EJLT European Journal of **Law and Technology**

A Similarity Assessment in Copyright Works: The Insertion of Intelligent Technology to Provide Certainty to Rights Holders and the Public Sector

Jesus Manuel Niebla Zatarain[1]

Abstract

Intelligent technology is currently being used in the creation of artistic works. These devices offer new operational approaches such as the emulation of human cognitive processes required to develop a specific form of art along with the capacity to operate law compliantly. Nevertheless, there are situations where these implementations may find relevant material whose legal status cannot be properly defined. In the present article, a potential solution for this is explored through the use of statistical and syntactical tools, which will increase the legal operative efficiency presented by these devices.

Keywords: Copyright, Artificial Legal Intelligence, Similarity Assessment

1.1 Introduction: Law compliant behaviour in automated creators.

The increasing use of artificial intelligence in the creation of artistic works has reshaped the relation between technology and copyright law.[2] From a technical perspective, these devices operate by gathering features from existing works to use them in the generation of new materials. This has led to the development of a new operational feature: law compliant behaviour based on the legal status of the acquired work. In this scenario, artificial creators are capable to operate by adapting their behaviour to the legal status of the gathered material thus, avoiding copyright violations. Additionally, to increase legal accuracy the insertion of cognitive processes such as "rule of thumb" assessments is proposed in this work. This consists in detecting legally relevant environmental cues, which allows the device to establish the legal status of a work without implementing unnecessary reasoning processes. Additionally, should the situation require full legal reasoning, the device is capable to perform this operation.

Overall, lawful operation of automated devices depends on two elements: the rights contained in the electronic license related to the work. Second, environmental cues that deliver legal operative guidance in relation to a particular work. Nevertheless, intelligent creators have the potential to add a third element, one based on similarity approximation. This approach will be delivered throughout this article.

1.2 Developing Compliance for Automated Devices.

Automated creators operate on predefined instructions, which allow them to detect a particular work and gather specific features relevant for the creational process.[3] After this is established, the next stage is to provide these materials with the capacity to operate according to the legal status of the work, whether they are located on the digital or real life scenarios. In the second environment, the device operates based on the legal elements gathered from the physical setting. This can be delivered by implementing aestheticodes,[4] these are elements contained within art works that deliver operative instructions in machine-readable language. To illustrate this, a brief description of the computational representation is delivered in the following lines:

```
int output_index;
void draw(){
 background(255);

 char character = output[output_index];
 Letter l = character == 32 ? letters[26] : letters[character - 97];

 for(int i=0; i < num_panels; i++){
  fill(l.colors[i]);

  rect(left_margin, (screen.height - (panel_size)) / 2, panel_size, panel_size);
  translate(spacing + panel_size, 0);
 }

 output_index++;
 if(output_index >= output.length) output_index = 0;
}
```

As it is presented here, legal directives result in compatible with computational code. These provide operative instructions that allow the device to operate according to the rights contained on the license of a particular work. From a legal logical perspective, this approach can be translated into ontological[5] terms:

```
(input-found
  (if (license ?access and ?use)
  Letter l = character == 32 ? letters[26] : letters[character - 97];
    (license_description= access and use)
        (allow access input-found)
  else
    (exclude input-found))
```

In this example, the device detects a material located in an art exhibition. Before processing it, its legal status is defined by the rights contained in the electronic license of the work. Here, the assessment process allows the device to understand permits for such material, which in this case are access and use only, excluding any other types. This scenario illustrates how the rights contained in a particular work are delivered to the device, providing a first step towards lawful

access and management of copyrighted material. Furthermore, law compliant operation is presented through two specific actions, which leads to an adaptation of the behaviour of the device.

Nonetheless, there will be situations where given the relevance of a specific material its use is indispensable, regardless of the fact its legal status cannot be defined. For these types of situations, law compliance is based on gathering only inspirational elements. To illustrate this operation from a technical perspective, it will be addressed from the position of text based works.

1.3 Similarity in British Case Law

As mentioned above, this article proposes the implementation of a third method to establish the lawful use of artistic works in the creation of new materials. This is based on a similarity assessment process. While the concept is clear, its application, however, is more complex. Implications of taking advantage of it will differ depending on the jurisdiction that it is being addressed from. To facilitate its comprehension, this work will implement the position provided by the UK legislation. Here, the existence of illegal copying is defined by the amount of the expressed idea gathered from a particular material. Consequently, literary works deserve special attention since their reproduction may not necessary require the totality of an original work to be considered illegal. In relation to text-based works, these materials are composed by elements such as incidents and episodes, whose elements can later be used to form new works, this is referred as "pattern of the work".[6] This presents a major challenge in defining copyright similarity: the sequence of events contained in a protected material can be expressed differently simply by changing words. Additionally, this can be added to the context of a recently created work, increasing the difficulty to detect illegal use. Consequently, the adoption of intelligent technology in the creation of literary works needs to be complemented with the capacity to adapt to this new context in order to operate law compliantly.

To achieve this, the expansion of the cognitive legal module to include a similarity assessment process is proposed. This would allow not only to use a material according to its legal status but also to create a work that is sufficiently original to constitute an independent material. Moreover,

works whose legal status cannot be defined can still be used for inspirational purposes by acquiring only unsubstantial portions of them. This points out the importance that the emulation of human cognitive processes has, from an operation and legal perspective. In this sense, it delivers the grounds for a new type of collaboration between the law and intelligent technology, where legal compliance is seen as a design feature rather than a complementary one.

