SUI GENERIS RIGHT FOR TRAINED AI MODELS

A Paper Presented by:
Intellectual Property Owners Association
Artificial Intelligence and Emerging Technologies Committee

November 2, 2020

TABLE OF CONTENTS

I. INTRODUCTION

Artificial Intelligence (AI) is a broad term used to describe a multi-aspect discipline in computer science that includes technologies that replicate functions ordinarily attributed to human intelligence. Unlike ordinary software, AI possesses capabilities for learning; when an AI system is exposed to more datasets, the results achieved are generally improved.[1]

In one's daily life, a person may interact with numerous systems and devices that make use of AI, including those that rely on deep neural networks, natural language processing, image data classification schemes, and AI-tailored hardware. For example, navigational systems may use AI to find the best route to a destination; chatbots may service customers by answering or providing further information to questions through the use of AI-based natural language processing; online retailers may use AI-based analytics to inform a customer's preferences and likely next purchases; and self-driving vehicles may use AI together with a combination of sensors, cameras, and radar systems to travel without a human operator.

AI is everywhere, and it is a key driver in the Fourth Industrial Revolution, a technological revolution that will fundamentally alter the way people live, work, and relate to one another.[2] In a study by PWC, AI is projected to add $15.7 trillion dollars to the global economy by 2030, boosting US GDP by 14.5%.[3] Emerging AI technologies are predicted to transform the global economy and have significant implications for America as a global innovation leader.

Increased investment in AI-based innovations requires careful assessment of the intellectual property protections afforded to such innovations. Certain aspects of AI may be protectable by traditional IP protection rights, such as patents, copyrights, and trade secrets. For example,

> Patents may protect elements of AI inventions, *e.g.*, the structure of the database on which the AI will be trained and will act, the process of training the AI model using one or more datasets, the AI algorithm as applied to a technical problem, the results or insights of the AI invention through an automated process, the policies/weights to be applied to the data that affects those results or insights;

---

[1] *Frequently Asked Questions: AI and IP Policy*, World Intellectual Property Organization, https://www.wipo.int/about-ip/en/artificial_intelligence/faq.html (last visited Aug. 4, 2020).

[2] Klaus Schwab, *The Fourth Industrial Revolution: What it Means How to Respond*, WORLD ECON. FORUM (Jan. 14, 2016), available at https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

[3] *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*, PwC, https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html (last visited Aug. 4, 2020).

Copyright protection may apply to the expression or implementation of an AI algorithm or process, such as, in the form of software, the selection of data as embodied in a copyrightable compilation, the output provided by an AI-based system, such as the artwork, the musical production, or journalistic articles; and

Trade secrets may protect secret AI-based information not commonly known or available to the public, that derives economic value from not being generally known, such as AI training data sets.

One important aspect of AI is the model itself, *i.e.,* the formulas, processes, and methods for handling data in an AI-based system. This paper explores how aspects of an AI model may be protectable by patents, copyrights, or trade secrets. It also explores how traditional IP-protection mechanisms may be unavailable or inadequate to provide adequate legal protection for the AI model itself, despite the fact that such AI models may warrant legal protection.

Additionally, this paper considers the creation of a *sui generis* right for trained AI models. The term "*sui generis*" means "of [its] own kind; in a class by itself; unique."[4] A *sui generis* right is created legislatively to protect a very specific type of work that may not be adequately protected by traditional forms of IP. Examples of current *sui generis* IP protections include integrated circuit layouts, ship hull designs, fashion designs, databases (EU), and plant varieties.

In considering the creation of a *sui generis* right for AI models, this paper addresses numerous topics should such a protection right be created, including, *inter alia,* policy considerations, the subject matter covered, rights conferred, registration requirements, fair use, marking, the duration of such a right, and litigation issues. While the issues evaluated herein are done so with a US-jurisdictional focus, some concepts presented may be applicable elsewhere.

II. ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND AI MODELS

Artificial intelligence ("AI") is an umbrella term for a class of different computer technologies which focus on replicating one or more functions attributed to human intelligence. One of the biggest fields today are algorithms that replicate human learning.

Machine learning ("ML") is focused on providing computers with the ability to learn without being explicitly programmed. Typically, with ML, a large data set is digested to produce a trained model that can, in deployment, receive the same types of inputs and output a decision, classification, etc.

[4] *Sui generis*, Wikipedia, https://en.wikipedia.org/wiki/Sui_generis (last visited Aug. 4, 2020).

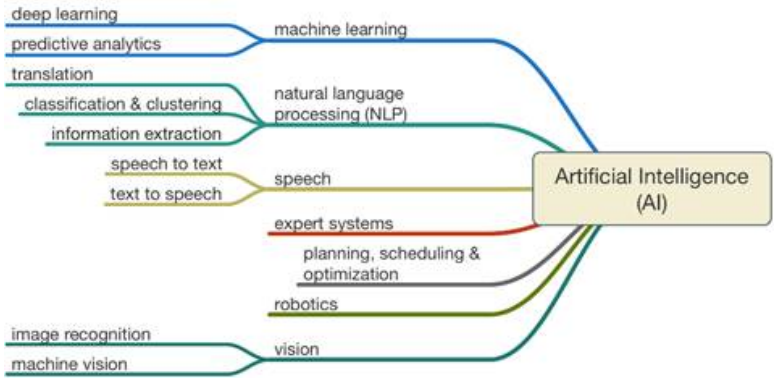

Figure 1 Different computer technologies under the umbrella term "AI." From "Artificial Intelligence in Law – The State of Play in 2015?" by Michael Mills, 2015, Legal IT Insider (https://legaltechnology.com/artificial-intelligence-in-law-the-state-of-play-in-2015/). Copyright 2017 WordsandVision Limited.

Typically, the AI step is the training; the trained model is the result. After the AI step, a trained model is often thought of as a black box. In truth, a set of coefficients for use in one or more mathematical functions is created. These coefficients need to be used wherever the trained model is to be deployed.

How an AI model trains data is exemplified as follows: (1) a data set labeled input and output examples are fed to an algorithm; (2) then, the algorithm runs on the data with its parameters adjusted until it reaches a satisfactory level of accuracy; (3) from this analysis, the algorithm creates a function that can predict future outputs. Suppose, for example, the AI model is given pictures of cats that are labeled as "cats." The model is then trained on the labeled data of "cats" until it can recognize the patterns in the images of cats. As a result, the model would be able to predict if later images are showing "cats" or "not cats" by responding to the previously recognized patterns.[5]

A lot of effort goes into the "AI step," including, data acquisition, data curation, data labeling, model architecture selection, training algorithm selection and configuration, the training itself, validation, performance evaluation, tweaking, and iteration. These efforts result in a collection of coefficients; with the coefficients in hand, the above efforts can henceforth be skipped.

After putting in this hard work, one must consider, what are the best strategies to protect the fruit of the such efforts, even when only textbook (non-inventive) methods and architectures are used?

---

[5] *Example from* Roshan Adusumilli, Artificial Intelligence and Its Application in Finance: How AI will drive the future of the industry, Towards Data Science (Nov. 4, 2019), https://towardsdatascience.com/artificial-intelligence-and-its-application-in-finance-9f1e0588e777.

