers to a “person”, and taking into consideration that, under Article 60 of the EPC, the rights conferred by the grant of the European patent are vested in the inventor, the inventor must have legal capacity.

For the Board, under current regulations, inventions of an AI system are not excluded from patentability. Although an AI system cannot be designated as an inventor, one possible solution is that a user or owner of an inventive device can designate themselves as an inventor. They ultimately suggest a solution similar to that of the UK Court of Appeal, which said that the applicant would have been more successful if he had named himself as inventor. However, this does not appear to be Dr Thaler's position.

Whether AI systems can be rights holders or have legal personality is a regulatory policy question that goes beyond the remit of patent offices. It is also clear that the patent system as it stands prevents the attribution of creations and inventions or non-humans, making it impossible to designate an AI as a creator or inventor. This is evidenced by the overlapping decisions on this application in the United Kingdom, the United States, Germany and New Zealand.

4.2.1.2.3.4. In Germany

On 31 March 2022, the German Federal Patent Court issued a decision similar to that of the EPO Board of Appeal when it ruled on the fulfilment of the formal requirements for the grant of the patent applied for by Dr Thaler. Although the court was unwilling to interpret the term “inventor” as defined in the German Patent Act in a way that extended to non-humans, it recognised that the obligation to truthfully designate the inventor, which the court said was “undoubtedly” DABUS, determined Dr Thaler's right to grant a patent as the rights holder of DABUS. However, the Court recognised that the designation of Stephen L. Thaler as inventor as the creator of DABUS was a good option for a successful application, provided that the applicant (Thaler) had claimed that an AI had generated the invention. The German Patent Office is appealing this decision.

4.2.1.2.3.5. In Australia

On 30 July 2021, the Federal Court of Australia became the first and only court in the world to rule that an AI can be designated as an inventor in a patent application. In that decision, Justice Beach acknowledged the importance of AI for the present and future of innovation and showed a level of pragmatism heretofore unknown in relation to the various decisions arising from the patent applications filed by the “smart inventor project” team.

In Beach’s words, arguing that AI can be an inventor under the Australian Patents Act 1990 “reflects the reality like many other patentable inventions where, applying common sense, a human being cannot be said to be the inventor”. Furthermore, Justice Beach also found that Dr Thaler was a derivative rights holder by virtue of his ownership of DABUS, his ownership of the copyright in the source code of DABUS and his ownership and possession of the computer on which DABUS was executed.

However, on 13 April 2022, the Federal Court unanimously overturned the first instance decision, concluding that only a natural person can be an inventor under Australian law and, therefore, that an invention conceived solely by an AI (an AI-generated invention) could not receive patent protection. While the Plenary Session of the Federal Court acknowledged the need for legislators to consider the questions raised by artificial intelligence in the field of patentability, it did not consider that the legislation, as currently drafted, was open to an interpretation that would include AI within the term “inventor”.

On 16 May 2022, Dr Thaler and his defence applied for special leave to appeal the decision to the High Court of Australia.

4.2.1.2.3.6. In New Zealand

On 31 January 2022, the Intellectual Property Office of New Zealand declared a patent grant application filed by Dr Thaler invalid on the grounds that the application did not identify a natural person as the inventor. The Assistant Commissioner of Patents found that the application did not comply with section 22(1) of the New Zealand Patents Act 2013 which provides that a patent may only be granted to a person who (i) is the inventor, (ii) owns the invention by assignment from the inventor or (iii) is the representative of a deceased person referred to in (i) and (ii).

The New Zealand Patents Act defines the term “inventor” as “the actual creator of the invention”. Although the definition does not refer to natural persons, the Assistant Commissioner considered it “intrinsic” to the proper interpretation of the New Zealand Patents Act that the inventor be a natural person. Even if DABUS could be considered an inventor, the Assistant Commissioner understood that Dr Thaler was not a derivative rights holder of DABUS’ invention because, as a non-human being, DABUS was not capable of being an original rights holder as it lacked the necessary legal personality.

This decision has been appealed to the High Court of New Zealand.

4.2.1.2.3.7. In other jurisdictions

Elsewhere in the world, on 19 August 2021 Taiwan's Intellectual Property and Commerce Court (IPCC) upheld a decision by the Taiwan Patent Office that an inventor must be a natural person. Dr Thaler's application has also been rejected by the Korean Patent Office and the Israeli Patent Office. The Indian Patent Office recently issued a first examination report stating that the application could not be granted because it did not name a natural person as the inventor. The decisions of the Taiwanese IPCC and the Korean and Israeli patent offices have also been appealed.

To date, South Africa remains the only jurisdiction in the world to have granted a patent naming DABUS as inventor. However, South Africa, unlike other jurisdictions, does not conduct a substantive examination of the application. The sole reason for the grant was that the application passed the necessary formal requirements, which is the only condition for granting a patent in that jurisdiction.

4.2.1.2.3.8. The decision regarding "A Recent Entrance to Paradise"

As mentioned above, Dr Thaler has not only applied for a patent on DABUS inventions. He has also applied to register an AI-generated image known as "A recent entrance to Paradise".

On 3 November 2018, Thaler filed an application to register this image with the U.S. Copyright Office. The named author on the application was "Creative Machine" and the applicant was Dr Thaler.

In the application, Dr Thaler explained that the work had been created autonomously by an algorithm running on a machine (hardware) and that the work had been created as "work made for hire" by the owner of the creative machine.

In a letter dated 12 August 2019, a Copyright Office registration clerk refused to register the application as it "lacked the human authorship necessary to support an application for copyright protection".

By communication dated 23 September 2109, Thaler requested that the office reconsider its initial refusal to register the work, arguing that "the human authorship requirement was unconstitutional and not supported by statute or case law".

The office, taking into consideration Thaler's arguments and in a communication dated 30 March 2020, again concluded that the work "lacked the human authorship necessary to underpin an application for copyright protection". The office set out that Thaler had "failed to provide evidence of creative input or the intervention of a human author in the creation of the work". The office also stated that it would not "abandon the established and unanimous interpretation of copyright regulations".

By communication dated 27 May 2020 and pursuant to 37 CFR § 2025 (c), Dr Thaler submits a new application with the same arguments as the first request and incorporating new regulatory policy. In Thaler's opinion, the office should allow the

registration of creations generated by AI because it is the way to achieve the fundamental goal of copyright regulations: to encourage creation. He states that “there is no regulation or court decision that excludes copyright for computer-generated works”, and that in fact they already allow non-human entities (legal persons) to be considered as authors when the creator (worker) has an employment relationship with them or when a “made for hire” contract has been formalised between the author and a company.

The Board accepts as true a fundamental factual issue: that an AI created the work autonomously and without any creative contribution from a human. As a general rule, the U.S. Copyright Office accepts as true the facts stated in the registration materials²⁵. However, in the Board’s perspective, copyright law only protects “the fruits of intellectual labour” that “are based on the creative capacities of the [human] mind”²⁶.

In accordance with § 3132 in the third Compendium of U.S. Copyright Office Practices, the office will not register works “produced by a machine or by a mere mechanical process” that operates “without any creative input or intervention by a human author” because, by virtue of the law, “a work must be created by a human being”.

Under these regulations, the Register understands that Thaler has two options: (i) provide evidence that the work is the result of a human creative contribution or (ii) convince the office to depart from a century of copyright case law. From the Register’s perspective, it has done neither.

Since Thaler neither states nor claims that the work was created with the contribution of a human author, the Board focuses on the second option: the possibility of changing long-standing copyright case law. Mr Thaler argues on this issue that the human authorship requirement is unconstitutional and not supported by case law. In its response, the Register reminds Thaler that the application is not admissible under § 306 of the Compendium (Third Edition), which states that “the office will refuse to admit an application if it concludes that a human being did not create the work”.

The Register also argues that, in accordance with 17 USC § 102(a), copyright protection is granted to those “original works of authorship” embodied in any tangible medium of expression now known or hereafter developed such that the works may be perceived, reproduced, or otherwise communicated either directly or with the aid of a machine or device.

The phrase “original work of authorship” was deliberately left undefined by Congress in order to “incorporate without change the standard of originality established by the courts by virtue of copyright regulations”. The term has a broad, significant, but limited scope. Congress chose to express itself in this way in order to allow the regulation to be extended to creative works other than those initially envisaged²⁷.

²⁵ U. S. Copyright Office, Compendium of U. S. Copyright Office Practices § 6024(C) (3d ed. 2021) (“Compendium (Third)”).

