



HIGHER COUNCIL OF LITERARY AND ARTISTIC PROPERTY (CSPLA)

ARTIFICIAL INTELLIGENCE AND CULTURE MISSION

Final report
27 January 2020

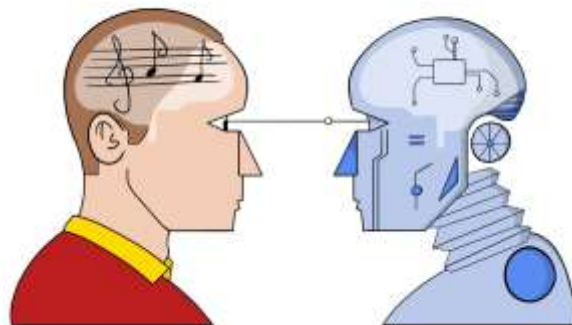
Alexandra BENSAMOUN, Professor at University of Rennes 1
Joëlle FARCHY, Professor at University Paris 1 Panthéon-Sorbonne

CSPLA Qualified Specialists
Chairs of the Mission

Paul-François SCHIRA, Auditor to the French Council of State

Mission Rapporteur

Report translated by Atenao



Artificial Intelligence and Culture

The content of this report, the result of reflection based on hearings, informal contacts and numerous readings, is the sole responsibility of its authors.

Parts 1 and 4 have been developed in-depth in *La culture des données* (Data culture), Farchy, Denis, Presses des Mines, to be published, 2020. For parts 2 and 3, the reader may also refer to the "Intelligence artificielle et propriété intellectuelle" (Artificial intelligence and intellectual property) chapter, in *Droit de l'intelligence artificielle*, dir. A. Bensamoun and G. Loiseau, LGDJ-Lextenso, coll. Les Intégrales, 2019, p. 235.

Contents

Summary	5
General introduction.....	7
1. Definition of AI.....	7
2. AI and culture, what are the challenges?.....	8
Part 1. – The use of AI in the cultural sector at different stages of the value chain.....	11
1.1. The place of AI in data science	11
1.2. The place of cultural data in AI advances	12
1.2.1. Types of cultural data.....	12
1.2.2. The promises of AI as regards automatically producing and extracting data	13
1.2.3. The potential of machine learning for culture	13
1.3. Consumption: exchanging with users and recommending content to them	15
1.4. Investment in production: knowing the market in order to take appropriate decisions.	17
1.4.1. Music: gaining insight to promote future successes.....	17
1.4.2. Audiovisual: gaining insight to limit production risks.....	18
1.4.3. The art market: gaining insight to better purchase and invest.....	19
1.5. Creation: the limited emancipation of the machine	19
1.5.1. Accompanying the (human) process of creation	21
1.5.2. Emancipating itself (a little) from human creation	21
Part 2. – The status of cultural products generated by "creative" AI	24
2.1. The copyright possibility.....	26
2.1.1. Existing protection	26
2.1.1.1. Non-protection of the algorithm.....	26
2.1.1.2. Protection of other AI components	27
2.1.1.3. The need to protect AI-generated "products"	28
2.1.2. The renewed application of copyright.....	29
2.1.2.1. Are AI-generated products "creations"?.	30
2.1.2.2. Are AI-generated "works" "original"?.	32
2.1.2.3. Do AI-generated "works" have an author?.....	33
2.1.3. The feasibility of applying copyright to AI-generated products (ownership issues)	34
2.1.3.1. First hypothesis of ownership: targeting the AI creator	35
2.1.3.2. Second hypothesis of ownership: targeting the AI user	37
2.1.3.3. Legally imposing a regime of ownership	38
2.2. Alternatives to traditional copyright law.....	39
2.2.1. Creating special copyright.....	39
2.2.2. Creating copyright in the style of a "related right"	40

2.2.3. Creating a <i>sui generis</i> right	41
2.3. The absence of private rights	43
2.4. So, which solution?.....	45
Part 3. – The regime of works enabling the production of cultural creations via AI.....	47
3.1. Does copyright apply to the works on which creative AI is based?	47
3.2. The difficult balance between the objectives of AI-related innovation and those of authors' legitimate protection	50
3.2.1. Possible exceptions to the copyright monopoly	50
3.2.1.1. The inapplicability of the exceptions currently in effect.....	50
3.2.1.2. The new data mining exception	51
3.2.2. The licensing mechanism as a suitable solution?.....	55
Part 4. – Improving data sharing to meet the challenges of AI development	57
4.1. The challenges of data access and sharing for the cultural sector.....	57
4.1.1. Use data: transparency and customer relationship control challenges	57
4.1.2 Metadata: content qualification challenges	59
4.2. Future approaches to sharing.....	60
4.2.1. Making cultural data essential facilities: an attractive but unsuitable proposal	60
4.2.2. The use data portability approach	63
4.2.3. The metadata of general interest.....	63
List of people who contributed to the mission	67
Written contributions.....	70

Artificial Intelligence and Culture Report

Summary

Artificial intelligence (AI) has clearly taken over cultural sectors, through algorithms and large corpus of data which fuel them and already offers a host of applications developed by research centres, large companies and specialized start-ups. Depending on the objectives pursued, the techniques as well as the types of data used are not the same. Although the question of AI is today closely correlated with that of "data", this generic term actually covers very heterogeneous realities. For different historical reasons, certain categories such as personal data, public data and "works" data are subject to very precise legal qualification. This is not so for other **categories of data** (metadata, use data, etc.) which correspond, in cultural industries, to professional practices.

Thanks to this data, **an increasing number of concrete applications are emerging throughout the value chain, from the creation stage through to the production stage and on to the consumption stage.** As regards consumption, AI and algorithms are massively used to recommend content to Internet users. Although algorithm-based personalized recommendation has been the subject of debate since the first "filter bubble" works appeared, given the risks of consumers being locked into their habits, in actual fact, a host of means of recommendation exists. Thanks to algorithms, the use of data also revitalizes the ambition to make appropriate investment decisions and to support, or even replace, the usual human intuitions and expertise with supposedly objective analyses of the determinants of the success of a work or an artist. In addition to analysing market trends, one of the promises of artificial intelligence is to compare, based on the use of historical data, contents which have been successful with those which are currently being produced so as to analyse the keys to success, and possibly to anticipate it. In terms of creation, among the various experiments rolled out in cultural sectors, not all have the same degree of maturity; some, which merely accompany the human process of creation, are largely present in the audiovisual and publishing sectors; others, which strive to emancipate themselves from it, are found more so in music and the art market.

From a legal point of view, AI intervenes in the **artistic creation** phase, which raises questions as regards copyright. Whilst the art market receives creations announced as being AI-generated, the **question arises as to the qualification of these new productions.** Are they intellectual works, and as such protected by copyright? If so, who is the author and the rightholder? **A renewed analysis of the conditions of access to protection (creation, originality, author) could enable these cultural works to fall under copyright.** But other solutions are also proposed (special right, absence of private protection, etc.). As such, it is important to test positive law and to be ready to intervene if ever a possible need for

regulation arises in the future. In any case, the approach should be carried out within an international, or *at least*, European framework.

Moreover, in a learning process, creative AI works by ingesting works which are deconstructed and analysed so as to identify common characteristics. This process enables the creation of an inference model whose implementation leads to the generation of an algorithmic creation. As such, the creation of the Edmond de Bellamy portrait was made possible thanks to the development of a training base of nearly 15,000 classical portraits from the 14th to the 20th century. Are these upstream acts to be considered as acts of use giving rise to copyright? The introduction of a "data mining" exception dedicated to AI uses, in Article 4 of Directive 2019/790 of 17 April 2019, seems to validate a positive response. However, this new, very broad limitation also provides for the possible exercise of an opt-out by rightholders, which enables a return to reservation. Consequently, other solutions must be put forward to facilitate the use of protected content whilst ensuring the protection of rightholders. In this context, voluntary general licences could ensure a balance.

Finally, **insofar as the quantity and quality of the data which can be called on to fuel the development of AI becomes a factor of competitiveness, the specific issues of data sharing and movement for cultural initiatives should be questioned.** Although data movement and sharing issues as regards use data and metadata are not always directly related to intellectual property concerns, they are nonetheless significant. Beyond the issue of transparency for a fair distribution of revenues for the benefit of rightholders, it is actually all the balances within the sector which are likely to be called into question by access to use data or, on the contrary, by the loss of control over the customer relationship. For metadata, it is basically a question of qualifying content so as to accompany the processes of artificial creation and to renew the means of forecasting and recommendation proposed by cultural industries. A regulation taking the specificities of each sector into account could usefully accompany the access of operators to the large masses of data held by others.

Artificial Intelligence and Culture Report

General introduction

1. Definition of AI

The AI concerned today has ancient origins since the idea of "intelligent" machines is rooted in science fiction, scientific thinking dates back to the 1950s and most current algorithms were invented in the 1980s. The expression "artificial intelligence" was first formalized in John McCarthy's presentation at Dartmouth College's Summer Camp in 1956. In his work carried out ten years after the appearance of the first computer, hand-in-hand with his colleague John McCarthy, the American mathematician Marvin Minsky assumes that it can be used to better understand intelligence and defines AI as "the science of making machines do things that would require intelligence if done by humans"¹.

The handful of guests at the 1956 conference would dominate research in the following years, which multiplied around this new concept of artificial intelligence. Over the last few decades, AI has as such experienced periods of more or less intense research, alternately dominated by two approaches, the symbolic or cognitivist approach and the connectionist approach. Symbolism refers to deductive reasoning, based on rules and logic, whilst connectionism corresponds to inductive reasoning, based on experience. The turn of the 2010 decade was marked by the triumph of deep learning, after 60 years of successes and disappointments stemming from the connectionist approach. Although machine-learning algorithms can use highly-different mathematical tools, within this vast ensemble, the focus is on artificial neural networks and deep learning. The term neuron (neural) here refers to a functional analogy with the behaviour of real neurons in biology. The new neural network architectures are called "deep" because they are composed of more layers than the traditional neural networks proposed in the 1960s. As such, deep learning mimics human brain functioning more effectively.

AI, a multifaceted discipline, is therefore back in the spotlight with deep learning, whose success corresponds to the combination of three factors:

- the improvement in the computing capacities of machines. Combining a large number of formal neurons in the form of a "deep" network was basically a theoretical construction as long as no machine was powerful enough to calculate and simulate the functioning of these networks;
- the progress in research on more sophisticated Convolutional Neural Networks (CNN), developed in the 1980s, which sparked a new wave of interest from 2012 onwards, triggered by some successful practical applications;
- the increase in the mass of varied and relevant data available. As regards the machine-learning models which are currently experiencing the greatest developments, AI performance is actually directly correlated to the data collected.

¹ CNIL (French Data Protection Authority), *How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence*, Paris, Dec. 2017.

The promise of "strong" AI, capable of equalling human intelligence, as pioneers had wished for, is confronted with AI which is weak, unconscious and specialized in a particular field, which can only solve the problems for which it was designed and trained. The AI on the market today, integrated in particular into consumer products, actually only corresponds to this second category. In a very schematic way, the currently-existing key AI systems (weak AI) can be divided into two main categories stemming from symbolic and connectionist trends:

- Deductive systems, based on predefined rules, from the simplest to the most sophisticated (formerly called "expert systems"), are capable of solving precise, predetermined problems based on known facts and rules. These AI systems can be described as deterministic, insofar as the answer they provide to a question can be precisely determined on the basis of input data and objective facts and rules integrated in the system.
- Inductive systems, based on machine learning; the term "machine learning" refers to the ability of machines to learn from training data without being explicitly programmed. It is no longer a question of specifying each action in the program, but of designing a program capable of tailoring its behaviour to its environment. These systems, which are based on probabilistic models, are therefore capable of solving complex problems using algorithms configured based on learning data.

Advances in machine learning have made predefined-rule-based systems obsolete in many areas, as machine-learning-based systems are generally much more efficient for the same task. Machine learning, which is therefore the major development pathway for contemporary AI, marks the transition from programming logic to self-configuration logic. In practice, AI business applications often combine these two systems, and the connectionist and symbolic approaches are more complementary than competitive.

2. AI and culture, what are the challenges?

On 25 April 2018, the European Commission unveiled its strategy via a Communication, *An Artificial Intelligence for Europe*. The group of experts appointed by the Commission proposed common actions: increase investment, make more data available, foster talent and develop an AI ethic to ensure trust. Moreover, following the submission of Cédric Villani's report, France decided to support the development of AI through an investment plan, whilst the European Commission plans to invest nearly 2.5 billion euros between 2021 and 2027. In the recommendations adopted by the OECD on 22 May 2019 as regards AI and in the European Commission's communication of 8 April 2019, some major principles for regulating AI, such as the absence of discrimination, transparency, explicability, respect for privacy and human autonomy, etc. exist.

Although the Villani report highlights a few sectors, focusing on those in which France can develop the most important competitive advantages, it notes that AI actually affects all areas, including those where it is not necessarily expected, such as sport and culture. There are a host of challenges for AI in the cultural sector. The general public has been discovering artificial intelligence through exhibitions such as the 2019 interactive exhibition at the Barbican Centre in London, *AI: More than Human*, which explores the creative and scientific innovations of artificial intelligence and the *Artists and Robots* exhibition at the Grand Palais in Paris in 2018.

Initiatives aimed at encouraging reflection and recommendations in the public sphere are also multiplying. In particular, UNESCO has taken an interest in the question through the prism of the diversity of cultural expressions² and, in 2019, the OIF (International Organization of La Francophonie) organized a series of encounters on the challenges and opportunities of AI for the cultural and creative industries sector³. The subject of artistic creation and its status is also a key focus for many authorities and institutions. For example, at the end of September 2019, WIPO reported the organization of a "Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)"⁴, which in particular addressed the legal issues studied in the mission report (but as the conversation was "informal", no written document has been published for the moment except the programme itself). Reflection is also ongoing in the United States⁵. The European Parliament (Legal Service) is also considering the question of the protectability of creations produced by AI within the framework of drafting its future work programme⁶, and the European Commission raised the subject of intellectual property in its two communications on AI⁷. In October 2019, the Global Forum on AI for Humanity, under the high patronage of the President of the Republic of France, proposed a workshop entitled "Rethinking cultural and ethical issues in AI"⁸. In France, a French-Quebec mission is underway on the "discoverability" aspects of AI use in the cultural sector⁹.

Taking note of this growing interest in AI and culture, the CSPLA entrusted Professors Alexandra Bensamoun and Joëlle Farchy with a mission on the legal and economic issues of AI in the cultural sectors, to identify all the issues raised (see mission letter of 1st July 2018¹⁰). **As planned from the start of this mission, this report is regarded as a simple step in the reflection, which will undoubtedly have to be continued over the medium term.** Its main aim is to provide an understanding of the technique and its legal and economic consequences, and to draw up an inventory of solutions in light of the state of the technique, which is itself bound to evolve.

² O. Kulesz, *Culture, platforms and machines: The impact of Artificial Intelligence on the diversity of cultural expressions*, Intergovernmental Committee for the Protection and Promotion of the Diversity of Cultural Expressions, 12th session, Dec. 2018:

https://en.unesco.org/creativity/sites/creativity/files/12igc_inf4_en.pdf/.

³ <https://www.francophonie.org/node/170/>.

⁴ https://www.wipo.int/meetings/en/details.jsp?meeting_id=51767/.

⁵ <https://www.uspto.gov/about-us/events/artificial-intelligence-intellectual-property-policy-considerations/>. – See the part "Copyright: Who is the author of AI-generated content? Are such works copyrightable? What policy implications arise from the use of copyrighted works for the purposes of machine learning?". The mission established contacts. It is envisaged to transmit the report.

⁶ Hearing in the European Parliament, seminar *Legal consequences of the technological revolution*: A. Bensamoun, "Droit d'auteur et intelligence artificielle : un nouvel enjeu pour la création" (Copyright and artificial intelligence: a new challenge for creation), 6 Dec. 2018, Brussels.

⁷ See, most recently, the European Commission's communication, *Artificial Intelligence for Europe*, COM(2018) 237 final, 25 Apr. 2018, pt 3.3. and note 52.

⁸ See the round table 3 "Cultural politics in the era of AI: creativity and discoverability", with V. Guèvremont, "The new measures in favour of discoverability"; O. Kulesz, "Artificial intelligence and the cultural sector: opportunities and challenges"; A. Bensamoun, "The protectability of AI creations: a legal and ethical issue"; P.-L. Déziel, "Use personal data in a context of valuing cultural diversity". – <https://gfaih.org/>.

⁹ V. Guèvremont et al., *Agir en faveur de la découvrabilité des contenus culturels dans l'environnement numérique : arguments, réflexions et pratiques émergentes* (Acting to promote the discoverability of cultural content in the digital environment: arguments, reflection and emerging practices), report commissioned by the Quebec Ministry of Culture and Communications a part of the *French-Quebec mission on the discoverability of online French language cultural content* launched 3 April 2019, to be published.

¹⁰ <http://traduction.culture.gouv.fr/url/Result.aspx?to=en&url=https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle/>.

Although the question of AI is today closely correlated with that of "data", this generic term actually covers very heterogeneous realities. For different historical reasons, certain categories such as personal data, public data and "works" data are subject to very precise legal qualification. This is not so for other categories of data (metadata, use data, etc.) which correspond, in cultural industries, to professional practices. Through algorithms and large corpus of data which fuel them, AI intervenes throughout the value chain of cultural industries. However, the issues at stake are not the same in legal and economic terms.

From a legal stance, it is at creative level that the issues of literary and artistic property, developed in the report hereof, arise. In economic terms, it is not only at the creative level, but also at consumption level, as well as production, that AI plays a role. One of the major challenges to address over the coming years will be to reconcile compliance with major regulatory principles, of which intellectual property is only one aspect, with business models which make data processing and enhancement the core of their business and which impose great movement of such data.

In order to grasp all the issues, the report first develops the current uses of AI in cultural sectors at the various stages of the value chain. AI already offers a host of applications developed by research centres, large companies and specialized start-ups. Cultural industries and the media, which experienced dematerialization early on, act as a laboratory for emerging business models mobilizing AI (Part 1).

From a legal stance, the works themselves constitute incoming data which fuels algorithms and AI produces new artificial "creations". Two key questions therefore come into play at creation level. The first - and, without a doubt, the most topical -, downstream, is to determine the status of the production generated by a creative AI. Is it intellectual work, and as such protectable by copyright? And, in assuming so, who is the author? (Part 2).

The second legal question, upstream, consists in identifying the regime applicable to the works which fuel the AI enabling the production of these artificial creations (Part 3). In the machine-learning creation model, AI is fuelled with data, from the simplest to the most sophisticated, of which intellectual works are a part. In addition to the question of the status of new creations, there is moreover the question of apprehension, through copyright, of this phenomenon of analysis, of use and of deconstruction of protected works.

Finally, insofar as the quantity and quality of the data which can be called on to fuel the development of AI becomes a factor of competitiveness, the specific issues of data sharing and movement for cultural initiatives and future avenues for development should be questioned (Part 4).

Part 1. – The use of AI in the cultural sector at different stages of the value chain

Part 2. – The status of cultural products generated by "creative" artificial intelligence

Part 3. – The regime of works enabling the production of cultural creations via AI

Part 4. – Improving data sharing to meet the challenges of AI development

Part 1. – The use of AI in the cultural sector at different stages of the value chain

Whilst machine learning always requires large amounts of data, conversely, data analysis mobilizes a wide variety of disciplines and fields of research which do not always call on machine learning (1.1). Depending on the objectives pursued, the techniques as well as the types of data used are not the same. (1.2). In addition to "works" data, other categories of data, which do not correspond to precise legal qualifications, such as use data and metadata, used by professionals, and concrete applications, throughout the value chain, appear to be greatly increasing, at consumption (1.3), production (1.4) and creation (1.5) levels.

1.1. The place of AI in data science

The different techniques developed do not pursue the same objectives and are not aimed at the same communities.

The explanatory purpose of quantitative techniques

Economists have been using quantitative data analysis techniques for a long time. These techniques, developed around the principle of causal inference (clarifying the relationship linking y to the x variables), imply *a priori* modelling in which the data validates assumptions and intuitions. In this approach, the interpretation of each explanatory variable is preponderant.

The performative purpose of machine learning

Alongside the tools developed by economists, the performance of automated processing, based on machine learning, has enjoyed great advances; these are inductive processes (see above) which are intrinsically conservative, which do not anticipate evolution but reproduce the past in the present or the future to produce a result, trends and/or estimates. The break from traditional quantitative techniques is based on the fact that modelling stems from the data itself. The performance of the model does not depend on the explainability of the results but on its ability to provide a rapid and effective response to the issue in question. The more complex the machine-learning algorithm is, the less the process that led to the result can be expressed in terms which are understandable to humans.

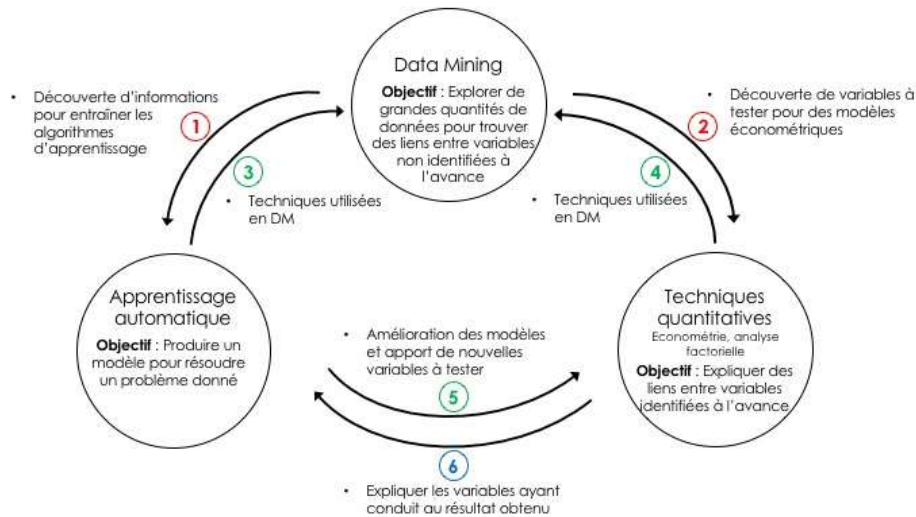
The exploratory purpose of data mining

The term data mining emerged in the late 1970s at the initiative of the marketing community. Over time, data mining has become not only a technique for searching for mass information, but also for highlighting structures and associations, *a priori* difficult to detect and to imagine, between data. The data mining approach uses both quantitative analysis and machine-learning techniques. It focuses on exploratory modelling and the discovery of previously unknown relationships between data.

Quantitative techniques, machine learning and data mining, which historically differ in their objectives, their communities and their methodologies, now tend to become complementary and are used, in turn, by the new data scientists based on the type of question to be resolved. Machine learning produces operational forecasts which are all the more precise as the system

is fuelled with large quantities of data. Data mining discovers totally unexpected correlations between data, which can be used as training for machine learning, and/or can be tested by quantitative techniques. The latter emphasize causal links beyond simple random correlations and provide building blocks for long-term strategy for companies and public decision-makers.

The diagram here below synthesizes the links between the various scientific approaches to data analysis.