III. Traditional IP Protections

A. Copyrights

Copyright, is a form of protection in creative works that grants an owner of that copyright the exclusive rights to reproduce, display, perform, and create derivative works of such creative works.[6] To qualify for copyright protection, a work created must be original to the author and possess some degree of creativity.[7]

1. Machine-Created Works are Not Copyrightable

Turning to whether the works created by *machines* are copyrightable, the USPTO and other intellectual property offices throughout the world have provided guidelines that only works created by *humans* can be protected by copyright.[8] According to the Compendium of U.S. Copyright Office Practices, "[c]opyright law only protects works of authorship that are created by human beings. Works made through purely mechanical processes or with an automated selection and arrangement are not eligible for copyright protection. *The U.S. Copyright Office will refuse to register a claim in a work that is created through the operation of a machine or process without any human interaction*, even if the design is randomly generated."[9] In a recent federal copyright case, known as the *monkey selfie dispute*, the court of Appeals of the Ninth Circuit held that an animal is not entitled to a copyright.[10] Furthermore, since copyright protects the expression of ideas and not the ideas itself, the software code underlying the AI models and training datasets, used to train the AI models, may be protected as a literary expression under copyright laws.[11] Copyright protection, however, is unlikely to be available for the functionality of AI models in

---

[6] 17 U.S.C. § 106.

[7] *Feist Pubs., Inc. v. Rural Tel. Svc. Co., Inc.*, 499 U.S. 340, 111 S. Ct. 1282, 113 L. Ed. 2d 358, 1991 U.S.

[8] Andreas Guadamuz, "Artificial Intelligence and Copyright," WIPO Magazine (2017).

[9] *Compendium: Visual Art Works*, United States Copyright Office, https://www.copyright.gov/comp3/chap900/ch900-visual-art.pdf (last visited Aug. 4, 2020).

[10] *Naruto v. Slater*, No. 16-15469 (9th Cir. 2018); *Naruto v. Slater*, Justia, https://law.justia.com/cases/federal/appellate-courts/ca9/16-15469/16-15469-2018-04-23.html (last visited Aug. 4, 2020) (The panel held that the monkey lacked statutory standing because the Copyright Act does not expressly authorize animals to file copyright infringement suits).

[11] *Copyright Registration of Computer Programs*, United States Copyright Office, https://www.copyright.gov/circs/circ61.pdf (last visited Aug. 4, 2020).

automating a decision process, the underlying algorithms, and output of the AI models.[12] It is hard to argue a modicum of creativity in a collection of numbers generated by a machine.

2. Machine-Created Works Must be Creative

The U.S. Copyright Act requires the work to be creative, meaning the work must be the product of a creative spark in order to be copyrightable. Works lacking creativity are not afforded copyright protection.[13] The U.S. Supreme Court has shed light on this creativity requirement and ruled that compilation works, such as a database must contain a level of creativity in order to be copyrightable. Accordingly, just listing names, phone numbers, and addresses lack creativity.[14] Under Section 313.2 of the Compendium of the U.S. Copyright Office, works produced by a machine or mere mechanical process that operates randomly or *automatically without any creative input* or intervention from a human author will not be afforded copyright protection.[15] Machine-generated AI models replicate a decisions process that enables automation and understanding. These AI models are trained using data and human expert input to automate the decision process.[16] Merely automating a process by an expert would most likely be considered to lack creativity, and therefore, would not be afforded copyright protection.

3. Enforcement and Infringement Considerations

The AI field, namely AI system, is a new and constantly evolving technology, with minimal clarity on the working and processes of the AI system, outside of the AI developer community.[17] Such uncertainty has created issues in regards to establishing and proving infringement, as well as enforcement issues on the legitimacy of using third-party copyright-protected content to train AI

---

[12] *See supra* note 11.

[13] *Creativity Requirement*, US Legal, https://copyright.uslegal.com/enumerated-categories-of-copyrightable-works/creativity-requirement/ (last visited Aug. 4, 2020).

[14] *Feist Publications, Inc., v. Rural Telephone Service Co.*, 499 U.S. 340 (1991) (holding that facts are not original. The first person to find and report a particular fact has not created the fact; he has merely discovered its existence. Facts may not be copyrighted and are part of the public domain available to every person.)

[15] *Introduction to the Third Edition of the Compendium of U.S. Copyright Office Practices*, United States Copyright Office, https://www.copyright.gov/comp3/docs/compendium.pdf (last visited Aug. 4, 2020).

[16] *AI/ML Models 101: What Is a Model?* Osprey Data, https://www.ospreydata.com/2020/02/24/ai-ml-models-101-what-is-a-model/#:~:text=In%20AI%2FML%2C%20a%20model,when%20provided%20that%20same%20information (last visited Aug. 4, 2020).

[17] *Understanding the AI Skills Gap*, Technative, https://www.technative.io/understanding-the-ai-skills-gap/ (last visited Aug. 4, 2020); Brigitte Vézina and Diane Peters, *Why We're Advocating for a Cautious Approach to Copyright and Artificial Intelligence*, Creative Commons (Feb. 20, 2020) https://creativecommons.org/2020/02/20/cautious-approach-to-copyright-and-artificial-intelligence/.

models with varying laws in different jurisdictions.[18] For example, current UK copyright laws and the EU Copyright Directive (to be implemented in the EU Member States before June 2021) permits "text and data analysis for non-commercial research."[19] However, Section 107 of the U.S. Copyright Act provides a statutory framework for the "Fair Use Doctrine" that permits the use of the third-party content to train AI systems, if the use is sufficiently transformative and does not compete with the original works.[20] The Fair Use Doctrine, therefore, creates an exception to infringement in the U.S. that is not applicable in other jurisdictions.

Based on the above described limitations of copyright protection, the extension of such traditional copyright protection to machine-created works, namely AI models, is unlikely at best. However, due to the large investments in AI technology, there are legitimate business concerns associated with the IP protection of AI models.[21] Such concerns have served as a catalyst to exploring alternate considerations, such as *sui generis* rights, or even other forms of traditional IP protection (*i.e.*, patents) to provide IP protection to AI models.

B. Patents

Given that AI models fundamentally rely on mathematical constructs, the Supreme Court's decision of *Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208 (2014), regarding subject matter eligibly pursuant to 35 U.S.C. § 101, can create unique challenges for AI related inventions. In particular, inventions claiming AI models are at risk of being found "directed to" an abstract idea without "significantly more."

Since *Alice*, however, the Federal Circuit has repeatedly found software-related inventions, which AI related inventions pertain, as patent eligible where such software-related inventions demonstrate an "improvement" to the functionality of a computer.[22] For example, most recently,

---

[18] Yohan Liyanage and Kathy Berry, *INSIGHT: Intellectual Property Challenges During an AI Boom*, Bloomberg Law (Oct. 29, 2019) https://news.bloomberglaw.com/ip-law/insight-intellectual-property-challenges-during-an-ai-boom.

[19] *See supra* note 18; Copyright *laws and artificial intelligence*, American Bar Association (Dec. 2017), https://www.americanbar.org/news/abanews/publications/youraba/2017/december-2017/copyright-laws-and-artificial-intelligence/; *Directive on Copyright in the Digital Single Market*, Wikipedia https://en.wikipedia.org/wiki/Directive_on_Copyright_in_the_Digital_Single_Market (last visited Aug. 4, 2020).

[20] *See Graham v. Prince*, 265 F. Supp. 3d 366, 370-73 (S.D.N.Y. 2017); Sarah Ligon Pattishall, *AI Can Create Art, but Can It Own Copyright in It, or Infringe?*, LexisNexis (Feb. 28, 2019), *available at* https://www.lexisnexis.com/lexis-practice-advisor/the-journal/b/lpa/posts/ai-can-create-art-but-can-it-own-copyright-in-it-or-infringe.