²⁶ Compendium (3d ed.) § 306 (where Trade-Mark Cases are cited, 100 U. S. 82 94 (1879)).

²⁷ [1909] Copyright Act. H. R. Rep. No. 94-1476, p. 51 (1976).

Courts interpreting the Copyright Act, including the Supreme Court, have limited copyright protection to the creations of human authors. For example, in *Burrow-Giles Lithographic Co. vs. Sarony*, a copyright defendant argued that photographs could not be protected by copyright because the law at the time protected certain types of creations of an "author or authors" and a photograph cannot be considered an "author's production" because it is simply "a reproduction on paper of the exact characteristics of some natural object or person". 111 U. S. 53 56 (1884).

The court rejected the argument holding that an author is "one to whom anything owes its origin, originator, maker, one who completes a work of science or literature" and that, therefore, the photographs are "original intellectual conceptions of [an] author"²⁸. In its opinion, the "authors" must be "human beings"²⁹.

It defines copyright as "the exclusive right of a man to the production of his own ingenuity or intellect"³⁰, and bases this on English case law in which judges describe "author" as the "person" who is "the cause of the image produced" or "the man" who creates or materialises the idea in the work.

For the court, the nexus between "human mind" and "creative expression" is important and is considered a "prerequisite" for copyright protection.

In *Mazer vs. Stein*, the court cited *Burrow-Giles* to sustain that a work "must be original, that is, the tangible expression by the author of his ideas"³¹. The court again cited *Burrow-Giles* to argue that "while an "author" may be seen as an individual who writes an original composition, the term in its constitutional sense has been interpreted to mean an "originator", one to whom anything owes its origin"³². The Register therefore follows Supreme Court precedent, which makes human authorship a bedrock requirement of copyright protection.

Lower courts have also repeatedly rejected attempts to extend copyright protection to "non-human" creations. Thus, the Ninth Circuit held that a book containing words written by "non-human spiritual beings" can only obtain copyright protection if there is "human selection and systematisation of the revelations"³³. They argue that for the book to be copyrightable "there has to be some element of human creativity because copyright laws are not intended to protect the creations of divine beings". The same applies to the *Naruto* case: a monkey cannot claim copyright on the photographs it takes with a camera because the Copyright Act refers to "children", "widows", "grandchildren" and "widowers". All these terms imply humanity

²⁸ See *Id.*, pp. 57–59.

²⁹ See *Id.*, p. 58.

³⁰ See *Id.*, pp. 60–61.

³¹ 347 U.S. 201 214 (1954) *California*.

³² 412 U.S. 546 561 (1973).

³³ *Kristen Maaherra*, 114 F. 3d 955 957–59 (9th Cir. 1997).

and exclude animals³⁴. Or the *Satava vs. Lowry*³⁵ case, which concludes that depictions of jellyfish are not protected by copyright because “material first expressed by nature is the common heritage of mankind, and no artist can use copyright regulations to prevent others from depicting it”.

Unsurprisingly, federal agencies have adhered to the case law of the courts. Even the 1970’s, questions about the impact of computer technology on copyright law led to the creation of the “National Commission on New Technological Uses of Copyrighted Works” (CONTU). One of CONTU’s objectives was to study “the creation of new works through the use of machines (including computers) or automatic reproduction systems”³⁶.

Following its analysis, CONTU decided that the judicial interpretation of the term “original works of authorship”, establishing the prerequisite of human authorship, guaranteed the protection of works created with the use of computers and that, therefore, no amendment to the copyright law was necessary³⁷.

For the Commission, “copyright protection does not depend on the device or devices utilised in its creation, but on the existence of a minimum human creative effort at the time of production of the work”³⁸.

Furthermore, the U.S. Copyright Office (the Register) also has a well-established opinion on the matter. A decade before the passage of the Copyright Act of 1976, in its 1965 annual report published in 1966, the Copyright Office already considered the copyright implications of the then incipient computer and also asked the question we all ask today: are works created by a computer copyrightable? And they reached the conclusion that human authorship was required for protection.

In short, as early as 1966, the U.S. Copyright Office was already considering when works were essentially authored by humans —where computers were instrumentally involved— and when the traditional elements of authorship of a work (literary, artistic or musical expression, or elements of selection, arrangement, etc.) had actually been conceived and executed not by man but by a machine³⁹.

On the conception of the work, the second Compendium of U.S. Copyright Office Practices already foresaw that for the work to be eligible for copyright protection it had to have originated from a human being. Materials produced solely by nature, plants or animals are not copyrightable⁴⁰.

³⁴ *Naruto vs. Slater*, 888 F.3d 418 426 (9th Cir. 2018).

³⁵ *Satava vs. Lowry* 323 F.3d 805 813 (9th Cir. 2003).

³⁶ National Commission on New Technological Uses of Copyrighted Works, Pub. L. 93-573, § 201(b)(2), 88 Stat. 1873 1873 (1974).

³⁷ CONTU, Final Report in 1 (1978).

³⁸ *Id.*, pp. 45 and 46.

³⁹ P. 5 of the U.S. Copyright Office, Sixty-eighth Annual Report of the Register of Copyrights for the Fiscal Year Ended 30 June 1965 (1966).

⁴⁰ Available at <https://www.copyright.gov/history/comp/compendium-two.pdf>.

For these reasons, the Compendium of U.S. Copyright Office Practices has long considered human authorship mandatory for registration, specifically since the promulgation of the Copyright Act of 1976. It was then that the compendium was updated to incorporate the human authorship requirement for the register⁴¹.

The current compendium retains this requirement and over time has applied it to multiple circumstances where non-human conception has posed different challenges. Thus, in the third Compendium of U.S. Copyright Office Practices we find §§ 709.1 (machine translations performed by computers); 803.6(B) (sound recordings made by purely mechanical processes); 805.4(C) and 806.4(C) (human performance required for choreography and pantomime); 808.8(E) (human selection of colour in coloured film); 906.8 (machine-produced expression in visual works of art, such as linoleum flooring); 909.3(B) (x-rays and other medical imaging); 1006.1(A) (hypertext markup language if created by a human being "rather than website design software").

Although none of the compendiums explicitly address artificial intelligence, the Board concludes that office policy and practice make human authorship a prerequisite for copyright protection.

The Register's position is supported by a recent report by the United States Patent and Trademark Office (USPTO) that addresses the intellectual property issues raised by AI. The USPTO requested public comment on whether, under the current regulations, "a work produced by an algorithm or by an AI process, without the participation of any natural person, can be considered a "work of authorship"⁴².

For the vast majority of those who commented, "current law does not allow a non-human to be an author [and] this should remain the law"⁴³.

In relation to the second of Thaler's arguments, that artificial intelligence can be considered an author because "the work made for hire doctrine" allows "non-human persons such as corporations to be considered authors", it is likewise rejected. For the Register, the work is clearly not made "for hire" as defined in the Copyright Act. There are two possibilities for a work to have been made for hire (i) it may have been created by "an employee" or (ii) it may have been the work of one or more contractors with whom it has been expressly agreed in writing that it is made for hire work⁴⁴. In both cases, the work is derived from a binding contract—an employment or work contract—. Since Dr Thaler's "creativity machine" has no legal personality and therefore cannot enter into binding contracts, it cannot possibly meet the requirements of 17 U. S. C. § 101.

⁴¹ U.S. Copyright Office, Compendium of U. S. Copyright Office Practices § 202.02(b) (2d ed. 1984) (Compendium [Second]).

⁴² United States Patent and Trademark Office, "Public Views on Artificial Intelligence and Intellectual Property Policy" p. 19 (2020), available at https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf.

⁴³ *Id.*, pp. 20 and 21.

⁴⁴ 17 U. S. C. § 101.

Furthermore, the Register reminds Dr Thaler that the work for hire doctrine only concerns the assignment of ownership of a work, not whether it is protected by copyright. Consequently, for the Register, the work is not made for hire because it is neither a “work of authorship” nor has it been created “for hire”.

For the Copyright Office, Dr Thaler’s second argument is nothing more than a political argument in favour of the legal protection of works created exclusively by artificial intelligence. He does not cite any case law or other precedent that challenges the Register’s interpretation of the Copyright Act.

Dr Thaler’s legal representatives filed an appeal against this refusal with the United States District Court for the District of Washington D. C.