Farchy, Denis, 2020

1.2. The place of cultural data in AI advances

In cultural industries, in addition to "works" data, a type of data mobilized by the players leads to a distinction between use data, metadata and hybrid data. Thanks to the processing of all this data, AI improves various tasks which are applied in new practical ways in cultural sectors.

1.2.1. Types of cultural data

Use data is produced when ONE user and ONE work come together. Data relating to the use of cultural and recreational content has become decisive "input" which makes it possible to analyse users' tastes and habits and to propose them a tailored offer. Data relating to listening and reading time and frequency, to purchasing preferences and purchases made, amongst others, may be considered as personal data if it leads to establishing the precise profile of an individual.

Metadata is defined herein as all information describing specific content or a creator based on objective or more or less determined characteristics. This metadata, whether visible to end users or not, is central to many systems. It may be classified, making a distinction between data which is legal (rightholders), descriptive (title, term) and enriched (genre, semantic analysis of the language used in a scenario, rhythm of a piece of music, etc.).

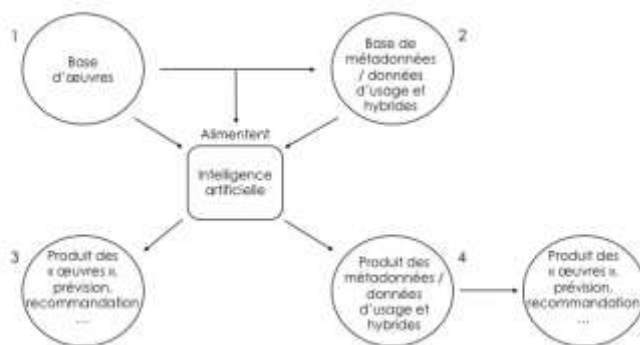
Finally, where use data and metadata intersect, data is formed which we describe as **hybrid**: information produced by comparing and/or aggregating use data and/or metadata (analysis of Internet user exchanges on social media associating terms significant to given content).

1.2.2. The promises of AI as regards automatically producing and extracting data

Even if AI does not necessarily revolutionize tasks, it extends and automates data production and extraction possibilities, as such leading to research which, in turn, leads to an increasing number of practical applications. AI actually mobilizes data in two ways.

First of all, upstream, machine-learning algorithms "dip into" databases of works which they deconstruct or analyse (1) or into pre-existing metadata bases (2). Based on the situations, AI is fuelled by metadata on the works, or, in more specific cases, directly by the works themselves (similarity of one musical work to another, for example) thanks, in particular, to advances in deep learning. In this second case, AI capable of "understanding" all aspects of the work (for a film both image and sound) is required, which leads to a high cost for training algorithms on a considerable amount of data. Moreover, mobilizing enriched metadata bases rather than works directly increases the explanatory power of the results proposed by the AI in terms of forecasting and recommendation. Depending on the objectives sought, the joint processing of metadata and use data fuels various applications.

Then, downstream, AI automatically produces artificial creations (3) and new metadata, which deliver increasingly precise content qualification work (4) and/or new hybrid data, to offer various services to Internet users and to professionals. The diagram below summarizes these complex circuits.



Farchy, Denis, 2020

1.2.3. The potential of machine learning for culture

One of the great promises of artificial intelligence is therefore the extraction and analysis of large unstructured databases such as works and the production of new metadata which can be understood by humans.

Amongst the various tasks improved by deep learning, many are applied in practical ways in cultural sectors. The following examples are worth mentioning.

Natural Language Processing (NLP), which refers to a better understanding of natural language and to a better restitution of results in natural language, has been greatly improved by AI. Firstly, on social media, for example, natural language data processing makes it possible to monitor not only the volume of discussions focusing on specific content but also the emotions expressed in these discussions. Secondly, the automatic generation of simple texts in natural language (Natural Language Generation - NLG) is intended for developing computer programs capable of producing texts in a language which can be understood by humans.

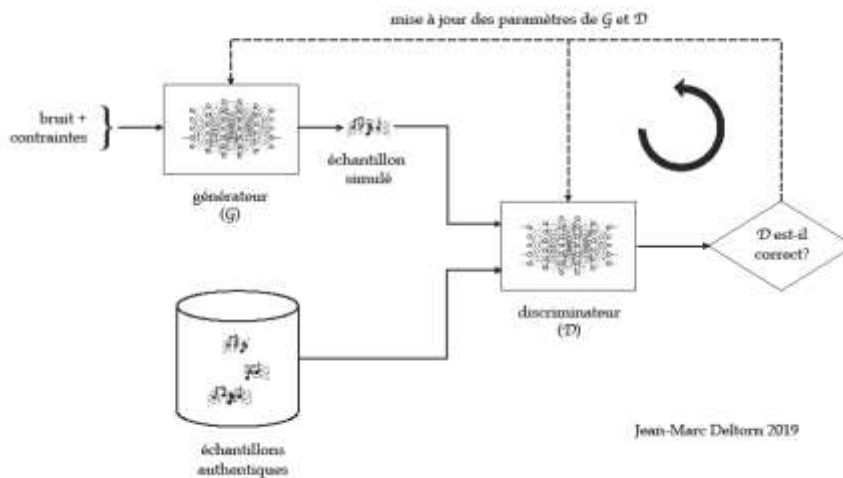
In the field of music, an academic research discipline in its own right emerged at the end of the 1990s, **MIR (Music Information Retrieval)**, for the purpose of collecting a host of information in order to qualify musical content (also known as music indexing). MIR techniques have been developed to solve issues such as classifying genres, identifying artists and recognizing pieces of music *in situ*.

In the audiovisual field, **automatic video cataloguing**, which has become an important research topic, originated in the 1990s when the amount of images available exploded. The aim of video cataloguing is to structure videos and extract semantic information to ensure that content can be quickly retrieved. Efficiently qualifying and cataloguing individual content encourages longer lifecycles and increases enhancement opportunities. In addition, deep learning brings new tools capable of real-time perceiving the video content offered. For example, when a viewer watches a live sports event, it would be possible to choose to watch the competitor of their choice, who has been automatically identified by artificial intelligence using a facial recognition and automatic labelling technique.

On the art market, new "**computer vision**" techniques aim to automatically obtain an image description and make effective search. Some research work has proposed algorithms which combine object recognition and language models to generate image descriptions based on natural language. The description of image content using natural language is particularly useful for presenting the creation of captions and for performing content-based searches based on text queries.

Finally, artificial content generation has been renewed by the Generative Adversarial Networks (**GAN**) models, which enable images and sounds to be modified and new, highly-realistic ones to be produced. GANs pit two neural networks; the first (Generator - G here below -), a sort of "forger", generates content which attempts to mimic the training data delivered to the AI (for example, take on the aspect of one of Van Gogh's works if Van Gogh's paintings were provided for the training); the second (Discriminator - D here below -) attempts to distinguish between the training data (one of Van Gogh's works) and the output provided by G (content resembling a Van Gogh painting). Iteratively-speaking, the information is analysed with each feedback, up to the optimal point where D is no longer able to distinguish whether the new content proposed is derived from the training data or artificially generated. Whenever all the training data is sourced in human-produced works, D will be able to distinguish between these human works and the artificially-generated output.

A generative adversarial network training sequence



These projects are essential, as shown by the many possible applications of AI in the sector, whether, as we shall see, they involve artificial creation processes, strategic investment decisions or directing Internet users towards specific content.

1.3. Consumption: exchanging with users and recommending content to them

As regards the general public, one of the most widespread AI applications is chatbots; the system is capable of recognizing and interpreting the text typed by the Internet user to extract the characteristics of their question and provide them as input data for a search engine. In addition to this direct exchange with users, AI and algorithms are extensively used to recommend content to Internet users. In a world of abundant information where consumer interest is the rare resource, recommendation refers to the set of systems aimed at directing the Internet user towards particular content or set of particular contents. From an economic stance, recommendation facilitates matching supply and demand on prototype markets where great information asymmetries exist. Depending on the perspective adopted, various types of recommendation may be proposed, which partly overlap.

Depending on the nature of the processing performed, a first distinction pits recommendations based on purely human processing (such as editorial recommendation) against recommendations based on automated algorithmic processing. The latter may be customized using use data or non-customized using semantic analyses of content similarities. Although for market analysis and forecasting, players only need aggregated consumption data, for customized recommendation tools, each user's use data must be known.

Moreover, amongst automated algorithmic processing, we can distinguish three types of recommendation based on the nature of the incoming data:

- Filtering based on the comparison between the precise description of the characteristics of content (metadata) and the preferences of each user (it's an action film, I like action films, so I'm going to like this film).
- Collaborative filtering does not rely at all on the intrinsic characteristics of content but on those of groups of users with similar tastes; by implementing a principle successfully developed by the Amazon platform, recommendations

are made by comparing the Internet user's profile with those of other users who have purchased the same content: "Customers who bought this item also bought...".

- Hybrid forms are often implemented in order to improve recommendation performance. Content-based filtering is actually not very effective when the metadata lacks precision. However, qualifying content often requires some form of human intervention to check quality in addition to the automated processing. Collaborative filtering, on the other hand, is inoperative when there is not enough existing data (algorithm cold start) and not very effective for unpopular content for which trends cannot be detected.

In the **audiovisual sector**, recommendation tools have different objectives depending on whether the aim is to attract the viewer to a cinema for a specific film at a given time, or to build recurring loyalty amongst viewers to a channel or a subscription video service. The Netflix platform made a name for itself through the power of its customized algorithmic recommendation model. However, beyond the marketing argument, the system Netflix proposes is more hybrid than it seems as it combines an automated algorithmic recommendation based on use data and intense human intervention to qualify content to produce relevant metadata; in-house, individual content is viewed from one end to the other by one or two humans and is subject to a hundred or so tags to be able to propose categorization which can be cross-referenced with user behaviour. As such, the power of the platform's business model is not linked solely to the technical strength of its algorithm, but to the intensity and quality of the data collected on users and content, which only a market leader with a huge database is capable of implementing.

To develop relevant recommendation systems, television channels suffer from a double handicap in relation to "native digital" platforms: access to use data and the lack of in-house skills to carry out an activity which is not their core business. Applications and the home screen of smart TVs, tablets and OTT boxes facilitate better data collection, a prerequisite for recommendation services, but this data is not always shared by equipment manufacturers or by telecom operators and requires setting up complex contractual relationships. To compensate for the lack of in-house expertise, pure players have positioned themselves on the audiovisual recommendation market. In France, two players, Cognik and Spideo, offer turnkey recommendation systems for TV channels and VoD publishers.

The **online music** market makes extensive use of recommendation tools and user knowledge has been greatly enriched by the use of context data. Many studies have shown that listeners are complex and multi-faceted beings who wish to listen to music which corresponds to the context they are in at a given moment in time. Moreover, MIR (Music Information Retrieval) research, for the purpose of collecting a host of information in order to analyse musical content in an increasingly more detailed way, has found a major scope of application in recommendation services. Beyond recommendation based on descriptive metadata, audio content analysis strives to model the similarity between songs with elements such as tempo, rhythm and melody. Recent advances in audio signal processing provide information which can be incorporated into recommendation systems to complement collaborative filtering methods; whilst the latter inherently favour the most popular content over the "long tail", the audio signal-based recommendation helps to promote musical discovery.

Market-leading streaming platforms have no choice but to offer increasingly efficient recommendation systems to differentiate themselves from rivals. Although most of them

initially tested forms of recommendation which were exclusively editorialized (Deezer) or based on past use (Spotify), in the end they moved towards a hybrid approach. As such, they analyse user behaviour as well as blog and social media data through natural language processing, context data and/or intrinsic metadata (audio signal) to create their recommendation systems.

On the **art market**, visibility on the web has become a key challenge for auction houses as well as galleries, historical operators and new intermediaries which position themselves by initiating contact between collectors and galleries or directly with artists. Recommendation tools are becoming increasingly hybrid to combine the effectiveness of customized recommendations based on the Internet user's use and a precise qualification of the content proposed. The strategies adopted vary from one company to another. The dominant platforms have the user base and the financial means required to develop recommendation tools in-house. Others draw on existing databases, rely on external developers or the takeover of pure players.

Although algorithm-based personalized recommendation has been the subject of debate since the first "filter bubble" works appeared, given the risks of consumers being locked into their habits, in actual fact, a host of means of recommendation exists. When faced with homogenizing trends, recommendation algorithms can also be used to unveil an unparalleled diversity of content.

1.4. Investment in production: knowing the market in order to take appropriate decisions

As the culture-based economy is a prototype economy, where uncertainty as regards demand is particularly high, identifying decision-making tools to limit risk-taking has always been a key focus of interest for professionals. A small number of productions attracts the bulk of demand according to the well-documented model of the Economics of Superstars. Thanks to algorithms, the use of data also revitalizes the ambition to make appropriate investment decisions and to support, or even replace, the usual human intuitions and expertise with supposedly objective analyses of the determinants of the success of a work or an artist. In addition to analysing market trends, one of the promises of artificial intelligence is to compare, based on the use of historical data, contents which have been successful with those which are currently being produced so as to analyse the keys to success, and possibly to anticipate it.

Be it for analysing market trends or for forecasting the future, the decision-making tools used by professionals are based on mechanisms specific to each sector. In the music industry, strongly reshaped by digital technology, professionals have many real-time trend tools on hand and success largely depends on streaming platforms. In cinema, where trend data is less accessible, a great deal of academic work on forecasting exists and provides multiple indications to professionals so they may develop their tools. Finally, on the art market, the key purpose is to enable an objective assessment of the performance and price of works as a support for purchasing and investment decisions.

1.4.1. Music: gaining insight to promote future successes

Analysing musical works or tracks based on their objective characteristics, using machine-learning algorithms, has become a field of scientific research in its own right through the development of MIR systems (see above). However, despite advances, there is no business

application mature enough to predict with certainty a likely "hit" based solely on the analysis of its own characteristics, and those who have tried this have not been very successful. Other researchers have attempted to predict the success of a track by no longer relying only on intrinsic metadata, but by integrating initial use data to anticipate a possible success, to accompany it in particular on social media and to make it last.

For majors, labels and artists, the real-time restitution of use data is as such a precious piece of information to help them make informed decisions. Some offer dashboards and reports to view trends and the evolution of artists on different platforms for a subscription fee; Soundcharts is one of the leaders on this market. Others are purely in-house systems or accessible only to music companies and artists distributed on a platform (Spotify for Artists, Deezer Backstage, etc.).

1.4.2. Audiovisual: gaining insight to limit production risks

When a film is still in a phase where relatively little money has been invested, producers are tempted to assess the commercial potential of a new script by taking the box-office performance of films deemed to be similar as a reference. Based on the qualifications initiated, successful film scripts are compared with those pending production. Several start-ups (Vault, ScriptBook, StoryFit) have positioned themselves on this market. However, the results of the script analysis, a currently-trending activity, appear to be mixed. The undeniable advances made in natural language processing has led to the multiplication of commercial offers designed to (semi-) automatically analyse a script solely on the basis of metadata; however, the quality of the results provided has not led to sound appraisals. Moreover, many professionals are reluctant to use these new tools, fearing that their work will be formatted. In the short term, it therefore seems more promising to use data modelling, not to analyse the quality of the script itself, but to quantitatively and qualitatively determine the potential audience for a specific script.

Beyond comparing a film with others which are similar, studios, distributors and operators use forecasting tools integrating use data. Apart from cinemas, for which audience data is available and enables using AI's potential for marketing purposes, as regards the rest of the audiovisual industry, the tools, developed during the analogue era, are not always tailored to digital opportunities. Unlike the music industry, professionals have only a limited range of use data they can rely on; this is either aggregated and of low quality, or not shared and not easily accessible by all professionals.

As in music, however, operators with customer relations at the lower end of the value chain are able to process larger amounts of data to provide effective forecasts. On the SVoD market, where Netflix has made its mark, use data is not well known. Initiatives taken by specialized companies, developed during the analogue era, are struggling to tailor to on-demand, customized and multi-support offers. As for the Netflix platform itself, after having chosen not to offer any tools accessible to professionals, it has evolved and now communicates general data on audiences for broadcasted content only to rightholders.

Finally, given the technological convergence of television and Internet, smart TV providers and telecom operators offering OTT (Over The Top) boxes and services have new capacities to real-time collect use data. However, in France, only SFR (Altice Group) offers the SFR Analytics Live service, which delivers real-time audience data for an annual subscription from

television channels. Other operators, such as Orange, deliver aggregated use data in accordance with limited contractual agreements concluded with television channels.

1.4.3. The art market: gaining insight to better purchase and invest

The boom in the art market and the arrival of new collectors have accelerated the demand for transparent and organized information, also on an international scale, which in turn has triggered a proliferation of rankings and other indicators. Over the last ten years, several platforms, taking advantage of scientific advances and the material, which some benefit naturally from given their activity as data aggregators, have created their own market indicators. The proliferation of indicators does not make it easy for buyers and sellers of works of art to understand this information. To enlighten them in their choices, operators have as such positioned themselves on the price estimation market. Others strive, through algorithms, to suggest which works it is advisable to invest in. On a particularly-speculative art market, even more than elsewhere, the boundaries between general information, estimation of a specific work and forecasting are proving to be porous, all the more so as the lack of transparency in classification methodologies fuels forms of speculation, price inflation and self-fulfilling predictions, works ranked as the most expensive are assumed to be those of the most highly-rated artists, which mechanically increases the artists' ratings.

As such, attempts to perfectly model the success of a book, a script or a painting using algorithms have, to date, failed to uncover the keys to this so-sought-after secret. Data analysis only provides success probabilities and algorithms are fundamentally conservative. Because they learn by analysing what has worked in the past, they are unable to take into account changes which will take place in the future. Moreover, a cost is incurred when implementing forecasting models, particularly for collecting and formatting non-structured data, which cannot be conceived in the all-content-based economy. Finally, attempts to automate success carry the seeds of a form of internal contradiction, as new productions based largely on comparing the results of past productions so as to copy success criteria can lead to absurd situations since similar content eternally offered to consumers and works populated with individually-tailored heroes and ideas will inexorably end up impoverishing it.

Beyond this largely-unattainable dream of global modelling, advances in machine learning in the field of visual recognition, natural language processing and audio variable analysis, as well as the opportunities at-hand for collecting and processing use data, have however shaken up professional practices. Algorithms are already widely-used to spot market trends, facilitate decision-making for publishers, producers and collectors, and even to specify a distributor's business strategy so as to target, broaden the potential audience and optimize the presence of a work or an artist in line with the dissemination media. Algorithms, as complementary indicators of professional experience, help to better position a project and anticipate the reactions of the public. The art and music markets favour the efficiency of forecasting, whereas in the audiovisual sector, the explanation of the combinations of success variables remains a major challenge.

1.5. *Creation: the limited emancipation of the machine*

Machines have long been central to artistic creation. Some tools are designed not just for accompanying but also for boosting creators' individual imagination by implementing, via AI, a stage of manipulation and appropriation and even integration of the world of other artists, thanks to use data and reactions of the public. Most of the leading digital players have

integrated the importance of culture as a scope of application for AI and have developed dedicated research programmes: IBM with its "home-made" Watson AI, Microsoft with its Next Rembrandt experiments and other editing tools, and Google with its Google Brain research team which proposes tools to musicians (Magenta) and artists in the art market (DeepDream). In 2018, music streaming leader Spotify inaugurated the Creator Technology Research Lab in Paris to develop tools to help artists in their creative process.

In order to precisely understand the legal aspects of this creation (Part 2), it is essential to have a good understanding of the **object elaboration process by creative AI** which is technical. First of all, exogenous starting data, in the form of a training base (or training set), fuels the AI. In the case of creative AI, these sources (or inputs) will be, in particular, intellectual works, which are deemed a form of external inspiration. A selection effort is sometimes made, with a more or less human-supported level of intervention, but it can also be performed by a robot as machine aggregation. This training set takes on the form of a database. It is injected into an algorithmic process (expressed in software form) which will identify correlations, correspondences and similarities in terms of structures. This deconstruction by the algorithm also makes it possible to represent the initial data in a machine-readable way.

Based on the objective identified, an objective utility function must be defined: here, for example, it is a question of indicating which characteristics should be highlighted. This function represents the effectiveness of the solution produced by the algorithm, i.e. measuring the gap between the expectations and the result of the algorithmic process. The learning function will as such be to decrease the gap, to minimize the error between the product of the algorithm and the results expected. The result of this minimization work is to determine the parameters of an inference model¹¹, which comprises the set of values characterizing the neural network in this type of model, i.e. the values characterizing the links between neurons. These values may be organized in the form of a database, which is automatically managed.

This inference model crystallizes fundamental elements of the ordering party's wishes: choice of the training base (the inference model depends on the initial matrix), choice of the model and the utility function. It is this object, this stabilized model (once the training has been completed), which will lead to production by fulfilling its function. As a learning repository, it is as such the starting point for creation. The inference model can then be used, even by a third party who has not participated in its design. The user only needs to "click" to generate the production of a creation¹². The inference model can also be modulated downstream, either by penalizing the output elements considered as bad and by promoting those, on the contrary, appraised as good¹³, or by reinjecting the products created into the training base¹⁴.

Amongst the various experiments carried out in cultural sectors, not all have the same level of maturity; some, which settle for accompanying the human process of creation, are largely present in the audiovisual and publishing sectors; others, which strive to emancipate themselves from it, are more so found in music and the art market. In many cases, it is not a question of replacing the human process of creation, but of facilitating the work and

¹¹ V. J. Deltorn, "Quelle(s) protection(s) pour les modèles d'inférence ?" (What protection(s) for inference models?), *Cahiers Droit, Sciences & Technologies*, 7/2017, p. 127, spec. p. 133: inference models are "made up of a set of parameters determined during the learning phase and organized into a data structure."

¹² This was the model proposed by Jukedeck (musical AI), which proposed (for a fee) an inference model to create musical jingles based on minor selections such as style and emotional characteristics.

¹³ This selection and assembly work was notably carried out by François Pachet for the production of *Daddy's Car*, a Beatles-style track produced using AI.

¹⁴ This is how David Cope creates classical music.

inspiration of the creator by reducing constraints. More ambiguous are works which aim to mimic, to create "like" other artists, or even to initiate completely new works which greatly limit human intervention.

1.5.1. Accompanying the (human) process of creation

In the **audiovisual sector**, as the production of a film or series is complex and costly, artificial intelligence tools intend to simplify all the stages and to streamline decision-making during pre-production, shooting and post-production (crowd simulation, production of special effects and animations). AI is also used in producing special audiovisual productions such as ads, trailers and clips for social media. For other types of productions, professionals are more reticent. In 2016, a Kickstarter campaign for a horror film entitled *Impossible Things* financed the production of a feature film whose script was co-written with a machine. However, the machine's part in this creation must be put into perspective, since the actual writing was carried out by a human. On the contrary, the same year, *Sunspring*, a short film created by director Oscar Sharp and artificial intelligence researcher Ross Goodwin for the *48-Hour Film Challenge* at the London Sci-Fi Film Festival, was written entirely by the machine. However, the experiment did not produce the results expected; after implementing a dozen or so science fiction scripts in an artificial intelligence machine nicknamed *Benjamin*, the machine did indeed propose a complete but completely incoherent script, with implausible dialogues and scenes and without any narrative logic.