[21] *How tech giants are investing in artificial intelligence*, Tech Advisor (Nov. 8, 2019) https://www.techadvisor.co.uk/feature/small-business/tech-giants-investing-in-artificial-intelligence-3788534/

[22] It is noted that demonstrating an improvement to the functionality of a computer is but one way to demonstrate eligibility under the Supreme Court's *Alice* jurisprudence. *See, e.g.,* M.P.E.P. § 2106. Also, it is recognized that

in *Uniloc USA v. LG Elec. USA*, the Federal Circuit provided that "[o]ur precedent is clear that software can make patent-eligible improvements to computer technology, and related claims are eligible as long as they are directed to non-abstract improvements to the functionality of a computer or network platform itself." No. 19-1835, slip op. at 10 (Fed. Cir. Apr. 30, 2020).[23]

The Federal Circuit has specifically recognized the importance of AI and the danger of applying *Alice* in a way to impede innovation: "the danger of getting the answers to [the] questions [raised by *Alice*] wrong is greatest for some of today's most important inventions in computing, medical diagnostics, artificial intelligence, the Internet of Things, and robotics, among other things." *Smart Systems Innovations, LLC v. Chicago Transit Authority*, 873 F.3d 1364, 1378 (Fed. Cir. 2017).

In addition, the USPTO has signaled that AI inventions are patentable. For example, in the U.S. Patent Office (USPTO)'s 2019 Revised Patent Subject Matter Eligibility Guidance (January 7, 2019), the UPSTO provides an example AI-based method for training a neural network for facial detection, that is patent eligible under *Alice*.[24]

In general, in view of *Alice*, when considering AI inventions for patent protection, a patentee should focus on the unique technical features that may be identified in an AI-invention workflow and related components. These may include, for example:

- The pre-processing of training data (*e.g.,* preparing unique datasets for input into a particular AI algorithm);
- The training process (*e.g*., improvements or adjustments to a machine-learning or neural network algorithm);
- The application of trained models (*e.g.,* to control machines or to provide unique results); and
- The hardware that executes a trained AI model, including any improvements to the hardware or its deployment in a given technical field.

In addition, under *Alice*, an important test is whether the claims of a computer or software related invention recite a specific improvement to "the functioning of a computer itself" or "any other technology or technical field." *See Uniloc USA,* No. 19-1835, slip op. at 5-10. With an AI-related

---

demonstration of patent eligibility of general software-related inventions remains challenging as the Federal Circuit has heretofore issued numerous decisions on the subject but where such decisions lack any clear standard or rule for achieving eligibility for software-related inventions.

[23] *Uniloc USA, Inc., et al. v. LG Electronics USA, Inc., et al.*, No. 2019-1835 (Fed. Cir. (N.D. Cal.) Apr. 30, 2020).

[24] *Subject Matter Eligibility Examples: Abstract Ideas*, United States Patent Office (Jan. 7, 2019), available at https://www.uspto.gov/sites/default/files/documents/101_examples_37to42_20190107.pdf.

invention, this can be described and claimed where an AI model improves the underlying functionality of a system, *e.g.*, an autonomous vehicle, where the AI model enhances, *e.g.,* the safety, fuel efficiency, or power usage of the vehicle.

On the other hand, while the Federal Circuit has generally recognized AI as an important technology (*see Smart Systems,* 873 F.3d at 1378), it has yet to address the eligibility of an AI related invention. Given this, it is possible that an AI related invention, especially one with claims having features typically found ineligible (*e.g.*, claims directed to fundamental economic practices) may be found ineligible despite incorporating AI aspects. Accordingly, in such circumstances, other legal protections, such as those found herein, may need to be considered.

C. Trade Secrets

Although patenting is an option, there are also challenges to patent protection on AI models as described above. As an alternative, trade secret can protect information where the owner "has taken reasonable measures to keep such information secret" and the information "derives independent economic value, actual or potential, from not being generally known" to other persons. *See* 18 U.S.C. § 1839(3) (Federal Defend Trade Secrets Act, definition of "trade secret"). Protection is available at the federal and state level.

A benefit to using trade secret to protect AI models is the relatively low cost. Unlike patents, trade secrets involve no registration and have immediate effect. Furthermore, trade secrets do not require compliance with formalities or public disclosure. If AI models are likely to be continuously revised or retrained, then this low-cost benefit can make trade secrets an appealing option.

On the other hand, trade secrets may be an inefficient means of protection in software development. For example, if the model is deployed on a product, reverse engineering becomes possible. A trade secret is protected as long as it remains a *secret*. Even if reverse engineering is prohibited by contract, the value of one's crown jewels may be significantly diminished if it is already exposed publicly. The algorithm can be moved to a server to better protect against reverse engineering, but that may not always be practicable. Additionally, collaborative scientific research is intrinsic to the development of new technology, leading to parties engaging in joint ventures, subcontracting relationships, and/or licensing opportunities.[25] While parties can take steps to

[25] Karthika Perumal, *How Do We Thrive Together: Understanding the "Yours, Mine, and/or Ours" of Intellectual Property Rights in Joint Development Agreements*, Lexology (Oct. 22, 2019) ("Collaboration is a key strategy for growth of many companies in certain sectors, such as energy, technology, and life sciences, where research and development is critical to maintaining a competitive advantage, but can be incredibly expensive to undertake. By working together on R&D, companies can both contain costs and bring together the best ideas and information from both parties.")

protect their trade secrets, such as through nondisclosure agreements and restricted access, difficulties in maintaining confidentiality continue to persist.[26]

Some industries also make trade secret protection impracticable because solutions need to be proven or disclosed before acceptance. For instance, a greater demand for AI systems to be transparent to other technology producers or consumers in the general public will make it difficult to keep valuable AI models a trade secret.

D. Open Contribution – Journals, Open Source, Etc.

Open contribution, specifically open source contribution or journals, is a form of free contribution to an AI model or other aspect of an open source software.[27] In open contribution, particularly for open source software, members of the public may work collaboratively by providing comments, or edits to various aspects of a project.[28] The owner of the open source software, or AI model, may review and determine whether to accept or reject the edits, as well as whether to modify the project based on the commented provided. In general, open contribution of software reduces development time, community supported development and code review, and platform adoption. To maintain IP integrity, various forms of IP protection may be implemented, such as customized or standard open source licensing agreements, or a hybrid approach in which traditional IP protection (*e.g.,* copyrights and patents) are still sought in parallel to seeking open contribution.[29]

When traditional IP protection, namely patents and/or copyrights, are sought with open contribution, traditional IP protection should be sought prior to open contribution and should be broad enough to include any substantive modifications implemented on the AI model.[30]

---

[26] Steven R. Daniels and Sharae Williams, *So You Want to Take a Trade Secret to a Patent Fight? Managing the Conflicts between Patents and Trade Secret Rights*, American Bar Association (Aug. 5, 2019) https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2018-19/july-august/so-you-want-take-trade-secret-patent-fight.

[27] *How to Contribute to Open Source,* Open Source Guides, https://opensource.guide/how-to-contribute/ (last visited Aug. 4, 2020).

[28] *See supra* note 25; Adrian Bridgwater, *Smart Artificial Intelligence Needs An Open (Source) Classroom*, Forbes (Jun. 26, 2020) https://www.forbes.com/sites/adrianbridgwater/2020/06/26/smart-artificial-intelligence-needs-an-open-source-classroom/#34b970651785; C. Titus Brown, *A framework for thinking about Open Source Sustainability?*, Living in an Ivory Basement (Jul. 2, 2018) http://ivory.idyll.org/blog/2018-oss-framework-cpr.html [hereinafter Brown].

[29] Gideon Myles, *How to Participate in Open Source While Maintaining IP Integrity*, IP Watchdog (Sep. 13, 2017) https://www.ipwatchdog.com/2017/09/13/how-to-participate-in-open-source-while-maintaining-ip-integrity/id=87858/

[30] *See supra* note 27.