4.2.1.2.4. *Conclusions regarding the DABUS case*

It is clear from the above analysis that Dr Thaler’s ambition is unfeasible today. Legal systems have been drafted by and for humans. The attribution of any right to a machine would imply the automatic recognition of legal personality to it, which is currently not on the regulatory agenda of countries. This is logical. The disruption is such that regulatory decisions of this magnitude should not be rushed, especially considering that we are still at the dawn of artificial intelligence.

Accepting this important issue—and consequently discarding the attribution of authorship to AI—homogeneous solutions will have to be adopted to encourage progress in AI and other disciplines.

In order to do so, it is first necessary to assume that it is the AI who are the genuine authors of the intellectual or industrial creations whose creation must be encouraged.

I am not alone in making this assumption. From the above analysis of administrative and judicial decisions regarding the authorship of AI, we see that the assumption is widespread. No one questions the creative autonomy or the material authorship of AI such as DABUS. Firstly, because the decisions analysed must take on board the declarations of the applicant (Dr Thaler), who has stubbornly maintained his position on the authorship of the AI, ignoring the requests for remedy by the administrative bodies responsible for granting protection.

Secondly, because from Mr Thaler’s explanations of how his “creative machine” works, it seems difficult to infer human creativity of any kind. The randomness and opacity of the creative process *de facto* preclude any possibility of “human conception”.

Now, assuming the material authorship of AI over protectable creations and the impossibility of granting it to them, two regulatory positions emerge:

1. That the creation is in the public domain.

The first is to understand that AI creations must enter the public domain and the second is to grant ownership of the creation of the AI to whoever holds the rights to it (the legitimate rights holder). It is from the U.S. and has much to do with the requirement that the creation must have a “human conception”. Since the promul-

gation of the Copyright Act of 1976, the Compendium of U.S. Copyright Office Practices has considered human authorship mandatory to apply for registration. "Materials produced solely by nature, plants or animals are not copyrightable"⁴⁵.

2. That the creation belongs to the AI model rights holder.

The second possibility is that ownership is held by the author of the computer code that governs the AI model or by whoever, according to the regulations in force, is its legitimate rights holder and originates in the UK. The UK legislator's argument is that AI is already utilised to create copyrighted works (music and works of art). In most cases, the use of AI is as a tool, and human creativity remains a fundamental part of the creative process. Indeed, some argue that there is currently no situation in which a copyrighted work can be created without any human involvement. Humans are involved in the design of AI or in their training. Furthermore, AI learns from copyrighted human creations. They also question the supposed total autonomy of these AI. They understand that humans can direct the work produced by the AI, for example, choosing what kind of song an AI should produce, what instruments should be utilised, what the song should sound like and what the tempo should be.

In short, they understand that to the extent that a work is made with the help of AI but includes human creativity, it should be protected like any other work. Copyright will protect the work to the extent that it is the result of a "human conception". They understand that in these cases AI can be considered to act as a tool that allows a human to express their creativity.

For this reason, the following provisions are incorporated in the Copyright, Designs and Patents Act 1988 (CDPA)⁴⁶:

Article 9.3): "In the case of a literary, dramatic, musical or artistic work generated by computer, the author shall be deemed to be the person who makes the arrangements necessary for the creation of the work."

Article 178: "Computer-generated", in relation to a work, means that the work is computer-generated in such circumstances that there is no human author of the work.

Thus, the UK protects computer-generated works that do not have a human creator (Section 178 CDPA). The law stipulates that the author of such a work is "the person who makes the arrangements necessary for the creation of the work" (Article 9.3 CDPA). The protection lasts for 50 years from the date on which the work is carried out (Article 12. 7 of the CDPA).

The astonishing thing is that the modification, which was proposed in 1987, was intended to protect "materials" that were the result of executing expert systems such as weather maps and, consequently, to some extent, AI-generated work⁴⁷.

⁴⁵ Available at <https://www.copyright.gov/history/comp/compendium-two.pdf>.

⁴⁶ U.K.: Copyright, Designs and Patents Act 1988 – CDPA.

⁴⁷ Lord Young of Graffham, a member of the House of Lords at the time, said it was "the first copyright legislation anywhere in the world that attempted to deal specifically with the advent of artificial intelligence".

The problem with the British proposal is that the conception of “originality” has evolved over time. In 1988 a work was “original” when it was the result of the “skill, labour or judgement” of its author. But the current approach is that an original work must be “the author’s own intellectual creation”. That is, it must be the result of the author’s free and creative choices, thus incorporating their “personal touch”. Consequently, it seems that the current conception of originality has evolved towards a more human linkage. Moreover, this is not a trivial issue considering that literary, dramatic, musical and artistic works are only protected by copyright if they are original.

By accepting that a human is the author of an AI-generated work, the British approach separates authorship and creativity. The creator of the original work is the AI, but the “author” according to the law is a person who has not made any creative contribution and who hardly fits the modern conception of originality, where, as has been said, creativity and authorship go hand in hand.

In order to circumvent the originality requirement, so-called “entrepreneurial works” (sound recordings, films, broadcasts, typographical arrangements, computer code...), which do not need to fulfil the originality requirement, will have to be utilised. Rights belong to their producers, manufacturers and publishers, regardless of their creative input. Equating AI creations with entrepreneurial works means assuming *de facto sui generis* protection of these creations as a way of circumventing the originality requirement.

This protection is *de facto* granted to AI-generated creations, without the need for a specific provision. However, it is less extensive than the protection granted to original works. Protection is not granted in the same way to the one who accesses authorship by creation (the original author) as to the one who accesses this status in a derivative way. For example, the copyright owner of a musical work can prevent any reproduction of their work, but the copyright owner of a sound recording can only prevent the copying of that particular recording.

In addition, this false attribution of authorship means a new breach of U.S. doctrine, which is blunt in its test to determine when a work has been “made for hire”, limiting it to two assumptions: (i) an employment relationship and (ii) an express commercial agreement that it is made for hire and at the expense of the contracting party. In U.S. doctrine it is understood that, in the absence of legal personality, AI is neither legitimate employees nor assignors of any rights, and therefore the “work made for hire” doctrine lapses.

The International Association for the Protection of Intellectual Property (AI-PPI), an international organisation that promotes the development of intellectual property laws, recently asked its members whether AI-generated works should be protected. The responses received highlight the different approaches to this issue. The UK group suggested that AI-generated works could be protected by a new right, lasting 25 years, that recognises the investment that AI developers make in this technology (protection of AI by a *sui generis* right). But other respondents

argued that copyright protection should be granted only to human creativity. The final AIPPI Resolution on this issue emphasises the need for human intervention and originality.

Regardless of whether such works are protected by copyright, it may be difficult in the future to determine whether a work was generated by a human being, a machine or both. This may require technological solutions to help determine authorship of works and ensure that the correct type of protection is applied.

4.3. PROTECTION OF PROTECTED CREATIONS UTILISED TO "TRAIN" AI. COPYRIGHT AND TDM

Having analysed the first two controversies, on the protection of the AI model itself and on the protection of AI creations, we focus on the last one: on the possible infringement by AI creators of the intellectual property rights of the owner of the information supplied to these systems. We have already referred to this issue when we reflected on the possibility of applying the *sui generis* right of makers of databases as a way to protect the input that feeds AI.

Indeed, if you recall, artificial neural networks emulate biological ones: they obtain information (dendrites), process it (soma) and send it on to other neurons (axon). A process similar to machine learning with which the "machine" is trained. The whole process starts with the acquisition of information by the AI.

AI has a significant appetite for data. The more data, the better. Their statistical calculations become increasingly more precise. With that in mind, neural networks need access to huge volumes of information (data), including large amounts of works that are often protected by intellectual property rights.

And where does this data come from? To put this in context, it is estimated that more than 90% of the data in the world today has been generated in the last two years, almost 16 quintillion bytes of data every day. It is also estimated that growth in the coming years will continue to be exponential (doubling in volume every two years). The widespread use of the internet and the devices connected to it (internet of things) or the decreasing cost of information storage ensure the increasingly rapid generation and processing of huge amounts of information.

Some of this information is composed of protected creations, another part is "off the market" (not commercialised), and some is orphaned (no known author or author who cannot be traced). One might think that orphaned data is an exception, but the figures from those who have analysed their collections belie this. For example, the British Library estimates that 40% of its potentially copyrightable collections are orphan works. Also illustrative is the BBC's statement, assuming that one million hours of programmes in its catalogue could not be utilised because of the impossibility or disproportionate cost of attributing those rights to individual authors with the subsequent legal risk of infringement of orphan rights.