Advances in "natural" language processing have been applied in various practical ways in **publishing**, leading to the emergence of tools for automating certain repetitive tasks and for generating simple texts, particularly in the press. Beyond that, producing literature by mimicking human skills has been one of the intentions of AI since its very beginnings. The WASP (Wishful Automatic Spanish Poet) poetry generator, the SPAR (Small Poem Automatic Rhymer) system, the Raconteur the real events narrator and the PropperWryter narrative generator, used in the musical *Beyond The Fence* are a few examples of possible applications. Researchers are also working to make it write a book "in the style of", by reproducing a famous author's universe. The finalist of the Nikkei Hoshi Shinichi Literary Award in Japan presented a short novel produced in collaboration with AI. *Just This Once* (1993), written in the style of Jacqueline Susann by AI in collaboration with its programmer, Scott French, sparked reflections on the sharing of the creativity part between the machine and the other players involved. Currently, natural language generation (NLG) is effective mainly for creating simple, informative press articles and releases. Being capable of composing literary works which hold the reader's attention over hundreds of pages, through the use of skilfully-handled language and an overall orchestration of events, is not yet within the reach of artificial intelligence.

1.5.2. Emancipating itself (a little) from human creation

Machine-associated creativity experiments have been undertaken for a very long time in graphic arts; as early as 1973, Harold Cohen developed the evolutive AARON program, which combines a software AI with automated painting devices to enable the machine to automatically produce paintings in a given style. The systems which use deep learning and neural networks, a far cry from the primitive robot-artist Harold Cohen initiated, are now attracting attention as regards their applications for artistic images. The J. Walter Thompson Amsterdam advertising agency initiated *The Next Rembrandt* project on behalf of Dutch ING Bank with technological support from Microsoft. This "in the style of" artificial intelligence

creation, unveiled in Amsterdam on 5 April 2016, was created using the database extracted from the entire collection (346 portraits) of Rembrandt's work.

Other programs seek to reveal new creations. As such, Google has developed an artificial intelligence program, DeepDream, which is in line with the CNN (Convolutional Neural Networks) neural algorithms. It led to the production of new creations, exhibited in galleries in San Francisco in 2016 and sold for up to \$8,000. Other types of neural algorithms, **GANs**, have been experimented with to produce new artworks. The artistic collective **Obvious**, founded by three young Frenchmen, integrated data from almost 15,000 classical portraits from the 14th to the 20th century into an algorithm and obtained new creations representing an imaginary family, the **Belamy** family. In this series, one of the paintings, entitled Edmond de Belamy, was sold for over \$430,000 in October 2018 by Christie's auction house in New York¹⁵. In actual fact, the novelty does not lie so much in the technical or artistic prowess, or even in the sale (other portraits of the Belamy family had previously been sold at more modest prices) as in the art market's ability to communicate about the first-ever artificial intelligence-created painting sold at an auction by auctioneers. The high price therefore more so illustrates the market's fascination with this new creative technique. Yet, the members of the collective themselves acknowledge that their experience is nothing original. The GANs technique had been developed by Ian Goodfellow and his team at an earlier stage; moreover, the algorithm source code was taken from the open source-available one created by Robbie Barrat who, incidentally, was subsequently touched by the situation given the project's commercial use.

AI researchers deem that this work's fundamentals have already been proven. However, the media frenzy caused by this sale shed light on research such as that of the artist Mario Klingemann, one of the pioneers of the use of neural networks and GANs, who developed "neurography" for creating new art forms. On 6 March 2019, a real-time-evolving work by this artist was sold at auction in London by Sotheby's for £51,000 and became the second AI-originating creation to be sold at auction.

Like graphic arts, the process of **creating music** was linked to the machine at an early stage. As early as the 1980s, the American composer and scientist David Cope had his EMI (Experiments in Music Intelligence) tool produce hundreds of productions in the style of Bach, Mozart and Rachmaninov. Today, most systems operate on the basis of deep learning networks and identify, among large amounts of data, rhythmic and melodic patterns so as to produce new tracks. Although the various tools are capable of generating a "finished" musical composition, they do not all offer humans the same opportunities in the creative process. Some tools require no prior knowledge (Amper, Jukedeck, etc.), whilst others rely on some knowledge of computer coding and music writing (IBM, Google).

Some commercial applications provide "**turnkey**" solutions which enable the neophyte, or the "rushed" creative, to obtain a **ready-to-use composition**, such as for jingles, drafts, music for news reports, etc. They can be used, on a simple self-service platform, to create compositions corresponding to understandable and user-defined parameters which, for example, reflect an emotion (joyful, disturbing, exotic) or a genre (rock, funk, etc.). The businesses concerned attract customers with compositions which, although not always of high quality have commercial value, by facilitating access to low-cost music which may be a cause for concern as regards substituting copyright-protected compositions. However, whilst it is highly likely

¹⁵ To view the portrait: <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>.

that AI will occupy a share of the market (lift music, music "to drive to", etc.) in the coming years, it will also lead to boosting the creativity of authors in more qualitative music segments which AI is still unable to produce.

Other tools developed by large groups or research laboratories **accompany "seasoned" users** (musicians, composers, programmers) by enabling them to intervene on a large number of parameters and, as such, to obtain a sort of enhanced complexity / customization of the work. AI pursues this process of deconstructing a piece by isolating each characteristic much more precisely. As such, IRCAM researchers have been developing customized tools and algorithms for each artist who can choose to adjust the algorithm's parameters in addition to using the training data set. The computer scientist, who is involved in the choice of parameters and also has musical knowledge, plays an important role in the creative process. Flow Machines, an artificial intelligence project created in 2012 by a team led by François Pachet, then director of Sony Computer Science Laboratories Paris, is presented as a tool for optimizing musical composition. In 2016, the Flow Machines developers unveiled two pieces "in the style of": Bach-style *DeepBach* and Beatles-style *Daddy's Car*¹⁶. In the end, the aim is not just to mimic a style, but to create a new work which will be the result of human-machine work.

With these assumptions in mind, AI is a **tool** - admittedly a very sophisticated one - in the hands of the author. The principle must therefore continue to be one of **human pre-eminence**. It is, therefore, not necessary to modify legal qualifications for what can be called *AI-assisted* creations.

From a legal stance, reasoning should change if the creation is in this case *generated* by AI without the possibility of identifying an author, in the classical sense of the term.

¹⁶ To listen: https://www.youtube.com/watch?v=LSHZ_b05W7o.

Part 2. – The status of cultural products generated by "creative" AI

Legal experts' interest in the intersection between computer science and copyright is not new¹⁷, in a discipline which is strongly influenced by technology.

The assumption adopted here is that of an *AI-generated* creation, for which the classical identification of a natural person as author-creator is made complex by the creative process. As such, we can speak about "**creative AI**", as AI which enables the creation of a cultural production which *resembles* a work. This creative AI (i.e. the tool) is itself a **hybrid** legal object (undoubtedly also a complex work, as multimedia creation can be). The qualification of this multi-faceted element will only be addressed rapidly (and partially), as it is located upstream of the creation produced for which the status is sought.

Downstream, clearly, the **boundary** between *assisted* creation and *AI-generated* creation is difficult to draw¹⁸, all the more so as the use of a machine to create is without consequence on the qualification, in the spirit of the principle of genre indifference¹⁹ and also of the more general principle of technological neutrality. The unfortunate experience of legislative errors as regards the protection of photography should serve as a lesson. Whether a creator uses a brush, a chisel, a camera or a high-tech instrument does not change the reasoning. Also, as long as AI remains a tool in the hands of an author-creator, a kind of **principle of human pre-eminence** must be agreed upon, which removes the machine and enables copyright rules to be applied in a classical manner. Consequently, whenever a human being intervenes creatively in creation, it makes little difference which tool is used.

It is only when human intervention departs from the classical approach given autonomous AI generation that the proposed reasoning will become applicable.

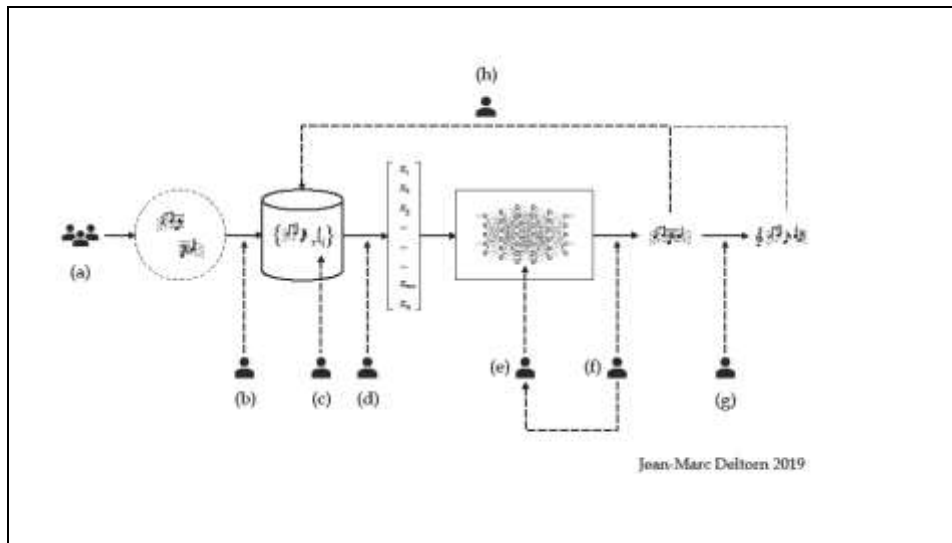
Caution is even more important since many players are likely to intervene at the various stages of the algorithmic creation chain, upstream, by selecting the training base, downstream, during the post-production and/or curation phase, as well as in the intermediate stages for selecting the signal representation method and the choice of model, as shown in the diagram below.

¹⁷ See *Computer and Copyright*, Proceedings from the 57th ALAI Congress, Éd. Yvon Blais Inc., Canada, 1990.

¹⁸ Y. Gaubiac, "Objet du droit d'auteur. Œuvres protégées. Œuvres créées avec un ordinateur (CPI, art. L. 112-2)" (Copyright subject matter. Protected works. Computer-created works (French Intellectual Property Code, Art. L. 112-2), J.-Cl. Literary and artistic property, fasc. 1164.

¹⁹ French Intellectual Property Code, Art. L. 112-1.

Diagram illustrating the production of an AI-generated artistic creation



Legend (human intervention):

- (a) Creating the initial works
- (b) Selecting the works (for developing a training base)
- (c) Possible label assigning²⁰ (li) to each of the entities selected (in a supervised learning approach)
- (d) A signal representation method is selected (the training data is written in a format that is easily read by the machine, in this case a parameter vector "x1" ... "xn")
- (e) The choice of model type (in this case a neural network), the details of its architecture and its parameters are chosen, as well as the training phase terms (cost function, etc.)
- (f) During training, the entities produced can be observed, selected and used to modify either the model or the learning characteristics
- (g) A "post-production" stage can be used to select and modify the entities produced by the generative model.
- (h) Some of the entities produced can possibly be reintroduced into the training base (to guide subsequent learning. David Cope worked in this way for example, see *above*).

Admittedly, not all of these players are systematically involved in the machine-learning creation process. However, this diagram illustrates the major interest of revealing **the number and variety of potential human players** in the creative chain. In addition to these programmers, trainers, curators, we could add investors, ordering parties, users and licence holders... Becoming emancipated from human creation therefore remains relative. Yet, can these creative players - or at least some of them - be considered as authors, as regards copyright?

Several questions as such arise, for example, this production's status and its protection, identifying an author. In copyright law, these questions often interconnect insofar as the definition of the subject matter of copyright (the work), its protectability (originality), its subject (the author) and sometimes even its owner are mixed together. It is worth mentioning that the circular structure of French copyright law, in which the key notions refer to each other, prompts this.

A priori, the personalist French conception seems to exclude the production of Ai-generated cultural content from protection, to the point, moreover, that some theorists reject its application in this instance (2.3). Notwithstanding, the appropriation of these products by

²⁰ Labels are keywords, metadata, for training the model as part of a supervised learning approach.

means of copyright has been suggested by the European Parliament, which proposes to define "criteria of "own intellectual creation" applicable to protectable works created by computers and robots"²¹. This justifies us firstly examining the possibility of copyright protection (2.1). Yet, the analysis is not straightforward and other means of appropriation could be brought to light (2.2). This diversity of options leads us to ask ourselves which solution should be favoured (2.4).

2.1. The copyright possibility

Although AI can be protected as a tool, these upstream protections do not impact downstream production (2.1.1). Hence the need to consider copyright protection, provided that the conditions of access to protection can be adapted to the artificial creation, with a new interpretation (2.1.2), and that an initial owner can be identified (2.1.3).

2.1.1. Existing protection

The more sophisticated artificial intelligence becomes, the more the investments to which it gives rise bring it within the reach of a large number of potential users, and the more important it is to separate the regime of artificial intelligence, on the one hand, from the regime of the products it is likely to generate, on the other hand.

However, the protection currently provided only concerns the matrix, i.e. the tool or innovation which artificial intelligence itself constitutes - and not the products it generates for the consumption of all. Some components are protected by an intellectual property right (2.1.1.2) whilst others are excluded (2.1.1.1).

2.1.1.1. Non-protection of the algorithm

The algorithm itself, as "a description of a finite and unambiguous sequence of steps (or instructions) to obtain a result from input material"²² in machine-readable language, is not protected as such by copyright. It is assimilated to a **method** or a **simple idea** which is, "intrinsically and through its purpose free to be used" (Desbois). In this respect, Article 9.2 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) states that "Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such". The rule is repeated in Article 1.2 of Directive 2009/24/EC²³, which states that "Ideas and principles underlying any part of a computer program, including those underlying its interfaces, shall not be protected by copyright under this Directive"²⁴. Both French²⁵ and European²⁶ jurisprudence have also pointed out the

²¹ European Parliament report with recommendations to the Commission on Civil Law Rules on Robotics, 27 Jan. 2017, 2015/2103(INL), p. 32.

²² CNIL (French Data Protection Authority), *How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence*, Dec. 2017, p. 15.

²³ Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (codified version).

²⁴ *Adde* Recital 11, excluding "the ideas and principles underlying logic, algorithms and programming languages".

²⁵ 1st Civ. Ct. Cass., 13 Dec. 2005, not. JCP 2006, I, 103, obs. C. Caron; JCP E 2006, 1895, § 1, obs. M. Vivant, N. Mallet-Poujot and J.-M. Bruguière; P. Belloir, "L'exclusion de la protection des fonctionnalités d'un logiciel par le droit d'auteur" (Exclusion of copyright protection of software functionality), RLDI 2006/14, No. 396; 1st Civ. Ct. Cass., 14 Nov. 2013, No. 12-20.687, RIDA Jan. 2014, p. 493 and p. 399, obs. P. Sirinelli; LEPI 2014, 174, obs. S. Chatry; Intell. Propr. 2014, p. 56, obs. A. Lucas.

exclusion of **functionalities** from the protection of ordinary law; however, the functionality of a software program is expressed by an algorithm, which is the logical statement. Likewise, the algorithm, as a simple mathematical method, is not, as such, patentable²⁷.

2.1.1.2. Protection of other AI components

Article 10.1 of the TRIPS Agreement states that "**Computer programs**, whether expressed in source or object code, shall be protected as literary works under the Berne Convention". The formalization of the algorithm in coded expression of a program is thus reserved by a special right within copyright law, notably stemming from Directive 2009/24²⁸ and already introduced in France by the Act of 11 July 1985²⁹. Artificial intelligence can as such benefit from this protection, but without it being able to extend to the method itself (the algorithm) or to the underlying ideas (due to the existence of a reverse engineering exception, open to the legitimate user, for accessing the program's functionalities³⁰).

Some artificial intelligences will be potentially patentable as "a new invention involving an inventive step and capable of industrial application" (Art. L. 611-10 of the French Intellectual Property Code). It is true that computer programs are excluded from patentability. But recent international preliminary examination reports under the Patent Cooperation Treaty have suggested that a number of applications developed by Deep Mind, a company acquired by Google, may be eligible for protection under patent law³¹ – they are currently under consideration by the European Patent Office (EPO). Generally-speaking, machine-learning algorithms may be deemed innovative technical solutions, included in a broader and rapidly expanding category which the EPO calls "**computer-implemented inventions**" (CII)³².

Likewise, it is not disputed that **databases** used by artificial intelligence may be eligible for specific protection. Directive 96/9/EC of 11 March 1996³³ establishes a dual protection regime: copyright for the contents of the database (its organizational principles and

²⁶ CJEU, 22 Dec. 2010, *BSA*, case C-393/09, see pts 42 to 44, RIDA Jan. 2011, p. 417 and 213, obs. P. Sirinelli; Electr. Comm. com. 2011, com. 42, note C. Caron; Intell. Propr. 2011, p. 205, obs. V.-L. Benabou; D. 2011, p. 2364, obs. C. Le Stanc; LEPI 2011, 35, obs. A. Lucas. – CJEU, gr. ch., 2 May 2012, *SAS Institute*, case C-406/10, v. spec. pt 32, RIDA Jul. 2012, p. 341 and p. 181, obs. P. Sirinelli; Electr. Comm. com. 2012, com. 105, note C. Caron; D. 2012, p. 2836, obs. P. Sirinelli; A & M 2012/6, p. 547, note B. Michaux; Expertises 2012, p. 265; RLDI 2012/83, No. 2768, obs. C. Castets-Renard; Europe 2012, com. 284, obs. L. Idot; Industr. Propr. 2012, com. 61, note N. Bouche; Intell. Propr. 2012, p. 423, obs. V.-L. Benabou; RTD com. 2012, p. 536, obs. F. Pollaud-Dulian.

²⁷ French Intellectual Property Code, Art. L. 611-10, 2, and EPC, Art. 52 (2).

²⁸ Afore. codified version of Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs.

²⁹ Act No. 85-660 of 3 July 1985 relating to copyright and the rights of performers, producers of phonograms and videograms and audiovisual communication companies.

³⁰ This special right explicitly authorizes the legitimate user of the software to observe, study or test the functioning and safety of the software "*in order to determine the ideas and principles underlying*" the software (Art. L. 122-6-1, III of the French Intellectual Property Code, incorporating Article 5.3 of the aforementioned Directive).

³¹ See for example the application W02018048934 – "*Generating Audio Using Neural Networks*". <http://ipkitten.blogspot.com/2019/01/full-speed-ahead-for-deepminds-ai.html>.

³² EPO, *Guidelines for Examination*, Nov. 2018, G-II, 3.6: "Computer-implemented invention" is an expression intended to cover claims which involve computers, computer networks or other programmable apparatus wherein at least one feature is realised by means of a computer program." - See the examples given by J.-M. Deltorn, "The examination of computer implemented inventions and artificial intelligence inventions at the European Patent Office", Industr. Propr. No. 3, 2019, dossier 4.

³³ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

architecture, in particular)³⁴ and the *sui generis* right for its content³⁵. As such, AI used in a creative context will mobilize two types of databases³⁶: the training set composed of collected and assembled data and the encoding of the inference model itself, as a collection of values, which allows the AI to be configured.

2.1.1.3. The need to protect AI-generated "products"

All the aforementioned protections – copyright, *sui generis* right, patent law – say nothing about the legal regime applicable to the products which are generated by the intelligent machine. We could be tempted, by invoking the model of plural works, to grant the status of authors of the AI-generated product to the various upstream creators. However, this approach is fragile because the various elements are not found in the final creation³⁷ (whereas this is the case, for example, for the complex work of a video game). Moreover, these upstream protections **are not sufficient** to protect and secure the investments in artificial intelligence³⁸, which can naturally be guided by the profit made from the use of the machine.

In addition, the absence of protection of artificial intelligence-generated products would leave the **interpreters** of these AI creations without rights, since the protection conferred by the regime of related rights implies, for these players, the existence of a copyright on the interpreted work (Art. L. 211-1 of the French Intellectual Property Code).

Finally, this lack of protection for AI achievements would undoubtedly be the cause of **circumvention**. Circumvented protection could then, in practice, develop, as there is no requirement to reveal the creative processes. As such, a natural person could conceal that a creation was generated using an AI tool and benefit from the presumption of authorship³⁹ by attaching their name to the creation. Likewise, a business, obtaining the qualification of collective work (enabling it to access the initial ownership of the rights)⁴⁰ or acting in infringement for the benefit of the jurisprudential presumption of ownership⁴¹, could claim the rights by keeping silent on the exact manner in which the works in question were produced.

³⁴ Dir. 96/9/EC, Art. 3.1: "In accordance with this Directive, 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."

³⁵ Dir. 96/9/EC, Art. 7.1: "Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database."

³⁶ V. J.-M. Deltorn, "*Disentangling deep learning and copyrights*", AMI - tijdschrift voor auteurs-, media- en informatierecht 2018/5, p. 172.

³⁷ See in this respect the position of the French AIPPI group, "Copyright in artificially generated works", pres. E. Treppoz, Apr. 2019. – For an international analysis, see the Summary Report – 2019 – Study Question: Copyright in artificially generated works.

³⁸ In this respect, on the WIPO website, A. Guadamuz, "Artificial Intelligence and Copyright", https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html

³⁹ French Intellectual Property Code, Art. L. 113-1.

⁴⁰ French Intellectual Property Code, Art. L. 113-2, par. 3 (definition) and L. 113-5 (regime).

⁴¹ 1st Civ. Ct. Cass., 24 March 1993, *Aréo* (2nd judgement), RTD com. 1995, p. 418, obs. A. Françon; *JCP G* 1993, II, 22085, note F. Greffe; *RIDA* Oct. 1993, p. 191, obs. A. Kéréver: "[in] the absence of any claim on the part of the natural person(s) who took the photographs, these acts of possession were such as to give rise to a presumption vis-à-vis third party infringers that [the company] was the owner of the author's incorporeal property right in these works, whatever their classification". – Formulating the rule in a recital of principle: 1st Civ. Ct. Cass., 9 Jan. 1996, *RIDA* Jul. 1996; *D.* 1996, summ. p. 285, obs. J.-J. Burst; *JCP* 1996, II, 2643, note X. Daverat.

All the more so as the boundary between AI-assisted creations and AI-generated creations is porous: it is more of a *continuum* than a clear separation.

For these reasons, it is important to test **copyright, which could, with a new interpretation⁴², include AI-generated creations.**

2.1.2. The renewed application of copyright

French copyright law is humanistic: through copyright, it is also the natural person of the author who is protected. This **personalist** conception has drawn an extremely flexible field of protection, where the aesthetic merit or novelty of the protected work is absent but where the person of the author is central.

Article L. 111-1 of the French Intellectual Property Code as such provides that "the author of an intellectual work enjoys an exclusive incorporeal property right over this work, *by the mere fact of its creation*, which is enforceable against all" and Article L. 112-1 imposes a general principle of non-discrimination between works⁴³. However, French legislation does not lay down any positive definition of an intellectual work – unlike the provisions relating to patent, trademark and design law. In this respect, it does not deviate significantly from the very cautious definitions given in the Berne Convention ("The term "literary and artistic works" shall include every production in the literary, scientific and artistic domain, whatever the mode or form of its expression"). No one doubts, however, that an intellectual work is, in French law, a **creation in an original form**. Although the essential criterion for protection is, in France as in international law, the originality of the creation, there is no statutory definition either.

This elliptical nature of the concept has enabled its **flexibility** – and hence a malleability of law – which explains the success of copyright and justifies the consideration of using it to provide protection for creations which, until then, were not likely to exist. Although copyright may turn out to be a malleable legal category, the idea of incorporating the artificial intelligence-generated products into its regime could nevertheless come up against several obstacles, caught up in criteria that would then have to be newly interpreted: creation (2.1.2.1), originality (2.1.2.2) and the author (2.1.2.3).