Following this, an important element is to identify and extract the cognitive processes implemented by human legal operators in courts to define similarity in litigation. To achieve this, a series of law cases regarding copyright infringement of literary works are analysed. As part of this, in *Ibcos Computers Ltd. v. Barclays Finance Ltd*[7] the court provided relevant parameters through which a detailed literary or artistic expression could be protected. It established that if the idea was expressed in general terms, even if it is contained in a work its use will not constitute an infringement of copyright. In *Ultra Marketing (UK) Ltd and another v Universal Components Ltd*[8] it was provided that if an idea was copied in a general and abstract way, it was less likely to produce copyright infringement. Nevertheless, the accumulation of such ideas and their disposition in a particular way can be considered substantial, and thus have copyright infringement consequences, as it has been mentioned in *Designers' Guild Ltd v Russell Williams*[9]. In *Baigent and another v Random House Group Ltd*[10] two of three authors of the book "The Holy Blood and the Holy Grail" stated that six chapters of this work were used to create the Da Vinci Code by Dan Brown. In this case, copyright infringement was dismissed since the work presented by Brown did not contain a chain of events like those depicted on the work of the claimant. It was also stated that regardless the level of similarity found within Brown's work, it was distributed on six chapters with no direct relationship between them. This brought the conclusion that they did not constitute infringement. In other situations, evaluation of similarity becomes a complex task, such as in *Allen v Bloomsbury Publishing Plc.*[11] Here, it was claimed that the book **"**Harry Potter and the Goblet of Fire" contained a high volume of similarities with an existing publication called "Willy the Wizard". The court searched for elements such as "the nature of the extended copying, the quality and importance of what had been taken, the degree of originality of what had been taken and whether a substantial part of the author's skill and labour in creating the original had been appropriated".[12] Again, the court ruled that the similarities founded were a product of similar ideas and they did not constitute sufficient evidence to

establish that the use of the expression had been taken. Nevertheless, this case mentioned heavy similarities that deserved to be pondered in a mini trial. Another relevant position can be found in *Hodgson v Isaac*[13], where the claimants stated that the defendant had used without their authorization a substantial part of an autobiography for a film script. This case measures the extension of copying by establishing that it was not based on volume but reflected the expression of intellectuality of their author. This position was reinforced when it was established that the script employed the same interpretations of the events than the book, which were situations that portrayed the personal life of the author of the work. Consequently, the judge stated that the film contained a substantial part of the original work thus, constituting illegal copying. An example of substantial copying can be found in *Ravenscroft v Herbert and New English Library Limited*[14] where the defendant claimed that he used the plaintiff's work "The Spear of Destiny" only as a source of inspiration and that the elements that were taken from such work were only historically thus, not violating copyright. For this case, it was held there had been substantial copyright copying covering from incidents, language of the characters and other relevant aspects. To decide whether illegal copying occurred, evaluation of evidence provided by the parties was needed, which confirmed the plaintiff's assumption. The next step was to measure to what extent was the original work used. It was found that more than sufficient volume of material that reflected the skill and labour of the plaintiff could be found in the defendant's work. Consequently, the court stated that it was not used as an inspiration but rather as a source of material that was later transferred to the work in question. Based on this, it was decided that the defendant had used substantial parts of the plaintiff's work to create the prologue of the material. Additionally, relevant incidents and description of the story can be found throughout the work. An interesting position by the Judge can be found here:[15]

> "In the case of works not original in the proper sense of the term, but composed of, or compiled or prepared from materials which are open to all, the fact that one man has produced such a work does not take away from anyone else the right to produce another work of the same kind, and in doing so to use all the materials open to him. But as the law has been precisely stated by Hall V.C. in Hogg v. Scott, `the true principle in all these cases is that the defendant is not at liberty to use or avail himself of the labour which the plaintiff has been at for the purpose of producing his work, that is, in fact, merely to take away the result of another man's labour or, in other words, his property'".

In *Harrison and Harrison*[16] two editions of a book entitled "How to Avoid Paying Care Charges" were assessed. Here, the claimant who wrote the first edition stated that a substantial part of his work had been used without his authorization in the second edition. In order to evaluate whether there had been an infringing use, both works were compared and it was decided that the author of the second edition had taken a sufficient portion of the first volume (11% verbatim) to constitute infringement. A similar position can be found in *JHP Ltd v BBC Worldwide Ltd.*[17] In this case, the author had created a character originally for a TV show; he later had an agreement with a production company to create three books based on it. Before the author's death, he kept on collaborating with both parties, and when he died, the publishers looked for an agreement with the TV company to create a new book based on this character. Nevertheless, a license was obtained by the defendant by estoppel from the author's estate. This provided sufficient permission to keep working with the author's character without infringing copyright. However, the court also mentioned that even in the event that such permission was not granted, the extension of the material used was not large enough to constitute copyright infringement. This was because the defendant either used little portions from the books, or already has this information in the form of TV scripts. By this, the court stated that the claim for copyright infringement was not sustained.

Related to the use of technology to access copyrighted material, *Newspaper Licensing Agency Ltd v Meltwater Holding BV*[18] provides a relevant position. Here, the claimant stated that by using media monitoring devices provided by a third party to access headlines and extract data from newspaper websites, copyright infringement was caused and that an end-user license was needed to receive such service lawfully. In this case, a technological device would extract specific information from a newspaper's electronic pages to deliver precise information to its users. The defendants monitored media services from a number of websites through a process that extracted relevant data that matched the parameters established by their clients. In this case, three elements were assessed by the court: first, whether a newspaper headline was capable of being free-standing original literary work, second, if the information extracted was to be considered a "substantial part" of the work and third, if a public relations company and its members were required to have a web end-user license to use and receive the service. Here, the claimants claimed that the creation of newspaper headlines demanded sufficient labour and skill

to be considered subject of copyright protection. It was also stated that the extracts gathered from the text were too specific to be considered general and that once the customers had received them they might proceed to generate unauthorized copies, which constitutes copyright violation. Finally, the court ruled that copyright subsists in headlines and short extracts of articles can be considered a "substantial part" of an artistic work. Following this, in the case of *Public Relations Consultants and Association Ltd v Newspaper Licensing Agency Ltd*[19] a decision ([2011] EWCA Civ 890, [2012] Bus. L.R. 53) that demanded a web-end user license to use and receive content from headlines news related site was appealed. The appellant stated that the service he provided was based on monitoring news on behalf of their clients. This was delivered by a group of companies that used automated software to operate based on specific details and the user's preference. Then, a notification related to the findings was sent to the client, either by email or through the company's web page. This brought the question of whether this temporary storage of copyrighted material constituted an infraction, the court ruled that it does not, based on the notion that temporary copies made for the purpose of browsing did not constitute infringement.