4.3.1. What is text and data mining (TDM)?

To approximate the meaning of “text and data mining” (TDM), imagine that you are able to access a large number of scientific articles on a particular subject, read them and make careful connections between the information in all of them. Imagine also that you are able to establish patterns of various nature that emerge from the analysis of the whole (every article). The truth is that the approach is utopian. Humans are not capable of reading and assimilating such detailed and interrelated scientific articles. However, there is a computational technique capable of doing so: TDM (text and data mining). It is able to distil the information contained in articles to establish patterns that enable the acquisition of “new knowledge”. The only condition for this feat to materialise is that the algorithms have the huge amount of data needed to make the statistical patterns sufficiently robust.

Directive 2019/790 on copyright and related rights in the Digital Single Market defines TDM as “any automated analytical technique for analysing text and data in digital form in order to generate information including, but not limited to, patterns, trends or correlations”.

And how is the miracle achieved? With computational processes. Mining itself essentially involves two sub-processes: (i) pre-processing of relevant text and data and (ii) extraction of structured data.

Regarding pre-processing, the extracted text may be unstructured, causing the data being processed to generate “noise” and causing errors in the computational process. In such cases, the extracted data will need to be “cleaned”. Data cleansing is necessary because data is often generated and stored with human assimilation in mind. When the analysis is computational, the structuring does not necessarily coincide with that required for human analysis and assimilation. Typically, this sub-process includes:

1. The elimination of irrelevant information. For example, imagine analysing hundreds of thousands of hours of TV programmes (the data) and deciding to remove advertising as irrelevant to the analysis.
2. The use of tables, formulas, figures or in general any element that serves to add or link the content.

The second sub-process is the extraction of structured data. To achieve this, data processing includes:

1. Tokenisation work⁴⁸ to provide documents with homogeneous structures and, if possible, to eliminate certain characters such as punctuation. In languages with completely different characters from those we are used to, such as the Chinese alphabet characters, homogenisation is difficult without tokenisation.

⁴⁸ It is a technology that allows the “securitisation” of data or datasets. A term widely utilised in the crypto world to refer to a set of data whose transfer could have legal relevance. They are utilised to represent “transferable value” through distributed records under blockchain technology.

2. The identification of synonyms through various linguistic resources such as lemmatisation (obtaining the lemma in the structure of the word), the application of linguistic rules (stemming) or through the application of learned statistics (embedding).
3. The generation of attributes, which involves the transformation of the text. The text is represented by a "bag-of-words" based on the words incorporated in the text itself and the frequency of their use in the text.
4. Identification of equivalent classes of texts or of the tokens into which texts are grouped.
5. Giving importance (influence) to certain texts or tokens over others.

When we talk about "predictive data mining", where supervised machine learning techniques are utilised, the extraction stage is followed by the identification (and verification) of patterns and the results obtained. In this way, TDM techniques allow the discovery not only of concepts, but also of the relationships between these concepts, as well as between these concepts and patterns.

4.3.2. **What are the expected uses of TDM?**

Although the use of text and data mining and especially AI is still in its infancy and it is risky to categorically establish the uses to which it can be put, it is expected to be utilised not only for research, but also for a multitude of social or business activities such as banking, marketing, anthropology, insurance, optimisation of natural resources, identification of bank customers, analysis of their creditworthiness or analysis of traditional industries and sectors. It is also utilised for visual monitoring of information dissemination to determine the relevance of published information, or in criminal investigation, generating patterns of illegal behaviour and associating them with certain personality types. They are also utilised to help anthropologists draw conclusions about cultural or social phenomena by mining information posted on social networks. For example, a recent study by Cornell University, whose research team had access to 100 million photographs taken from Instagram, established patterns of how clothing varied around the world and was able to determine patterns of frequency of wearing certain clothes and colours in different geographical locations, establishing differences in patterns according to date and location and reaching conclusions about the favourite colours in each area at a given time. To give another example, IBM's Watson Explorer has been utilised to date, among other things, to improve productivity in the workplace, efficiency in public health management (in Italy), and to achieve medical diagnoses as in the case of the oncology department of the Memorial Sloan-Kettering Cancer Centre in New York, or for the prevention of cybercrime and hacks, or to create new culinary recipes or even to guess (correctly) who was going to win the San Remo music festival.

4.3.3. What is the controversy surrounding TDM?

The controversy arises in the first of the three layers: the acquisition of the information that will then be conveniently “distilled” through the execution of the corresponding algorithms. What happens is that the operation of AI is the result of the execution of computer code, which feeds its “variables” with data provided to the system —called input in computer jargon—. Recall that computer code not only executes itself; it also “refines” its algorithm based on input data. Furthermore, exclusive rights are granted over the exploitation of the whole (the software and the data), which is the one that obtains the result —the output—. However, it is possible and even frequent that the ownership of the software and the ownership of the data do not belong to the same natural or legal persons.

It will be necessary to distinguish when the owner of the software is also the owner of the data or works that feed it and when they are not. If the owner of the AI model is also the owner of the data or works or they are in the public domain, there is no barrier to their incorporation into the AI model. As the owner of these creations, you can also prevent third parties from extracting or re-utilising the information.

Moreover, to the extent that the selection of the data incorporated into the system is original (which will sometimes be the case due to the intimate relationship between the selection of the data and the algorithm that manages them), you will hold a classic intellectual property right of those contemplated in our current consolidated text on intellectual property (selection or arrangement of materials). To understand this, let us reflect on the work of a Microsoft AI known as “The next Rembrandt” which, after distilling data from 326 works by the Flemish painter, asked the trained AI model to create a new work in the painter’s “style”. The project was done after authorisation was granted to access his works. But imagine if a living painter decided to do the same. They would feed the “model” with data (images) of their works and after training the model they would obtain new pictorial works in their style. Moreover, in principle, they could prohibit third parties from “distilling” their works to train an AI model to create new works in their own style. This may be more easily understood if we provide an example with literary works by two authors (for example, Cervantes and a living author). If an AI model is implemented that writes a new literary work in the author’s style, in the first case, the use of Cervantes’ works is free because they are in the public domain. The living author, however, could prohibit the “extraction and/or re-utilisation” of their works for this purpose. Consequently, the problem arises when the owner of the software uses works that are not in the public domain and in which they do not hold intellectual property rights. The controversy is over the use of protected creations. Is the natural or legal person who selects, extracts and reproduces information for text and data mining authorised? Can the rights holders of any intellectual creation prevent AI creators from incorporating their creations into the TDM process to train AI?

We find different positions on both sides of the Atlantic. In the USA, protection has never been granted to “raw” data. Therefore, the collection of event data from live or previously recorded broadcasts, which can subsequently be commercially exploited, cannot be prevented⁴⁹.

However, in Europe (UK) in the case *Football Dataco Ltd vs. Sportradar GmbH*, the court found that there can be a re-utilisation of sports broadcast data as long as there is also processing of the data because such processing requires effort, discretion and considerable expertise, which ultimately involves a clear intellectual input⁵⁰. The UK court’s decision is a foretaste of the EU’s position on the ownership of data obtained through text and data mining, which I will discuss in more detail below.

4.3.4. Text and data mining in the USA. The fair use doctrine

Focusing on intellectual property regulations, the problem is that TDM requires the copying and extraction of protected information. Indeed, in most cases, text and data mining activities require the prior copying of the information to be supplied to the AI model. However, not all copying infringes copyright. If the TDM procedure only requires the “temporary copying” of text or other data and the result of the computational analysis are mere “facts” extracted from the text or data, there is no infringement of any intellectual property right. The reason is that the copyright holder has the exclusive right to reproduce items in “copies” and that many jurisdictions, including the USA, define “copy” as a reproduction that “remains”, as distinguished from other reproductions that have an ephemeral or temporary duration. Thus, a U.S. Court of Appeal ruled that a copy made in the internal memory (buffer), which lasted one to two seconds, was excluded from the application of copyright regulations⁵¹.

In any case, this does not seem to be an absolute truth. Temporary copies are not an infringement of the rights holder’s rights as long as they are instrumental copies; those that are technically necessary for the performance of lawful activities. Since, in my opinion, the real substance of the authorisation to copy in the “cache memory” is not its duration, but the technical necessity of doing so and the lawfulness of the purpose pursued, it must be understood that the exception for temporary copying will depend on these circumstances and not on the duration of such copying in the memory of the computer in which it is stored.