Otherwise, the owner of the AI model may be precluded from seeking any patent or copyright protection for an AI model.

On the other hand, by implementing a non-traditional approach for IP protection, open contribution is more economically affordable than the more traditional copyrights or patent methods of IP protection. The cost of filing a non-provisional patent application far extends the nominal, if not free, cost of seeking open source contribution from a community of developers, as well as working with a larger pool of talented information technology professionals to improve the AI models at no additional cost, since there is zero cost associated with open contribution.[31] Therefore, open contribution may lead to an abandonment or significant reduction in copyright registrations and patent filings for AI models.

Furthermore, in instances, where previously sought IP protection no longer protect the final AI model, or broader IP protection is sought in light of the constantly evolving nature of AI industry, open source licenses may be sought to protect the IP rights associated with the AI model, instead of seeking traditional IP protection. Depending on the modifications to the AI model, namely the software code, functionality, output and algorithms, prior traditional IP protection may no longer be adequate to protect the final AI model and its associated features after open contribution has been completed. The open source license may be crafted to provide IP protection to the final AI model, such as AI output, functionality, software code, algorithms, that were excluded from the traditional IP protection of copyrights and/or patents.[32] Additionally, the AI industry is constantly evolving and changing in complexities, challenges, and functionality. In light of the endless series of unknown future advancements or changes in AI technology, there exists minimal understanding and clarity on the processes that enable AI systems to work, outside of the AI developer communities. Since open contribution is largely dependent on the community of AI developers, an owner may seek open contribution. However, due to the ongoing nature of open contribution in which the code, AI output, or other aspects of the AI model may significantly change during the course of open contribution. IP protection in the form of open source licenses, rather than traditional IP protection, may be utilized as a preferred method to protect IP rights associated with the AI model. The open source license to grant broader IP protection for the AI model. Some popular standard open source licenses that are widely used, or have strong communities, include Apache License 2.0, MIT license, FreeBSD license, and Common Development and Distribution License. In addition, new open source datasets have developed to provide a large volume of contracts, including open source licenses, that may be useful for granting IP protection for the AI model, for example, Atticus Project. Many of these open source datasets include permissible license terms and free to the public.[33] Some of the source licenses may be customized may include

---

[31] Michael Tiemann, *The (awesome) economics of open source*, Opensource.com (Sep. 13, 2018) https://opensource.com/article/18/9/awesome-economics-open-source; *see supra* note 26 on Brown.

[32] *See supra* note 27.

[33] https://www.atticusprojectai.org/white-paper.

adding a custom patent license to an existing open source license, or crafting a termination clause that includes specific grounds for terminating the open source license related to the specific concerns or circumstances connected to the open source community.[34]

Lastly, other non-traditional forms of IP protection, such as *sui generis,* have been compared to open contribution as a vehicle to obtain broad IP protection to AI models. However, unlike *sui generis*, open contribution is considered to require less legislative involvement and possibly less economic investment, depending on the *sui generis* protection process enacted by the legislature.[35] With open contribution, no legislative involvement is necessary to craft special legislation for IP protection associated with AI technology, namely AI models. But rather, open contribution merely requires free contribution from one or more communities of developers. Additionally, the openness provided by open contribution seems to promote competition and socially beneficial for the common good.[36]

In conclusion, with open contribution, the IP protection for AI models may largely depend on the developers or open source community complying with the terms of the open source license, and/or the licensor (owner of the AI model) seeking traditional IP protection, as necessary, in parallel with or prior to engaging in open contribution.[37]

E. Existing *Sui Generis* Rights

In their present form, *sui generis* rights are generally a targeted form of protection than traditional forms of intellectual property. However, the rights are also conferred through less rigorous procedures than are often required. *Sui generis* rights differ in that they are not restricted to products of creativity or particular forms of expression and are instead rewarded to protect the investment and labor required to produce certain works or products.[38] The intent of the protection is to limit unauthorized extraction and re-utilization of the works in order to protect the underlying investment necessary to produce or compile the works. Examples of *sui generis* IP protections have included integrated circuit layouts, ship hull designs, fashion designs, databases (EU), and

---

[34] *See supra* note 27.

[35] Dam, Kenneth W., *Some Economic Considerations in the Intellectual Property Protection*, (Coase-Sandor Institute for Law & Economics Working Paper No. 26, 1994), available at https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1457&context=law_and_economics.

[36] Nick Bostrom, *Strategic Implications of Openness in AI Development*, Global Policy (2017), available at https://www.nickbostrom.com/papers/openness.pdf.

[37] *See supra* notes 27 and 34.

[38] Mauritz Kop, *Machine Learning & EU Data Sharing Practices*, Transatlantic Antitrust and IPR Developments, Stanford University, 7 (Issue No. 1/2020).

plant varieties. We review two examples of *sui generis* protections with respect to the topic of trained model rights – masked works and databases.

1. Mask Works

Sui generis protection on masked works cover images that represent a pattern for the layers of a semiconductor chip. Under the Semiconductor Chip Protection Act of 1984, the associated circuit layout becomes legally protected upon registration similar to Copyright registration and is processed by the US Copyright office. If successfully registered, the owner/applicant receives exclusive rights to reproduce the mask itself, to import or distribute chips incorporating the mask, or induce others to perform the preceding. The exclusion prohibits only copying, not independent creation and allows for reverse engineering for instructive purposes and subsequent incorporation into new original works; the rights lasts 10 years.

In contrast with the database rights described in the following section, mask works require some level of originality. Accordingly, while the protection of mask works is specialized to suit a specific category of invention, the qualifications for protection maintain aspects of originality or novelty associated with conventional IP rights. More specifically, a mask work must not only be original in terms of independent creation, but also must not be "commonplace, or familiar in the semiconductor industry, or variations of such designs, combined in a way that, considered as a whole, is not original."[39] In contrast, other forms of *sui generis* rights withdraw the requirement for originality entirely, and instead recognize the value and investment associated with a work as opposed to the creative aspects.

2. Databases (EU)

Recognized in a small number of jurisdictions such as the European Union and United Kingdom, *sui generis* protection on databases covers collections of data arranged in a systematic way and individually accessible by electronic means. The developer or owner must "show that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents."[40] Once shown, the right is created automatically and does not require registration.[41] The protection affords the owner the right to prevent "extraction and/or re-utilization of the whole or of a substantial part … of the contents of that database."[42] These rights last 15 years, and no registration requirement is mentioned in European Directive.

---

[39] 17 U.S.C. § 902(b)(2).

[40] Directive 96/9/EC, Article 7, paragraph 4.

[41] Directive 96/9/EC, Article 10, paragraph 1.

[42] Directive 96/9/EC, Article 7, paragraph 1.

Advocates for database rights posit that the legal uncertainty surrounding database and related information systems creates legal uncertainty and shifts away from dissemination toward trade secrets, thus stifling innovation.[43] The critics of database rights often align with proponents of open source coding who argue that that the rights create an "unnecessary obstruction for the access to and reuse of information...."[44] Additionally, critics argue that the alleged obstruction is not warranted because the value has not been shown to incentivize the database market in the EU. These arguments suggest that innovation would be better served by allowing databases without creative elements to be substantially reused as reference materials for future innovation and collaboration. It is similarly suggested that property rights are not necessary to promote the use and development of databases. Critics also warn that the term of such rights may be perpetual if "substantial investment" in the database occurs with later public disclosure.