From these cases, the process of measuring similarity in literary works can be summarised as follows: first, to analyse the volume of material gathered, quality is considered more important than quantity. Second, the volume of material is used in a way that resembles the expression of the original work. Third, there has to be the intention of using this material in an identical way as the original one.[20]

From the analysis of the cases, it can be inferred:

- Copyright protects expressions, rather than mere ideas,
- Ideally speaking, copying an idea is permitted,
- Inspiration of an existing work is permitted, only if it provides general and unsubstantial inspiration,
- Nevertheless, if that idea has been expressed already, a high level of resemblance can lead to copying, which avoids creational labour, and
- The law provides general guidelines about what copying is, but it does not offer the parameters to perform an effective measurement.

From this, it can be stated that UK case law does not provide a definitive method to perform a similarity assessment, it merely leaves it to the particularities of each case, allowing the legal operator to implement the process he considers proper according to the characteristics of the case. This creates a gap where key legal terms such as "*substantial copying*" are never defined in terms of quantity, varying greatly from case to case, leaving legal technological approaches without a proper method to apply.

Under this situation, judges tend to rely on the content of previous cases, basing their operations in finding patterns within already established decisions. This method however, is not suitable to copyright enforcement efforts, among the reasons for this there are:

- Most cases do not get published, given the fact they are settled outside the court.
- Those cases that are published present the outcome delivered by the judge. Nevertheless they do not provide the process implemented to elaborate such outcome.
- Court decisions usually state only whether there was illegal copying or not.
- In the potential scenario where machine-based methods are used to find similarity, both works need to be present.

This creates a situation where the development and implementation of a similarity assessment process will not only benefit private users but it will also assists human judges in cases related to it.

2. The Role of Inspiration in the Creation of Artistic Works

The quantification of copying has become an element, which by its own nature, is very difficult to define. The inspirational aspect allows a person to lawfully use the material of another author to create a new work. This gives the notion of idea a prevalent position: through its assessment originality can be established between similar works. Nevertheless, the law does not provide a process to define illegal copying. This leads to scenarios where high resemblance could be

solved by the adoption of the method proposed on this research. To illustrate this, take the example of a a story of a wealthy man who is happy using his money to help people in a particular city and time. Another author can practically have the same idea, but changing drastically the notion by adding, "he is not happy" (idea-inspiration). This example provides that, when closeness between terms is very high, defining similarity becomes a more complex process. In this example, it evolves from being a mere process of counting words, to one that requires syntactical elements to define the intended meaning of the material (idea-context).

2.1 Similarity and Asymmetry in AI, Computing Text Similarity

To this point, the benefits offered by the adoption of a similarity assessment process to the automated creation of artistic works have been provided. In this sense, lawful management of materials whose license cannot be established is important. This is provided through the implementation of a method that allows gathering general, yet relevant features of a particular work. The described process requires the capacity to distinguish between idea and expression, a method that varies given the extension of the material. The proposed component is, in summary, a similarity module capable of delivering the level of resemblance between two text works.[21] This is performed by prioritizing the semantic meaning among words, whilst similarity between generic concepts is relegated.[22] Furthermore, this method is capable of detecting the intended meaning of a word based on the context in which it is being used.

Due to the nature of the approach presented in this work, the use of Latent Semantic Analysis (LSA)[23] is proposed. This method operates by assessing the content of text-based works, delivering resemblance at context level rather than mere word-by-word meaning. Lastly, this delivers the level of resemblance between materials that appear unrelated at a superficial level.

In relation to its operative performance, LSA is capable to operate in the following situations:[24]

1. LSA was assessed as a predictor of query-document topic similarity judgments.

2. LSA was assessed as a simulation of agreed-upon word-word relations and of human vocabulary-test synonym judgments.

3. LSA was assessed as a predictor of text coherence and resulting comprehension.

4. LSA was assessed as a simulation of word-word and passage-word relations found in lexical priming experiments.

In a case scenario, this method provides control of syntactical features such as management of mimic synonym, antonym, singular-plural, and compound-component word relations, aspects of some classical word-sorting studies, to simulate aspects of imputed human representation of single digits; and to replicate semantic categorical clustering of words found in certain neuropsychological deficits.[25] Consequently, LSA delivers a statistical outcome, in which the resemblance of words contained in different materials is used to calculate the degree of similarity, thus allowing to state the potential existence of copying. This is achieved by implementing only raw text distributed into words defined as unique character strings and separates them into meaningful passages or samples such as sentences or paragraphs. Then, matrixes containing word counts from paragraphs (shown as rows and columns) are created from large pieces of text and by using singular value decomposition it is possible to add words and present similarities. The next step is to implement singular value decomposition (SVD), which then decomposes the original matrix that contains the elements of the original text into the three new ones.[26]

Fundamentally, this method decomposes words contained in paragraphs, measuring them with the context and volume presented in the literary work.[27] After this, similar words are detected, their context analysed, and based on this, the level of likeness among terms is defined. The implemented approach is based on a corpus-based method,[28] which delivers the capacity to evaluate sections of the work rather than only large portions, without compromising the accuracy of the outcome. Landauer, T. K *et al* (1998) explain this process: "One component matrix describes the original row entities as vectors of derived orthogonal factor values, another describes the original column entities in the same way, and the third is a diagonal matrix containing scaling values such that when the three components are matrix multiplied, the original matrix is reconstructed." Essentially, the corpus-based method proposed on this work delivers a framework that accurately measures the grammatical and syntactical resemblance of the

expression contained in two pieces of text. This feature allows distinguishing between potentially unwanted similitudes and anticipated ones, such as parody.

Consequently, this provides legal certainty for both, the rights holders and the users of these materials. For the rights holders, it assures that the work will be used according to the rights provided on the license. From the perspective of the user of automated technology, it guarantees that the legal status of relevant materials will be used as operative guidance.

