media LAWS
Law and Policy of the Media in a Comparative Perspective

Law and Media Working Paper Series

no. 3/2021

GIULIA MAZZINI

DEEP LEARNING WORKFLOW: A NEW CHALLENGE FOR EUROPEAN INTELLECTUAL PROPERTY

SUMMARY: 1. Introduction – 2. What is Deep Learning? A brief explanation for legal experts – 2.1. GANs (Generative Adversarial Networks) and GPT-3 (Generative Pre-trained Transformer) – 3. Machine - generated data: which kind of legal protection should be accorded? – 3.1. Copyright: a Romantic idea – 3.2. Deep learning-generated works: not purely data? – 4. Conclusion

1. *Introduction.*

For years now, in the broader field of machine learning, data scientists and programmers have developed intelligent computational systems which, starting from a collection data, run software that, whit the additional support of hardware extensions such as robots, emulates human cognitive functions like vision, prediction, speech recognition, automated language translation, artistic works and so on.

One of the most promising techniques in this arena is represented by deep learning, a subcategory of machine learning. The essential idea behind DL[1], is to enable machines to learn from thousands of examples within a particular context and thereafter build specific "internal model" from these examples, which in turn will be used to produce output when confronted with new imputs[2]. More specifically, DL as a mode of machine learning that utilises training algorithms, allow the machine

[1] Deep Learning, from here on out.

[2] Theodoros Chiou, *Copyright lessons on machine learning: what impact on algorithmic art?* 10 (2020) JIPITEC- Journal of Intellectual Property, Information technology and E-Commerce Law, 398 para 1.

Law and Media Working Paper Series

Codice ISSN: 2531-484X

EDITORE: Oreste Pollicino - COORDINATORE EDITORIALE: Marco Bassini

SEDE: Via Roentgen, 1 – 20136 Milano e-mail: submissions@medialaws.eu

to learn through the analysis of structured corpora of (big) data and, through their continuing use, improve their performance over time without being specifically programmed to so do: hence our conceptualization of "intelligent machines". While this method may appear complicated to the layman, researches may assert that it is easier to program a computer to *learn to be* intelligent than program a computer to be intelligent from the outset[3].

The development of DL systems has called into question the fundamentals of Intellectual Property Law in the European States, particularly in the context of copyright law. Indeed, especially within civil law jurisdictions, copyright law is concerned primarily with the protection of the author's personality – an entirely human conception – insofar as it informs his or her creative works, rather than the content of those works themselves. Consequently, this conception raises several issues. DL algorithms allow machines to learn how to produce artistic output, under the auspices of so-called algorithmic art, a genre encompassing everything from musical compositions and paintings to translated text and novels[4]. In so doing, algorithms are trained on given datasets created by humans, consisting of the type of works relevant to each project such as paintings or lyrics. As many scholarly articles persuasively argue, these training works which function as the starting point for any machine learning systems, are protected by copyright[5].
Nevertheless, the main issue pertaining to automated creativity has been whether copyright protection, or neighbouring rights, may be granted to DL-generated works, that is to say, the output. The question then follows of whether the "romantic author", that unique expression of individual talent and human creativity, is the only entity that may be afforded protection by copyright law, or whether – and if so, under what conditions- it may in fact extend to new expressions of creativity, built by automated tools fashioned by intelligent machines[6].

2. *What is deep learning? A brief explanation for legal experts.*

At very basic level, DL is a machine learning technique that permits the automated entity to learn from examples. It teaches a computer to filter inputs through layers to learn how to predict and classify information which can take the form of texts, sounds or imagens. By way of illustration, let us imagine the cognitive act of grasping a conceptual notion: first, we learn it and, immediately after, we are exposed to another. Our brains collect the input of the first and elaborate upon it with the aid of the second to which may be added a third and a fourth, the original notion being further abstracted with each interaction[7]. Scientifically speaking, it is correct to assume that the deep learning action constitutes the learning of data that are not furnishes by the programmers but rather, through the use of statistical computing – that is to say, algorithms. The scope of the latter is designed to emulate the human brain and its means of interpreting. The biological neuron is the computational paradigm that feeds deep learning with artificial neural networks. Finally, the hierarchical architectures of the data are read in the context of its previous experience, there after providing new and advanced levels of input/output.