⁴² See not. J. Larricau, "Le robot et le droit d'auteur" (The robot and copyright), in *Mélanges A. Lucas*, LexisNexis, 2014, p. 465; B. Michaux, "Singularité technologique, singularité humaine et droit d'auteur" (Technological singularity, human singularity and copyright), in *Droit, normes et libertés dans le cybermonde* (Law, norms and freedom in cyberspace), *Mélanges Y. Pouillet*, Larcier, 2018, p. 401; P.-Y. Gautier, "De la propriété des créations issues de l'intelligence artificielle" (Ownership of creations resulting from artificial intelligence), *JCP G* 2018, 913; J.-M. Deltorn, "Droit d'auteur et créations des algorithmes d'apprentissage" (Copyright and the creation of learning algorithms), *Intell. Propr.* 2016, No. 58, p. 4; F. Macrez and J.-M. Deltorn, "Authorship in the age of machine learning and artificial intelligence", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261329; A. Cruquenaire, A. Delforge, J.-B. Hubin, M. Knockaert, B. Michaux, T. Tombal, "Droit d'auteur et œuvres générées par machine" (Copyright and machine-generated works), in *L'intelligence artificielle et le droit*, dir. H. Jacquemin, A. de Steel, Bruxelles, Larcier, 2017, p. 189; A. Bensamoun, "Les créations générées par une intelligence artificielle ou la tentation du droit d'auteur" (Creations generated by artificial intelligence or the temptation of copyright), in A. Bensamoun and F. Labarthe (dir.), *Culture et numérique, Rencontre franco-québécoise*, Mare et Martin, coll. PUS, 2020, to be published; A. Bensamoun, "Intelligence artificielle et propriété intellectuelle" (Artificial intelligence and intellectual property), in *Droit de l'intelligence artificielle*, dir. A. Bensamoun and G. Loiseau, LGDJ-Lextenso, coll. Les Intégrales, 2019, p. 235.

⁴³ "The provisions of the code hereof protect the rights of authors on all intellectual works, irrespective of their genre, form of expression, merit or purpose."

2.1.2.1. Are AI-generated products "creations"?

The first question which arises is whether it is possible, when talking about artificial intelligence-generated products, to describe them as "creations": is there really a creative process at work when artificial intelligence works? To answer this question, we need to take an interest in the notion of creation – perhaps even to the point of reinventing it?⁴⁴

Creation, defined as a transformation of what is real, what exists, traditionally refers to the notion of **consciousness**. In this respect, we consider that creation must be conscious, as implying "a minimum of intellectual mastery of the creative process"⁴⁵, which excludes the creations of mentally ill and *infans* from protection. In this respect, we can state without hesitation that AI has no conscience. And we could stop the demonstration there and simply deny AI's creative capacity⁴⁶.

However, the criterion of consciousness of creation is not unanimously accepted⁴⁷, as some prefer to assess the work from its result and not from the process which led to its existence⁴⁸. Moreover, anonymous and pseudonymous works⁴⁹ are well protected (French Intellectual Property Code, Art. L. 113-6), whereas it is impossible to probe the level of consciousness of their authors. Moreover, the judge has always disregarded such questions. Arguments based on the fact that the work was created in a state of hallucination or by accident were never accepted, likewise for the argument where the work was created under the influence of a "*psychic agent*", a sort of spirit summoned by a medium⁵⁰. The argument is not without impact: as Jacques Larrieu explains, it would not occur to anyone to distinguish Van Gogh's works based on whether they were painted at different stages of his life, when his level of consciousness of the world around him was not the clearest (remember the self-portrait with the ear cut off...). The same could be said of Baudelaire's *Flowers of Evil*, some passages of which were written under the influence of "artificial paradises".

Although consciousness is not an overwhelming requirement, creation still refers to the notion of the **creator as a physical person**. The act of creation protected by copyright is pitted against the mechanical nature of machine-generated content "production". As scientists

⁴⁴ M. Vivant, "Intelligence artificielle et propriété intellectuelle" (Artificial intelligence and intellectual property), *Electr. Comm. com.* 2018, study 18, No. 10.

⁴⁵ A. Lucas, A. Lucas-Schloetter, C. Bernault, *Traité de la propriété littéraire et artistique* (Treatise on Literary and Artistic Property), *op. cit.*, No. 57 (and No. 59 on AI-generated creations).

⁴⁶ L. Devillers, *Des robots et des hommes, Mythes fantasmes et réalités* (Robots and Humans: myths, fantasies and reality), Plon, 2017, spec. p. 79.

⁴⁷ M. Vivant and J.-M. Bruguière, *Droit d'auteur et droits voisins* (Copyright and related rights), Dalloz, 4th ed., 2019, No. 119; J. Larrieu, "Le robot et le droit d'auteur" (The robot and copyright), in *Mélanges A. Lucas*, LexisNexis, 2014, p. 465 s., spec. p. 467, and B. Michaux, "Singularité technologique, singularité humaine et droit d'auteur" (Technological singularity, human singularity and copyright), in *Droit, normes et libertés dans le cybermonde* (Law, norms and freedom in cyberspace), *Mélanges Y. Pouillet*, Larcier, 2018, p. 401 s., spec. p. 410.

⁴⁸ Not. N. Enser, *Conscience et création en droit d'auteur* (Consciousness and creation in copyright law), thesis, Paris-Sud/Paris-Saclay University, No. 373: "creation can only be understood as a result which can be witnessed, without any need to focus on the process which led to its creation".

⁴⁹ French Intellectual Property Code, Art. L. 123-3.

⁵⁰ International examples exist: see A. Bridy, "*The Evolution of Authorship: Work Made by Code*", 39 *Colum. J.L. & Arts* 395 (2016), p. 18 s., citing the judgements *Bell v. Catalda* 191 F 2d 99 (2d Cir 1951); *Penguin Books v. New Christian Church*, 55 USPQ 2d 1680 (SDNY 2000); *Cummins v. Bond* [1927] 1 Ch 167; *Leah v. Two Worlds Pub'g* [1951] 1 Ch 393, concerning works supposedly created by spirits, voices or a God.

explain, software will be forever "devoid of emotion"⁵¹, which is inherent to the artist who operates the AI and selects and develops the models it offers. There will systematically be as such a human being behind cultural creation.

Legally-speaking, in France, this requirement stems from the very nature of the humanistic and personalist copyright law, and the spirit of the law. In this respect, it may be noted that the law refers to the author even before specifying the work (French Intellectual Property Code, Art. L. 111-1 and -2). Moreover, the attribution of rights to a natural person is mentioned for certain types of works⁵² and it is hard to understand why the individual would depart from this point. For its part, the Court of Cassation ruled that "a legal person cannot have the status of author"⁵³, implying *a contrario* that only a natural person can be an author. This link, between creation and natural person, moreover appears to be shared at international level⁵⁴. In the United States, the registration of a work at the Copyright Office is only permitted if it has been produced by a human being⁵⁵. In Australia, the Supreme Court refused protection to a database automatically generated by AI⁵⁶.

Creation therefore appears to be the preserve of human beings, the fruit of human imagination. However, it is not a "magic act": it complies with rules, even when it has been decided to deviate from them. AI is by nature **biomimetic** – it simulates, mimics the human; it can therefore also mimic the arbitrariness of human choices, in terms of creative activity. Therefore, the rules of creation could probably be computer-coded⁵⁷, at least in part.

Nevertheless, a creator, natural person needs to be identified, but certainly from a more distant stance from creation than in the traditional approach to law, for example a human being whose imagination may have been coded. The act of creation could therefore be characterized as long as the idea of **different creation** and a **more indirect creator** is accepted (see 2.1.2.3 below).

⁵¹ H. Bersini: <https://lactualite.com/techno/lintelligence-artificielle-ne-sera-jamais-artiste/>

⁵² French Intellectual Property Code, Art. L. 113-2 for collaborative work, Art. L. 113-7, par. 1st, for audiovisual work, Art. L. 113-8, par. 1st, for radio work.

⁵³ 1st Civ. Ct. Cass., 15 Jan. 2015, No. 13-23.566, D. 2015, p. 206 and p. 2215, obs. C. Le Stanc; RTD com. 2015, p. 307, chron. P. Gaudrat; Légipresse 2015, p. 223, note N. Binctin; Electr. Comm. com. 2015, No. 19, obs. C. Caron; Industr. Propr. 2015, No. 3, com. 25, obs. N. Bouche.

⁵⁴ A. Dietz, "The concept of authorship under the Berne convention", RIDA Jan. 1993, No. 155, p. 3, spec. p. 11-13; S. Ricketson, "People or Machines: The Berne Convention and the Changing Concept of Authorship", 16 Columbia-VLA Journal of Law & the Arts, 1 (1991).

⁵⁵ *Compendium of U.S. Copyright Office practices*, § 306 and § 313.2.

⁵⁶ Australian Supreme Court, *Telstra Corporation Limited v Phone Directories Company Pty Ltd* [2010], FCAFC 149: "Following the High Court's reasoning in the 2009 decision of *IceTV v Nine Network*, the Full Federal Court recognised that to be original, a work must:

1. Not be copied;

2. Originate from a human author; and

3. Be the result of independent intellectual effort.

Secondly, the Full Court found that the extraction phase effort did not originate from an individual or group of individuals, but rather from a computerised process of storing, selecting, ordering and arranging the data to produce the directories in their published form. The Court emphasised that it was the extraction phase work that constituted the relevant effort for the purposes of originality."

⁵⁷ See M. A. Boden's demonstration, "Computer Models of Creativity", Association for the Advancement of Artificial Intelligence, 2009, p. 23.

2.1.2.2. Are AI-generated "works" "original"?

Above all, copyright protection presupposes that the creation for which protection is sought is deemed to be "original". Can AI-generated creations meet this criterion?

In France, there is traditionally a **subjective** conception of originality, defined since Desbois as the **imprint of the author's personality**. Jurisprudence and doctrine have adopted an anthropocentric conception of copyright, which protects the creative work of the human being. The work must therefore be traceable back to its author, who is its *originator*, who has left **traces** of their person in it – even if personal performance is not a prerequisite for this⁵⁸.

Yet, this classical approach is doubly questioned: in national law, by a certain **objectification of the originality criterion**, such as the "imprint of intellectual input"⁵⁹; in European law, since the CJEU has made it an **autonomous concept of Union law**, understood as "an intellectual creation specific to its author"⁶⁰, even if the concept seems to retain a certain degree of subjectivity in European law⁶¹.

Criticism also comes from the doctrine. Some authors as such consider that the value attributed to the work by the public must be taken into account in the assessment of originality: this objectification would take copyright away from the person of the author and would make it possible to envisage that any product which would have been recognized, in measurable terms, as being intellectual work by its public could receive copyright protection⁶². Other authors suggest a more objective approach to the condition – perhaps also

⁵⁸ 1st Civ. Ct. Cass., 13 Nov. 1973, No. 71-14.469; Civ. Bull. I, No. 302; D. 1974, jur. p. 533, note C. Colombet; JCP G 1975, II, 18029, note M.-Ch. Manigne; see also B. Edelman, "La main et l'esprit" (The hand and the mind), D. 1980, chron. p. 7.

⁵⁹ 1st Plen. Ass., 7 March 1986, No. 83-10477, *Babolat v/ Pachot*, see not. JCP E 1986, II, 14713 and JCP G 1986, II, 20631, note J.-M. Mousseron, B. Teyssié and M. Vivant; JCP E 1986, I, 15791, obs. M. Vivant and A. Lucas; D. 1986, p. 405, conclusions from Advocate General Cabannes, and p. 412, note B. Edelman; RIDA, Jul. 1986, No. 129, p. 136, note A. Lucas; RTD com. 1986, p. 399, obs. A. Françon.

⁶⁰ CJEU 16 Jul. 2009, *Infopaq*, case C-5/08; 22 Dec. 2010, *BSA*, case C-393/09; 4 Oct. 2011, *Premier League*, joined cases C-403/08 and C-429/08; 1st March 2012, *Football Dataco*, case C-604/10; 2 May 2012, *SAS Institute*, case C-406/10; 13 Nov. 2018, *Levola*, case C-310/17.

⁶¹ In the aforementioned judgements, certain jurisprudential expressions could lead one to believe that the court had favoured the French subjective conception. However, when the court referred to creative choices or personal touch, it was referring to photographs, for which this subjective approach is present in the text (see recit. 16 of Directive 2006/116/EC of 12 Dec. 2006 on the term of protection of copyright and certain related rights - codified version). We could as such consider that the subjective approach was confined to photography. However, this analysis can no longer be upheld insofar as the court now uses these subjective expressions for other types of works: CJEU, 12 Sept. 2019, case C-683/17, *Cofemel*, recit. 54: "Secondly, it is certainly true that considerations of an aesthetic nature form part of creative activity. However, the fact that a model generates an aesthetic effect does not in itself make it possible to determine whether that model constitutes *an intellectual creation reflecting the freedom of choice and personality of its author*, and thus satisfying the requirement of originality referred to in paragraphs 30 and 31 of this judgement" (emphasis added); CJEU, 29 Jul. 2019, case C-469/17, *Funke*, recit. 19: "As established jurisprudence shows, for subject matter to qualify as a 'work', two cumulative conditions must be met. On the one hand, the subject matter in question must be original, in the sense that it constitutes an intellectual creation of its author. In order for an intellectual creation to be regarded as the author's own, it must reflect the author's personality, which is the case if the author was able to express their creative abilities in the making of the work by making free and creative choices (see, in this respect, judgement of 1st December 2011, *Painer*, C-145/10, EU:C:2011:798, paragraphs 87 to 89)."

⁶² Assumption put forward by S. Yanisky & L. A. Velez-Hernandez, "Copyrightability of Artworks produced by creative robots and originality: the Formality-Objective Model", *Minnesota Journal of Law, Science & Technology*, Vol. 19, p. 18

more in line with judicial reality – proposing, for example, that originality should be understood as "novelty in the world of forms"⁶³.

In any case, should the originality of the work be sought more so *in* the work itself, and not by reference to its author? Moreover, how can this be demonstrated when the author is deceased or anonymous? Also, the criterion of originality should remain **intrinsic to the work**, as a kind of minimum creative illustration⁶⁴ inscribed in form. We could specifically imagine that an intrinsically-original form should somehow pass the **Turing test**⁶⁵ and **be able to mislead as to its origin**. It would then be a creation *assimilable* to a work of art, in so far as it resembles it and fulfils the same functions.

Even without going as far as this renewed interpretation of the notion of originality, we could also accept to seek out a **more indirect**, further removed personality, that of the AI designer in particular, which delimits the algorithmic creation framework by shaping the inference model⁶⁶. Here, it is not a matter of focusing solely on the software – it is obvious that the originality of the software is not reproduced at the stage of the work produced – but at the AI stage as a complex work whose characteristics – the biases – constrain the creation generated.

If we accept a certain reinterpretation of the criteria for accessing protection, then artificial intelligence-generated products may be included within copyright. The original connection to an author still needs to be checked.

2.1.2.3. Do AI-generated "works" have an author?

Contemporary conception which largely predominates in positive law, even that which is far removed from continental humanism (such as in the United Kingdom or New Zealand), essentially implies that the author of a work can only be a **human person**. Hence the need to identify a person in connection with the AI-generated creation, and this is the third issue posed when including AI-generated products under the umbrella of copyright.

Should we then consider that AI can be an author? The idea of a **legal personality for artificial intelligence** was already put forward in a European Parliament Resolution in 2017⁶⁷ and some authors have championed the idea⁶⁸. In its Resolution, the Parliament referred to "creating a specific legal personality for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying

⁶³ M. Vivant and J.-M. Bruguière, *Droit d'auteur et droits voisins* (Copyright and related rights), Dalloz, Précis, 4th ed., 2019, No. 271. – Adde J. Larrieu, "Le robot et le droit d'auteur" (The robot and copyright), *op. cit.*; P.-Y. Gautier, *Propriété littéraire et artistique* (Literary and artistic property), PUF, coll. Droit fondamental, 11th ed., 2019, No. 35.

⁶⁴ Comp. US Supreme Court, *Feist Publications, Inc. v. Rural Telephone Service*, 499 U.S. 340, 345 (1991) "Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity".

⁶⁵ A. Turing, "Computing machinery and intelligence", *Mind*, Oxford University Press, Volume LIX, Issue 36, Oct. 1950, p. 433, <https://academic.oup.com/mind/article/LIX/236/433/986238>

⁶⁶ See the following points.

⁶⁷ European Parliament Resolution of 16 Feb. 2017 with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL).

⁶⁸ A. Bensoussan, *Droit des robots* (Robotic technologies law), Larcier, 2015, p. 41 s. and "La personne robot" (The robotic person), D. 2017, p. 2044; M. Willick, "Artificial Intelligence: Some legal approaches and applications", (1983) 4 :2 AI Mag 5.

electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently"⁶⁹.

The report hereof does not intend to dwell on this option, which appears to be largely **impracticable**⁷⁰: apart from the ethical and philosophical difficulties which it will inevitably address, granting legal personality to artificial intelligence would not solve any of the concrete problems posed by AI. On the one hand, such an upheaval would imply defining which AIs would be eligible for legal personality and the other AIs which are not, an extremely complex qualification operation; on the other hand, this option would come up against the identification of an autonomous patrimony into which copyright would be paid – and which would be liable to be pledged⁷¹; moreover, attributing a legal personality would not solve the question of the authorship linked, in France in particular, to the existence of a natural person – and not just a legal personality; finally, it would risk relieving the AI producers of their own responsibility for the results which AI functioning could lead to⁷². As such, an AI cannot be qualified as an author, or as an inventor, as the European Patent Office has indicated⁷³.

Furthermore, if the identification of a human-creator is necessary under the law, it should perhaps to be sought elsewhere, in a novel creative relationship, an **intermediated relationship**. The author would as such be more distant than in a traditional analysis, but they would still be the basis of choices which influence creation. Ownership options should therefore be assessed.

2.1.3. The feasibility of applying copyright to AI-generated products (ownership issues)

Linking AI-generated creations to copyright implies, once the principle has been accepted, checking their legal feasibility. As such, **by taking an objective approach to the criteria for accessing protection, the copyright route seems to be open, unless a natural person is identified as the author**. The difficulty then lies in the fact that the creative process is obscured by the use of AI, distancing the link between the author and creation, where this link – sometimes seen as an "umbilical cord" – is one of the characteristics of "French-style" copyright.

Several persons could be concerned, from the user (2.1.3.2) to the designer of the AI (2.1.3.1). A legal identification could also be proposed, based on the English law model in particular (2.1.3.3).

⁶⁹ Afore. Resolution, 2015/2103(INL), § 59 f).

⁷⁰ V. G. Loiseau, "La personnalité juridique des robots : une monstruosité juridique" (The legal personality of robots: a legal monstrosity), JCP G 2018, 597; A. Bensamoun and G. Loiseau, "L'intégration de l'intelligence artificielle dans l'ordre juridique en droit commun : questions de temps" (The integration of artificial intelligence into the legal order in ordinary law: questions of time), Dalloz IP/IT, Apr. 2017, p. 239.

⁷¹ R. Pearlman, "Recognizing Artificial Intelligence (AI) as Authors and Inventors under US Intellectual Property Law", 24 Rich. J. L. & Tech, no. 2, 2018; P. Samuelson, "Allocating Ownership Rights in Computer-Generated Works", 47 U. Pitt. L. Rev. 1185 (1985).

⁷² G. Courtois, "Robots intelligents et responsabilité : quels régimes, quelles perspectives ?" (Intelligent robots and responsibility: which regimes, which perspectives?), Dalloz IP/IT 2016, p. 287; M. Bacache, "Intelligence artificielle et droits de la responsabilité et des assurances" (Artificial intelligence and liability and insurance rights), in *Droit de l'intelligence artificielle*, dir.A. Bensamoun and G. Loiseau, LGDJ-Lextenso, coll. Les Intégrales, 2019, p. 69.

⁷³ Refusal of two European patent applications in which an AI, "DABUS", was identified as inventor: epo.org/news-issues/news/2019/20191220.html (reasoned decision announced for January 2020).

2.1.3.1. First hypothesis of ownership: targeting the AI creator

Copyright could, in the first case, be traced back to the AI creator⁷⁴ (i.e. the person who makes the inference model which establishes the framework for the creation). This creator defines the field of possibilities in which the AI will then evolve. In this case, the AI ultimately only executes the creative framework established by the creator. This assumption has the advantage of identifying a human person who will have maintained very close links with the machine, who will have been able to **inject their own biases** into it and as such leave a certain imprint. As P. Samuelson points out, in many cases, AI would have generated the same product regardless of the personality of its human user⁷⁵. It is then the author of the AI who seems best placed to receive the rights to the creations generated by the machine they have created. As an **"indirect" creator**⁷⁶, the AI creator guides, as it were, the algorithmic implementation.

Identifying the AI "creator" as the author of the products it generates does not, however, remove all uncertainties.

One of them lies in **identifying this status**: is it the programmer of the algorithm, which establishes the machine's own space through predefined rules? Is he its "trainer", i.e. the person who corrects the AI according to its spontaneous reactions to randomness, who adjusts its behaviour by critical choices in the learning phase without having to explain these choices through predefined rules⁷⁷? Is he the person who selects the data, the "training set" or "training corpus" and who injects their own selection bias into the mass of data they make available to the AI⁷⁸? Is it, more generally and more systematically, the person who holds the copyright on the AI software itself? Or should we, in this plural creation, be more interested in the person who supervised the creation? With reference to the collective work regime, provided for under Article L. 113-5 of the French Intellectual Property Code, it would then no doubt be relevant to identify the person who initiated the project and supervised it.

A second difficulty lies in identifying the **legal mechanism** by which copyright in computer-generated creation would pass to the "creator" of the infrastructure generating these creations. Theorists, particularly in the United States, have indicated their preference for an analysis in terms of derived right: the AI code would as such be watermarked, albeit invisibly, in the AI-generated creations⁷⁹; in this case, the person holding the rights to the AI would necessarily be the author of the AI-generated work. However, the proponents of this analysis admit that it is

⁷⁴ In this respect, see not. B. Michaux, "Singularité technologique, singularité humaine et droit d'auteur" (Technological singularity, human singularity and copyright), *op. cit.*; T. Lebrun, "L'apprentissage machine est une appropriation" (Machine learning is an appropriation), *Les Cahiers de PI* 2018, vol. 30, No. 3, p. 895; Y. Gaubiac, *JCl. Propriété Littéraire et Artistique* (Literary and Artistic Property), fasc. 1164, which reiterates the importance of the role of the AI creator in the final results it produces.

⁷⁵ P. Samuelson, "Allocating Ownership Rights in Computer-Generated Works", *op. cit.*

⁷⁶ B. Michaux, "Singularité technologique, singularité humaine et droit d'auteur" (Technological singularity, human singularity and copyright), *op. cit.*, p. 413-414; T. Dreier, "Creation and investment: artistic legal implications of computer-generated works", *International Computer Law Adviser*, No. 3-4, vol. 5, 1991, spec. p. 16.

⁷⁷ This is one of the hypotheses advocated by J.-M. Deltorn, "Droit d'auteur et créations des algorithmes d'apprentissage" (Copyright and the creation of learning algorithms), *PI* January 2016, No. 58, p. 4.