2.2 Analysing Similarity through Technology: Corpus Based Method

Following the position provided so far, the insertion of an efficient and accurate assessment method in automated creators of literary works becomes a priority. In this scenario, the core of this process is performed through Latent Semantic Analysis (LSA), which addresses key semantic elements such as synonymy and polysemy.[29] These two features have a direct effect in the development of text similarity assessment systems.

Summarizing, LSA mechanics is based on four steps:[30]

1- *Term-Document Matrix:* This represents a large collection of text. Rows and columns are contained here; individual cell entries will mention the frequency with which a term appears in a document. Alongside, sections of text considered large are assessed in order to define whether they contained a high volume of resemblance that may lead to illegal copying.

2- *Transformed Term-Document Matrix:* The terms contained in the previous section are transformed. To enhance the performance of this process, frequencies are cumulated in a sublinear fashion. Essentially, instead of counting every time a word appears in a text, a value is allocated that is later used assess similarity between works.

3- *Dimension Reduction:* Singular Value Decomposition is implemented to retain the largest singular values. This provides a proper reproduction of relevant rates contained in the

original matrix. Here, a vector represents each document and term. After this, relevant elements are independently separated and later used to provide tendencies to illustrate term similarity in corpus-based text.

4- *Retrieval in Reduced Space:* In this stage, only those elements with the highest similarity levels are used, discarding those with the lowest levels. Those terms are used to create a matrix that contains only relevant elements. Here, evaluation methods based on precise features are performed such as term-term or term-document. Additionally, these elements can be used to develop new assessment parameters based on, for example, location of the text or distance between words. This method is important to assess similarity according to specific terms or locations of the text.

At this point, the implementation of similarity assessment through LSA has been presented as a solution to deliver legal certainty related the use of literary works, even in those situations where the license status is unknown. The reasons for this are twofold: first, the growing tendency of the literary industry to use artificial creators to develop their materials. Second, from an operative perspective, these devices process large volumes of potentially protected works, which requires the capacity to manage them according to their legal status.

Relevantly, this approach is compatible with the cognitive features presented in artificial creators, which facilitates its insertion to these devices. Latent Semantic Analysis will be complemented with a two-mode factor analysis, based on Singular Value Decomposition (SVD).[31] In this scenario, large volumes of data are processed to detect semantic likeness without direct human intervention. This method aims to decompose the matrix created through LSA into smaller ones containing the most relevant findings thus, facilitating the assessment process. This increases the accuracy of the outcome even delivering the area of the corpus where it is located. Further assessment processes are performed based on the cohesion of words that remained on the paragraphs, thus defining the degree of similarity between the analysed texts. This allows to focus on the actual construction of the assessed work, giving importance to the conformation of the literary description contained there.

Due to the above, the combination of LSA with SVD was adopted as the method to perform similarity assessment processes in text-based works.[32] This approach is suitable of being inserted to the digital environment, given its capacity to operate as an element in context-based queries.[33]

3. Delivering an Assessment Method to Automated Creators of Text-Based Works

Automated creators of literary works[34] require an accurate assessment process that allows automated creators to extract relevant yet general elements from material considered relevant in a law compliant way. To avoid similarity related issues, the implementation of an approach based on Latent Semantic Analysis (LSA)[35] is presented in the following form: [36]

A first step consists in analysing the pieces of text that are submitted to the semantic space, choosing those that are likely to present certain degree of resemblance. Here, material from the same area of interest should be preferred. The following step addresses the number of factors, which ideally must be from 50 to 300 to produce an accurate outcome. Once the previous is established, a matrix comparison method is performed. At this stage of the process, relevant terms between the two texts are represented in the main matrix. Then, through Singular Value Decomposition (SVD), such matrix is decomposed into three smaller ones, providing the accurate role of each relevant term. To illustrate this, three paragraphs from Hansel and Gretel[37] are used:

- Text 1: Hansel and Gretel are children whose father is a woodcutter. When a great famine settles over the land, the woodcutter's abusive second wife proposes to take the children into the woods and abandon them there, so that she and her husband will not starve.

- Text 2: Hansel and Gretel is the story of two German children whose father is a woodcutter and they discover a house made of confections in the woods near their house. The children are taken into the woods to be abandoned.

- Text 3: Hard by a great forest dwelt a poor wood-cutter with his wife and his two children. The boy was called Hansel and the girl Gretel. He had little to bite and to break, and once, when great dearth fell on the land, he could no longer procure even daily bread so he took them to the woods.

The evaluation of these three texts is provided in the following chart:

Document	Text 1	Text 2	Text 3
Text 1	1	0.74	0.83
Text 2	0.74	1	.84
Text 3	0.83	0.84	1

The results provided by this chart deliver a high level of resemblance between Text 2 and Text 3. Under these conditions, coincidental likeness is very unlikely to occur. In the case of Text 1 and Text 2, the resemblance can still be considered high, however it is not as definitive as the on first case. Ideally, this operation will provide a range between 0 and 1, considering similarity ranges from 0.7 to 1 as too close to be inspirational. [38]

Finally, this method compares two text materials and provides a syntactical analysis through which close resemblance is analysed. Furthermore, this method presents a statistical method that measures similarity between pieces of text, providing a suitable tool to establish whether plagiarism has occurred or not.

3.1 Implementing Latent Semantic Analysis into Technological Legal Devices

As mentioned earlier, legal certainty has become an important requirement in the design of automated generators of text-based works. This provides legal operative compliance by using relevant sections of a work according to the rights contained in its license. If no electronic tag is found the device proceeds to use the work only for inspiration purposes. As is possible to infer, this method implies an adaptation of the behaviour of the device, which is defined by the legal

status of the work. Consequently, the combination of legal reasoning and similarity assessment delivers a new and relevant solution to copyright infringement related to the use of literary material by automated creators.