IV. BUILDING THE *SUI GENERIS* RIGHT FOR TRAINED AI MODELS

A. Policy Considerations

As discussed above, there are challenges to protecting AI models using existing IP protections. Given the inapplicability or inadequacy of these protections, a mechanism may be needed to reward AI developers or owner's rights to the "sweat of the brow." As explained earlier, training AI models can often times involve numerous steps from data acquisition to the iterative tweaking of the model. AI models play a crucial role in, not just providing an outcome, but are constantly assessed to correct for error and adapt for evolving data. As society places more emphasis on eliminating bias in AI systems, it is not just the data itself that requires handling but also the way the model is applied to curating the data.[45] A high cost to create with a low cost to misappropriate (copy) afterward could cheapen the value of AI models or deter the advancement of newer or more robust models.

It can be helpful to reconsider a few limitations of existing IP. As mentioned above, patents are limited to constructions which are novel and non-obvious. Patents are also presently entangled with the development of subject matter eligibility law, which is of considerable concern to the nature of AI models. Copyrights are not intended to protect the underlying idea, but rather the

[43] *See supra* note 36.

[44] Policy Paper in reaction to the public consultation on the Database Directive, Communia (Aug. 2017).

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

original expression of such an idea, and thus, they may miss the core mark with AI models. Trade secrets may be generally less preferred within the community.[46]

Yet, the potential shortcomings of the existing IP protections alone may not necessarily justify the creation of a new medium in *sui generis* rights. For example, many commentators predicted that patenting of software inventions would stifle independent creation. But those concerns may not have come to fruition in the digital age thus far.[47] Additionally, the availability of anti-trust protections might be considered to balance the threat of over-constraining the marketplace.

More specific to AI models is that the underlying premise that misappropriating another party's AI model is simple or low cost. Indeed, implementation of a trained model for a similar, but non-identical process, may still require considerable investment in implementation, including time, personnel, and supporting infrastructure (*e.g.*, data storage, processing, etc.). In practice, it is unclear that such AI model misappropriation could so readily occur that would undermine the creator's investment.

Assuming misappropriation can practically occur, if adequate protection of AI models requires extensive IP protections across multiple realms (patent, copyright, and trade secret) simultaneously and with some remaining doubt about their adequacy; at what point does this patchwork approach suggest overall inadequacy? Moreover, existing IP can be zero-sum, such that establishing one right can deteriorate or eliminate the other, for example, in comparing the disclosure requirement of patents with the secrecy requirement of trade secrets. Although not necessarily determinative, the potential for unintended impacts of *sui generis* rights on existing IP protections should not be ignored.

A balancing of factors may better reveal the incentives and disincentives for such *sui generis* rights. For example, a significant disproportion in the investments of different parties may be informative. Hypothetically, if an originating party of an AI model attributed a significantly greater level of investment to develop the original AI model than another party as a potential misappropriator of the original AI model, then a comparative framework exists to consider the basis of the *sui generis* right, and the associated policy both for and against such a right.

For the time being, it seems that the concerns which merit consideration in the creation of *sui generis* rights for AI models stem from the particularly high investment costs for development. Accordingly, comparative demonstration of the low barriers for misappropriation may favor such a new right. For example, significantly low resource requirements, low cost, few personnel requirements with low level of technical expertise may indicate disparate investment levels

---

[46] Michael R. McGurk and Jia W. Lu, *The Intersection of Patents and Trade Secrets*, 7 Hastings Sci. & Tech. L.J. 189, 196-200 (2015).

[47] *See generally* John C. Phillips, *Sui Generis Intellectual Property Protection for Computer Software.* Geo. Wash. L. Rev. (1992).

sufficient to justify such a new right. Consideration of investment disparity might likewise provide a benchmark for either or both the establishment of rights and/or enforcement of established rights.

Of course, one counterargument to *sui generis* rights in general is that the existing law is itself quite flexible. Often, in the evolution of humanity, an old rule (whether statutory or judicial) is applied in a modified way to address shortcomings in facing new instances. For example, before the computer age, the U.S. considered adopting *sui generis* rights for computer programs.[48] However, Congress ultimately adopted a framework clearly including computer programs to be copyrightable as literary works under the Copyright Act of 1976.[49] Among other justifications, advisors cautioned against the economic impact of excluding computer programs from copyright protections wherein the programs themselves are "expensive to make and cheap to copy," echoing the intangible inequity of the issue.[50] Yet, "[i]t has proven difficult to apply traditional copyright doctrine derived from cases involving novels, dramatic plays, and fabric designs to protect non-literal aspects of computer programs, which has caused courts to develop special tests for software copyright infringement."[51] Commentators have observed U.S. courts to essentially create a *sui generis* sub-form of copyright protection to fit computer programs into copyright law.[52]

In one practical instance of data protection itself, U.S. case law has denied compensation to sports leagues for data collected from their events.[53] Historically, raw data has not been protected under copyright law and third-parties cannot be prevented from collecting event data from live or recorded broadcasts, which could then be commercially exploited.[54]

However, as entities contract with data companies to collect data from their events, the problem of data reutilization evolves.[55] In the United Kingdom, the landmark decision of *Football Dataco Ltd v. Sportradar GmbH* found that a data company's *sui generis* database right can be violated by third-party publication if the data is real-time, processed sports data rather than merely fixed and easily obtainable characteristics because it requires "considerable skill, effort, discretion,

---

[48] Pamela Samuelson, *Evolving Conceptions of Copyright Subject Matter*, 78 U. Pitt. L. Rev. 17, 66 (2016).

[49] *Id.*

[50] *Id.* at 67.

[51] *Id.* at 69.

[52] *Id.*

[53] *See* Aaron Feld, Note, *Gambling on Sports Data: Protecting Leagues' High-Level Data from Sportsbooks*, 2020 U. Ill. L. Rev. 341, 366 (2020).

[54] *Id.*

[55] *Id* at *367.*

and/or intellectual input."[56] In comparison, the *sui generis* database protections afforded in *Dataco* would appear to have required significant manipulation of traditional copyright protections to distinguish real-time data from other so called "fixture lists" by measures other than the effort invested. Consequently, the level of investment remains a guidepost whether in rethinking existing law or analyzing a *sui generis* form of protection.

Nevertheless, new forms of IP face intense obstacles while patchwork approaches can be costly and require time to develop case law, if at all. With the evolution of technology and business practice in general, IP systems are incentivized to periodically evaluate the role of *sui generis* rights and to determine whether a new approach or a revitalized old approach is appropriate. This is true even if the outcome merely yields consideration of the issues in existing protections, further defining and evolving the nature of the IP system.

B. Subject Matter Covered

The scope and qualifications required to qualify for protection may be broadly considered in relation to content, investment, and originality. Concerning content, the scope of the subject matter afforded protection may be broadly considered. The subject matter can include any mathematical terms embodied as computer data for use in a mathematical function derived automatically by an ML method, as well as any mathematical function or algorithm embodied as computer instructions incorporating such mathematical terms. Accordingly, the considerations for the protection of trained models may, in some ways, be similar to those afforded to software under copyright law. However, as proponents for intellectual property protection for software have argued, copyright protection "only prevents another from copying the particular expression used by the copyright holder in conveying the underlying idea; the underlying idea itself remains unprotected and free for public use."[57] The question then becomes whether the protection of a particular form of expression is adequate to protect works generated by ML and AI. For example, should the expression of a computer-generated model or algorithm be the focus of protection or the underlying operation and resulting capability?

The apparent deficiencies in copyright protection lead to one of the more contentious issues - originality. To avoid this issue, one option is to remove it entirely as in the European model for database protection. Another option is to require originality similar to mask works, which must not only be original in terms of independent creation, but also must not be "designs that are staple, commonplace, or familiar in the semiconductor industry, or variations of such designs, combined in a way that, considered as a whole, is not original."[58] However, variations in mask works may

---

[56] *Id.*

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

[58] 17 U.S.C. § 902(b)(2).

be visibly discernable and supported by a list of distinguishing components or aspects. Such variations that are typically discernible may not be effective if considered in relation to trained models because the differences in trained models may not even be readily distinguishable to their developers. This lack of understanding is related to the primary benefit of trained models - the developers only enabled the system to create the model and do not program the solution as in conventional software development.