⁷⁸ Hypothesis also evoked by J.-M. Deltorn, *ibid.* See also F. Macrez and J.-M. Deltorn, "Authorship in the age of machine learning and artificial intelligence", *op. cit.*, which uses the example of *Emily Howell*, AI created by David Cope based on a corpus of data chosen from classical compositions.

⁷⁹ Hypothesis considered by A. Bridy, "Coding Creativity: Copyright and the Artificially Intelligent Author", *afore.*, p. 25.

subject to the judge being able to identify, in the work generated, a recognized expression of the code itself⁸⁰, which may prove complicated, if not impossible (at the very least novel; the method would have to be invented). From this stance, the AI-generated product cannot be qualified as a composite or derived work⁸¹. The hypothesis actually comes up against the impossibility of finding elements of the first work (the AI itself) in the second work (the AI-generated creation)⁸².

Another basis could be drawn from property law, specifically from the mechanism of **accession by production**⁸³, which enables the owner to acquire the accessories produced by their thing, increasing the basis of the main property. As such, the fruit-bearing thing, AI, generates works which are its fruits and could then, by accession, become the property of the holder or holders of rights on the AI⁸⁴. Jurisprudence has already applied these rules to determine the fate of royalties derived from the use of a patent⁸⁵ or protected trademarks⁸⁶. This opening would be all the more justified since the French Intellectual Property Code itself admits, in particular cases, that copyright is attributed to persons other than the natural person who ensures the creation. This is the case, for example, for posthumous and collective works. Consequently, it can be considered that the modes of appropriation in the French Intellectual Property Code are not closed and leave room for the attribution of rights on intelligent creation by way of accession.

In addition to this practical and theoretical question of implementation, identifying the author of the AI (its creator) as the author of the AI-generated creations may give rise to remarks.

First of all, the distinction between the machine itself and the product created by the machine is becoming increasingly clear. As Anne-Marie Bridy reminds us, one of the commercial arguments of AI creators is precisely related to the promised **interactions** between the machine and its users, interactions which aim to provide the latter with the product which best corresponds to their personality. The mechanism is likely to emancipate the tool from its initial programming. As such, it distances the creator from the daily operation of their algorithm. This is all the more true since the AI creator always has the choice not to commercialize their program and to pass on the generated works as their own. Moreover, as they have already received remuneration in exchange for the sale or licensing of their program, and have already, in the majority of cases, a copyright on AI (to be considered as a complex work), they are not, according to some, the most legitimate person to claim rights on the creations it has generated.

Moreover, the **automatic** nature of authorship has been criticized as making it possible to dispense with an analysis of the originality of the work generated, which cannot by definition

⁸⁰ V. P. Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, *op. cit.*

⁸¹ Art. L. 113-2 and Art. L. 113-4 of the French Intellectual Property Code.

⁸² WIPO, "Recommendations for settlement of copyright problems arising from the use of computers for access to or the creation of works – Creation of protected works by means of computer systems", Copyright 1982, 115, 9.

⁸³ Fr. Civ. Code Art. 546: "Ownership of a thing, whether movable or immovable, gives a right to everything it produces, and to what is incidentally connected to it either naturally or artificially. This right is called the right of accession". – In this respect, see P.-Y. Gautier, "De la propriété des créations issues de l'intelligence artificielle" (Ownership of creations resulting from artificial intelligence), JCP G 2018, 913.

⁸⁴ This is the SACEM's position.

⁸⁵ Cass. com., 28 Apr. 2004, No. 02-21585, Civ. Bull. IV, No. 80.

⁸⁶ CA Orléans, 10 Jul. 2003, No. RG 02/01964, PIBD 2003, No. 773.

be characterized on a case-by-case basis⁸⁷. Notwithstanding, if we accept to assess originality intrinsically, i.e. in the work, by noting whether or not it is creative, the opposition could be lifted (by the judge in this case, as is the case for all works). There would therefore be a case-by-case analysis.

Yet, automaticity is also in itself open to criticism, in that it involves fictitious reasoning which consists in acting as if AI already potentially contained all the works to be created, without the creator even having any idea of the immensity of the outlets for their own creation⁸⁸. This approach is, at the very least, unprecedented in copyright law.

The objection also has concrete implications: in particular, how can the AI creator ensure that their rights are respected when the instrument is in the hands of a third party? In practice, the contractual agreement will undoubtedly settle matters. Moreover, how can the legitimate desire of the user to "disseminate" the creations they have helped generate be reconciled with the creator's monopoly? Is it relevant to grant full moral rights, which could moreover indirectly increase the creator's monopoly of use?⁸⁹ How can the user be permitted to use the instrument as they so wish, to modify the output, without damaging integrity? And should the right of disclosure be presumed? Furthermore, how should post-mortem productions, i.e. creations generated by AI when the AI creator is dead, be managed? Finally, what would the starting point for protection be? The "publication" of the work, in line with the logic of special works such as anonymous and pseudonymous works or collective works?⁹⁰

But these are questions which could arise in practice and which jurisprudence could partly address. In any case, it would seem that **identifying the AI creator would be the most copyright-respectful solution**. However, for the purpose of completeness, other proposals should be considered.

2.1.3.2. Second hypothesis of ownership: targeting the AI user

The AI user is the person who physically controls the tool and who designates the result as a work. Identifying the AI user as the author of the AI-generated creations has the advantage of **simplicity** as it abolishes the distinction between assisted creation and computer-generated creation. Some also see it as a sort of **financial legitimacy**, as the user will often have purchased a licence to use it⁹¹. This solution is still in line with technical feasibility: it is the user who has the actual **custody** of the program and who will disclose the work; they therefore play a vital role⁹². These advantages justified the fact that such a hypothesis has received the approval of part of the doctrine.

The question then arises as to the criterion for linking the author to the work. The first possible criterion is the existence of a **minimal creative choice**: the user stops the process and declares the AI-generated product generated as a work. But is this enough to make an author?

⁸⁷ Problems pointed out in particular by P. Samuelson, *op. cit.*

⁸⁸ Arguments raised by Y. Gaubiac, G. Azzaria and A. Bridy in their aforementioned articles.

⁸⁹ See questions from V.-L. Benabou, "AI, moral right and adaptation right", *EU copyright, quo vadis? From the EU copyright package to the challenges of Artificial intelligence*, European Copyright Society Symposium, Brussels, May 2018: <https://europeancopyrightsocietydotorg.files.wordpress.com/2018/06/benabou-presentation-ai-moral-rights1.pdf>.

⁹⁰ French Intellectual Property Code, Art. L. 123-3.

⁹¹ Arguments reiterated by P. Samuelson, "Allocating Ownership Rights in Computer-Generated Works", *op. cit.*

⁹² R. C. Denicola, "Ex Machina: Copyright Protection for Computer Generated Works", 69 Rutgers UL Rev., 251 (2016).

Other authors, in particular American, evoke a second relevant criterion by invoking the doctrine of "*work made for hire*". This American theory⁹³ provides that in certain cases, ownership of the rights of a protected work created by a commissioned agent, the real author, goes back to the commissioner, the fictitious author, as soon as it is the latter who commissioned it and it is in their name that it was made. When applied to A.I., this doctrine would as such require that the A.I. be the "author by fact" of the work, but that only the user be its "author by law" as it is at the user's request that the A.I. has produced the work for which copyright protection is sought⁹⁴.

However, this hypothesis presents **principled opposition**. Firstly, it could be **fictitious** in many cases. If the user makes creative choices downstream, there is no doubt that they could be qualified as an author (or even as a co-author of the generated collaborative work, if we also remember that the creator has left their mark upstream). But if they do not make any creative choices, and just settle for pressing a simple button, the solution should not be chosen. Copyright would lose its strength if it were to "reward" the mere act of commissioning a creation or any act ancillary to the creation itself. It would as such flout all principles. As the Court of Cassation pointed out, "the status of author cannot be granted to a person who has limited themselves to providing an idea or a simple theme"⁹⁵. Secondly, attributing any merit whatsoever to the declaratory act by which an AI consumer judges the product spontaneously generated by the machine to be "a work" is, here too, fictitious: the qualification would be virtually meaningless and the courts would merely act as a registration body. Such a solution would moreover contravene the law, insofar as it would lead to automatic protection, which would take the genre of the work into consideration⁹⁶.

2.1.3.3. Legally imposing a regime of ownership

In order to remove existing doubts, legally fixing authorship and ownership has also been envisaged.

Also, some theorists propose to draw inspiration from **English law**, which establishes a derogatory regime for "computer-generated works", defined as those made "in circumstances such that there is no human author"⁹⁷. The *Copyright, Designs and Patents Act* as such establishes a legal fiction under which the author shall be taken to be the person by whom the "arrangements necessary for the creation of the work are undertaken"⁹⁸. Many consider that the intended beneficiary is therefore the user, although there is legitimate doubt about this⁹⁹. Moreover, the doctrine remains relatively divided on the identification of this reputed author.

⁹³ See Art. 201 of the US Copyright Act.

⁹⁴ See A. Bridy, "*The Evolution of Authorship: Work Made by Code*", 39 COLUM. J.L. & ARTS 395 (2016); S. Yanisky-Ravid and L. Antonio Velez-Hernandez, "*Copyrightability of Artworks produced by creative robots and originality: the Formality-Objective Model*", Minnesota Journal of Law, Science & Technology, Vol. 19, p. 18; M. E. Kaminsky, "*Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*", University of Colorado Law Legal Studies Research Paper No. 17-26, vol. 51, 2017.

⁹⁵ 1st Civ. Ct. Cass., 8 Nov. 1983, Civ. Bull. I., No. 260.

⁹⁶ Article L. 112-1 of the French Intellectual Property Code prohibits taking gender, form of expression, merit or purpose into account for access to protection.

⁹⁷ *Copyright, Designs and Patents Act* (CDPA), 1988, Art. 178 (b).

⁹⁸ CDPA, Art. 9 (3).

⁹⁹ A. Guadamuz, "*Do Androids Dream of Electric Copyright? Comparative analysis of originality in artificial intelligence generated works*", Intellectual Property Quarterly, 2 (2017): "[if] the artificial agent is directly started by the programmer, and it creates a work of art, then the programmer is clearly the author in accordance to s. 9 (3) CDPA. However, if a user acquires a program capable of producing computer-generated works, and uses it to generate a new work, then ownership would go to the user."

Only one judgement makes a brief mention of it, seeming to identify the programmer of an image generation algorithm in a video game as the person who "made the arrangements necessary" to create the work, and not the user. Nonetheless, this judgement was handed down on different grounds, which do not enable a satisfactory interpretation to be drawn from it¹⁰⁰.

The legal formulation therefore gives rise here to an imperfectly resolved solution, simply misplaced in reality. Instead of seeking to qualify the author, one seeks to qualify the person who takes the necessary steps by identifying them as the author...

Moreover, such a solution, once again takes obvious liberties with the principles, since it proposes to invest the quality of author in a person who is content to "order" the creation, to organize its generation, far removed from any idea of creative choices. How could the commissioner of a work be the author? The frescoes in the Sistine Chapel are, of course, by Michelangelo and not by Pope Julius II, as Jane Ginsburg provocatively writes¹⁰¹.

Finally, English law has a **special protection regime**. As such, the term of economic protection is reduced to fifty years, instead of seventy years in principle, and the starting point is the creation of the work and not the death of the author¹⁰². With regard to moral rights, exceptions are made to the right of authorship¹⁰³ and the right of respect for the integrity of the work¹⁰⁴.

In conclusion, even though the copyright possibility is uncertain on certain points, it is nonetheless attractive. It would imply, to a large extent, the judge's interpretation, in a role which is however traditional in copyright law, i.e. that of adapting the law. Another solution could be to lay down an *ad hoc* solution in the law.

2.2. Alternatives to traditional copyright law

Some existing regimes could usefully serve as models for determining a protection regime for AI products. These include special copyright regimes (2.2.1 and 2.2.2) and the *sui generis* right (2.2.3).

2.2.1. Creating special copyright

If we remain on the boundary of copyright, we could first draw inspiration from specific cases. The first model which comes to mind is undoubtedly that of **software** (or computer program), on the one hand, for its characteristic closeness – a computer creation – and, on the other hand, for the model for adapting the regime. The particularity of software has imposed not only a new jurisprudential definition of originality (in the aforementioned *Pachot* judgement), but also a totally derogatory legal regime. As such, the content of the monopoly of use differs (French Intellectual Property Code, Art. L. 122-6); the exceptions are specific (French Intellectual Property Code, Art. L. 122-6-1); the moral right is reduced to a mere

¹⁰⁰ *Nova Productions Limited v. Mazooma Games Limited & Others*, [2006] EWHC 24 (Ch) §105, 20 Jan. 2006, IIC04C02882.

¹⁰¹ J. C. Ginsburg, "People Not Machines: Authorship and What It Means in the Berne Convention", 49 International Review of Intellectual Property and Competition Law, No. 2, 131 (2018).

¹⁰² CDPA, Art. 12 (7).

¹⁰³ CDPA, Art. 79 (2) (a).

¹⁰⁴ CDPA, Art. 81 (2).

pittance (and to the minimal expression of the Berne Convention - "prejudicial to honour or reputation"; French Intellectual Property Code, Art. L. 121-7); ownership rules organize a legal "devolution" to the employer (French Intellectual Property Code, Art. L. 113-9), thus making an exception to the principle of indifference of the creative framework (French Intellectual Property Code, Art. L. 111-1, par. 2). Ultimately-speaking, the special software right is copyright which actually departs from copyright in many respects.

As regards AI-generated creation, the choice is clearly **political**. If the political will is expressed (as was the case for software), this could then lead to the creation of a new right, attached to copyright (because of the apparent closeness of the result created), but with specific rules. The model for reflection (that of software) could therefore inspire this special right, in particular as regards the rules of ownership and moral rights (in fact, to the extent that the link to the author is expanded, an adjusted regime of moral rights could be justified).

The **collective work**¹⁰⁵ model could then be used, specifically because it represents a more economic vision of copyright, less centred on the creator. As such, it derogates from the classical rules of ownership, as rights are established, independently of any transfer, vis-à-vis the initiator who, in a vertical creative process, manages the creation and publishes it under their own name. It would therefore be permissible to add a fourth paragraph to Article L. 113-2 of the French Intellectual Property Code to define the work in question as "the creation generated by an artificial intelligence and to the accomplishment of which no natural person contributed". As for the associated regime, it could be integrated into Article L. 113-5 of the French Intellectual Property Code, along with that of collective work. The new Article L. 113-5 would read as follows: "The collective work and the AI-generated work are, unless proven otherwise, the property of the natural or legal person under whose name they are disclosed. This person is vested with author's rights". However, although this solution defines a first owner, it does not make it possible to identify an author, which runs counter to the rigour of the reasoning. Yet a re-interpretation of the criteria for accessing protection would make this possible (see 2.1.2 above). Moreover, in the most recent copyright jurisprudence, it is clear that identifying the authors of a collective work is not a priority, to the extent that judges sometimes seem to characterize such a work with an owner but without an author¹⁰⁶...

2.2.2. Creating copyright in the style of a "related right"

Another option could be to create a specific right, in copyright, but also on the boundary with a related right.

We could draw inspiration from the **posthumous work**¹⁰⁷ regime, which grants a right of use to the owner of the material media of the work who ensures its publication, even though the term of economic protection has expired under ordinary law¹⁰⁸. As such, like the owner of the posthumous work, whoever discloses an AI-generated creation would benefit from prerogatives of an economic nature. The event giving rise to this protection would be the

¹⁰⁵ As regards this proposal, see N. Enser, *op. cit.*, No. 612.

¹⁰⁶ V. A. Bensamoun, "La personne morale en droit d'auteur : auteur contre-nature ou titulaire naturel ?" (The legal entity in copyright: unnatural author or natural owner?), D. 2013, p. 376.

¹⁰⁷ Making this proposal, N. Enser, *op. cit.*, No. 613.

¹⁰⁸ French Intellectual Property Code, Art. L. 123-4, par. 3: "If disclosure is made at the end of this period, it shall belong to the owners, by succession or other title, of the work, who initiate the publication or cause it to be initiated."

publication, also known as disclosure¹⁰⁹ or, in the text of the "Term" Directive, communication to the public (lawfully made)¹¹⁰. Even when incorporated into copyright law - because it relates to a work - the monopoly granted to the publisher can be analysed, according to the doctrine, as a special "related right"¹¹¹, based on the media of the work. Such a special monopoly could therefore be granted to the person who assumes responsibility for the communication to the public of an AI-generated creation.

It would not, however, be a question in this instance of creating a monopoly on raw information or ideas, without any other condition. We are talking here about a copyright, relating to a work. As such, the form generated would have to be **intrinsically original**; it would have to pass the Turing test¹¹², so to speak, and be able to mislead as regards its origin. It would be a creation *assimilable* to a work. Reference is made here to the demonstration undertaken on adapting copyright criteria.

Moreover, it would also be possible, if the political will was aligned, to add a condition of **involvement** or **investment** (material, human or financial) to this framework. More broadly-speaking, we could draw inspiration from English law for identifying the owner.

The provision could then be incorporated into Article L. 123-4¹¹³ of the French Intellectual Property Code, *in fine*, in these terms: "Whoever takes the necessary steps in terms of investment to communicate a creation in a form generated by artificial intelligence and assimilated to an intellectual work to the public enjoys a right of use for a term of X¹¹⁴ years from the date of communication". In order to enable more flexibility, negotiated contractual solutions could be favoured here by adding at the beginning of the provision "unless otherwise stipulated"¹¹⁵.

2.2.3. Creating a *sui generis* right

The private right over the creative productions of an AI could then consist of a ***sui generis* right**, on the model of the right granted to the database producer¹¹⁶, for the stated purpose of supporting and protecting the investment¹¹⁷.

¹⁰⁹ The term disclosure and the term publication should be understood here in the broad sense of communication to the public.

¹¹⁰ Dir. 2006/116/EC of 12 Dec. 2006 on the term of protection of copyright and certain related rights (codified version), Art. 4.

¹¹¹ See not. A. Lucas, A. Lucas-Schloetter and C. Bernault, *Traité de la propriété littéraire et artistique* (Treatise on Literary and Artistic Property), *op. cit.*, No. 673; P.-Y. Gautier, *Propriété Littéraire et Artistique*, PUF, coll. Droit fondamental, 11th ed., 2019, No. 401; F. Pollaud-Dulian, *Le droit d'auteur* (Copyright), Economica, 2nd ed., 2014, No. 2508; C. Caron, *Droit d'auteur et droits voisins* (Copyright and related rights), LexisNexis, 5th ed., 2017, No. 351; M. Vivant and J.-M. Bruguière, *op. cit.*, No. 978.

¹¹² A. Turing, "Computing machinery and intelligence", *Mind*, Oxford University Press, Volume LIX, Issue 36, Oct. 1950, p. 433, <https://academic.oup.com/mind/article/LIX/236/433/986238>.

¹¹³ This linkage to the copyright section would make it possible to follow the logic of the law on posthumous works. This option serves as a reminder that these creations can be assimilated to works.

¹¹⁴ The choice of a term is a political choice and it would then be up to the legislator to decide. By comparison, the new related right of press publishers has a term of 2 years from the date of publication. The initial 2016 version proposed a term of 20 years.

¹¹⁵ V. A. Bensamoun, "Intelligence artificielle et propriété intellectuelle" (Artificial intelligence and intellectual property), *op. cit.*

¹¹⁶ See Dir. 96/9/EC, of 11 March 1996, on the legal protection of databases; French Intellectual Property Code, Art. L. 341-1 and s.

¹¹⁷ A. Bensamoun and G. Loiseau, "L'intégration de l'intelligence artificielle dans certains droits spéciaux" (The integration of artificial intelligence into certain special rights), *Dalloz IP/IT*, May 2017, p. 294, spec. p. 297.

The protection of the database by this special literary and artistic property right is conditioned to making a financial, material or human investment¹¹⁸ for obtaining, establishing, checking and/or presenting the contents of the database, to the exclusion of investments for the creation of the data¹¹⁹. The monopoly then enables the producer to prohibit the extraction and/or reuse of all or a substantial part of the database. The notion of producer is recognized from literary and artistic property law¹²⁰. Following similar logic, Article L. 341-1 of the French Intellectual Property Code refers to "the person who takes the initiative and the risk of the corresponding investments".

When applied to AI, this *sui generis* right would enable the person who takes the risk a return on investment, thwarting attempts to appropriate value and thereby encouraging investment in the field. The proposal could be integrated into the first part of the French Intellectual Property Code, devoted to literary and artistic property, following the database right described above¹²¹, and impose the following protection¹²²: "The producer of an artificial intelligence enabling the generation of creations assimilated to intellectual works¹²³ benefits from protection on these creations whenever they result from substantial financial, material or human investment".

We could also decide to limit the right of use attributed to the producer – reproduction right and right of communication to the public – to uses for profit¹²⁴, so as to reduce the scope of reservation to purely parasitic behaviour. Admittedly, French law does not take commercial use into account when defining the scope of the monopoly. But European law is not insensitive to it¹²⁵. Some authors also insist on the need to take the function of intellectual property law into account so as to determine its scope¹²⁶.

Such an option would, on the one hand, have the advantage of being simple and, on the other hand, would be coherent. It would be unlikely to weaken copyright law by seeking to extend its legal categories. It would enable a regime to adapt to the specificities of computer-

¹¹⁸ As regards the notion, see P. Gaudrat and F. Sardain, *Traité de droit civil du numérique* (Treatise on Digital Civil Law), v. 1, *Droit des biens* (Property law), Larcier, 2015, No. 1085 and s.

¹¹⁹ CJEC, 9 Nov. 2004, 4 judgements, cases C-203/02, *The British Horseracing Board Ltd and Others v William Hill Organization Ltd*, Rec. CJEC 2004, I, p. 10461; Electr. comm. com. 2005, comm. 2, note C. Caron; Industr. Propr. 2005, comm. 7, 1st judgement; Industr. Propr. 2005, comm. 22, 2nd judgement CJEC, 9 Nov. 2004, case C-46/02, *Fixtures Marketing Ltd v Oy Veikkaus AB*, Rec. CJEC 2004, I, p. 10396. CJEC, 9 Nov. 2004, case C-338/02, *Fixtures Marketing Ltd v Svenska Spel AB*, Rec. CJEC 2004, I, p. 10532, pt 29. CJEC, 9 Nov. 2004, case C-444/02, *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP)*, Rec. CJEC 2004, I, p. 10590. On these judgements, see not. S. Lemarchand and S. Rambaud, *Intell. Propr.* 2005, p. 99; F. Pollaud-Dulian, *RTD com.* 2005, p. 90; M. Vivant, *RLDI*, March 2005, No. 104.

¹²⁰ See producers of phonograms (French Intellectual Property Code, Art. L. 213-1) and of videograms (French Intellectual Property Code, Art. L. 215-1).

¹²¹ Perhaps in a new Title V of Book III?

¹²² V. A. Bensamoun, "Intelligence artificielle et propriété intellectuelle" (Artificial intelligence and intellectual property), *op. cit.*

¹²³ Once again, the reference to the work must be made objectively.

¹²⁴ On a related issue, see J. Lapousterle and A. Latil (dir.), *L'usage commercial des biens intellectuels* (The commercial use of intellectual property), Mare et Martin, PUS, 2020, to be published.