3.1.1 Increasing Legal Certainty: The Potential Implementation of CBR Technology

In order to provide an additional element of certainty, legal cognitive processes are complemented with case-based reasoning technology (CBR). This method provides a reasoning scheme with the capacity to present conclusions from comparing relevant legal cases with the input provided. [39]

In relation to the composition of CBR, three operative elements are found: [40]

- *Retrieve:* It gathers cases from the legal database that share similar features and that offer a potential solution for the problem. Here, elements related to illegal copying deliver relevant aspects of similarity assessment.

- *Reuse:* The closest findings gathered from cases are analysed to be used as a potential solution. Due to the nature of the law, complete correlation is unlikely to be found in copyright cases. Still, this can provide relevant elements to construct a relevant outcome.

- *Revise:* It performs a final assessment between the characteristics of the current case and those gathered from the legal database to guarantee matching.

- *Retain:* The new outcome is stored for future use, enhancing the accuracy of the legal database. In parallel, this will improve the quality of copyright law related operations, whilst at the same time operational efficiency is achieved through a better management of computational resources.

Additionally, CBR[41] is capable of managing large concentrations of data, which bonds perfectly with the structure of case law. In this scenario, relevant material requires to be gathered from voluminous datasets, which can be obtained through the CBR's operational features.

Furthermore, this technical approach adopts the specifications of the case and uses them as the parameters on which the search process is based.[42] Consequently, it will provide a legal assessment method based on the particularities of the situation and its potential impact on the legal framework. This is by extracting legal knowledge for weighting purposes: legal reasoning will allow the device to decide according to the relevance of the features of the case (resemblance between works), which is the optimal course of action.[43] In this scenario, the interpretation of the outcome is complemented by elements contained in relevant legal cases. This provides a solid position to define the legal status of the recently created work. Finally, it can be stated that the inclusion of CBR technology functions as a complement to the assessment process performed by the device.

4. Conclusions

Intelligent technology is currently being used by several industries in the creation of artistic works.[44] Its method of operation replicates the cognitive processes implemented by human authors to create a specific type of material. Consequently, these devices are capable of providing works that resemble the quality and expression of those created by human authors, making them practically indistinguishable for the average consumer. This has led to the reconsideration of the role that technology has in relation to the generation of art.In this sense, the technical operational implemented by these devices makes the inclusion of law compliance operation a requirement that needs to be present during the design of the device. This will allow these devices to accommodate their behaviour according to the legal status of the detected material. However, there will be situations where this becomes a complex task, specifically when a relevant work does not contain a license. In this scenario, the addition of a similarity assessment method is proposed. This will allow measuring the volume of data acquired from the work and also the amount of expression contained thus, avoiding copyright infringement. Consequently, it means s legal compliance is achieved from a perspective that does not involve sacrificing computational efficiency.

Finally, this method will allow the use and dissemination of automated technology to create artistic works law compliantly, reducing the possibility of copyright violations. Overall, the

positions presented in this work will have an undoubtedly positive effect on both the public and the private sectors.   It not only avoids unnecessary litigation and legal procedures, but Sit also provides a platform that operates by understanding the actions an author has authorised in relation to his work. This also includes situations where no license can be found, using them only as inspirational sources.

REFERENCES

Books

Forsythe, G. E., Moler, C. B., & Malcolm, M. A. (1977) *Computer Methods for Mathematical Computations* (Prentice Hall).

Kolodner, JL(1993), *Case-Based Reasoning* (Springer Science Business Media).

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013) *Handbook of Latent Semantic Analysis* (Psychology Press).

McCorduck, P, (1991) *Aaron's code: Meta-art, Artificial intelligence, and the Work of Harold Cohen*. (Macmillan).

Marshall, J., Grimm, J., Grimm, W., Grimm, J., & Grimm, W. (2005). *Hansel and Gretel*(Weston Woods Studios).

Reiter, E., Dale, R., & Feng, Z. (2000) *Building Natural Language Generation Systems*(Cambridge: Cambridge university press).

Conference Proceedings

Berman, D. H., & Hafner, C. D. (1993), 'Representing teleological structure in case-based legal reasoning: the missing link'. In Proceedings of the 4th international conference on Artificial intelligence and law. ACM. June 1993.

Brüninghaus, S., & Ashley, K. D. (2001), 'The role of information extraction for textual CBR' In International Conference on Case-Based Reasoning. Springer Berlin Heidelberg.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988), 'Using latent semantic analysis to improve access to textual information' In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM. May 1988.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., & Ruppin, E. (2001), 'Placing search in context: The concept revisited', In Proceedings of the 10th international conference on World Wide Web. ACM. May 2001.

Gabrilovich, E., & Markovitch, S. (2007), 'Computing semantic relatedness using Wikipedia-based explicit semantic analysis' In International Joint Conference of Artificial Intelligence (Vol. 7). 1606-1611, 1992.

Hofmann, T. (1999), 'Probabilistic latent semantic indexing' In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 50-57, August 1999.

Laham, D. (1997). 'Latent semantic analysis approaches to categorization'. In Proceedings of the 19th annual conference of the Cognitive Science Society. 979. 1997.

Mihalcea, R., Corley, C., & Strapparava, C. (2006). 'Corpus-based and knowledge-based measures of text semantic similarity'. In *AAAI* (Vol. 6, pp. 776) July 2006

Rissland, E. L., & Skalak, D. B. (1989), 'Case-based reasoning in a rule-governed domain' In Proceedings of the Fifth Conference of Artificial Intelligence Applications, March 1989.

Rissland, E. L., Valcarce, E. M., & Ashley, K. D. (1984), 'Explaining and arguing with examples'. In Proceedings of the Fourth AAAI Conference on Artificial Intelligence. AAAI Press, August 1984

Schone, P., & Jurafsky, D. (2000), 'Knowledge-free induction of morphology using latent semantic analysis' In Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning-Volume 7. Association for Computational Linguistics. September 2000.

Shen, X., Tan, B., & Zhai, C. (2005), 'Context-sensitive information retrieval using implicit feedback'. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM. August 2005.