---

[3] Take a look at Viktor Mayer-Schonberger researches collected in Oxford University articles.
[4] Theodoros Chiou, *Copyright lessons on machine learning: what impact on algorithmic art?* 10 (2020) JIPITEC- Journal of Intellectual Property, Information technology and E-Commerce Law, 398 para 1.
[5] Thomas Margoni, *Artificial Intelligence, Machine Learning and EU copyright law: who owns AI?* in CREATe Working Paper 2018/12, University of Glasgow.
[6] Giovanni Sartor, Francesca Lagioia, Giuseppe Contissa,*The use of copyright works by AI systems: art works in the mill.*
[7] *Cos è il deep learning? in https://www.intelligenzaartificiale.it*

To adopt more technical language, deep learning makes use of multi-layered artificial neural networks, a collection of trainable mathematical units – neurons – which collaborate to execute a complex function, in a manner loosely comparable to the human brain[8]. These neurons are organized in layers, stacked in a hierarchy of increasing complexity and abstraction, in which each layer learns from the layer below it. In this manner, as information ascends the ladder, acquainting the neural networks with patterns, then patterns to patterns and so forth, the data's sematic density is increased exponentially. In its uppermost layers, the networks have the capacity to understand quite abstracts concepts[9], ad algorithms gradually learn how to deploy additional data from their environment, adjusting their operation accordingly so as to create new algorithms and reach new autonomously derived outcomes.

This, then, is the resounding leap forward in the realm of "intelligent machines": the computer shall not be programmed but, rather, trained. The programmer's role is merely to feed the machine with new data; it is the province of the machine to learn, through its iteration with this data, new models, rules and norms that will inform and direct the machines' responses – even in the face of entirely novel data.

2. ***GANs (Generative Adversarial Networks) and GPT-3 (Generative pre- trained Transformer)***

There are several DL techniques in the play at time of writing, but the biggest stirs can be attributed to GPT-3 and GANs, as it is becomes increasingly evident that AI- enabled machine learning has the capacity to produce artistic works whose true creators are not humans, but machines[10].

The first example is the third iteration of the Generative Pre-trained Transformer, created by the artificial intelligence research laboratory OpenAI. GPT-3 is a language – generation model capable of producing, on demand, human – like texts of such quality that it is often difficult to differentiate from that written by humans. A language model is an artificial intelligence system that has been trained by an enormous corpus of text – GPT-3's algorithms study the statistical patterns in almost a trillion words collected from the web and digitized book – to the extent that, with enough text and sufficient processing, probabilistic connections between words are recognised, leanr and stored for future reference[11]

GPT-3's resulting data is significantly better than previous GPT models: for example, Microsoft's Turing NLG model can generate text at character – level accuracy on a test set of Wikipedia articles, but requires an enormous amount of training data to so do[12]. OpenAI claims that GPT-3 can achieve this level of performance without any additional training data after its initial pre-training period. After that, the machine learns by itself. Indeed, the training data does not include any information on what is a "right" or "wrong" response: all of the requirements is gathered from the training texts themselves. Let's imagine, for example, that during training, the algorithms

---

[8]Juergen Schmidhuber, *Deep Learning in neural networks: an overview, neural networks,* volume 61 pp 85-117, January 2015.

[9] Daniel Schonberger, *Deep Copyright and downstream questions related to AI and ML* in Droit d'auteur 4.0 / Copyright 4.0, DE WERRA Jacques (ed.), Geneva / Zurich (Schulthess Editions Romandes) 2018, pp. 145-173.

[10] Derrick De Kerckhove, *L'arte sottile del dottor macchina,* in Corriere Innovazione, november 2020.

[11]Theodore F. Claypoole, *New AI tool GPT-3 ascends to new peaks but prove show far we still need to travel,* in the National Law Review.

[12]Miguel Grinberg, *The ultimate guide to OpenAI's GPT-3 Language model,* August 8, 2020, in http:// www.twillio.com.

encounter the phrase "the cat plays with a X". It then scans all of the text in its training data to determine which word should be used to recreate the original phrase. Its initial results are likely to be erroneous potentially millions of times over, but eventually it will identify the right word. By caching its original input data, it will come to "know" that it has produced correct output and assign "weight" to the algorithmic process that generated the correct answer. The scale of this dynamic weighting process is what makes GPT-3 the largest artificial neural network ever created. The number of weights dynamically held in its memory and used to process each query is 175 billion – ten times more than its closest rival, produced by Nvidia[13]. The computing time on which this achievement depends is said to have cost OpenAI $4.6 million.