In almost all cases, researchers will be empowered to conduct data-mining activities on protected creations that are “accessible” —that have been published or otherwise disclosed—. They will be covered by the fair use doctrine if those generating the AI

⁴⁹ Aaron Feld, Note, *Gambling on Sports Data: Protecting Leagues’ High-Level Data from Sportsbooks*, 2020 U.Ill. L. Rev. 341 366 (2020).

⁵⁰ *Id.* p. 367.

⁵¹ *Cartoon Network LP, LLLP vs. CSC Holdings, Inc.*, 536 F.3d 121 (2d Cir. 2008).

(i) have lawful access to the copy and (ii) are not bound by a contract that restricts their fair use rights (including a specific opt-out for TDM activities). In such cases, mining activities may be carried out to the extent that the results do not make the whole of the text (protected creation) or substantial parts of it accessible to the public.

The fair use doctrine has long been a critical element in settling copyright disputes. It is a common legal principle that was finally incorporated into the discipline in the Copyright Act of 1976. It was not expressly codified because Congress understood at the time that in a period of rapid technological change it did not seem appropriate to regulate for specific purposes⁵². Nevertheless, Congress encouraged the courts to continue to apply the fair use doctrine as an equitable and technologically sensitive rule.

It is regulated in 17 U. S. C. 107 which states: “Notwithstanding the provisions of articles 106 and 106A, 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 intellectual property rights”.

The courts, for their part, have emphasised the importance of applying the doctrine in a flexible manner⁵³. In determining when the use made of a protected work is fair, factors to consider include:

1. The purpose and nature of the use made, including whether it is of a commercial nature or is for non-profit research purposes.
2. The nature of the protected work.
3. The amount or substantiality of the original work utilised.
4. The effect that the intended use may have on the normal exploitation of the protected work.

In order to understand the application of the rules defining use as “fair”, it should be noted that the four elements are considered as a whole, without the fulfilment of all or any of them being necessary for the use to be considered fair. Courts will have to consider all of them together and make decisions in the specific circumstances of the case. Special consideration should be given to the flexibility of this doctrine in which one or more of the four factors mentioned above may be decisive in the judicial decision. The flexibility of the doctrine allows it to accommodate unforeseen changes that may occur as a result of technological breakthroughs. When applying this doctrine, U.S. courts have ruled on the use of TDM for a variety of purposes:

In 2003, the case of *Kelly vs. Arriba-Soft*⁵⁴ is the first case to rule on this issue. The court understands that there is fair use when images are incorporated into a “search engine”. The Arriba-Soft search engine is utilised as a tool to facilitate

⁵² 3 H. R. Rep. No. 94-1476.

⁵³ *Cartoon Network LP, LLLP v. CSC Holdings, Inc*, 536 F.3d 121 (2nd Cir. 2008).

⁵⁴ 336 F.3d 811 (9th Cir. 2003).

indexing and improve access to images accessible from the internet. The search results include thumbnails of the images that serve to automatically link—in-line link—to the publisher's website.

In *Authors Guild vs. HathiTrust*⁵⁵, the defendant markets a tool to facilitate the identification and location of information sources. To achieve this, it needs to digitise and incorporate into a database of protected works.

In *White vs. West*⁵⁶, two legal publishers, Westlaw and LexisNexis, created an interactive legal text search tool. To do this, they added metadata to each of the texts in the database. Search results include access to the full version of the legal texts.

In *Fox vs. TVEyes*⁵⁷, the latter records all the programmes broadcast by television and radio stations and then, by applying technologies that transcribe all this content, generates a textual search engine of everything that is broadcasted. The search results include transcripts of parts of the programmes.

In *Authors Guild vs. Google*⁵⁸, it was also considered "fair use" when Google scanned and digitised the collections of various bookshops and incorporated the result into a database for students and researchers to consult. The search results include fragments of the text. In its ruling, the Southern District Court of New York says that Google's project "transforms, through text and data mining, the text of the book into data for the purpose of research in new areas, thus opening up new avenues of research". It states: "The words in the books are being utilised in a way that has never been done before".

In *A.V. vs. iParadigms*⁵⁹, the company iParadigms created a database that allowed teachers to compare their students' work to determine the extent to which it had been plagiarised from material published on the internet or from other work submitted to the system. Although the service provided was of a commercial nature, the court found it to be "highly transformative".

In *Perfect 10 vs. Amazon*⁶⁰, the latter incorporated thumbnails of protected images into its search engine, including an in-line link that allowed the searcher to access the images in the size in which they had been published by redirecting to the website of the claimant who had published them. The Ninth Circuit also found the tool to be very transformative.

In *Field vs. Google*⁶¹, the Californian company provided copies of the specific content of a website on a specific date. This type of cache links are typically utilised for archiving purposes, as well as for identification and accreditation of the publication of specific content on a given date. The court also found that this was "fair use".

⁵⁵ 755 F.3d 87 (2d Cir. 2014).

⁵⁶ (S. D. N. Y. 2014).

⁵⁷ (S. D. N. Y. 2014).

⁵⁸ 770 F.Supp.2d 666 (S. D. N. Y. 2011).

⁵⁹ LLC (4th Cir. 2009).

⁶⁰ 508 F.3d 1146 (9th Cir. 2007).

⁶¹ 412 F.Supp.2d 1106 (D. Nv. 2006).

4.3.4.1. *Analysis of court rulings to conclude on fair use, depending on the factors considered*

U.S. courts have emphasised the transformative nature of TDM. They have also taken the view that the application of the fair use doctrine by authorising these activities is unlikely to have an adverse impact on the market for protected works. Analysing by factors:

1. *Purpose and nature of the use of the protected works*

As discussed above, U.S. courts have repeatedly ruled that search engines are transformative in nature and, consequently, this factor has been found to be a determining factor in considering their use to be “fair”. If we recall the rulings analysed, in *Authors Guild v. HathiTrust*, the court justifies the creation of a search engine on whole texts because the result of a search for a word in a search engine is different in purpose, nature, expression, meaning and message from that of the page (and the book) from which such word has been extracted. Thus, even if the search engine has a commercial use, its transformative nature tips the balance towards accepting the use of protected works as “fair”⁶².

2. *The nature of protected work*

This is not a factor that has attracted either consensus or particular relevance in judicial decisions. Of the eight analysed, two have found that this was a neutral factor or have favoured the defendant carrying out the TDM activities and others have found in favour of the claimant/rights holder.

3. *The amount and/or substantiality of the original work utilised.*

The TDM needs to incorporate into the process the full verbatim copy of the text or work it uses. If it does not, the researcher will make ineffective use of this technology because potentially necessary portions of the protected work would not have been analysed. For this reason, U.S. courts have concluded that verbatim copying of the entire protected work is reasonable and fair.

4. *The effect that the intended use may have on the normal exploitation of the protected work*

Decisions to date suggest that this factor tends to favour the one who copies and incorporates the protected work into their database. They understand that the transformative nature of TDM makes its use unlikely to have an adverse effect on the natural market for the protected work for the simple reason that, given the nature of this technology, the protected work is unlikely to be replaced by the derivative.

⁶² *Fox News vs. TVEyes* CITE (S. D. N. Y. 2014).

This, however, is a position I do not share. To understand that an AI model that distils knowledge from books is not going to affect the publishers who publish them is questionable. If a student is given the choice between (i) searching the appropriate bibliography, reading the relevant books, assimilating their content and then writing down the conclusions drawn or (ii) requesting the "work" from an AI model, it is reasonable to understand that a high percentage of students will opt for the easy way out, causing obvious harm to the rights holders of the literary works.

4.3.4.2. *Conclusions. "The right to read is the right to mine" doctrine*

The combined weighting of the four factors leads to the conclusion that U.S. courts have understood TDM to have a highly transformative use, the result of which is not a substitute for the original works, even when a full copy of the original works has been made. In short, it seems that decisions on the transformative status of TDM incorporate "the right to read is the right to mine" doctrine.