Turney, P. (2001). 'Mining the web for synonyms: PMI-IR versus LSA on TOEFL'. In Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001).

Watson, I., & Gardingen, D. (1999), 'A distributed case-based reasoning application for engineering sales support' In International Joint Conference of Artificial Intelligence 1999.

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (2004). 'Latent semantic analysis'. In Proceedings of the 16th International Joint Conference on Artificial Intelligence. November 2004.

Cases

United States

Mazer v. Stein, 347 U.S. 201 (1954); *Peter Pan Fabrics, Inc. v. Martin Weiner Corp.,* 274 F.2d 487 (2d Cir. 1960).

United Kingdom

Allen v Bloomsbury Publishing Plc [2010] EWHC 2560 (Ch); [2010] E.C.D.R. 16; Official Transcript; Ch D; 14 October 2010.

Baigent and another v Random House Group Ltd - [2007] EWCA Civ 247

Designers Guild Ltd v Russell Williams Textiles Ltd [2001] 1 All ER 700, [2000] 1 W.L.R.

Harrison v Harrison [2010] EWPCC 3; [2010] E.C.D.R. 12; [2010] F.S.R. 25; Official Transcript; PCC; 19 March 2010.

Hodgson v Isaac [2010] EWPCC 37; [2012] E.C.C. 4; Official Transcript.

Ibcos Computers Ltd. v. Barclays Finance Ltd [1994] F.S.R. 275

Jarrold v. Houlston (1857) 3 K & J. 708.

JHP Ltd v BBC Worldwide Ltd [2008] EWHC 757 (Ch); [2008] F.S.R. 29; [2009] Bus. L.R. D1; Official Transcript.

Newspaper Licensing Agency Ltd v Meltwater Holding BV [2011] EWCA Civ 890; [2012] Bus. L.R. 53; [2012] R.P.C. 1

Public Relations Consultants and Association Ltd v Newspaper Licensing Agency Ltd [2013] UKSC 18; [2013] 2 All E.R. 852

Ravenscroft v Herbert and New English Library Limited [1980] R.P.C. 193

Ultra Marketing (UK) Ltd and another v Universal Components Ltd - [2004] EWHC 468 (Ch), Transcript.

Journal Articles

Ashley, K. D. (1992), 'Case-based reasoning and its implications for legal expert systems' *Artificial Intelligence and Law*, 1.

Bench-Capon, T., & Sartor, G. (2003), 'A model of legal reasoning with cases incorporating theories and values' *Artificial Intelligence* 150.

Berry, M. W. (1992), 'Large-scale sparse singular value computations' *International Journal of Supercomputer Applications*, 6(1).

Boroughf, B. The Next Great Youtube: Improving Content ID to Foster Creativity, Cooperation, and Fair Compensation'(2015). *Albany Law Journal of Science and Technology*, *25*, 95.

Burgess, C., Livesay, K., & Lund, K. (1998). 'Explorations in context space: Words, sentences, discourse' *Discourse Processes*, 25.

Chafee, Z. (1945). 'Reflections on the Law of Copyright: I'*Columbia Law Review*, 45(4).

Colton, S. (2012), 'The painting fool: Stories from building an automated painter', In *Computers and Creativity*. Springer: Berlin Heidelberg.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990), 'Indexing by latent semantic analysis', *Journal of the American society for Information Science*, 41(6).

De Mantaras, R. L., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S. & Keane, M. (2005), 'Retrieval, reuse, revision and retention in case-based reasoning', *TheKnowledge Engineering Review*, *20*(03), 215-240.

Dumais, S. T. (2004). 'Latent semantic analysis', *Annual Review ofInformation Science and Technology*, 38(1).

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001), 'Eigentaste: A constant time collaborative filtering algorithm', *Information Retrieval*, 4(2).

Hahn, U., & Chater, N. (1998), 'Understanding similarity: a joint project for psychology, case-based reasoning, and law', *Artificial Intelligence Review*, 12 (5), 404.

Hofmann, T. (2004), 'Latent semantic models for collaborative filtering',ACM Transactions on Information Systems (TOIS).

Hofmann, T. (2001), 'Unsupervised learning by probabilistic latent semantic analysis' *Machine Learning*.

Ip, H. H., Law, K. C., & Kwong, B. (2005, January). 'Cyber composer: Hand gesture-driven intelligent music composition and generation'. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*. IEEE.

Komuves, D., Niebla, J., Schafer, B., & Diver, L. (2015). 'Monkeying around with copyright: animals, AIS and authorship in law'. *Jusletter-IT*, 26, RZ1-RZ27.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998), 'An introduction to latent semantic analysis',*Discourse Processes*, 25.

Landauer, T. K., & Dumais, S. T. (1997). 'A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge', *Psychological Review*, 104(2).

Leake, D. B. (1996) 'CBR in context: The present and future', In D. B. Leake (Ed.), *Case-based reasoning: Experiences, lessons, & future directions*. Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press.

Nimmer, M. B. (1969), 'Does copyright abridge the first amendment guarantees of free speech and press' *UCLA L. Rev.,* 17.

Pirolli, P. L., & Anderson, J. R. (1985), 'The role of learning from examples in the acquisition of recursiveprogramming skills',*Canadian Journal of Psychology/Revue canadienne de psychologie,* 39(2).

Skalak, D. B., & Rissland, E. L. (1992), 'Arguments and cases: An inevitable intertwining',*Artificial intelligence and Law*, 1(1), 3-44.

Wang, A. (2006),'The Shazam music recognition service', *Communications of the ACM*, *49*(8).

Watson, I. (1999) 'Case-based reasoning is a methodology not a technology', *Knowledge-based Systems* 12.5, 304.

Xenakis, I., (2001), 'Formalized Music: Thought and Mathematics in Composition' (Harmonologia Series No.6), Hillsdale, NY: Pendragon Press.