¹²⁵ See, for example, jurisprudence on hyperlinks. – See generally J. Lapousterle, "L'usage commercial en droit d'auteur : indifférence véritable ou déni ?" (Commercial use in copyright law: genuine indifference or denial?), in *L'usage commercial des biens intellectuels*, *op. cit.*

¹²⁶ M. Vivant, "Intellectual property rights and their functions: determining their legitimate 'enclosure'", in H. Ullrich, P. Drahos and G. Ghidini (dir.), *Kritika: Essays on intellectual property*, vol. 3, Edward Elgar Pub., 2018, p. 44.

generated creation and to general interest objectives. It would not prevent a particular creation from being qualified as an intellectual work, when the traditional criteria are met – particularly in cases related, especially, to *assisted* creation. Finally, it would make it possible to preserve the incentives for investment in AI by granting protection limited to economic rights and for a shorter period than that provided for by copyright. The solution could therefore be an interesting way forward, but only if the need was proven and the future revealed that copyright could not apply. All the more so as the approach of a *sui generis* right has the disadvantage of denying the obvious kinship between traditional and intelligent creations.

Generally speaking, if a new right were to be created, we would have to question its specific scope. It might be appropriate, in particular, and in order to preserve the value of human creativity, to grant less extensive rights than those offered by traditional copyright. Moreover, attention should be paid to the rights of performers, as their monopoly is based on the performance of an intellectual work.

2.3. *The absence of private rights*

First and foremost, in this instance we should remember that as soon as a human intervenes in a sufficiently creative manner, AI must be understood as a **tool**¹²⁷, regardless of its sophistication or technicality.

Apart from this assumption, some theorists doubt the need to protect AI-generated creations, noting in particular the **existence of protection upstream** of the chain (see above). Some also advocate their position by stating that cultural value is found more so in the "**educated**" **algorithm** than in the creation it generates. The creative act is to be found more so in AI programming than in its functioning. In their opinion, the result of AI is the pure product of chance or mechanical planning: as such, it expresses very little of any creativity specific to the human person which would deserve copyright protection. However, the very purpose of copyright protection is to promote "human communication"¹²⁸.

Moreover, as the AI programmer is remunerated by the sale of the algorithm and/or by a **licence** for software use, copyright would therefore not have to be called on to protect AI creations, at the risk of being misleading.

Thereupon, the law would assume an absence of intangible protection. Only a mere tangible property right over the tangible material produced by the AI would, where applicable, remain, which would probably be attributed to the person who holds the title to use the AI itself. This regime would also be intended to be complemented by a competitive component to avoid any misappropriation of the public domain. The theory of unjust enrichment could also be invoked to tackle the illicit use of unprotected creations¹²⁹, although jurisprudence on this subject is not very extensive.

¹²⁷ We can take *Daddy's car* as an example.

¹²⁸ V. S. Ricketson, "*People or Machines: The Bern Convention and the Changing Concept of Authorship*", 16 Colum.-VLA J.L & Arts 1 (1991-1992); C. J. Craig and I. R. Kerr, "*The Death of the AI Author*", March 2019: <https://ssrn.com/abstract=3374951> and <http://dx.doi.org/10.2139/ssrn.3374951>.

¹²⁹ A. Lebois, "Quelle protection juridique pour les créations des robots journalistes ?" (What legal protection for the creations of robot journalists), CCE 2015, Study 2.

For its part, the French AIPPI group¹³⁰ rejects the idea of a monopoly considering that the conditions are simply not met in the **absence of a human author**. This finding prohibits the implementation of such protection which would as such contravene the foundations of copyright. Moreover, the risk of creation **massification** and creative field saturation would weaken the position of the natural person creator.

Furthermore other commentators put forward tried and tested arguments, such as the **infringement of freedom of expression** which any exclusivity may constitute¹³¹. Moreover, at a time when any intellectual property reservation is questioned by public opinion, these authors consider that it is to be feared that the introduction of a new right would be analysed as an excess of protection.

Others favour flexibility and support the path of **contractual settlement** between the AI creator and the user or protection by technology, through the possible application of **technical measures**. The latter approach would imply modifying the law as, for the time being, circumvention is punishable in so far as the measure aims to protect subject matter eligible for copyright or related rights (French Intellectual Property Code, Art. L. 331-5).

Moreover, the absence of private rights does not necessarily mean the absence of protection. The AI creator can as such protect the creative process through the **business secrecy** mechanism (French Commercial Code, Art. L. 151-1, transposing Directive 2016/943 of 8 June 2016), which will enable the former, by taking reasonable material and legal measures, to oppose the unlawful disclosure of their know-how and the appropriation of their investment. They may also, on the basis of the ordinary law of extra-contractual civil liability (French Civil Code, Art. 1240 et seq.), take action against acts of **unfair competition** or **parasitic acts**.

Lastly, according to a doctrinal analysis, these works without author should be considered as commons by nature – commons by design –, integrated into a default public domain¹³². This extensive qualification of the public domain would be subject to two conditions: that AI plays a predominant role in the expression of the work and that human intervention is consequently residual, making it impossible to imprint creative choices, which are as such the result of AI. Another author proposes basing integration into the public domain, not on the absence of an author, but on **ecological reasons** (ecological cost of these mass productions) and on the risk of **cultural impoverishment** linked to these productions. This would justify maintaining only a contractual valuation in a legislative policy choice¹³³.

¹³⁰ See the report from the French AIPPI group (Association Internationale pour la Protection de la Propriété Intellectuelle (International Association for the Protection of Intellectual Property)), 2019 – *Question Study – Copyright in artificially generated works*: https://aippi.org/wp-content/uploads/2019/06/2019_FR_2019_Study_Question_Copyright_in_artificially_generated_works_2019-05-24.pdf.

¹³¹ In this respect, not. J. Cabay, "Droit d'auteur et intelligence artificielle : comparaison n'est pas raison" (Copyright and artificial intelligence: comparison is not an argument), Feb. 2019, <https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/285602/Holdings>.

¹³² See not. G. Azzaria, "IA et création non appropriable" (AI and non-appropriable creation), <https://www.lesconferences.ca/videos/mise-en-commun-numerique/> and "Intelligence artificielle et droit d'auteur : l'hypothèse d'un domaine public par défaut" (Artificial intelligence and copyright: the default public domain hypothesis), *Les Cahiers de PI*, 2018, vol. 30, No. 3, p. 925. Comp. D. Gervais, "Can Machines be Authors", <http://copyrightblog.kluweriplaw.com/2019/05/21/can-machines-be-authors/>, Mai 2019; "The Machine As Author", *Intell. Propr.* 2019, No. 72, p. 7.

¹³³ In this respect, A. Billion, CSPLA hearing.

Yet this approach, which has the advantage of being simple by choosing to exempt purely AI-generated creations from private law, poses other issues. As aforementioned, it **leaves candidates for the ownership of related rights unprotected**, specifically the interpreters of artificial creations. For some, it would also favour substitution and parasitism effects, multiplying the risks of litigation, and would risk limiting the incentive to invest in the field of AI¹³⁴. However, these aspects would deserve empirical economic validation to be substantiated. Moreover, the approach does not resolve the issue of distinguishing between computer "assisted" creations, which would be eligible for traditional copyright, and "generated" creations, which would fall into the public domain – although this distinction could be left to the discretion of lower courts. Finally, in practice, it would be rather impracticable and would encourage **misappropriation behaviour**: the AI user would have no incentive to reveal that their 'work' is computer-generated and that it falls into the public domain.

2.4. *So, which solution?*

The choice of a status for the cultural productions of AI appears to be an almost **philosophical** subject. It is also clearly a matter of **legislative policy**. The last word will be that of the legislator. Above all, the legislator will have to decide whether or not it is appropriate to protect AI productions by a private right. Currently, it would appear that the question is settled **contractually**. Specifically, whenever AI is made available to create, two formulas are proposed¹³⁵: free use for private use of the creative results and paid use for their commercial use. As such, legislative intervention would need to highlight the need to create a private right, by assessing these contractual solutions as well as by carrying out an impact study.

It is therefore important to remember that copyright must remain linked to a **human being**¹³⁶. The implementation of copyright cannot do without the human presence (even if the link to creation could be expanded more). Moreover, as soon as AI becomes a tool in the hands of a creator who is a natural person – who chooses the inputs of the training base upstream, carries out the curation downstream... –, classical copyright should then apply. This is the most common situation today (as seen with *Daddy's car* and *Edmond de Belamy*). It is then just a question of assessing the involvement of the natural person who creates, either upstream by programming and coaching the machine, or downstream by "thwarting" production through downstream actions¹³⁷.

Nonetheless, we cannot ignore the fact that technology is developing and that the tool is becoming more autonomous, reducing the role of the human being. This ongoing evolution

¹³⁴ Challenging this incentive effect idea: J. Cabay, "Contre la protection du droit d'auteur pour les productions de l'intelligence artificielle" (Against copyright protection for artificial intelligence productions), Nov. 2018, <https://orbi.uliege.be/handle/2268/232795> ; <http://hdl.handle.net/2268/232795>.

¹³⁵ This is the framework in particular chosen by Aiva Technologies.

¹³⁶ In this respect, see the 2019 resolution – *Study Question – Copyright in artificially generated works*, Sept. 2019: https://aippi.org/wp-content/uploads/2019/10/Resolution_Copyright_in_artificially_generated_works_English.pdf.

¹³⁷ V. J. C. Ginsburg and L. A. Budiardjo, "Authors and machine", Columbia Public Law Research Paper No. 14-597; Berkeley Technology Law Journal, Vol. 34, No. 2, 2019. Available on SSRN: <https://ssrn.com/abstract=3233885> and <http://dx.doi.org/10.2139/ssrn.3233885>. But the authors consider that in order to be an author, the person must have control over the creation and use of the output.

could moreover justify that the legislator does not rush to intervene¹³⁸. It would then be urgent to wait...

In this respect, it is important to note that **copyright seems flexible enough to include these creations** and that the attribution of rights to the AI creator appears likely to provide relevant solutions. This is what a **Chinese court** seems to have decided, to the benefit of the Tencent company¹³⁹. As such, unless jurisprudence highlights a new need or refuses to receive these creations, **positive law should, for the time being, be able to be applied, through a new interpretation of the criteria for accessing protection.**

If, in the future, a shortfall in positive law and a need for protection were reported, then legislative intervention would be required. In this respect, it is likely that the first regional regulation or decision could influence other normative frameworks¹⁴⁰. For which content and on which scale?

An interesting approach could then be, given the different positions and analyses, to create a **special copyright** (to show the obvious kinship between these creations and classical works), based on classical criteria and through a new interpretation. By taking into account the fact that the link to the author, in the classical sense of the term, appears to be more expanded and that the content of the protection may prove to be unsuitable (in particular from a moral right aspect), the special right could permit adjusted protection.

In any case, attention should be paid to the rights of performers, to ensure that their case would be settled when they perform such works.

In any case, even if a national solution is possible – considering that the field is non-harmonized – it would be desirable that the solution adopted should be imposed at **international level**¹⁴¹, at least at **European level**. As regards this point, copyright protection would make it possible to take advantage of the normative tools already in effect, in particular the Berne Convention, which would have the advantage of applying the rule of national treatment to intelligent creations.

The European legislator is aware that the subject sometimes requires its intervention. In this sense, the question of the regime of incoming works, which enable the inference model to be established, has been (imperfectly) dealt with in the "Digital Single Market" Directive, through the "data mining" exception dedicated to AI uses (see 3.2.1.2 below). If the legislator has intervened upstream, it should be able, if the need was proven, to regulate downstream.

Clearly, the subject is one of interest and the issue is making headway. Therefore, and as was mentioned by the CSPLA President at the beginning of the mission¹⁴², the reflection should be developed by the institution. This is a first milestone which could encourage the CSPLA to reassess the issue – which is constantly evolving – in the coming years.

¹³⁸ Considering that technology is still developing, in its resolution, the AIPPI considers that it would be premature to adopt positions on the relevance of new protection.

¹³⁹ <https://venturebeat.com/2020/01/10/chinese-court-rules-ai-written-article-is-protected-by-copyright/>.

¹⁴⁰ We are aware that the national treatment of the Berne Convention implies that even if AI-produced creations were protected in the US, they would not necessarily be protected in Europe.

¹⁴¹ In this respect, see the conclusions of the AIPPI *Summary Report, 2019 – Question Study – Copyright in artificially generated works*, Jul. 2019: "A clear majority of the responding Groups considers that harmonization regarding the protection of artificially-generated works is desirable": https://aippi.org/wp-content/uploads/2019/07/SummaryReport_COPYRIGHT-DATA_London2019_final_160719.pdf.

¹⁴² The President at the time was P.-F. Racine.

Part 3. – The regime of works enabling the production of cultural creations via AI

As we have seen, the very principle of machine learning is based on the massive use of large amounts of data, whether or not they have been selected beforehand. Machine training focuses on the way in which the data is processed, calculated and the result extracted. This extraction implies an analysis of the data present in the machine, so as to identify its characteristics, including aesthetic characteristics.

Should the deconstruction and analysis of ingested works¹⁴³, enabling them to be represented in a space intelligible to the machine and to reveal links and similarities by reconstruction, be considered as an **act of use giving rise to copyright**? How should we comprehend the **new exception for data mining** permitted by Article 4 of Directive 2019/790 and widely open to AI players, but which also preserves the possibility of a technical opt-out by the rightholder? Clearly, this limitation, adopted under pressure from digital players to allow easier use of AI, leads to a very specific compromise. Firstly, this text, although posing a limitation, indirectly constitutes the recognition of an extended monopoly. Secondly, it enables the author's exclusivity to be waived. Finally, it enables leaving the exception and returning to a reservation. There is no doubt that the linkage will not be an easy one.

As such then, does copyright apply to the phenomenon of analysis, use and deconstruction of protected works (3.1)? How can a balance be found between AI-related innovation objectives and the legitimate protection needs of rightholders (3.2)?

3.1. Does copyright apply to the works on which creative AI is based?

As we know, the monopoly of use belonging to the author is particularly broad in French law, where the legislator has adopted a synthetic and open approach, in the form of two categorical prerogatives which embrace all the acts of use of the work: the right of representation (or right of communication to the public) and the right of reproduction. But the works "ingested" by AI are not classically used: they are deconstructed, broken down, to develop a model based on their common specificities. As such, they are not used in terms of works themselves, but for their **informational value**. The work is, in a certain sense, "**datafied**". What AI uses in its creative process is not so much the primary work but more so the characteristic features it has identified. The subsequent process of reconstruction will therefore not focus so much on the work itself as on the mixing and interaction of the data which AI will have extracted. Should we see use of the incoming works in this instance?

According to some, each ingestion by artificial intelligence would consist in an "**appropriation**" of the work¹⁴⁴, likely to violate the monopoly of use. The author of the works mixed by AI may even be able to influence the fate of the work generated. In keeping with this reasoning, it would be possible to find, in the work generated by the machine, certain characteristics of the primary works extracted during the deconstruction and reconstruction

¹⁴³ For example, moving from an image to a set of intensity values or from a sound sequence to a set of frequencies.

¹⁴⁴ T. Lebrun, "L'apprentissage machine est une appropriation" (Machine learning is an appropriation), Vol. 30, No. 3, Les Cahiers de propriété intellectuelle, 2018, p. 895.

process¹⁴⁵. In this case, the AI-generated creation would, at least in part, be analysed as a work derived from the works processed by artificial intelligence during its creative process.

In reality, however, the case would rarely exist where the final creation reflects essential and recognizable components of the initial (first) work, which is a determining criterion for qualifying a work as a derivative work¹⁴⁶. Of course, we might wonder whether the lower courts' assessment could not be adapted to the specific features of the process of creation by IA so as to set a recognition threshold low enough to allow the qualification of a derivative work: several decisions, without necessarily concerning AI-generated creation, have already marked a tendency towards a lax assessment of the criterion of recognition of the first work¹⁴⁷. In Canada, an artist recently sued another artist for using AI to process works derived from the first to produce new (non-commercial) works. The case is pending¹⁴⁸ and it will be interesting to see how similarities will be assessed.

But the approach seems rather unconvincing. It will be difficult to consider, in the case of AI-generated works, that the process of deconstruction followed by that of reconstruction can be considered as a representation of the first work to the public. Actually, the first work will most of the time be made unrecognizable by the AI creative process (sometimes even "anti-plagiarism software" is added to the intelligent system); it is then impossible to consider that it is communicated to the public. As such, the essential criterion for determining whether copyright and its exclusions are intended to apply could be whether or not the work contains "**detectable material elements**" of a prior work, and not whether or not this prior work was an "intangible" source of inspiration for the subsequent creation¹⁴⁹.

If it appears doubtful that the right of communication to the public is implemented, on the other hand, it appears that the work digested by AI is very often **reproduced** beforehand. Is this reproduction as such an act of use subject to the author's consent?

Some refute this however, considering that reproduction is in this instance a **technical** step and not an autonomous act of use. However, a distinction must be made between technical and legal copy, as the former does not necessarily characterize the latter¹⁵⁰, even if the legislator tends to establish exceptions in these areas¹⁵¹. As regards AI-generated creation,

¹⁴⁵ See in particular F. Macrez and J.-M. Deltorn, "*Authorship in the age of machine learning and artificial intelligence*", *op. cit.*

¹⁴⁶ The agreement of the author of the first work is only sought if the elements found are borrowed from the first work. This is not the case when the similarities between two songs may result from the simplicity of the melody and their common inspiration (Caribbean rhythm): CA Paris, 25 Apr. 1972, RIDA Jul. 1972, p. 221.

¹⁴⁷ See in particular *Bridgeport Music, Inc v. Dimension Films*, 410 F. 3d 792 (6th Cir. 2005); Bundesgerichtshof, *Kraftwerk et al. v. Moses Pelham, et al.*, No. I ZR 112/06, 20 Nov. 2008; Paris High Court, 5 Jul. 2000, CCE mars 2001, comm. No. 23, obs. C. Caron.

¹⁴⁸ *Chamandy v/ Basanta* (2018) case: <https://www.theglobeandmail.com/arts/art-and-architecture/article-artist-faces-lawsuit-over-computer-system-that-creates-randomly/>.

¹⁴⁹ J.-M. Deltorn, "'In the style of...' - deep learning, style transfer and the limits of copyright protection. A European perspective", in *Artificial intelligence and intellectual property, Annali italiani del diritto d'autore, della cultura e dello spettacolo* (AIDA), vol. XXVII, May 2019, p. 337.

¹⁵⁰ A. Lucas, A. Lucas-Schloetter and C. Bernault, *Traité de la propriété littéraire et artistique* (Treatise on Literary and Artistic Property), LexisNexis, 5th ed., 2017, No. 263, on the subject of provisional technical reproductions. See also S. Dusollier, "L'exploitation des œuvres : une notion centrale en droit d'auteur" (The use of works: a key notion in copyright law), in *Mélanges A. Lucas*, LexisNexis, 2014, p. 263, spec. p. 267.

¹⁵¹ See, for example, Article 5.1 of Directive 2001/29 on the temporary technical copy which aims to enable a lawful use or a transmission in a network.

reproduction only aims to extract the informational content of the "ingested" work¹⁵². And it is not guaranteed that the exclusive right is intended to capture this informational value of the work. Nonetheless, admittedly the law does not make such distinctions. Article 2 of Directive 2001/29 gives the author a **particularly broad prerogative**, formulated as "the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form whatsoever, in whole or in part". Under these conditions, the functional approach, which was adopted in trademark law¹⁵³, does not seem to work in copyright law¹⁵⁴.

Nevertheless, a recent **CJEU** decision could also influence the debate on the implementation of the reproduction right. In a *Pelham* judgement of 29 July 2019¹⁵⁵ on the subject of music sampling, the Grand Chamber of the Court decided that an unrecognizable sample is not a reproduction – in fact it is not covered by the exclusive right. Specifically, after having recognized, interpreting Article 2(c) of Directive 2001/29, that "the reproduction by a user of a sound sample, even a very brief one, of a phonogram must, in principle, be regarded as a reproduction 'in part' of that phonogram" (pt 29), it added, modifying the expected solution, that there is no reproduction if the sample is used, to create a new work, "in a modified form which is not recognizable on listening" (pt 31). Admittedly, the decision does not relate to copyright but to the producers' related right. And the solution undoubtedly deserves to be discussed. In this event, if transposed to creations generated (and even assisted) by AI, it could lead to considering exclusion from the scope of protection for reproductions upstream of the works (as training data) in the AI, insofar as these works are, in principle, **not recognizable in the final creation**.

At first sight, this analysis may appear to contradict the new text and data mining exception in Directive 2019/790. Basically, why make an exception if the act does not facilitate monopoly of use?

But it is possible that the contradiction is only apparent, even if it is difficult to draw certain conclusions from the *Pelham* decision. Although the judgement resolves the issue of the right of reproduction, it gives the impression that it is by reference to the right of communication to the public since the work is not recognizable when **listened to** (which constitutes a communication). The act of reproduction is, however, autonomous from that of communication to the public. Under these conditions, sampling should be considered as a particular hypothesis, mixing the two acts of use and which as such differs in nature from data mining, which only concerns the act of reproduction.

¹⁵² C. Bernault, *Open access et droit d'auteur* (open access and copyright), Larcier, 2016, No. 185-187. – P. Kamocki, "«Laissez fouiller!» – L'argument pour les "utilisations orthogonales" des œuvres de l'esprit dans le contexte du débat sur l'exploitation des données" (Allow mining! - The argument for "orthogonal uses" of intellectual works in the debate on text and data mining), RIDA Jan. 2016, p. 5-85, spec. p.55.

¹⁵³ J. Azéma and J.-C. Galloux, *Droit de la propriété industrielle* (Industrial property law), Dalloz, 7th ed., 2012, No. 1434-1435.

¹⁵⁴ A. Lucas and J. Ginsburg, "Droit d'auteur, liberté d'expression et libre accès à l'information (Etude comparée de droit américain et européen)" (Copyright, freedom of expression and free access to information (comparative study of American and European law), RIDA Jul. 2016, p. 5-153, spec. p. 107-109. – On the different functions on which intellectual property is based: W. Fisher, "Theories of intellectual property", <https://cyber.harvard.edu/people/tfisher/ipttheory.pdf>.

¹⁵⁵ CJEU, GC, 29 Jul. 2019, case C-476/17, *Pelham*.

3.2. The difficult balance between the objectives of AI-related innovation and those of authors' legitimate protection

By considering that the upstream act of ingestion by the AI facilitates monopoly, the consent of each rightholder on the works transmitted for the learning process and then for IA creation should be collected (in return for remuneration, if applicable). AI creation would most certainly be hindered.

Consideration should therefore be given to whether exceptions may apply in such a situation. It is also worth considering whether a licensing system could be implemented to take better account of the specific nature of AI creation.

3.2.1. Possible exceptions to the copyright monopoly

The exceptions in effect (Directive 2001/29, Art. 5) are inadequate for grasping the situation (3.2.1.1). This is why the EU has taken an interest in the matter, creating a new "data mining" exception which clearly targets the hypothesis of ingestion of works (3.2.1.2).