As discussed previously, originality in copyright protection requires that the work possess "at least some minimal degree of creativity."[59] As such, an inability to argue a modicum of creativity in AI models makes it nearly impossible to determine and distinguish originality, making an originality consideration similar to mask works unworkable.

Another qualification for trained model rights may be related to eligibility. If intended to avoid the complexities and challenges that prevent patent protection from effectively covering trained models (*e.g.*, inventorship and subject matter), the eligibility requirements for trained models may be relaxed. However, the subject matter covered may at least need to be distinguished from software to avoid abuse. One possibility would be distinguishing trained models from software by requiring applicants to demonstrate that a model was derived automatically from an ML method or the product of a computer. Such a requirement could be enforced through a requirement of documentation. The documentation requirements may provide a framework to distinguish trained models from software written or generated by human influence. Such a documentation can also provide some level of explainability to AI.[60] Even so, distinguishing the extent of human intervention necessary to preclude rights, or alternatively the level of autonomy of a computer, may be challenging.

Subject matter qualifications may also extend to investment similar to *sui generis* database protection (a "significant investment" model). That is, developers of trained models may need to demonstrate a substantial investment in either training or enabling the generation of a trained model. Such a requirement may be utilized as an alternative to a significant demonstration of originality. For example, rights could be afforded to developers who can demonstrate independent creation, which may be substantiated by documenting investment and labor related to supplying training data and/or equipment to process the trained data and optimize the coefficients of a model. In this way developers would have to be able to show that there was, qualitatively and/or quantitatively, a substantial investment in either the obtaining or verifying the model or its coefficients.

A qualitative showing could include the size of the investment in training the model. Efforts in procuring and curating data to train the model could be factored into the process of training. On

---

[59] See supra note 7, *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, 499 U.S. 340, 345 (1991).

[60] *Explainable Artificial Intelligence*, Wikipedia, https://en.wikipedia.org/wiki/Explainable_artificial_intelligence (last visited July 14, 2020).

the quantitative aspect, one could assess the value of the algorithm as compared to a certain functionality or to the overall AI system declared at the time of registration or notice.

C. Rights Conferred

Whatever the circumstances which qualify certain subject matter for protection under a *sui generis* AI right, such as the "significant investment" model outline above, or others which may be proposed in the future, the next logical question becomes what other rights are conferred upon the holder. The rights chosen will vary depending on the ultimate goals of the *sui generis* AI right, with broader rights likely to encourage development and implementation by industry, while narrower rights may instead focus on the protection on individual autonomy and data protection.

Drawing a comparison again to the *sui generis* database right, discussed above, one set of rights might prevent copying, or more specifically "the extraction or re-utilization of the whole or a substantial part, evaluated qualitatively or quantitatively" of the AI.[61] This right could apply to an AI in three contexts (as could any proposed right). First, as applied to the *input* used by an AI, the right would prevent competitors and others from determining the exact data from which an AI learned and developed its functional algorithm. Such a right, though, would ultimately cover a broad range of data, which would be difficult to track and enforce, since one would be required to monitor the data collecting practices of various other entities. Further, such a right may allow for protection of *personal* information, as competitors, who may or may not have explicit consent from the person or entity whose data is being used, cannot use that data. This protection, though, would likely be better covered by data protection laws, such as the EU's GDPR[62] or the US's CCPA,[63] rather than a *sui generis* right. While there may be advantages to applying a *sui generis* AI right to the input of an AI, there are likely better aspects to which it could be applied.

The *algorithm* or *function* used by the AI, though, might provide a better target for this set of rights. Starting from a base set of variables and developed through the use of input data and other parameters, the AI algorithm represents one of the functional products of the AI. As it is fed data, the algorithm is refined and better meets the goals of its creators, representing the intellectual product of its creators. If we consider a "significant investment" model as identifying the subject matter most qualifying for protection by a *sui generis* AI right, this algorithm represents the product of its creators' intellectual "investment," since they created the initial algorithm *and* chose the data on which to further train and refine the AI. One must also consider, though, the breadth of rights conferred. Considering the interrelated and foundational nature of much computer

---

[61] European Commission, Evaluation of Directive 96/9/EC on the Legal Protection of Databases, SWD (2018) 146 final at 57.

[62] Regulation (EU) 2016/679, On the *Protection of Natural Persons with Regard to the Processing of Personal Data and On the Free Movement of Such Data, and Repealing Directive 95/46/EC* (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[63] Cal. Civ. Code §§ 1798.100–.192 (West 2018) [hereinafter CCPA].

software, allowing for infringement based on extraction or re-utilization of a "substantial part" of an AI algorithm, may be overbroad. As such, a further refinement to the right might extend only to protection of the algorithm itself, or at least, the algorithm as initially created, before it is trained.

Finally, these rights could apply to the *output* of an AI, whether it be a song, a picture, data, or any of myriad possible output forms. One must obviously consider whether this output is (or could be) covered by another IP right, such as a patent, a copyright, or perhaps even another *sui generis* right. Given that copyrights and patents, at least are of limited use when protecting the output of an AI,[64] the right to prevent others from copying *specifically* the output of an AI might hold great appeal to various parties, including both large businesses as well as smaller actors.

Thus far, the focus has been only a single right, the right to prevent the extraction or re-utilization of the whole or a substantial part of an AI, as could be applied to the three primary facets of an AI, however this is certainly not the only conceivable right. Beyond a prohibition against copying elements of AI, one can conceive of other rights which might apply to these elements. For example, inherent in a right to prevent copying might be a right to prevent unlawful distribution of any medium which contains or stores data that embodies the subject matter. Further, beyond active copying or storage of this data, there may exist a right to prevent others from *inducing* others to engage in such behavior.

Each of these rights can be considered in the context of the specific elements which they would be purported to protect, with specific exceptions carved out where necessary (for example, an individual whos personal data was used as an input might logically never be held liable for reproducing that information elsewhere, subject to certain limits). Overall, the most important right would likely be the right against copying; this right, like the *sui generis* database right, would be able to cover a broad range of data and information. However, as discussed above, it might apply better to certain facets of an AI than other facets.

D. Registration

In order to fairly prevent others from copying or using an AI, it is important to ensure what cannot be copied is understood. Accordingly, some form of registration or notice system should be put in place to ensure that the public and competitors have access to the AI. Beyond notice, registration serves another essential function: quid pro quo.[65] In return for what is essentially an exclusive

---

[64] In re Application of Application No. 16/524,350, USPTO (Apr. 22, 2020), https://www.uspto.gov/sites/default/files/documents/16524350_22apr2020.pdf (requiring a natural person to be named an inventor for a patent); *Naruto v. Slater*, 888 F.3d 418 (9th Cir. 2018) (holding that only natural persons can be an "author" under the copyright laws of the US).

[65] See, *e.g.*, Elizabeth Pesses, *Patent and Contribution: Bringing the Quid Pro Quo into Ebay v. Mercexchange*, 11 Yale J.L. & Tech. 309, 320 (2009).

right to have and use the AI (and possibly elements thereof), the AI inventor must teach their AI to the public for its use and benefit once the term of exclusivity expires.

Though the specific details of such a system might vary, certain aspects of potential systems can be investigated. The first aspect, of course, is *what* should be registered? At a minimum, the most useful facet of an AI that could be disclosed, in terms of teaching the public, is the model itself and the algorithms from which it is composed. This would allow the public to both practice the model (when it is freely available) while also providing direct notice to competitors on what *not* to copy or use.