Phrase initially attributed to Peter Murray-Rust, an engineer and proponent of open data. It summarises a philosophy that, as we have seen, is being applied when examining the use of protected creations in machine learning (text and data mining) activities. Feeding AI with protected creations can be seen as a blatant infringement of the intellectual property rights of the owners of those creations. However, there is another view that seems to be gaining ground. When you acquire a protected creation, for example a book, and read it, you gain knowledge. With that knowledge you can develop new protected creations or establish an *ad hoc* business model, and yet you are not obliged to pay any amount to the rights holders of the works you acquired to gain the knowledge. Based on this reality, open data proponents understand that the distillation of the data contained in protected creations, the interrelation of this data with that of other creations and the obtaining of patterns are activities that are commonly carried out by humans to acquire new knowledge. The "automation" of the process of reading and understanding texts by computers equipped with algorithms that are refined according to the information fed into them is no different from human rumination, except that the machine's capacity for interrelation is much greater and, therefore, it is easier to acquire "new knowledge".

If we look at the decisions of the American courts which, in application of the fair use doctrine, have understood that the activity was very "transformative" and that it did not interfere with the usual exploitation of the creations with which the AI is fed, they seem to have incorporated the doctrine coined by Peter Murray-Rust.

4.3.5. **The exception for TDM in the UK**

The UK introduced the first exception to intellectual property rights for text and data mining in 2014. It did so by incorporation of section 29A into the Copyright, Designs and Patents Act 1988. The exception exclusively concerns copyright (not the *sui generis* right). Only those who mine text and data for non-commercial re-

search benefit from the exception. In short, it follows the criterion in force in the EU at the time of its transposition and more specifically in Article 5(3)(a) of Directive 2001/29/EC, known as the “InfoSoc directive”, which already accepted the exception for research purposes.

However, following a consultation by the UKIPO (UK Intellectual Property Office) in June 2022, it was decided to extend the exception beyond non-commercial research, to include any purpose (including commercial). In this way, the rights of the owners of protected creations are limited and, from that date and in the circumstances mentioned above, they can no longer circumvent the exception or require the formalisation of specific licences for the performance of TDM activities. However, the obligation of lawful access to protected creations in order to benefit from the exception was maintained.

In any case this is a highly controversial issue. There is an obvious clash of interests. While rights holders of classic creations perceive a high risk of providing the “fuel” necessary for AI to end their exploitation model, the technology industry and open data movements understand that closing protected creations to machine analysis will impede important advances in human development.

With that in mind, the latest news from the current British government on this issue implies a change of position. Swords are still at the ready.

4.3.6. TDM in the European Union. From the fair use doctrine to *sui generis* law

As mentioned above, with the generalisation of the exploitation of databases, it became clear that copyright protection was not sufficient for the proper protection of database owners. Recall that the problem arose when it became clear that many database owners, which under copyright regulations did not meet the originality requirement, were left unprotected against the unlawful extraction and re-utilisation of the “elements” making up the database. The rights holders who had invested time and money in creating them had no legal recourse to complain about the obvious unjust enrichment of those who extracted and re-utilised them.

For this reason, the European Parliament’s Directive 96/9/EC on the legal protection of databases incorporated what at the time was called the *sui generis* right of database makers. It took the form of a right independent of copyright that solved the problem of authorship by giving the database maker the right to prohibit and prevent the extraction or re-utilisation of all or a substantial part of the content of the databases, evaluated both from a quantitative and qualitative point of view and provided that the maker had made a substantial investment in order to achieve this.

The incorporation of the *sui generis* right into European Community regulations has obvious relevance to text and data mining by making it much more difficult to mine. In application of these regulations, AI developers are required to seek authorisations (licences) from the rights holders of the protected creations they wish to incorporate into their “AI model”.

The positions of rights holders of protected works differ from those held by AI developers. While the creators of AI systems embrace "the right to read is the right to mine" doctrine, which, remember, is essentially an understanding that, as long as you have lawful access to protected works, you have the right to draw conclusions from those works, both individually and as a whole. Proponents of this doctrine understand that it is unfair to limit the progress of knowledge and that the computational analysis of the data that make up the works is not substantially different from the human analysis of these works. They also argue that mining activities do not result in outcomes that harm or distort the natural market for protected works.

On the other side, some rights holders (mainly publishers and owners of protected creations) are sceptical about the possibility of authorising the mining of text and data by means of a general licence that derogates from the norm for this specific use. Press publishers understand that licensing should be on a case-by-case basis, including licences granted for non-commercial research to prevent abuse, loss of their archives and, above all, destruction of their business models.

Other publishers —such as academic publishers— have offered to include TDM activities in their licensing policies. For example, the publisher Elsevier offers all its journals and books in XML format⁶³. Users (licensees) can register and access an account on the developer portal which creates an interface between the publisher's and the developer's systems that allows the import of data and protected works in the above-mentioned XML format.

Other academic publishers, such as Oxford University Press and Springer, generally authorise (without the need for a specific licence) TDM activities as long as they are for non-commercial purposes and comply with appropriately published editorial policies.

Furthermore, where TDM activities are for commercial purposes, several publishers (for example, Oxford University Press, Springer Nature, Wiley, BMJ, The Royal Society of Chemistry, Taylor & Francis, SAGE, Cambridge University Press, American Diabetes Association, American Society for Nutrition and Future Medicine) participate in the Copyright Clearance Center's RightFind™ XML for Mining solution, which authorises those who wish to perform TDM activities for commercial purposes to obtain, on demand, the requested content in this format.

To analyse the impact of TDM activities on intellectual property rights, we also start with access to the content on which TDM techniques are to be applied. Agreeing with the Americans, a distinction must be made between open and non-open data and texts, taking into consideration that free access to content does not imply that it is free from legal restrictions. It is for this reason that the licensing modality is of particular importance.

But for a licence to be granted it is first necessary to know the identity of the rights holder as licensor. The Court of Justice of the European Union (CJEU) has

⁶³ It is a standardised file format that is suitable for TDM activities.

recently ruled that the performance of the main activities legally restricted to rights holders (authors or derivative rights holders) is conditional upon their prior consent. Thus, without the accreditation that the rights holders have accepted the licensing of the new modalities of use of their creations (use for TDM), third parties —such as Copyright Collective Organisations— are not authorised to take any decision on any use other than the one usually foreseen for these creations.

However, legal access to content, either because it is free or because it has been accessed as a licensee of the content, does not necessarily entitle the accessor to engage in TDM activities. This is because in order to carry out these activities it is often necessary to carry out other preparatory activities that include the extraction or copying of that content. These are activities reserved to the rights holder and for which a specific authorisation may be required.

In any case, in Europe, the general exception is also utilised when copies are “temporary” in accordance with Article 5(1) of Directive 2001/29/EC⁴⁴. In such cases, the authorisation of the rights holders is not required. In fact, the UK Supreme Court has had occasion to rule on this issue, stating that “not every viewing or reading constitutes an infringement of the rights holders’ rights”, provided that, among other requirements, access to the content has been lawful.

The CJEU has reiterated that the use will be considered lawful when it is authorised by the rights holder or when it has not been restricted by the applicable regulations. On the basis of this doctrine, it could be argued that the mining of websites, which requires the use of data mining techniques with potentially protectable information, will not always require the authorisation of the rights holders. Following the CJEU’s own case law in the *Svensson* case⁶⁴, it is understood that, when the creation has been freely disclosed on the internet without any requirement or condition restricting access to it, the rights holders of the AI will be entitled to copy and extract it.

However, if the TDM technique utilised involves acts of extraction or copying outside the above exception scenarios, the regulatory restrictions will apply. If the content has been extracted or copied from a database, potential claims could be pursued in two ways: for copyright infringement and for infringement of the *sui generis* right of the database owner. The two rights are independent of each other and can subsist together, at the same time and in the same database. The unlawful extraction or copying of the contents of a database can infringe both rights at the same time, even if the rights holders are different.

It should also be taken into consideration that the CJEU, in case law originating from the interpretation of Directive 96/9/EC⁴⁸ —on the legal protection of databases— has defined “database” in a broad manner, including in the definition the expression “in any format” and granting protection to literary, artistic, musical databases or to “other collections of works or other materials such as texts, sounds, images, numbers, facts and data”.

⁶⁴ Judgment of the 4th Chamber of the CJEU of 13 February 2014 (ECLI:EU:2014:76).