3.2.1.1. The inapplicability of the exceptions currently in effect

Three exceptions could be evoked in order to limit the monopoly of use in the case of AI-generated creation, but none of them seems to be able, as it stands, to provide a satisfactory solution to the hypothesis of ingestion of works by an AI. It should first be noted, as a reminder, that although the Court of Cassation has on several occasions seemed to enshrine a form of exception which did not appear in the texts, by exempting "accidental" or "incidental" reproductions¹⁵⁶, most of the doctrine considers this legal thinking to be uncertain. In any event, the reproduction of works in the AI creative process does not in any way present a fortuitous aspect (criterion of "accidental inclusion" required by Article 5.3(i) of Directive 2001/29), as, on the contrary, it constitutes a voluntary step in the very midst of the creative process envisaged.

The first exception to the copyright monopoly concerns **private copying**, preventing the author from prohibiting "copies or reproductions made from a lawful source and strictly reserved for the private use of the copyist", not intended for collective use (2° of Article L. 122-5 of the French Intellectual Property Code). It is clear that the reproduction of works intended to be poured into the AI algorithm for their use in the creative process cannot claim this exception, since it will not be carried out, in the majority of cases, for the strict private use of the copyist.

The second exception prevents the prohibition, by the author, of "**provisional reproduction of a transitory or incidental nature**, when it is an integral and essential part of a **technical process** and its sole purpose is to enable the lawful use of the work or its transmission between third parties by means of a network using an intermediary" (6° of Article L. 122-5 of the French Intellectual Property Code). Essentially, this exception for technical copies,

¹⁵⁶ See not. 1st Civ. Ct. Cass., 12 May 2011, D. 2011. AJ 1409, obs. J. Daleau; D. 2011. 1875, note C. Castets-Renard; RTD com. 2011. 553, obs. F. Pollaud-Dulian; CCE 2011, N° 62, note C. Caron; JCP 2011. 814, note M. Vivant; Légipresse 2011. III. 627, note Renault; RIDA Jul. 2011, p. 341, note P. Sirinelli; RLDI 2011/72, 2371, obs. A. Bensamoun. – 1st Civ. Ct. Cass., 12 Jul. 2012, RIDA 4/2012, p. 547 and p. 537, obs. P. Sirinelli; CCE 2012, comm. 91, note C. Caron; JCP E 2012, 1627, note J.-M. Bruguière; Intell. Propr. 2012, p. 405, obs. A. Lucas; RTD Com. 2012, p. 775, obs. F. Pollaud-Dulian; D. 2012, p. 2071, note C. Castets-Renard.

introduced by Directive 2001/29 of 22 May 2001, exempts copies made necessary by digital transmissions of works from authorization.

We might be tempted to draw a comparison with the situation under consideration - which is, moreover, what recital 9 of Directive 2019/790 does - insofar as reproduction could be analysed as an incidental process integrated into the technical operation and that it can be quite temporary (it is no longer necessary once the inference model has been developed). However, the exact meaning of the condition of lawful use is questionable. In particular, the reproduction made in the case of data mining does not have as its sole object the lawful use of the work: it is precisely the lawfulness of its use that is in question. As such, even if the Directive seems to indicate that certain hypotheses will be covered by this exception, it is difficult to envisage them and to understand the borderline with the new dedicated exception.

Finally, the **exception for research purposes** could be used as a justification for the upstream act of AI ingestion, but apart from being an optional exception (Art. 5.3, a)), it is ill-suited to the technology and user licences (particularly those offered to researchers) may exclude data mining¹⁵⁷.

3.2.1.2. The new data mining exception

Directive 2001/29 did not anticipate the difficulties that **text and data mining (TDM)** practices would create in terms of copyright. Nevertheless, the issue arose during the parliamentary debates that preceded the adoption of the act for a digital republic and during the slow institutional process that led to the reform of the Copyright Directive known as the "DSM"¹⁵⁸. This new practice, whose concept emerged in 1989 under the acronym KDD (Knowledge Discovery in Databases), could lead to "a revolutionary change in researchers' working methods", with the CNRS comparing data mining to "creating the microscope"¹⁵⁹. From mapping the human genome to pharmaceutical research and digital humanities, all areas of research were concerned¹⁶⁰. Yet, there was more to come. **Machine learning** involves TDM and the need therefore goes far beyond research itself.

The Conseil Supérieur de la Propriété Littéraire et Artistique (Higher Council for Literary and Artistic Property) has defined TDM as "an automated search process which deals with a set of digital data with the aim of discovering new knowledge or new ideas"¹⁶¹. The Villani report on AI speaks of a "set of computer processing operations consisting of extracting knowledge based on a criterion of novelty or similarity from texts or databases".

¹⁵⁷ In this respect, see recit. 10, Dir. 2019/790.

¹⁵⁸ See the reports transmitted to the European Commission: J.-P. Triaille, J. de Meeûs d'Argenteuil and A. de Francquen, *Study on legal framework text and data mining*, March 2014, and I. Hargreaves (dir.), *Standardisation in the area of innovation and technological development, notably in the field of text and data mining*, Jul. 2014. – Adde Reda report (*Report on the implementation of Directive 2001/29/EC of the European Parliament and Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society*), 2014/2256 (INI), 24 June 2015, draft No. 28.

¹⁵⁹ CNRS hearing, reported by C. Mélot, opinion No. 525 on behalf of the Senate Committee on Culture, Education and Communication on the Act for a Digital Republic, 5 Apr. 2016, p. 45.

¹⁶⁰ P. Kamocki, "«Laissez fouiller !». – L'argument pour les "utilisations orthogonales" des œuvres de l'esprit dans le contexte du débat sur l'exploitation des données" (Allow mining! – The argument for "orthogonal uses" of intellectual works in the debate on text and data mining), RIDA Jan. 2016, p. 5 s., spec. p. 7.

¹⁶¹ J. Martin and L. de Carvalho, CSPLA Mission on Data Exploration, July 2014, p. 9.

In **2016**, the Act for a Digital Republic provided an exception to copyright to encourage data mining¹⁶². As such, rightholders (authors as well as database producers) cannot prohibit "digital copies or reproductions" from a lawful source, for the purpose of exploring "texts and data included or associated with **scientific writings**" for the needs of **public research**, excluding any commercial purpose¹⁶³.

The French exception is limited in two ways: by its scope of application, which is reserved for the written word, as such excluding mining audiovisual, visual and musical works; and by its beneficiaries, as only part of public research is concerned, that which is not based on public-private partnerships. This approach may come as a surprise at a time when European law was about to evolve and insofar as Directive 2001/29, which, in its Article 5, contains an exhaustive list of possible exceptions within the EU, does not expressly provide for such a limitation.

Taking the view that the existing legislative framework already allowed Member States to permit text and data mining activities under the exception for scientific research (Article 5(3)(a) of Directive 2001/29), the United Kingdom also has an exception for text and data analysis (Article 29A of the Copyright, Designs and Patents Act 1988¹⁶⁴), introduced in 2014. The exception is limited to non-commercial uses. Notwithstanding, the beneficiaries of the UK exception are not only research organizations, but also individual researchers (any person who has lawful access to a work)¹⁶⁵.

Japan¹⁶⁶ also has a broad and flexible TDM exception, which allows any use, with or without commercial purpose, provided that the "information analysis (i.e. the extraction, comparison,

¹⁶² Act No. 2016-1321 of 7 Oct. 2016, Art. 38. – See Art. L. 122-5, 10°, and L. 342-3, 5°, of the French Intellectual Property Code.

¹⁶³ The 10° of Art. L. 122-5 refers to "copies or digital reproductions made from a lawful source, for the purpose of exploring texts and data included or associated with scientific writings for the needs of public research, excluding any commercial purpose. A decree lays down the conditions under which text and data mining is carried out, as well as the methods of conservation and communication of the files produced at the end of the research activities for which they were produced; these files constitute research data".

¹⁶⁴ "Copies for text and data analysis for non-commercial research:

(1) *The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that:*

(a) *the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and*
(b) *the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).*

(2) *Where a copy of a work has been made under this section, copyright in the work is infringed if:*

(a) *the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or*
(b) *the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.*

(3) *If a copy made under this section is subsequently dealt with:*

(a) *it is to be treated as an infringing copy for the purposes of that dealing, and*
(b) *if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.*

(4) *In subsection (3) "dealt with" means sold or let for hire, or offered or exposed for sale or hire.*

(5) *To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable."*

¹⁶⁵ N. Jondet, "L'exception pour le *data mining* dans le projet de directive sur le droit d'auteur. Pourquoi l'Union européenne doit aller plus loin que les législations des Etats membres" (The exception for data mining in the draft Copyright Directive. Why the European Union must go further than the legislations of the Member States), *Intell. Prop.* 2018. 25. – Comp.: There is also a TDM exception in Germany for scientific research and non-commercial use. (UrhG, § 60d).

¹⁶⁶ Art. 30-4, point *ii* of the Japanese Copyright Act (Act No. 48 of 6 May 1970, mod. 2018; <http://www.japaneselawtranslation.go.jp/law/detail/?id=3379>) :

classification, or other analysis of language, sound, image, or any other element of which a large number of works or a large amount of data are composed)" is not for the purpose of "enjoying" the work¹⁶⁷.

The European Union took up the issue as, henceforth, Article 2 of Directive 2019/790 of 17 April 2019¹⁶⁸ defines TDM as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations".

Mining operations often involves upstream intermediate reproductions of the works or databases concerned which must be stored, at least temporarily, and sometimes modified (format, cuts, mergers, compilations, etc.) to make them analysable¹⁶⁹: each of these operations may as such be in contradiction with the rules for reserving intellectual property, requiring the express agreement of their owners. Thereupon, the implementation of an exception indirectly implies that the monopoly was intended to apply in such a situation. As such, an exception was necessary. Nonetheless, according to the Directive, it applies as soon as a reproduction is made and the act implements the exclusive right without being excused by the compulsory exception for technical copying¹⁷⁰. The approach therefore suggests a certain amount of casuistry in the implementation of the relevant regime.

The new text proceeds in two stages.

Firstly, the DSM Directive imposes in **Article 3**¹⁷¹ a mandatory exception, which cannot be derogated from by contract (Art. 7.1¹⁷²), for the benefit of "research organisations and cultural

"It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation: (...)

(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data; the same applies in Article 47-5, paragraph (1), item (ii) (...)".

¹⁶⁷ "Enjoying the work" means "satisfying the desire to know or having intellectual fun listening to or seeing these works": see M. Nagatsuka, "Lettre du Japon. La réforme de la loi sur le droit d'auteur en 2018 et à venir" (Letter from Japan. The reform of the copyright act in 2018 and beyond), *Intell. Propr. Oct. 2019*, No. 73, p. 121.

¹⁶⁸ Dir. 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 17 Apr. 2019.

¹⁶⁹ E. Rosati, "Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity", *Asia Pacific Law Review*, Sept. 2019, <https://ssrn.com/abstract=3452376>.

¹⁷⁰ Dir. 2019/790, recit. 9.

¹⁷¹ "Article 3 - Text and data mining for the purposes of scientific research

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

2. Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results.

3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively."

heritage institutions in order to carry out, for the purposes of **scientific research**, text and data mining of works or other subject matter to which they have lawful access."

The Directive also provides, in **Article 4**¹⁷³, for another, broader TDM exception for **all uses of AI, including commercial uses** (with in all cases the limit of the triple test; Art. 5.5 of Dir. 2001/29), which may apparently be limited by the contractual agreement (no mention in Article 7). This new exception also offers a possibility of opt-out for rightholders¹⁷⁴. As such, it is a kind of presumed consent, requiring the author to explicitly express their refusal to grant it, which, in the case of content made available to the public online, will require a reservation "in an appropriate manner, such as machine-readable means in the case of content made publicly available online"¹⁷⁵. This opt-out possibility was introduced as a counterpart to the mandatory nature and the risk of parasitism due to the wide extent of the exception.

These two provisions (Articles 3 and 4 aforementioned) concern **all types of data**: texts, sounds, images, etc., and they require **legitimate access** to the work, which ensures the remuneration of the authors. They are an exception solely for the right of reproduction, not the right of representation: the TDM cannot be used as a Trojan horse for the dissemination of protected objects without the authorization of the rightholders. Finally, copies of protected content may be retained, in particular for subsequent verification of scientific research results but, in such cases, the copies should be stored in a secure environment¹⁷⁶. No specific provision is included however as regards automatic data deletion. Point 2 of Article 4 nevertheless mentions that reproductions may be retained "for as long as is necessary for the purposes of text and data mining". This reference to *necessity* must be interpreted as a requirement for deletion as soon as the copy is no longer useful for the operation.

By enshrining these exceptions, and precisely the one in Article 4 which is of particular interest to AI uses, the European legislator has **reinforced** the monopoly's scope in a place where its applicability could be doubted.

Nevertheless, questions remain as to the very **feasibility** of the planned opt-out mechanism. How can we consider that an author will have "appropriately reserved the rights of reproduction and extraction"¹⁷⁷? Should they expressly declare themselves, for example on a

¹⁷²"Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable."

¹⁷³ "Article 4 - Exception or limitation for text and data mining

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

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

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

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

¹⁷⁴ See above Article 4.3 of Directive 2019/790.

¹⁷⁵ Recit. 18.

¹⁷⁶ Recit. 15.

¹⁷⁷ Recit. 18 of the Directive states: "In the case of content that has been made publicly available online, it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service. Other uses should not be affected by the reservation of rights for the purposes of text and data mining. In other cases, it can be appropriate to reserve the rights by other means, such as contractual agreements or a unilateral declaration. Rightholders should be able to apply measures to ensure that their reservations in this regard are respected."

centralized file¹⁷⁸? By what technical means can the effectiveness of this reservation be ensured? On this point, the implementation of a **technical standard** would be of obvious interest, provided that it is correctly taken into account. Moreover, what happens in the event of a contradiction between different files and with what responsibility? We may also ask ourselves if the exception could not become a negotiating tool in the hands of AI players. Is it conceivable that a search engine would refuse to reference protected content for which the holder would have expressly prohibited TDM, having exercised their opt-out faculty, or for which there would be doubt? Unless this exception, which is met with disapproval by some rightholders, simply leads them to exercise the opt-out, as a precaution/default, to avoid massive looting.

These uncertainties as regards Article 4 of Directive 2019/790 indicate difficulties in transposition and, consequently, relative harmonization within the EU.

Beyond these questions, this text will most certainly require the assistance of major research platforms and tools to ensure, through technical means, its correct application and the monitoring of the various reservations expressed by authors whose works are made available to the public online.

3.2.2. The licensing mechanism as a suitable solution?

The mechanism of exceptions to the monopoly of use is, as we have seen, too uncertain to be able to ensure both the legitimate protection of rightholders and the development of the AI sector, whose growth must be encouraged.

Rather than a "soft" (and uncertain) regulatory exception mechanism, it could be envisaged to resort, on a **voluntary** basis, to a **"general" licensing** mechanism, on the model of the general performance agreement, "by which a professional authors' organization confers on a performance contractor the right to represent, for the term of the agreement current or future works constituting the repository of the said organization under the conditions determined by the author or the rightholders" (French Intellectual Property Code, Art. L. 132-18). In this respect, collective management organizations in particular could grant licences on their repositories (for their members) to enable mining. We can even ask ourselves if it would be possible to implement systems to enable not only use procurement but also access to useful "data" in an already structured format for TDM¹⁷⁹.

Last but not least, Article 12 of Directive 2019/790¹⁸⁰ offers Member States the opportunity to provide for mechanisms of "collective licensing with an extended effect", given "the nature of some uses, together with the usually large amount of works or other subject matter involved" (recit. 45; also see recit. 47), enabling "collective management organisations to conclude licences, on a voluntary basis, irrespective of whether all rightholders have authorised the

¹⁷⁸ But wouldn't this contravene the spirit of the very principle of copyright, which protects without any prior action?

¹⁷⁹ See in the USA the *RightFind XML for Mining* model: <https://www.youtube.com/watch?v=-gUhAkwZbVQ>; <http://www.copyright.com/business/xmlformining/>

¹⁸⁰ Also see Article 7 which provides for extended collective licensing for works which are not commercially available. – See in doctrine, F. Siiriainen, "Les licences collectives à effet étendu : entre gestion collective consentie, facilitation des autorisations et sécurité juridique (Commentaire de l'article 12 de la directive)" (Collective licensing with an extended effect: between granted collective management, facilitation of authorizations and legal certainty (Commentary on Article 12 of the Directive)), CCE Oct. 2019, dossier 5; C. Bernault, "Gestion collective – Licence collective étendue" (Collective Management – Extended Collective Licensing), IP Jul. 2019, No. 72, p. 59.

organisation concerned to do so" (recit. 46). Paragraph 3 of Article 12 lists appropriate safeguards to protect the legitimate interests of rightholders, including non-members of the collective management organization. Firstly, this organization must be "sufficiently representative"; secondly, there must be "equal treatment" of all rightholders; the latter must still retain the possibility of regaining control of their works by opposing the application of this mechanism "at any time, easily and effectively", this possibility of opt-out (of the licence) is subject to publicity measures being taken.

Such a mechanism could, in certain sectors, usefully accompany the development of AI creation, even if territoriality undoubtedly limits its interest.

However, such an option would undoubtedly ultimately oblige rightholders who are definitely opposed to the TDM to carry out the opt-out twice (to escape the TDM and to get out of collective management).

Part 4. – Improving data sharing to meet the challenges of AI development

Whenever artificial intelligence is presented as one of the keys to tomorrow's power in a digital economy, and whenever this economy feeds on data, the question, upstream, of access to data and critical asymmetries between players as regards access proves to be crucial. For "works" data, the compatibility between the development of AI-based business models and respect for intellectual property protection raises questions, as the previous section has just illustrated. For other categories, such as use data and metadata defined in the first part of this report, although data movement and sharing issues are not always directly related to intellectual property concerns, they are nonetheless significant. Reflection embraces a more general context of free movement of data, with the adoption of Regulation 2018/1807, of 18 November 2018, "establishing a framework for the free flow of non-personal data in the European Union"¹⁸¹, and the various European strategies unveiled to promote data sharing¹⁸².

As regards the cultural sector, after specifying the importance of the challenges of controlling this data in light of an increase in AI use (4.1), likely improvements in data access and sharing will be presented (4.2).

4.1. The challenges of data access and sharing for the cultural sector

The challenges involved in sharing this data are not the same in all cases. In the case of use data, the challenges are not only the classic question of transparency, but also that of controlling customer relationship (4.1.1). For metadata, it is essentially a question of qualifying content in order to accompany applications developed at all stages of the value chain (4.1.2).

4.1.1. Use data: transparency and customer relationship control challenges

The consumption of online content is accompanied by unprecedented production of use data. Yet, paradoxically, whilst the digital world is marked by the almost total and much richer potential traceability of all interactions with end users, use data is occasionally less accessible than in the physical world. In the physical world, the issue of transparency as regards the remuneration of rightholders is one of the central challenges which has motivated legislative intervention to reduce information asymmetries in favour of downstream users.

This is why, since 1948 in France, strict control of admissions to venues has been carried out by the CNC in order to reliably collect revenue for the benefit of rightholders. Likewise, the French Intellectual Property Code also contains bases of transparency requirements through the publishers' obligation to report, and extended to digital publishing, on the performance of shows, audiovisual production and the transfer of performers' rights to a phonogram producer (see Articles L. 132-21, L. 132-25 and L. 212-15 of the French Intellectual Property Code, respectively).

¹⁸¹ This text mainly prohibits restrictions on data location within the EU, imposes free access to data by national authorities and, through the development of codes of conduct, encourages data porting for the benefit of professional users.

¹⁸² Communication from the Commission, *Building a European Data Economy*, COM(2017) 09 final, 10 Jan. 2017; Communication from the Commission, *Artificial Intelligence for Europe*, COM(2018) 237 final, 25 Apr. 2018.

The transparency of production and operating accounts of cinematographic and audiovisual works is also organized by French Act No. 2016-925 of July 7, 2016 on the freedom of creation, architecture and heritage (LCAP) which introduces a new chapter on this subject in the French Cinema Code. This text systematizes the transmission of production and operating accounts and introduces the new feature of extending the obligations of regular repayment of revenue and costs, not only to the rightholders but also to the co-producers and to any person concerned by the revenue: these repayments are also transmitted to the financial partners of the film, as well as to the performers and technicians.

However, the legal obligations of transparency of remuneration data do not apply to subscription-based video-on-demand (VoD) services, even though in these models, where publishers' revenues depend on the number of pages read or the time spent, access to the data is decisive. Some players such as Netflix, whose business model does not require them to remunerate their content providers on a fee-for-service basis, have, after having provided very little data to rightholders, evolved: the SACD (French Society of Dramatic Authors and Composers) as such concluded agreements with Netflix, whilst a European association of audiovisual producers has initiated discussions with the platform. Notwithstanding, these commitments are still purely contractual. Moreover, all these provisions only deal with the transfer of rights and do not apply to the kinds of platforms which disseminate digital content without falling within this contractual scheme.

No instrument under French law therefore enables rightholders to participate in a data-driven economy, because they do not have access to use data held by downstream players: data which makes it possible to know, for example in publishing, how readers read, which passages they have highlighted, at what pace they read, when they give up, the pages they glance over, etc.

In Europe, the 2019/790 "DSM" Directive aims to guarantee the right to information for authors and performers by introducing transparency obligations in its Article 19 so that rightholders receive information on the use of their works, in particular on all the revenue generated and the remuneration due. However, "these provisions bear the hallmark of the analogue era, where all an author needed and could imagine knowing was the number of copies of their works sold and the revenue they received from them"¹⁸³. An ambitious interpretation of Directive 2019/790 could, however, argue in favour of sharing use data, in the combined interest of owners and users, as long as they consent to it, in order to respect their privacy.

Beyond the issue of transparency for a fair distribution of revenues for the benefit of rightholders, it is actually **all the balances within the sector which are likely to be called into question by access to use data or, on the contrary, by the loss of control over the customer relationship**. In a digital world of individualization and of increasingly detailed knowledge of consumer practices, control of this relationship has become crucial. In many cases, in the digital world, transparency alone, i.e. the publication of overall consumption data on the performances of content is no longer sufficient. For AI-required uses, this type of access to low-granularity data is, in most cases, disappointing. Other more qualified data becomes indispensable: performances of other works offered by rival providers for comparison/contrasting purposes, user profiles, contextual data on the presentation of works,

¹⁸³ V.-L. Benabou *et alii*, *Literary and Artistic Property Law, Data and Digital Content*, CSPLA Report, Sept. 2018, p. 124.

the origin of the link leading to the content, the number of times it is proposed by a recommendation algorithm, etc.

For authors, access to use data can play a potentially important role in the creative process, in "data-driven creation" logic, in particular to promote better interaction between the author and their audience. For publishers, producers and distributors, access to use data has also become essential. It is the use data which makes it possible to find out how the work is received, to attract an audience with a view to finer, more personalized distribution, and to adapt the distribution strategy to the platforms on which the content proves to be more effective. It makes it possible to enrich the range of services on offer by proposing customized recommendations, apprehending new trends, renewing the services on offer, supporting decision-making (motivating the production or purchase of content, and arguing in business negotiations). It is this data which enables targeted advertising facilitated by digital technology.