Generative Adversarial Networks, or GANs, which represent a shift in the architectural design of deep neural networks, are mainly used for generating images, though they also have the capacity to automatically create text. The generative model adopted by GANs creates new data instances which nonetheless resemble their training data and are thus of a broadly similar nature. Several advantageous features of distinguish this model from other DL techniques: the most salient being that it works and learns from a limited set of data – approximately 10% of the training data of other types of DL – thus greatly expediting the primary step of data collection by the machine's programmers[14]. This remarkable improvement derives from GAN's unique bifurcated structure, predicated on two neural networks – the generator and the discriminator – operating in competition whit one another. They interaction is essentially antagonist: the generator tries to fool the discriminator, while the discriminator tries to keep from fooled. In so doing, the generator learns to create apparently plausible data, which ultimately become negative training examples for the discriminator; meanwhile, the discriminator learns to distinguish the generator's fake data from the real, penalizing the generator for production implausible results. As training progresses, the generator becomes more adept at producing outputs that can fool the discriminator, while the discriminator's ability to distinguish the real from the fake deteriorates accordingly it starts to classify fake data as real, and its accuracy decreases. Therefore, the generator's output is connected directly to the discriminator's input and through backpropagation, a process by which the discriminator's classification produces a signal that is used by the generator to update its weight. Thanks to a well- balanced competition, both networks can improve their efficiency: the generator learns to develop more realistic data sets, while the discriminator learns to correctly identify both fake and real data[15].
It is notable that the popularity of GANs can be attributed primarily to their application beyond the arena of computer science. Three French students, operating under the collective moniker of Obvious, have created a much discuss artwork using precisely this kind of technology, *The Portrait of Edmund Belamy,* sold at Christie's auction housed in New York for $432,500 in October 2018. By the way of input, Obvious furnished the AI- enable machine with a dataset of 15,000 portraits painted between the XIV and XX centuries. What this example reveals in the extent to which the antagonist component constitutes the real turning point in the development of the DL works. In this instance the generator created a new imagine based on the input of the aforementioned human- made portraits which the discriminator subsequently tried to distinguish from those images creates by the generator. The goal is to deceive the discriminator into believing that the

[13] Bernand Marr, *What is GPT-3 and why is it revolutioninzing artificial intelligence",* in Forbes, October 5 2020.
[14] *Inside the GAN's architcture,* Packt_Pub in http://www.medium.com
[15] *GANs: il lato crativo dell'apprendimento automativo,* Digital Guide Ionos, by 1&1.

new computer- generated images ae in fact real portrait – the success of this ruse resulting in none other than *The Portrait of Edmond*[16].

3. *Machine- generated data: which kind of legal protection should be accorded?*

The *Portrait of Edmond* might be instrumental in the quest to answer the fundamental question of the property status attributable to machine-generated works by DL techniques. Who is the creator of the Portrait? The AI-enable machine? Or Obvious, the three French student's group? Though the issue may apparat convoluted to those not versed in mathematics or computer engineering, an understanding of the basic contours is intelligible to non-specialist and may in fact be easier to grasp than expected. The creator is the machine – a machine called, not surprisingly, *intelligence.* In fact, the machine reads the input, which is to say the training data (portraits in this case) and simultaneously implements the DL algorithms, step by step, in order to recognize and subsequently extract empirical observations, including those pertaining to technical and aesthetic elements such as lighting, colouration, brushstrokes and geometric patterns[17]. In this way, DL may culminate in a new set of rules, inferred and abstracted via the processing of training data, and subsequently augmented and fortified the processing of new internal models of the machine's own creation. This is new knowledge discovery, used by the machine to make automated, intelligent decisions regarding new and unknown future inputs which – crucially - were absent from the training data[18] furnished by the AI's human-programmers. This new knowledge, consisting of new- born inputs, is added to processed alongside the pre-determined input; together, these data inform the new set of rules saved by the machine and called upon to create new outputs, in this instance artistic works. If we follow the train of logic, we are left in little doubt that the machine is and intelligent entity capable of creating works of the imagination in its right.

Identifying a DL machine as the real creator of the *Portrait*, however, confronts us with a two-fold challenge. Should we consider the machine as the author, a designation granted by European Law legislation and thus accord the rights and protection that such a designation demands? Or, if not, should we then grant protection to DL works in a different way, thereby avoiding the possibility of such works ending up directly in the public domain?

3.1 *Copyright: a Romantic idea*

National Copyright laws, in both Europe and beyond, have been historically based on the centrality of the idea that authorship derives from individual human genius [19] [20]. This profoundly anthropocentric conception is structured around the belief that human beings are the sole source of

---

[16] Derrick De Kerckhove, *L'arte sottile del dottor macchina,* in Corriere Innovazione, November 2020.