The database owner is entitled to prohibit and prevent the performance of certain acts, including the reproduction —whether temporary or permanent— in any form and for any purpose, of all or part of the contents of the database. The only restriction to database owners exercising their rights relates to the creation of a “back-up copy” by the lawful user of the database, as well as any other acts that are necessary to access the content, as well as to a normal or customary use of the database. In relation to the *sui generis* right of the database owner, the regulation grants the “maker” who has made a substantial investment in obtaining, verifying or presenting the results (understanding the substantiality of the investment from both a quantitative and qualitative point of view) the right to prohibit or prevent the extraction or re-utilisation of all or a substantial part of the contents incorporated in the database, assessing the substantiality of the extraction or re-utilisation from both a quantitative and qualitative point of view.

The restrictions will remain in place even if the databases are not protected either by copyright or by the *sui generis* right of the database makers. In this regard, the CJEU has said that the owner of such a database is not subject to the limitations of the legal protection of databases directive. For this reason, the owner is free to contractually establish —as long as they respect the applicable national regulations— the terms of use of the database that they agree with the licensee in each case. Furthermore, parties with access to the content of databases that want to apply TDM techniques should also take into consideration other related rights, including, for example, the new right of press publishers enshrined in Article 11 of Directive 2019/790 on copyright in the digital single market.

However, the EU legislator cannot simply dismiss the fact that AI is an unstoppable phenomenon that may be the key to primacy and in the development of which Europe is lagging behind. Nor can the EU legislator forget that it lacks a regulatory instrument such as “fair use”, which makes it possible to adapt intellectual property regulations to the recurrent technological challenges. For this reason, Directive 2019/790/EC on copyright and related rights in the digital single market was promulgated in 2019.

The recitals of the directive justify the adoption of the exceptions in order to strike a fair balance between the rights and interests of authors and other rights holders, on the one hand, and users, on the other, and warn that they may only be applied in certain special cases which are compatible with a normal exploitation of works or other subject-matter and do not unreasonably prejudice the legitimate interests of rights holders.

Its main novelty is that it incorporates a legal framework to solve the problems raised by text and data mining by creating two legal exceptions to allow these activities to be carried out under certain conditions. It uses the whole of Title II to regulate this issue under the expressive title of “Measures to Adapt Exceptions and Limitations to the Digital and Cross-border Environment”.

The first, drafted in Article 3, restricts the rights of rights holders of protected creations provided for in Articles 5(a) and 7(1) of Directive 96/9/EC on the legal protection of databases, Article 2 of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society and Article 15(1) of Directive 2019/790/EC (which transposes the exception).

The new exception concerns extractions or reproductions of texts and data by research bodies or cultural heritage institutions for the purpose of text and data mining of works or other benefits to which they have lawful access and provided that the purpose of such extraction or reproduction is educational or research.

Beneficiaries are allowed to indefinitely keep the material reproduced or extracted for the verification of research results provided that copies of the protected material are kept with appropriate security measures, authorising the rights holders to take measures to ensure the security and integrity of the networks and databases in which the protected works are stored.

Article 4 refers to the second exception, which concerns the same articles of the Directives as the first exception, but which does not provide for any limitation on the beneficiaries, so that this second derogation must be understood as applying to intended uses for commercial purposes. Moreover, it applies both to copyright and to the *sui generis* right of database makers. It differs from the first exception in another important respect: it grants the rights holder of protected creations the right to prohibit the mining of texts and data by means of an express contractual provision (opt-out)⁶⁵ provided that the agreement complies with regulatory provisions, since Article 7 provides that contractual provisions contrary to the application of the exceptions are inapplicable.

⁶⁵ “A deeper look into the EU Text and Data Mining Exceptions: Harmonisation, data ownership and the future of technology”, Thomas Margoni, Martin Kretschmer. *GRUR International*, 71(8), 2022 685-701.

CONCLUSIONS

A first and important conclusion can be drawn from the above analysis. Things are far from clear. As we have seen, there is no consensus solution to the problems that have arisen. Neither how to solve them nor how to deal with them.

One of the reasons for the lack of consensus is undoubtedly that the current regulations are written by and for humans. This is a major issue that not only affects author's rights —copyright— but also all areas of law. It is not difficult to predict that this circumstance will prove to be significant in the coming years. There will come a time when the importance of the autonomy of “AI” will have such legal relevance that it will be impossible to avoid the debate. In a world with an increasing number of autonomous machines whose performance has legal significance, it will be necessary to establish clear rules of governance that take into account the non-human nature of this AI and the undoubted importance of their performance in the environment.

Focusing on copyright and by way of conclusion, I will now refer to the main questions that emerge from the reflection on the three main issues that arise.

The first question has to do with the autonomy of “AI”, which is their main characteristic and which fully affects the attribution of human authorship both of the AI model itself and of its creations. The autonomy of this AI is not only reflected in the result, but also in the composition of the algorithms that obtain it. As mentioned above, the attribution of human authorship to an algorithm that evolves autonomously and opaquely is questionable. Although the conception of the initial algorithm is human, the same does not happen with the final one, which is the one that is executed and with which the result is obtained. In short, it seems that we are facing blurred boundaries. It is not known when the training ends and the actual execution begins, with the obvious difficulties that this circumstance entails.

The autonomy of AI is correlative to the questioning of human authorship, both of the AI model itself (computer software) and of the result obtained after its execu-

tion. If there is no human authorship, we have a problem, because there are no regulations that attribute these rights to non-humans and, even if we were to attribute them to AI, they would have no legal personality to act against potential infringers.

The second question that arises is about the dynamic condition of the software that governs the AI model. Protection against plagiarism of a creation must be based on comparison; on the comparison between the rights holder's software and that of the alleged plagiariser. When the software, the algorithm and the execution result evolve, the comparative process becomes diabolical. Is the allegedly plagiarised code compared to the rights holder's initial code or any of its successive versions?

The third major unknown derives from the opaque nature of the neural computation on which AI is based. Since it is a "black box" it is difficult to attribute authorship to the person who has not devised either the final result after the execution of the code or the algorithm that achieves that result. Moreover, the human authors of the AI model (computer software) are incapable of justifying in a reasoned manner the modifications made to the initial code and algorithm, thus failing to comply with the essential obligation to logically document the obtaining of the result so that, once the period of exclusive exploitation has expired, someone with sufficient knowledge can replicate the creation.

The only plausible way to solve this issue is to use an atypical protection granted to the creator of the AI as the author of the AI and also of its creations, but replacing the requirement of human creativity with a model to protect the (human) investment. In this way, creativity and authorship are separated. In short, it is a matter of using the protection already afforded to entrepreneurial works and databases; separating creativity—which corresponds to the machine—from authorship—which is attributed to the human originator of the machine—. We are therefore talking about a false attribution of authorship to humans.

The impostor attribution of authorship to a human has a fundamental drawback. There are many who believe that this undermines the protection system. Granting exclusive rights to someone who has merely parameterised the options that appear in a drop-down menu and executed a computer code does not seem the best way to encourage creations of any nature. If we were to grant intellectual property rights over a literary work to someone who simply asks an AI model to write it, we would be putting both creations on an equal footing, thus devaluing those of legitimate human creators. I insist: would you pass a student who asks Chat GPT to author their dissertation?

One possible solution is to name the AI model as the author of the creation. Therefore, the jurists involved in the Artificial Inventor project have applied for protection for two patents and a computer-generated image in the registers of various jurisdictions. We have had the opportunity to analyse the grounds on which the competent administrative and judicial bodies have based their rejection of Dr Thaler's claims. The regulations in force only contemplate the possibility of attributing these rights to humans. Moreover, even if they were attributed to them, the AI would not have

the necessary legal capacity (legal personality) to (i) take action against those who violate them or (ii) assign rights derivatively to the humans who build or use them. At the insistence of the inventor of DABUS, Dr Stephen Thaler and his lawyer Ryan Abbott, upon filing applications naming DABUS as inventor or author, the registers responsible for processing the applications have urged applicants to name Dr Thaler as inventor and to reference in the application that an AI model had assisted the applicant. However, as we have said, Dr Thaler's position is that this false attribution of authorship to humans devalues the creations that are indeed human. In short, their desire is unfeasible today. Legal systems have been drafted by and for humans. The attribution of any right to a machine would imply the automatic recognition of legal personality to it, which is currently not on the regulatory agenda of countries.

With that in mind, there are two positions on the table in relation to this controversial issue. The first, from the U.S., understands that AI creations must enter the public domain and has much to do with the requirement that the creation must have a "human conception". Since the promulgation of the Copyright Act of 1976, the Compendium of U.S. Copyright Office Practices has considered human authorship mandatory to apply for registration. "Materials produced solely by nature, plants or animals are not copyrightable".