However, the creation of user databases by suppliers of terminals, operating systems, social media and content aggregators could lead traditional content publishers and distributors to lose control of the relationship with the end customer. As such, press and audiovisual service publishers do not have access to data on their own content circulating on social media. Telecom operators which propose boxes and services by bypassing (over the top service aka OTT) collect a myriad of use data by restoring it in a very limited way, in line with contractual agreements with the television channels which edit the programmes. In order to regain control over user data, television channels have as such become accustomed to requesting registration for their digital service, such as the French M6, which was the first to make registration for its video replay service compulsory in December 2015: it asks the Internet user to provide their gender and age, which may be the subject of user reluctance, whereas telecom and social media operators, which operate in already "logged" (identified) worlds, do not need to require it.

Use data, collected by the economic players which carry out their activities as close as possible to the end consumers of the works, is therefore crucial for the whole sector. The **concentration of data among these "downstream" players**, particularly the content delivery platforms, and the possible **absence of sharing** this data with upstream players, could be seen as real **hindrances for AI-permitted innovation**. Depending on its intensity, information asymmetry can create new power relations within the sectors, strong dependence of the upstream on the downstream and opportunistic behaviour of the downstream. Whilst upstream operators work "blindly" and are not able to develop innovative services from the information activity generated by the use of works, new professions are emerging which are specifically focused on the increasingly sophisticated analysis of user behavioural patterns.

4.1.2 Metadata: content qualification challenges

Metadata aimed at precisely qualifying content is multiplying and plays a key role in services; it accompanies the artificial creation processes and significantly renews the forms of forecasting and recommendation proposed in cultural industries.

However, among all the relevant metadata, the data which, even though basic, concerns the identification of each work is not always harmonized. Beyond the battle concerning identification, the way in which the multiple descriptive or enriched metadata is produced and made available for each work is not shared either. The information may be incomplete,

inaccurate or simply not accessible. Metadata may be generated or enriched at different stages of the value chain and the players in each sector tend to create metadata bases themselves corresponding to their own objectives, without concern for an overall vision or for their interdependencies.

Amongst professionals, the way in which metadata is viewed differs greatly from one sector to another. In the music industry, there is a real awareness of the issue and collaborations are multiplying. In the audiovisual industry, despite some attempts, metadata standards and norms are subject to very strong competition with each other, without any, whether public or private, managing to impose themselves and without the subject really being top of professional agendas. On the art market, late online commercialization has not led to any reflection on implementing metadata standards. Work reproduction bases, in particular photos, are widely circulated on the web, but according to rules of association with piecemeal metadata.

On this metadata market, AI is in great demand because it transforms the possibilities for analysing works, extracting and producing associated metadata. Although AI extends and automates metadata production and extraction possibilities, this automation is still costly and error-prone, as such still requiring a lot of manual intervention at this stage. The project for improving the quality of the metadata produced remains, as such, still largely incomplete.

4.2. Future approaches to sharing

To address the challenges, whose importance we have seen, of data access and sharing in the cultural sector, several avenues can be explored.

The first is, via a cooperative framework, to create management tools with shared governance, like those implemented, for example, in smart cities, based on the logic of **commons**, theorized by the economist Ostrom. Many examples of cooperative sharing exist in the cultural sector.

The second option is to let the business players agree among themselves, within the **market** framework, on the best way to organize data sharing. In business relationships, the reuse of data is already largely part of API logic whereby a private player proposes to others to take advantage of the use value of the data they hold. The limitation of this option is that in this case, sharing remains "in the hand" of the first operator which can always limit or refuse access.

This is why a third option, that of **regulatory arrangements for opening and sharing**, may prove to be complementary. We take an interest in this last option in this section, by first exploring the idea of essential facilities provided by competition law (4.2.1). We then discuss the sharing arrangements of the two categories of data identified in the cultural sector, use data (4.2.2) and metadata (4.2.3).

4.2.1. Making cultural data essential facilities: an attractive but unsuitable proposal

The essential facilities theory originated in American antitrust law in the early 20th century. It is not legislative, and its formalization depends on jurisprudence. Whenever the owner of a dominant position controls access to a resource and refuses to supply it, even though it is a condition for market entry, this may constitute an abuse of a dominant position. The resource is as such referred to as "essential facility", a notion which gives its name to a theory which

can lead to compelling a public or private business with or without a public service mission, which owns or controls an essential facility, to cooperate: "when access to a resource is indispensable to operate on a market, the owner of that facility may, in certain circumstances, be obliged to guarantee it to other operators, even if they are rivals"¹⁸⁴, in return for remuneration. In the event that this theory is applied, the regulator will then be able to require the owner of the facility concerned to make its facility available to other undertakings to operate on its market or on an upstream or downstream market, at reasonable price conditions (by referring to operating costs) and on a non-discriminatory basis.

The European Commission received this theory in European law for the first time in 1992, in relation to port facilities¹⁸⁵. It laid down the conditions for its application. For a facility to be considered essential, the business concerned must be in a dominant position on the market where the facility is located; the facility must be effectively indispensable for the business activity by claiming access; and it must be impossible for an operator, from a practical or reasonable point of view, to reproduce the essential facility: the fact that reproduction is difficult or simply costly is not sufficient.

It was in 1995 that the CJEU, in its *Magill* judgement¹⁸⁶, introduced this theory into European competition law on a question concerning intellectual property specifically, i.e. access to a licence. In this judgement, the Court held that while the refusal by television channels to provide licences to a newspaper presenting all their television programmes did not in itself constitute an abuse of a dominant position, in so far as copyright confers on them an exclusive right of reproduction, this could be the case in certain "exceptional circumstances" in so far as the refusal to provide access to the raw data would have the effect of preventing the appearance of a new product for which there was potential consumer demand, and without the newspaper being able to turn to a facility which was substitutable.

In France, the theory of essential facilities was received by the Autorité de la concurrence (France's national competition regulator) (Opinion No. 02-A-08 of 22 May 2002 relating to the referral procedure of the Association pour la promotion de la distribution de la presse (Association for Promoting Press Distribution)) and the Court of Cassation, which specified the conditions for its application.

Throughout jurisprudence, essential facilities may refer to a facility or equipment, a product or a service. There is nothing to prevent intangible assets from being considered as resources which can be qualified as essential facilities. However, it is difficult to oblige a business to provide access to a facility when it has no rights to it (in so far as data, as such, is not subject to property rights). The theory of essential facilities should therefore be reserved solely for access to appropriable databases as defined under intellectual property law, i.e. protected by the *sui generis* right under Article L. 341-1 of the French Intellectual Property Code or by copyright. An appropriable database may *a priori* be deemed to be an essential facility to which the business which holds these rights, if it is in a dominant position, should be able to be obliged to provide access to in order to enable other players to carry out their own business activities.

¹⁸⁴ C. Castets-Renard, F. Marty, see "*Facilités essentielles*" (Essential facilities), in *Dictionnaire des communs*, dir. M. Cornu, J. Rochfeld, F. Orsi, PUF, 2017.

¹⁸⁵ Decis. No.94/19 of the Commission, 21 December 1993, *Sea Containers v/ Stean Sealink*, Europe, March 1994, No. 115, obs. L. Idot.

¹⁸⁶ CJEC, 6 April 1995, *RTE and ITP v/ Commission*, joined cases C-241/91 P and C-242/91 P, Rec. I. 743.

Another solution would be to consider that since data is not appropriable, except in cases where it is structured in protected databases, it is basically then only access to this data which should be permitted - i.e. access to the files which "host" it. This would amount, in a way, to an **intangible "right of passage"** through the servers of the business in a dominant position in order to access data which it does not possess but which it has access to, for example by means of APIs.

The possession of data can as such increase the market power of some players and create the conditions for an abuse of a dominant position. However, such data must be considered as essential facilities, i.e. not reproducible or substitutable by reasonable business means, which will rarely be the case. Most data required for an activity can be obtained from different sources; data on musical tastes, for example, is likely to be known through purchases on commercial sites, streaming sites, generalist search engines and the personal pages of social media and will as such not fall within the scope of essential facilities. It is not the data as such that is not reproducible, but its cross-referencing with other data and its contextualization.

A German study commissioned by the Federal Ministry for Economic Affairs and Energy and undertaken by the Düsseldorf Institute for Competition Economics also raised the question of "**essential data**". It recommended that the possibility of reasonable access to essential data be explicitly mentioned as a relevant criterion to be taken into account when determining the market power of an operator. It as such proposed to make it possible to assess the "legitimate interests" at stake in the data market in the light of the competitive purpose when a business refuses access to its own data to another business. Notwithstanding, it also noted that the threshold to be crossed in order to qualify a refusal of access as abusive must necessarily be lowered from the threshold corresponding to more traditional cases of refusal of access to essential facilities. The report therefore specifies that this would be more easily the case when the refusal relates to raw, automatically generated use data which the business concerned only produces incidentally – i.e. without any particular investment on its part.

The essential facilities theory, because it leads to imposing a real obligation to contract on economic players operating in a market, remains a **cautious theory**. Very strictly understood by national and European competition authorities, it is not based on any particular text, but on a practice. In the sole hands of the regulators and, ultimately, the judges, it is only applied based on the circumstances of the case and variable casuistry. As such, it offers no direct leverage for action to make practices in this area more reliable, or for the deployment of an overall public policy. The doctrine generally considers that there is no "right of access", but only a right of access in the event of abuse in the refusal of access.

The question of whether the demanding qualification of essential facilities could be applied to big data, and as such to the masses of data held by powerful digital companies, continues however to generate intense debate amongst competition law specialists. According to Professor Béhar-Touchais¹⁸⁷, this qualification is not without discussion but, **as AI develops, the question will arise as to whether it is necessary to create licences, enabling operators to have paid access to large amounts of data held by others**. We could imagine that sectoral regulation would be useful to permit this access.

¹⁸⁷ M. Béhar-Touchais, "Intelligence artificielle et droit de la concurrence" (Artificial intelligence and competition law), in *Droit de l'intelligence artificielle*, dir. A. Bensamoun and G. Loiseau, LGDJ-Lextenso, coll. Les Intégrales, 2019, p. 171.

More than a general theory with difficult-to-define contours, it is therefore a different approach which should be followed in order to address, in culture, the challenges of data access and sharing for fuelling machine-learning algorithms. We need to go back to the very bases of the previous questions and initiate a segmented approach corresponding to the two categories of data identified in the cultural sector, i.e. use data and metadata.

4.2.2. The use data portability approach

To encourage use data access and sharing, one option would be to create a new right for the "portability" of this data to correct the informational asymmetry between cultural sector players. This right for the portability of use data, where those liable would not only be the operators as defined under the French Intellectual Property Code, but also all those which provide the public with intermediation services giving access to protected works and subject matter, is one of the proposals put forward in a previous CSPLA report¹⁸⁸. A compulsory contractual clause could as such be imposed in agreements between protected content rightholders and the operator which has obtained authorization to use it.

The aim would be to ensure that downstream operators in direct contact with the end consumer are aware of the use data relating to the works concerned. **Portability would not aim just for access and transparency but, in a "data-driven" world where businesses are evolving, the reuse and enhancement of a wider spectrum of use data in order to implement new services made possible by AI.**

The owners' claim to have detailed knowledge of users would imply the inclusion of personal data, pursuant to privacy provisions. In the cultural sectors, personal data concerns, as defined under the General Data Protection Regulation (GDPR), the use data which we have specified in the report hereof, whenever it is directly or indirectly related to an individual and concerns the characteristics, behaviour, habits, tastes and/or convictions of a specific person. The transparency and control of personal data by individuals are the focus of the aforementioned European Regulation, which nevertheless provides for a form of "sharing" through a right of portability; this form of citizen portability, which aims to enable Internet users to control their personal data, therefore remains solely in the "hands" of the latter.

The difficulty of introducing a new form of portability for use data also lies in the way of defining its scope and the elements which make up the compulsory clause. The legislator and the contracting parties will have to identify the use data concerned, its level of granularity and processing as well as its format, the frequency of updating and the period during which the data must be available. The compulsory clause should as such probably leave the contracting parties some leeway for establishing the conditions under which the use data would be reported (remuneration, data format, periodicity, etc.). This flexibility, which seems essential since it is impossible to foresee all possible scenarios, would however reduce the effectiveness of the law. This promising approach remains, inevitably, to be refined.

4.2.3. The metadata of general interest approach

In France, the open data movement has been organized for public data since 1978 by the Act providing various measures for improving relations between the administration and the public, Title I of which relates to "freedom of access to administrative documents" (known as

¹⁸⁸ V.-L. Benabou *et alii*, *Literary and Artistic Property Law, Data and Digital Content*, CSPLA Report, Sept. 2018.

the CADA Act). The 2016 Act for a Digital Republic strengthened and extended this movement beyond public data and single access logic. Public data is public information as defined under Article L. 321-1 of the Code of Relations between the Public and the Administration (CRPA), i.e. information contained in administrative documents, disseminated or communicable to all, not encumbered by third-party intellectual property rights. Over the last few years, the regulatory framework defining the terms and conditions for making public information available and re-using it has evolved considerably. The Act for a Digital Republic, enacted on 7 October 2016, introduces two new features as regards sharing.

The first concerns public data. The Act establishes the principle of open data by default: the opening up of public data, already practised, becomes the rule and no longer the exception. In the cultural sector, public data includes data held by institutions such as museums and libraries. However, the legal framework defining the conditions for making available and reusing data from public cultural institutions remains highly complex and not all issues have been resolved to date¹⁸⁹.

The second innovation is that the Act for a Digital Republic (DR) proposes a new concept of data of general interest (DGI) which must be accessible to all. In this case, its status is different as it is not public data; it is private data but whose openness may be justified on the grounds of general interest. Unlike personal data and public data, DGI does not have a stabilized definition in positive law; under this heading, the DR Act has grouped provisions relating to the opening up of data from public service grantees and subsidy recipients as well as to the access of public statistics to data from private businesses.

The Cytermann Report¹⁹⁰, which introduced the concept of DGI and part of whose work was taken up in the DR Act, highlights four key purposes justifying the "general interest" nature of this data and its openness: informing citizens, conducting public policies, research and economic innovation. The great novelty introduced by the notion of DGI is as such to organize the opening of specific data held by private entities. The Villani Report¹⁹¹, which advocates an offensive data policy aimed at favouring, depending on the sectors, the access, sharing and movement of specific private data held by businesses, is in line with this. However, whilst it is legitimate to promote the opening up of private data of general interest, there is also a tension between preserving the incentives to create new databases and opening them up, as much as possible, particularly to boost economic innovation. The issue of data movement and sharing for general-interest purposes, whilst respecting the freedom of enterprise is therefore central.

The question raised in the cultural sector is whether general-interest reasons could lead to specific metadata being subject to forms of mutualization and sharing. To address the challenges of qualitative metadata production, should this metadata circulate, be accessible and widely shared? Professionals state that content can be qualified in many ways and that the quality of this work, because it is a factor of differentiation, a buoyant market for many startups, is an element of competitive emulation; metadata bases should as such remain private.

¹⁸⁹ See in particular J. Farchy, M. de La Taille, *Free licence economy in the cultural sector*, 2018.

¹⁹⁰ L. Cytermann *et alii*, *Rapport relative aux données d'intérêt général* (Report on data of general interest), Sept. 2015.

¹⁹¹ C. Villani, *Donner un sens à l'intelligence artificielle* (Giving meaning to artificial intelligence), March 2018.

Various reports (Lescure, 2013¹⁹²; *Bearing Point*, 2015¹⁹³; *Berklee Institute for Creative Entrepreneurship*¹⁹⁴) have, however, advocated the creation of consolidated and open metadata bases, without so far having been followed up. We can, therefore, imagine that mutualizing metadata production and sharing would be beneficial for a large number of players, which separately, as the current situation illustrates, have no incentive to invest in this production. **The aim is not to reinforce the already profuse production of metadata but to improve its quality through various cross-references and pooling.** Not all metadata would be concerned but descriptive and legal metadata would be prioritized.

Moreover, as the efficiency of the content analysis and recognition algorithms used for the automatic generation of new metadata depends on the volume of databases used for the training stage of these algorithms, these tools are of particular benefit to already-established players which already have large volumes of metadata. As such, the gap is widening between the possibilities offered to public research and small operators and those implemented in an industrial setting by businesses which have a significant competitive edge thanks to metadata. The lack of complete, high-quality and accessible metadata bases is an obstacle for the research community which, with the progress of deep learning, has to train on considerable masses of metadata in order to hope to benefit from the effects of AI. **The development of metadata bases shared between players would most likely reduce these obstacles to entry and, from a cultural policy perspective, encourage the diversity of players implementing innovative services made possible by AI in this sector.**

To finance investments in "datafication", which is a source of efficiency gains for everyone but which no player is encouraged to handle alone, and to facilitate the sharing of metadata, it would be possible to promote "virtuous reassignment loops" of data and the development of "data facilities", following the example of what is promoted by the general data administrator for public data. Consideration should therefore be given to building an environment conducive to the reciprocity of the gains expected from the use of the data according to the investment made so as to enable each party to enhance its own service, based on a "coopetition" approach¹⁹⁵.

Acculturating the cultural sector to a data-driven economy...

Even when certain agreements and legislation provide for terms and conditions for accessing data, their analysis often comes up against the low level of acculturation of the players. As we have already mentioned, the value of data does not exist in itself and numerous investments are required to give it economic value, and which mobilize not only technical but also

¹⁹² P. Lescure, *Contribution aux politiques culturelles à l'ère numérique*, Mission "Acte II de l'exception culturelle" (Contributing to cultural policies in the digital age, "Act 2 of the cultural exception" mission), La Documentation française, May 2013.

¹⁹³ Feasibility study on the implementation of an open metadata register studying proposal No. 79 of the report by P. Lescure.

¹⁹⁴ Berklee ICE, *Fair Music: Transparency and Money Flows in the Music Industry*, 2015.

¹⁹⁵ H. Verdier, *La donnée comme infrastructure essentielle* (Data as an essential facility). Report from the data administrator on data in administrations, 2016-2017, La documentation française, 2018. It should be noted that the Verdier Report uses the notion of essential facility for data, in the political sense of the term, and not as defined under the aforementioned competition law.

organizational and managerial skills. Far from being a mere technical concern, real data governance is vital to analyse and integrate the effects of a data-driven economy.

The French Ministry of Culture could accompany this acculturation, especially for small players, so that, in this sector as in others, the potential of a data-driven economy and AI are expressed for collective interest.

List of people who contributed to the mission

➤ Authorities and official representatives:

B. Pailhès, National Coordinator for France's AI Strategy

HADOPI:

D. Rapone, President

P. Blassel, Secretary General

French Ministry of Culture:

A. de Nervaux, Head of the Legal and International Affairs Department

S. Jacquier, Head of Mission for the Head of the Legal and International Affairs Department

M.-V. Leroi, Digital Innovation Department

N. Orsini, Digital Innovation Department

A. Therwath, DGCA (Directorate General of Artistic Creation), Photography Delegation

C. Devillers-Sigaud, DGCA (Directorate General of Artistic Creation), BAJ (Office of Legal Affairs)

➤ Interested bodies:

ADAGP (French Society of Authors of Graphic and Visual Arts):

T. Maillard, Legal Director

Publishers/book sector:

V. Barthez, Les Editeurs d'éducation, Director

F. Piacentino, SNE (French National Publishers' Union), Digital Commission

M. Trauman, Deputy Director Hachette Innovation Program

A. Robert, Hachette Legal Director

J. Chouraqui, SNE (French National Publishers' Union) Legal Director

L. Bernard, SNE (French National Publishers' Union) Head of Mission for the Legal Department

-L. de Carvalho, Madrigal Chief Legal Officer

M. Benesty, Machine Learning Engineer, former lawyer, Dalloz

C. Blache, SNE (French National Publishers' Union) Head of International Institutional Relations

L. Bernard, Head of Mission for the Legal and Illustration Commissions, Data Protection Delegate

Publishers/press:

L'Alliance de la presse d'information générale (French general information press alliance)

S. Ouachtati, Head of the Social and Legal Division

F. Rimbart, Head of Digital Development

GFII (French Information Industry Association):

D. Berthault, LexisNexis, GFII (French Information Industry Association) President

M. Dessertenne, GFII (French Information Industry Association) General Delegate

A. Raulin, Director of Studies, Bureau van Dijk, GFII (French Information Industry Association) Administrator
F. Campana, Legal Officer, DILA, GFII (French Information Industry Association) Administrator
T. Parisot, Head of Institutional Relations CAIRN, Board Member

SACD (French Society of Dramatic Authors and Composers):

H. Tilliet, Director of Legal Affairs and Audiovisual Contracts

SACEM (French Society of Authors, Composers and Music Publishers):

D. El Sayegh, Secretary General

C. Bonin, Legal Director

SAIF (French Society of Visual Arts and Still Image Authors):

A. Defaux, Legal Director

SGDL (Société des gens de lettres):

G. Pelletier, Director

M. Bensimon, Chief Legal Officer

A. Morvan, Legal Expert

D. Couet-Lannes, Legal Expert

SNEP (French National Phonographic Industry Union)

A. Lasch, Director General

➤ **Businesses:**

Aiva Technologies:

Pierre Barreau, CEO

IBM:

J.-M. Leclerc, Government and Regulatory Affairs Executive

Groupe M6:

A. Cathala, Legal Expert for Copyright Litigation

Netflix:

Marie-Laure Daridan, Director of Public Policy France

Josh Korn, Innovation and Technology Public Policy

Rakuten France:

B. Moutte-Caruel, Director of Legal and Regulatory Affairs

L. Ach, Senior Manager Europe, Rakuten Institute of Technology

Spotify:

F. Pachet, Director of Spotify Creator Technology Research Lab

O. Régnier, Director of European Policy

TF1:

A. Level, Director of Digital Regulatory Affairs

A. Briffod, Chief Legal Officer

➤ **Experts and Qualified Specialists**

G. Azzaria, Professor, Université Laval-Québec

V.-L. Benabou, Professor, University of Aix-Marseille, CSPLA Qualified Specialist

J.-A. Benazeraf, Lawyer, Artlaw, CSPLA Qualified Specialist

A. Billion, PhD student, AI Ethics and Legal, IBM

J. Cabay, Professor, Université libre de Bruxelles

Raja Chatila, Professor, Pierre and Marie Curie University, Director of ISIR (Institute for Intelligent Systems and Robotics)

De Gaulle Fleurance et associés:

G. Courtois, Associate Lawyer

J.-S. Mariez, Associate Lawyer

J.-M. Deltorn, PhD in Physics, PhD Student in Law, examiner at the EPO

S. Dormont, Senior Lecturer, University Paris-Est Créteil

N. Enser, PhD in Law, Lawyer, Vercken-Gaullier Law Firm

IRCAM (French Institute for Research and Coordination in Acoustics/Music):

G. Beller, Director of the Research-Creation Interface Department

J.-L. Giavitto, Director of Research at the CNRS and Deputy Director of the Music and Sound Sciences and Technologies Laboratory (STMS)

P. Esling, Head of the Orchestration Research Axis

J. Nika, Human-Machine Musical Interactions: Research (PhD) & Computer Music Design for Performance & Composition

T. Lebrun, Legal Expert in Digital Law, PhD in Literature Université Laval-Québec

T. E. Synodinou, Professor, University of Cyprus

E. Treppoz, Professor, University Jean Moulin Lyon 3, President of the French AIPPI Group

C. Zolynski, Professor, University Paris 1, CSPLA Qualified Specialist

Written contributions

Hadopi

SACEM (French Society of Authors, Composers and Music Publishers)

SGDL (Société des gens de lettres)

De Gaulle Fleurance et associés