Electronic References

LSA @Boulder Last accessed August 17 2017. Retrieved from: http://lsa.colorado.edu/

Microsoft,'Compare and Merge Two Versions of a Document', last accessed February 22 2018. Retrieved from: https://support.office.com/en-us/article/compare-and-merge-two-versions-of-a-document-f5059749-a797-4db7-a8fb-b3b27eb8b87e

Aesthethicode.org(2012). Last accessed February 20 2018. Retrieved from: http://aestheticode.org/

Ings S. (2016, September 21st),'When art and Technology pull each other to bits', New Scientist. Retrieved,https://www.newscientist.com/article/2106665-when-art-and-technology-pull-each-other-to-bits/. Last accessed 16 February 2018.

---

[1] Professor and Researcher at the Faculty of Law, Autonomous University of Sinaloa (Campus Mazatlan) Mexico. Email: jmniebla@gmail.com

[2] For a relevant position about this see: Komuves, D., Niebla, J., Schafer, B., & Diver, L. (2015). 'Monkeying around with copyright: animals, AIS and authorship in law'. *Jusletter-IT*, *26*, RZ1-RZ27.

[3] For a view on the different approaches implemented to generate artistic works through automated devices see: Ings S. (2016, September 21st). When art and Technology pull each other to bits. New Scientist. Retrieved https://www.newscientist.com/article/2106665-when-art-and-technology-pull-each-other-to-bits/. Last Access 16th February 2018; Reiter, E., Dale, R., & Feng, Z. (2000). Building natural language generation systems. (Cambridge: Cambridge university press). Xenakis, I., (2001). Formalized Music: Thought and Mathematics in Composition (Harmonologia Series No.6). Hillsdale, NY: Pendragon Press; Ip, H. H., Law, K. C., & Kwong, B. (2005, January). Cyber composer: Hand gesture-driven intelligent music composition and generation. In Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International. IEEE; Colton, S. (2012). The painting fool: Stories from building an automated painter. In Computers and creativity Springer Berlin Heidelberg and McCorduck, P, (1991). Aaron's code: meta-art, artificial intelligence, and the work of Harold Cohen. (Macmillan).

[4] Aesthethicode.org (2012). Last access 16th February 20th 2018. Retrieved from: http://aestheticode.org/.

[5] The term *ontological* is used here to refer the description of a legally relevant action in logic terms, in this case, lawful access of an artistic work through.

[6] Chafee, Z. (1945). 'Reflections on the Law of Copyright: I' *Columbia Law Review*, 45(4), 513. A decision related to this can be found in *Designers Guild Ltd v Russell Williams Textiles Ltd* [2001] 1 All ER 700, [2000] 1 W.L.R. An extension of this position can be found in: Nimmer, M. B. (1969). 'Does copyright abridge the first amendment guarantees of free speech and press' *UCLA L. Rev.,* 17, 1190. An additional interesting early position can be found in the US case: *Mazer v. Stein*, 347 U.S. 201 (1954); *Peter Pan Fabrics, Inc. v. Martin Weiner Corp*., 274 F.2d 487 (2d Cir. 1960).
[7] *Ibcos Computers Ltd. v. Barclays Finance Ltd* [1994] F.S.R. 275

[8] *Ultra Marketing (UK) Ltd and another v Universal Components Ltd* [2004] EWHC 468 (Ch), Transcript.

[9] *Designers' Guild Ltd v Russell Williams (Textiles) Ltd* [2001] 1 All ER 700

[10] *Baigent and another v Random House Group Ltd* [2007] EWCA Civ 247

[11] *Allen v Bloomsbury Publishing Plc* [2010] EWHC 2560 (Ch); [2010] E.C.D.R. 16; Official Transcript; Ch D; 14 October 2010.

[12] *Designers' Guild Ltd v Russell Williams (Textiles)* [2001] 1 All ER 700

[13] *Hodgson v Isaac* [2010] EWPCC 37; [2012] E.C.C. 4; Official Transcript.

[14] *Ravenscroft v Herbert and New English Library Limited* [1980] R.P.C. 193

[15] Ibid p. 73

[16] *Harrison v Harrison*[2010] EWPCC 3; [2010] E.C.D.R. 12; [2010] F.S.R. 25; Official Transcript; PCC; 19 March 2010.

[17] *JHP Ltd v BBC Worldwide Ltd* [2008] EWHC 757 (Ch); [2008] F.S.R. 29; [2009] Bus. L.R. D1; Official Transcript.

[18] *Newspaper Licensing Agency Ltd v Meltwater Holding BV* [2011] EWCA Civ 890; [2012] Bus. L.R. 53; [2012] R.P.C. 1

[19] *Public Relations Consultants and Association Ltd v Newspaper Licensing Agency Ltd* [2013] UKSC 18; [2013] 2 All E.R. 852

[20] An early case about *animus furandi* on the part of the defendant was treated by Page-Wood V.C. in Jarrold v. Houlston (1857) 3 K & J. 708 and it was defined "as equivalent to an intention on the part of the defendant to take for the purpose of saving himself labour; fourthly, the extent to which the plaintiff's and the defendant's books are competing works".

[21] To see similar approaches implemented in other areas see: Wang, A. (2006). The Shazam music recognition service. *Communications of the ACM*, *49*(8), 44-48; Boroughf, B. The Next Great Youtube: Improving Content ID to Foster Creativity, Cooperation, and Fair Compensation' (2015). Albany Law Journal of Science and Technology,25, 95.

[22] Mihalcea, R., Corley, C., & Strapparava, C. (2006). 'Corpus-based and knowledge-based measures of text semantic similarity'. In *AAAI* (Vol. 6, pp. 776) July 2006 and Berry, M. W. (1992), 'Large-scale sparse singular value computations' *International Journal of Supercomputer Applications,* 6(1), 13-49.

[23] Landauer, T. K., & Dumais, S. T. (1997). 'A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge' *Psychological Review*, 104(2), 211. Other approaches are proposed by Hofmann, T. (2004), 'Latent semantic models for collaborative filtering'. ACM Transactions on Information Systems (TOIS), 22(1), 91 and Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001), 'Eigentaste: A constant time collaborative filtering algorithm' *Information Retrieval*, 4(2), 133-151.