The second possibility is that ownership is held by the author of the computer code that governs the AI model or by whoever, according to the regulations in force, is its legitimate rights holder and originates in the UK. The UK legislator's argument is that AI is already utilised to create copyrighted works (for example, music and works of art). In most cases, the use of AI is as a tool, and human creativity remains a fundamental part of the creative process. They understand that AI is another tool for the expression of human creativity. Indeed, some argue that there is currently no situation in which a copyrighted work can be created without any human involvement. Humans are involved in the design of AI or in their training. Furthermore, AI learns from copyrighted human creations. They also question the supposed total autonomy of these AI. They understand that humans can direct the work produced by the AI, for example, choosing what kind of song an AI should produce, what instruments should be utilised, what the song should sound like and what the tempo should be.

In short, they understand that to the extent that a work is made with the help of AI but includes human creativity, it should be protected like any other work. Copyright will protect the work to the extent that it is the result of a "human conception". They understand that in these cases the conception pertains to the human who has provided the necessary means for everything to happen and that, therefore, AI can be considered to act as a tool that allows a human to express their creativity. They understand, in short, that there is human conception when we create a machine to conceive. They understand that AI creations would not exist without the human work of programming and creation of algorithms, without the selection of the data sets and (human) works with which the machine is fed, without the verification work to check its progress and without the appropriate parameterisation of the machine to take it in the desired direction...

For this reason, section 9.3 of the Copyright, Designs and Patents Act 1988 (CDPA) states: “In the case of a computer-generated literary, dramatic, musical or artistic work, the author shall be deemed to be the person who makes the arrangements necessary for the creation of the work”, specifying in Article 178 that “computer-generated”, in relation to a work, means that the work is computer-generated in such circumstances that there is no human author of the work. The protection lasts for 50 years from the date on which the work is carried out (Article 12.7 of the CDPA).

In relation to the last of the issues raised, on the protection of the human creations that feed the AI model, an interesting debate arises as to whether AI creators are allowed to feed the AI model with protected works. We understand “feeding” as providing the AI model with the necessary information to be conveniently “distilled” through various computer processes that compare the information and extract patterns from it. These activities are known as “text and data mining” (TDM).

From a legal perspective, the controversy arises because TDM requires the copying and extraction of protected information, which, in principle, are activities reserved to copyright holders.

Visions for solving this problem differ on both sides of the Atlantic.

The U.S. has chosen to apply the fair use doctrine, which has long been a crucial element in settling copyright disputes. In determining when a use of a protected work is fair, factors to be considered include: (i) the purpose and nature of the use being made, including whether it is commercial in nature or has a non-profit research purpose; (ii) the nature of the protected work; (iii) the amount or substantiality of the original work utilised; and (iv) the effect that the intended use may have on the normal exploitation of the protected work.

In order to understand the application of the rules defining use as “fair”, it should be noted that the four elements are considered as a whole, without the fulfilment of all or any of them being necessary for the use to be considered fair. Courts will have to consider all of them together and make decisions in the specific circumstances of the case. However, U.S. courts have placed particular emphasis on two factors: the purpose and nature of the use made of the protected works and the effect that the intended use may have on their normal exploitation.

On the first factor, they have repeatedly pointed out that search engines—which use TDM—are transformative in nature and, consequently, this factor has been a determining factor in considering their use to be “fair”. If we recall the rulings analysed, in *Authors Guild v. HathiTrust*, the court justifies the creation of a search engine on whole texts “because the result of a search for a word in a search engine is different in purpose, nature, expression, meaning and message from that of the page (and the book) from which such word has been extracted”. Thus, even if the search engine has a commercial use, its transformative nature tips the balance towards accepting the use of protected works as “fair”.

Regarding the second factor, the decisions to date suggest that this factor tends to favour the one who copies and incorporates the protected work into their database. They understand that the transformative nature of TDM makes its use unlikely to have an adverse effect on the natural market for the protected work for the simple reason that, given the nature of this technology, the protected work is unlikely to be replaced by the derivative.

This, however, is a position I do not share. It is acceptable in the area covered by the court decisions analysed —search engines— but it is not so obvious when we extend the application of the doctrine to AI such as Chat GPT. To understand that an AI model that distils knowledge from books is not going to affect the publishers who publish them is questionable. As mentioned above, it seems difficult for a student who has to do a paper and can ask Chat GPT to do so to choose the path requiring more effort, which requires access to documentation for reading, analysis and drafting.

In short, it seems that decisions on the transformative status of TDM incorporate “the right to read is the right to mine” doctrine. Phrase initially attributed to Peter Murray-Rust⁶⁶, an engineer and proponent of open data. When someone acquires a protected creation (e.g. a book) and reads it, he or she obtains knowledge. However, this does not entail an obligation to remunerate for the revenues obtained from that knowledge. Proponents of open data argue that the statistical process performed by AI is, in essence, similar to human rumination. The only difference is that algorithmic rumination is more reliable than human rumination. This is why both the fair use doctrine and the EU exceptions make it a condition that there must be lawful access to the protected work.

The decisions of U.S. courts, in application of the fair use doctrine, when they have understood that the TDM was very transformative and that it did not interfere with the normal exploitation of the incorporated work, are in line with Peter Murray Rust’s approach.

The increasingly settled assumption of this “doctrine” invites reflection on the consequences of its implementation. Accepting the doctrine without further ado could imply the acceptance of a monopoly in rumination. If human rumination can be replaced by autonomous and increasingly infallible logical processes that bypass human frailties and imperfections, the human capacity to influence the future becomes diminishing and the Frankenstein complex is resurrected. The growing influence of algorithmics, which could lead to “mathematocracy”⁶⁷ —algorithm rule— is becoming increasingly suspicious. Algorithmic influence, which reports to the digital industry, is of concern to policy makers and calls for governance models that avoid positions of dominance and put human development and influence at the heart of legislative initiatives. When so many intellectually sound people of proven expertise are sounding the alarm on this issue and calling for pause and reflection,

⁶⁶ Wikipedia.

⁶⁷ “The mathematocracy”, *El Español* online newspaper, 13 January 2017.

perhaps we should react. With that in mind, accepting once again the infringement of the rights of protected creation rights holders under the pretext of encouraging the development of knowledge, but losing control of that knowledge to a large extent and favouring positions of dominance derived from its disproportionate concentration, does not seem to be the governance option that is in the interest of humanity.

Crossing the Atlantic, the incorporation of the *sui generis* right into European Community regulations has obvious relevance to text and data mining by making it much more difficult to mine. In application of these regulations, AI developers are required to seek authorisations (licences) from the rights holders of the protected creations they wish to incorporate into their “AI model”. However, if the TDM technique utilised involves acts of extraction or copying outside the above exception scenarios, the restrictions laid down in the regulations shall apply. If the content has been extracted or copied, potential claims could be pursued in two ways: for copyright infringement and for infringement of the *sui generis* right of the rights holder of the creation. The two rights are independent of each other and can subsist together, at the same time and linked to the same creation. Extraction or copying without legal protection can infringe both rights at the same time, even if the rights holders are different.

However, the importance already attributed to this revolutionary technology makes it advisable not to lag behind. The U.S. position, which encourages investment in AI, and the absence of flexible instruments such as the fair use doctrine, oblige the EU legislator to make a move. To this end, Directive 2019/790/EC on copyright and related rights in the digital single market was promulgated in 2019, which incorporates legislative developments to address TDM activities. To this end, it incorporates two exceptions in Title II.

The first, which is mandatory, authorises copying and extraction for TDM activities for scientific and educational purposes on the sole condition that the person carrying out these activities has legitimate access to the protected material. The second, which is voluntary (it can be circumvented by contractual provision), applies where there is legitimate access to the protected material or where this material has been made available to the public “online”, and to the extent that rights holders have not adequately reserved the rights to make reproductions and extractions for text and data mining purposes. This second exception has no limitation in relation to the beneficiaries carrying out the TDM activities and, consequently, these extractions could be utilised for commercial purposes.

In the case of content made publicly available online, only the use of machine-readable media, including metadata and the terms and conditions of a website or service, is subject to reservation of rights. Other uses should not be affected by the reservation of rights for text and data mining purposes. Where content is made available to the public offline, other means of reserving rights, such as contractual agreements or a unilateral declaration, may be utilised. Licensees must allow rights holders to take measures to ensure that the established reservation is respected.