[17] Andrea Guadamuz, *Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in the Artificial Intelligence Generated Works,* Intellectual Property Quarterly, 2017 (2). pp. 169-186. ISSN 1364-906X.

[18] Theodoros Chiou, *Copyright lessons on machine leaning: what impact on algorithmic art?* 10 (2020) JIPITEC- Journal of Intellectual Property, Information technology and E-Commerce Law, 398 para 1.

[19] Peter Jaszi, *Toward a theory of copyright: the metamorphoses of authorship,* 1991 Duke Law Journal 455-502 (1991).

[20] Michael Madison, *Beyond creativity: copyright as knowledge law.* Vanderbilt Journal of Entertainment and Technology Law, Vol. 12, p. 817, 2010, University of Pittsburgh Legal Studies Research Paper No. 2010-15.

the creativity manifested in original works that, as such, require the protection of copyright. Works are always generated by choices, intentions and expressions, in which the author infuses her own personality, creativity and, especially, her originality. Ate the moment, in which these ideas are expressed in a specific form, they become a right of the author, deserving to be granted protection under copyright law. Therefore, in light of this conception, and contrary to conventional wisdom, copyright may not in fact pertain to a human author's right of protection, but rather to the law of copyrightable works hinging on the idea-expression dichotomy: the proposition that copyright protection is accorded not to the idea itself, but rather its tangible manifestation[21] [22]. This could be the turning point in the development of the human conception of copyright.

In any case, starting from the aforesaid historical and traditional assumptions pertaining to the Romantic author, and recalling that European law does not furnish a clear definition of "authorship", it is hardly surprising that the European Court of Justice, in delivering the *Infopaq* [23] decision of 2009, interpreted "originality" and "creativity" as harmonised concepts deriving from "the author's own intellectual creation"[24]. This definition establishes a bi-univocal relationship between the act of creating a copyrightable work and a human acting as its creator[25]. Furthermore, the explanatory memorandum for the proposal of a Software Directive states: *"In common with all literary works, the question of authorship [...] is to be resolved in favour of the natural person [...] who created the work. Although the right to exercise exclusive rights may be assigned to another, the author will retain at least the unalienable rights to claim the paternity of his work".* It therefore appears impossible to apply conventional conceptions of copyright to DL generated works, in which the contribution of the human is limited to the initial phase of the creative process, namely, in preparation of the training data, the primary dataset input which is then augmented by machine itself[26].

As far as the challenge that DL-generated creative works poses to conventional notions of copyright is concerned, European law remains in a bind. At a national level, however, potential ways forward have been and are being identified. Indeed, UK law has adopted a solution that essentially grounds protection in the concept of deemed authorship, whereby the author of a computer-generated work is considered the person who *"makes the necessary arrangements"* for the creation of the work. In this way, depending on the circumstances, the owner of the copyright in machine-made such as those discussed here, may alternatively be the programmer or the user of the program. Again, section 178 of CDPA[27] contains a definition of the aforesaid works that can be adopted to establish a universal meaning of these for copyright purposes: "*the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken*."

---

[21] Massimo Maggiore, *Artificial Intelligence, computer generate works and copyright,* [2018] ELECD 878; in Bonadio, Enrico; Lucchi, Nicola (eds), "Non-Conventional Copyright" (Edward Elgar Publishing, 2018) 382.

[22] Daniela Simona, *Copyright or Copyleft: Wikipedia as a turning point for authorship,* (July 1, 2013). (2014) 25(1) Kings Law Journal 102, Available at SSRN https://ssrn.com/abstract=2330766.

[23] C-578 Infopaq International A/S v. Danske Dagblades Forening.

[24] Neither in Berne Convention

[25] Massimo Maggiore, *Artificial Intelligence, computer generate works and copyright,* [2018] ELECD 878; in Bonadio, Enrico; Lucchi, Nicola (eds), "Non-Conventional Copyright" (Edward Elgar Publishing, 2018) 382

[26] Fort the legal protection of the data collected by the machine's programmers, it could be said that they are protected, for the most part, by copyright or, specifically, by the directive 96/6/CE as a dataset. For a more detailed examination, however, please refer to the scholars as A. Guadamuz "*AI and copyright"* WIPO megazine; or Giovanni Sartor, Francesca Lagioia, Giuseppina Contissima *"The use of copyright works by AI system: art works in the data mill"*.