Beyond the algorithm/model itself, registration might also require the disclosure of certain input data used by the model. This disclosure could take various forms. A truly monumental requirement might require ongoing disclosure of all data that is used by the AI. Beyond the obvious privacy concerns that this would implicate, the sheer infrastructure alone required to store this much data makes it infeasible. On the other hand, requiring *no* disclosure might prevent the registration from properly teaching the public. If future users do not know what kind of data should be used by the AI, then they would be prevented from utilizing the AI without a certain degree of experimentation. Perhaps a more ideal compromise would be to require the categories of input data which were used to train the model (or which can be used by the model in the future). Such a system would teach later users what data should be input (perhaps ideally input, rather than extending protection to *only* those inputs) without impinging on various privacy concerns that could arise with registration. This disclosure might relate to either 1) the data set used just to train the model, or 2) the intended input data on which the AI is meant to further develop or produce. A similar category system might apply to the output of the AI.

As briefly touched upon above, any registration system related to the underlying data used by an AI raises questions of privacy and data protection, but there is another area of law that might be implicated: trade secrets. Whether it be the AI owner's own secrets or those licensed for use for particular purposes, any registration system should include specific carve outs and exceptions to allow trade secrets to be protected. This might entail initial disclosure to the registration agency, which then keeps the trade secret or personal data private or an exception which allows for a non-disclosure of particular "sensitive" materials.

A final matter to consider would be examination of an AI. Considering the scale of an AI algorithm which may encompass thousands of variables, coefficients, and other code, individual examination of each elements for uniqueness seems unfeasible. Rather, registration should include only a brief examination to ensure that the subject matter is eligible for registration and protection but otherwise allow for registration of any qualifying material. This would prevent the logistical nightmare involved in examining the minute details of every AI to be registered, while also ensuring that registrants can obtain the protection they desire for the subject matter created.

E. Independent Development v. Copying

An important consideration for infringement of any intellectual property right is the dichotomy between strict liability and intentional copying.[66] Under a strict liability regime, if a user produces a copy of (or otherwise infringes) a work, their intent is irrelevant. They do not need to *intend* to copy an existing work, so they may be liable even for accidental or incidental copying. On the other hand, intent-based regimes require the alleged copier to intend to copy a work, or at the very least know about the work which is being copied.

Each regime has its benefits, measured again against the incentives which it can create. A strict liability regime requires relatively little evidence to show infringement, and thus can be more favorable to patent owners, who need only prove copying of their work, not intent of the allegedly offending party. On the other hand, an intent model protects "innocent" infringers who had no knowledge of the work which they inadvertently copied. In order to prove infringement, an owner would need to meet the higher threshold of showing the alleged copier's intent. Alternatively, a slightly lesser threshold to meet might be a knowledge requirement. Rather than showing that an alleged infringer *intended* to copy an AI, infringement might only require that the alleged infringer knew or should have known of the AI they are accused of copying. This knowledge might be provided for (and even assumed) in various ways.

Registration, discussed above, might create constructive knowledge nationwide (or perhaps even globally). Not only would this further incentivize registration, but it would provide a relatively easily accessible database for interested parties to compare and contrast their own work with that which already exists. Another method of creating constructive notice might be by the use of notices sent directly to the offending party, similar to takedown notices use in copyright law.[67]

Regardless of the intent of an accused infringer, another question should be considered: how *much* copying should count as infringement? Pure identity of code (or input data) would be incredibly easy to bypass, whereas too loose a standard could sweep up a variety of innocent actors. Again, the *sui generis* database right can be informative on that point. That right prevents extraction of all or "a substantial part" of the contents of a database, which may be determined either qualitatively or quantitatively by the government body enacting the right.[68] A similar standard could be applied to a *sui generis* AI right, with governments given the freedom to fine tune the law

---

[66] *See* Christopher A. Cotropia & Mark A. Lemley, *Copying in Patent Law*, 87 N.C. L. Rev. 1421 (2009) (discussing the differences between infringement as a strict liability offense versus intentional copying).

[67] *See generally* Jennifer M. Urban & Laura Quilter, *Efficient Process or "Chilling Effects"? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 Santa Clara Comp. & High Tech. L.J. 621, 626 (2006) (outlining the effect of a takedown notice on the knowledge requirement of infringement).

[68] *See* Samuel E. Trosow, *Sui Generis Database Legislation: A Critical Analysis*, 7 Yale J.L. & Tech. 534, 562 (2004).

to determine the most helpful or relevant degree of similarity between a registered (or otherwise preexisting) AI and an allegedly infringing one.

F. Fair Use

Explicit carve outs for academic and instructive should be considered, while the creation of derivative works could still constitute infringement. It would be very easy to seed a model, train it on a bit of additional data, and produce slightly different coefficients. It is also common to *extend* a trained model to perform a different but related task.

Fair use may apply in several different contexts when considering AI.[69] One is whether AI can *use* copyrighted data in training or for other uses. Such use, within the U.S. at least seems allowable. *See Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015) (allowing unauthorized digital copying of copyright-protected works, creation of a search functionality related to those works, and display of snippets from those works as non-infringing fair uses). While the *Author's Guild* case suggests that the creation of a searchable/readable database qualifies as a fair use, it does not squarely address the issue of whether AI can use such collected data without implicating the same concerns.

In order to eliminate these concerns, one element of a *sui generis* AI right might be a broader interpretation or application of the fair use doctrine. For example, AI may be permitted to integrate copyrighted works into its training without infringement, so long as it does not retain or substantially copy the work. Of course, safeguards would need to be put in place to ensure that such an exception is not abused. Greater damages for infringing activity, or perhaps more strenuous marking requirements, could be used to either disincentive abuse or make it easier to detect and prevent. Regardless of the means use, it must be recognized that "use" by AI may not be equivalent to use by a natural person.

Another consideration for fair use is use of AI *by others*. Given the collaborative nature of AI and the fact that governments likely wish to incentive further development and growth, a fair use regime which allows for further training while also protecting a developer's investments is ideal. According to copyright law, fair use can be implicated when copying protected materials for a "transformative" purpose. Within the copyright sphere, such purposes can include criticism, parody, and commentary.[70] Within the AI context, different purposes may be desired. For example, use for the purpose of training AI might be desirable, not only by the developer who does not have access to the same data set but also by those who have data access but do not have access to the algorithms.

---

[69] *See generally* Joshua I. Miller, *Towards a Doctrine of Fair Use in Some of Patent Law*, 2 Am. U. Intell. Prop. Brief 56 (2011).

[70] *See Lewis Galoob Toys, Inc. v. Nintendo of America*, 964 F.2d 965 (9th Cir. 1992).

A criticism of this approach is that it would be overly time consuming while also easy to manipulate. Developers would need to carefully track the purposes behind their use of AI so that, if accused, they could present this information to show that they were only using AI for training purposes. Alternatively, "bad actors" might use a fair use exception as a way to wiggle out of their legal liabilities and use AI beyond the scope of what a proper fair use would cover, to their own commercial gain. But one can conceive of safeguards which would prevent this abuse while still allowing for the development and further training of persons and organizations acting in good faith. Safeguards might include non-commerciality of the use, as well as others which ensure that the user does not simply copy the underlying algorithms (even though such copying might be of limited use without the original training data). Any test of this nature will be, at best, relatively subjective, depending on the state of mind of the accused infringer/fair user and ultimately difficult to monitor. However, such subjective tests appear throughout the IP legal landscape and parties still manage to comply with these tests.