[24] Landauer, T. K., Foltz, P. W., & Laham, D. (1998), 'An introduction to latent semantic analysis' *Discourse Processes*, *25*(2-3), 270-271.

[25] Laham, D. (1997). Latent semantic analysis approaches to categorization. In Proceedings of the 19th annual conference of the Cognitive Science Society. 979. 1997.

[26] Another approach of LSA that does not implement SVD can be found in Hofmann, T. (1999), 'Probabilistic latent semantic indexing' In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 50-57, August 1999 and Hofmann, T. (2001), 'Unsupervised learning by probabilistic latent semantic analysis' Machine learning, 42(1-2), 177-196.

[27] *ibid*

[28] See Turney, P. (2001). 'Mining the web for synonyms: PMI-IR versus LSA on TOEFL'. In Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001).

[29] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990), 'Indexing by latent semantic analysis' *Journal of the American Society for Information Science*, 41(6), 391.

[30] Dumais, S. T. (2004). 'Latent semantic analysis' *Annual Review of Information Science and Technology*, 38(1), 192-193 and Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988), 'Using latent semantic analysis to improve access to textual information' In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM. 281-285, May 1988.

[31] Forsythe, G. E., Moler, C. B., & Malcolm, M. A. (1977). *Computer methods for mathematical computations.* (Prentice Hall) and Schone, P., & Jurafsky, D. (2000), 'Knowledge-free induction of morphology using latent

semantic analysis' In Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning-Volume 7. Association for Computational Linguistics. 67-72. September 2000.

[32] See: Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space: Words, sentences, discourse. Discourse Processes, 25(2-3), 249.

[33] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., & Ruppin, E. (2001), 'Placing search in context: The concept revisited', In Proceedings of the 10th international conference on World Wide Web. ACM. 406-414, May 2001. And expanded version of this method based on anterior queries to maximise effectiveness can be found in: Shen, X., Tan, B., & Zhai, C. (2005), 'Context-sensitive information retrieval using implicit feedback'. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM. 43-50. August 2005 and Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (2004), 'Latent semantic analysis' In Proceedings of the 16th International Joint Conference on Artificial Intelligence. 8. November 2004

[34] A commercial form can be found in: Microsoft. Compare and Merge Two Versions of a Document. . Last access February 22 2018. Retrieved from: https://support.office.com/en-us/article/compare-and-merge-two-versions-of-a-document-f5059749-a797-4db7-a8fb-b3b27eb8b87e

[35] Latent Semantic Analysis and Singular Value Decomposition are properly addressed later on this work.

[36] Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013), 'Handbook of Latent Semantic Analysis'. (Psychology Press). 56-59.

[37] Marshall, J., Grimm, J., Grimm, W., Grimm, J., & Grimm, W. (2005) 'Hansel and Gretel' (Weston Woods Studios).

[38] For an operative view of Latent Semantic Analysis visit: LSA @Boulder Last access August 17th 2018. Retrieved from: http://lsa.colorado.edu/

[39] Ashley, K. D. (1992), 'Case-based reasoning and its implications for legal expert systems' *Artificial Intelligence and Law*, 1(2-3), 114, Gabrilovich, E., & Markovitch, S. (2007), 'Computing semantic relatedness using Wikipedia-based explicit semantic analysis' In International Joint Conference of Artificial Intelligence (Vol. 7). 1606-1611, 1992. For a further introduction see: Brüninghaus, S., & Ashley, K. D. (2001), 'The role of information extraction for textual CBR' In International Conference on Case-Based Reasoning. Springer Berlin Heidelberg. 74-89, Leake, D. B. (1996). 'CBR in context: The present and future'. Case-Based Reasoning, Experiences, Lessons & Future Directions, 1-30. To see potential learning methods: Pirolli, P. L., & Anderson, J. R. (1985), 'The role of learning from examples in the acquisition of recursive programming skills'. *Canadian Journal of Psychology/Revue canadienne de psychologie*,39(2), 240.

[40] Watson, I., & Gardingen, D. (1999), 'A distributed case-based reasoning application for engineering sales support' In International Joint Conference of Artificial Intelligence. 600-605, 1999. De Mantaras, R. L., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S. & Keane, M. (2005), 'Retrieval, reuse, revision and retention in case-based reasoning', The Knowledge Engineering Review, *20*(03), 215-240. For furthers reading in each case see: Kolodner, JL. (1993), Case-Based Reasoning. (Springer Science Business Media), Watson, I. (1999). 'Case-based reasoning is a methodology not a technology' Knowledge-based systems 12.5, 304.

[41] For a detailed explanation on a functional CRB system see: Rissland, E. L., Valcarce, E. M., & Ashley, K. D. (1984), 'Explaining and arguing with examples'. In Proceedings of the Fourth AAAI Conference on Artificial Intelligence. 288-294, AAAI Press, August 1984 and Rissland, E. L., & Skalak, D. B. (1989), 'Case-based reasoning in a rule-governed domain' In Proceedings of the Fifth Conference of Artificial Intelligence Applications, 45-53, IEEE, March 1989. A method that includes social values to complement the legal approach see: Bench-Capon, T., & Sartor, G. (2003), 'A model of legal reasoning with cases incorporating theories and values' Artificial Intelligence 150.

[42] Hahn, U., & Chater, N. (1998), 'Understanding similarity: a joint project for psychology, case-based reasoning, and law'. Artificial Intelligence Review 12(5) 404.

[43] Skalak, D. B., & Rissland, E. L. (1992). 'Arguments and cases: An inevitable intertwining'. Artificial intelligence and Law, 1(1), 3-44. For a further reading on how the characteristics of each case affect the adaptation of the law see: Berman, D. H., & Hafner, C. D. (1993), 'Representing teleological structure in case-based legal reasoning: the missing link'. In Proceedings of the 4th international conference on Artificial intelligence and law. ACM. 50-59. June 1993.

[44] *Supra* note 2