[27] Copyright, Design and Patents Act 1988.

Nonetheless, this definition, for the all the reasons already explained and more besides, is not without fault as far as the output generated by deep learning machines is concerned.
The European legislature is evidently aware of this issue and is proceeding to address it, with draft legislation formulated by the EU Commission concerning Artificial Intelligence, including the protection of the new, machine-generated creative processes delineated above, due in early 2021.

6. ***Deep Leaning – generated works: not purely data?***

Having excluded the possibility of protecting a work on the basis that creativity and originality are the sole province of the human mind, and instead positing that the copyrightable entity is not an idea's its content but rather its form, which is to say, the external structure of the idea, we shall direct our attention to other kinds of European protection available to creative works.
Once again, an understanding of the basic operation of DL techniques may facilitate our comprehension. The starting point of these mathematical architectures is the input dataset, through interaction with which the machine produces additional information to add to the original dataset, before ultimately emitting the final results – the outputs. The latter are nothing more than new data, organized in the systematic, methodical form of a real, bona fide database. This conclusion is applicable both to outputs such as *The Portrait of Edmond,* and to the production of texts, melodies, and suchlike, to which it is perhaps more easily understandable.
The Database Directive, 96/6/CE concerns the legal protection of databases in any form. For the purpose of this legislative provision (art.1), a database is a "*collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means*." In accordance with article 3.1, databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright. No other criteria shall be applied to determine their eligibility for that protection. The author of a database shall be the natural person or group of natural persons who created the set or, where the legislation of the Member States so permits, the legal person designated as the right holder by that legislation. The notion of the author's own "intellectual creation" serves as a criterion for the determination of the object of protection under copyright law. No database is copyrightable if its structure does not reflect the intellectual creation of its author or, more accurately, the selection or arrangement thereof. This copyright does not extend to the contents of databases, but only the structure, which may be protected by its own independent copyright, like a contract, or be copyright free.
While considering deep learning-generated works as collections of data, the application of copyright protection is once again subject to the anthropocentric conception of the author as a creator of an original work. However, article 7 of the Database Directive provides the *sui generis database right*: a new form of European intellectual property right formulated to protect economic investment in database when the latter is not original in the sense defined in article 3. The sui generis right may be applied to a database if a substantial investment was made in obtaining, verifying and presenting its contents. For this purpose, a substantial investment constitutes a financial and/or professional investment which may consist in the development of financial resources and the expenditure of time, effort and energy in obtaining and collecting the content. It is important to underline that the investment is not made in the creation of the collection of included data, but only "*in the obtaining, verification or presentation of the contest*" of the database. Indeed, while the business of "obtaining…contents" does concern the seeking out of existing independent materials and collecting them in the database, it does not cover the resources used for

the creation of materials that make up the contents of database[28]. The "verification", meanwhile, relates to the checking, correcting and updating of data already extant in the database, while "presentation" involves the retrieval and communication of compiled data, such as the digitalization of analogy files, the creation of thesaurus or the design of user interfaces[29]. This kind of protection gives the right to prevent the temporary or permanent extraction of all or substantial part of the database, as well as the right to prevent a reutilisation: this is to say, any form of making available to the public all or substantial part of the database. Thus, when copyright protects the original structure of the database, the sui generis right protects the contents themselves where the aforesaid elements are present. For this reason, in European Member States, both rights may apply cumulatively if the prerequisites for both regimes are fulfilled.

6. *Conclusion*

Deep learning has the capacity to produce high value outputs. The creation of potentially massive amounts of content produced either partly or totally independent of human hands, is challenging the marketplace in fundamental and unforeseen ways. And again, deciding whether machine-generated artistic materials, such as literature, melodies and so on, should be protected by copyright or, better, the sui generis right, has a profound impact on the market of creative works. This issue not only involves the programmers/makers of the machines under consideration, but also the market, and consumers, in general. Indeed, if DL-generated works are copyright-free, machines will produce free goods, available in the public domain, which can compete with their paid- for counterparts – that is, works created by humans expecting a financial return, thus also distorting the precarious state of concurrency.

Considering the current European legislation pertaining to copyright and the protection of database, there are two alternative options by which to grant protection to works generated by deep learning techniques. The first adopts an exception to, or more accurately, a reformulation of, the traditional concept of authorship, following on the heels of UK copyright law. In this instance, the author is designated as the person making the necessary arrangements for this kind of work to occur, and the investment their creation necessitates. However, while this solution may seem to offer the faster and easier way, it implies at least a legislative intervention into European regulations, if not a radical change in jurisprudence which remains beholden, as far as copyright is concerned, to the anthropocentric conception of authorship previously described. However, the outputs/works generated by deep learning have been in the market for some time now.
For this reason, suggesting this amount merely to collections of data, processed by a machine, and thus no more than a database, might open the door for the application of the sui generis right, provided for in Directive 96/6/CE, article 7. In the application of this type of technology, the economic investment furnished by the programmers or makers is ongoing, required not only to support the primary collection of input that initiates the machine's "intelligence work", but also to allow it to continue its function. Evidently, the labour of these AI-enabled machines does not end with the creation of a determinate number of outputs, be they images, novels, songs or other works, but may continue for a long time, increasing the number of the output potentially ad infinitum. We may think, for example, of the translation of text, on demand, by GPT-3 technology,

[28] CJEU, C-203/202.
[29] Eleonora Rosati, *Originaly in the EU copyright, full harmonization thought case law.*

an endeavour which might well continue indefinitely, as long as speakers of different language exist and need to communicate.
It is true, moreover, that the sui generis right is granted not to the database in its totality, namely, the whole structure, but only to the content housed therein, apparently without a form, though this aspect might be conveniently bypassed by an extensive interpretation. Some scholars, have already supported this solution [30], especially in the light of the jurisprudence of the Court of Justice of European Union, which is inclined to extend the definition of "database" with a view to ensuring grater legal protection[31].
At the basis of this proposed application is, firstly, recital 3 of the of the Database Directive, which describes ""existing differences distorting the functioning of the internal market need to be removed and new ones prevented from arising, while differences not adversely affecting the functioning of the internal market or the development of an information market within the Community need not be removed or prevented from arising" and, secondly, recital 40, which claims that "the object of this sui generis right is to ensure protection of any investment in obtaining, verifying or presenting the contents of a database for the limited duration of the right; whereas such investment may consist in the deployment of financial resources and/or the expending of time, effort and energy."

By way of conclusion, while waiting for a specific contribution by the European legislator, It is possible to consider DL generated works as various collections of data, arranged in different databases and, consequently, apply the sui generis right to the maker of the DL machine, circumventing, in the meantime, the question of whether or not to ascribe a legal personality to an AI machine. As a result, it is possible to guarantee a very similar right, comparable to the raw copyright, to artistic works on the premise that they might be created by a human e not by a machine. Secondly, this present itself as the solution for pursuing the objectives provided by the recitals of the Directive: avoiding distortions of internal competition and, above all, defend both the economic value of data and the investment made by the creator of the machine.

***

GIULIA MAZZINI – giulia.mazzini@master.unibocconi.it

DATE OF PUBLICATION: 10 MAY 2021

ABSTRACT: If we consider data as the oil of the current millennium, machine learning, especially through techniques like deep learning, may be concomitantly compared to an oil well. Intelligent machines, with supercomputing capacity, can learn from a corpus of examples embodied in a dataset. Through this, the machine – the oil well of our analogy – working with learning and training algorithms, uses the given dataset – the oil – in order to autonomously extract new data and, in so doing, predict or emulate human behaviour and decision-making. Thus, by drawing on a vast array of examples (inputs), programmers can create machines with the ability to make autonomous intelligent decisions (outputs), through the application of ever more sophisticated

[30] Alexandre Cruquenarie may be an example.
[31] Case C-490/14 concerning geographic data.

algorithms – all without the intervention of human minds. In this way, artificial intelligence has the potential to supplant and surpass the abilities of its human creators to generate new, valuable knowledge – including, somewhat remarkably, such imaginative works as paintings, literary texts, and melodies.

Through the analysis of the most recent deep learning-based generative models, this article seeks to explore what kind of legal protection, in the context of European Intellectual Property legislation, should be accorded to deep learning-generated works. Particular attention will be given to artistic creations created by two very current techniques, GAN and GPT-3. In so doing, this article addresses the possibility that works of this nature, belong to the domain of copyright, thereby subverting the anthropocentric conviction that only ideas engendered by human originality and creativity are capable of producing copyrightable work. Alternatively, these works, in the form of input, could be safeguarded by neighbouring rights, such as the European sui generis right required by Database Directive 96/9/CE. The analysis that follows investigates this problem in the context of EU copyright law, considering the deep learning workflow as creative data.

KEYWORDS: Artificial Intelligence – Machine Learning – Deep Learning – Copyright – Sui Generis Right