G. Marking

In other IP contexts, marking provides various benefits. In the US, marking of patented products provides for constructive notice to potential infringers and can expand the damages timeframe to the maximum permitted time period in the event of infringement. On the other hand, the marking of trade secret materials is just one factor in determining whether those materials even qualify for protection.

While these considerations may be relevant for a *sui generis* AI right, there is another reason that marking may be important in an AI context: public awareness. As recognized by both the European Union (in the *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*) and the U.S. (in the OMB's *Guidance for Regulation of Artificial Intelligence Applications*), public trust in AI applications will be vital to ongoing development and implementation. As such, the EU and the US have both recognized that the public has a right (or at least, recognized the benefit) to know when they are interacting with an AI system. Not only would this help to reduce confusion that may be inherent in such interactions, but it also allows the public to make informed choices regarding what information to disclose to such systems.

As with any software, the marking of an AI might be difficult to achieve, but a variety of options do exist which can help facilitate this goal. First, the mark may be embedded within the software itself providing notice to anyone attempting to modify or otherwise use the AI for their own purposes. However, such marking would provide little in the way of notice to the average consumer.

With the consumer in mind, and given the digital nature of most AI, the most reasonable marking that would provide adequate notice might be a screen or pop-up that informs the user of the nature of the AI they are about to use. While not foolproof, such a screen would at least require the user to acknowledge that they are about to use or interact with an AI before they are allowed to continue, which would help to further the goals set forth above.

H. Duration

Thus far this paper has discussed, in brief, the scope of a *sui generis* right, registering such a right, important exceptions to the right, and what infringement might looks like. But there is one more important question that must be considered: duration. Considering the blistering pace at which AI inventions are both developing and spreading through the world,[71] the duration of a *sui generis* AI has to carefully balance a number of factors in order to remain relevant and useful. Too long a duration could stifle innovation, preventing new actors from building upon the strides and innovations of those who come before, or given the pace of technology, the AI may grow stale/irrelevant. At the same time, too short a duration could prevent innovators from obtaining a return on their investment.

Continued use of an innovation to maintain and extend protection is one possible avenue of determining duration. For example, in trademark law in the U.S., a declaration of use (or excusable nonuse) is required every ten years in order to maintain trademark protection.[72] This provides owners an incentive to continue using the right (*i.e.* so that they can continue to maintain protection), while also allowing competitors or other innovators to create their own marks within the same field.

Such a continued use duration, though, has downsides when applied to an AI. So many aspects of an AI invention, whether it in the input data, algorithms that comprise it, or output produced, can change throughout its "lifetime" that defining "use" to any single embodiment would be almost impossible, as that use could change often. In the trademark context, the mark (as well as the company to which it applies) will generally remain the same throughout the course of its use. A solution could be that each subsequent embodiment of an AI, developed over time, would be entitled to its own protection. However, such a scheme would require further administrative costs as each use would need registration (as outline above) as well as a determination of *when* the AI has changed enough to warrant further protection.

Another alternative could be a set term of years, as is usually used in copyright and patent regimes. Given the speed at which this field develops, a shorter period may be desired, on the order of perhaps 5 to 10 years. This duration would depend upon a careful study of how these technologies develop as well as the costs related to that development *and* potential for recouping that cost in the market.

Another alternative, used by the *sui generis* database right, is the option to renew the right for an additional term of years upon showing of "substantial investment." This could be attractive to AI developers, as in this way they can continue developing their AI, changing the base algorithms,

---

[71] See WIPO, Technology Trends 2019 – Artificial Intelligence (2019).

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

inputs, and outputs without fear of losing their protection. Of course, there is still the administrative cost of registering and tracking the AI changes, but in this case, they do not risk losing protection that they already have, but rather risk not gaining protection, if they fail to properly report and track their invention.

I. Litigation

Whenever a right is granted legal protection, a key feature is the ability to protect and enforce that right via litigation. In the IP sphere, ownership of patents, copyrights, and trademarks all carry the right to sue for infringement, and trade secret ownership carries the right to sue for misappropriation. A *sui generis* right in trained AI models would naturally be expected to carry with it a corresponding right of enforcement through litigation.

Several aspects of trained AI models pose particular challenges in the litigation context.

First, pleading rules typically require that enough specific facts be alleged to demonstrate that the plaintiff's claim is plausible. *E.g.*, *Ashcroft v. Iqbal*, 556 U.S. 662 (2009). This often requires some level of specificity in allegations that include a comparison between the right held and the allegedly infringing instrumentality (or, in the case of trade secret misappropriation, improper acquisition from the rights-holder and improper disclosure or use). Given that many trained AI models are "black boxes," without any human-readable or human-understandable output, there could be challenges in both articulating the contours of the rights-protected AI model, as well as the contours of the allegedly infringing AI model. The potential need to allege facts showing a black-box-to-black-box comparison for infringement along the lines of what is required for patent or trademark infringement may be especially challenging. By contrast, a pleading model that is more analogous to the elements of a claim for trade secret misappropriation may be more manageable, where the facts to be alleged focus on the defendant's access to, and improper taking and disclosure or use of, the protected AI model. This would be consistent with a *sui generis* model that protects against actual copying (but does not exclude those who have independently developed identical or similar models).

Second, proofs at trial require more detailed evidence to prove the facts alleged in the pleadings. To that extent, the same challenges and questions arise. Additionally, litigation often lasts several years from initial complaint to trial. And, assuming a statute of limitations of two to four years, a complaint may not be filed until several years after the wrongful conduct. Many trained AI models are continually evolving – as the underlying algorithms are tweaked and as more training data is added. Consideration must be given to what the proof of claim will look like (and the concomitant scope of rights protected), given the changes that come with the passage of time. Thus, when compared with the AI model as it existed at the moment of copying/misappropriation, by the date of trial, both the plaintiff's AI model and the defendant's AI model may have evolved considerably – perhaps in ways that magnify differences. There are several possible approaches to address this. For example, all that could be required is to show the act of copying, misappropriation or taking, and the match between the protected model and the defendant's model, at the time of misappropriation. Or there could be a "fruit of the poisonous tree" type analysis that allows proofs of the defendant's model as-evolved, with liability tied to its traceability to the misappropriated

version of the model. Each of these options raises issues about the interplay between the nature and scope of the *sui generis* right and the remedies available in court.

Finally, determination of remedies – both legal and equitable – may pose challenges. First is the challenge of articulating the AI itself. Many AI models are "black boxes," without a human-readable or human-understandable output. This could make it difficult to fashion effective injunctive relief; stating in a court order clear and enforceable boundaries of what conduct is prohibited could pose practical challenges.

Second is the variability, or continually evolving status of many trained AI models. Most other forms of IP repose rights in a fixed item – a logo (trademarks), a novel (copyright), a new drug that cures cancer (patent), or a unique formula (trade secret). This would tend to reinforce the discussion, above, concerning protection against derivative works, and could also impact the scope and articulation of injunctive relief.

Third is the challenge of proving damages. Much of the discussion above has focused on an economic rationale for protection – protecting the plaintiff's investment interest in developing the trained AI model. This would tend to suggest a damages model based on the plaintiff's actual cost of development. Such an analysis can be a factor in, for example, calculating patent damages, but has not been a primary determinant of damages for other forms of IP. Based on a rationale that seeks to place the infringing defendant in the position it would have been in had it (properly) conducted its own independent development of the AI model, a cost-of-development model of damages could be seen as fair and appropriate. However, it could also provide significant windfalls to the plaintiff if the full cost of development is awarded in damages: a successful plaintiff would continue to own the IP rights in the trained AI model and would have been fully reimbursed for its costs of development. A more traditional disgorgement model, whereby damages are based on disgorgement of the defendant's profits from its wrongful conduct, could provide a similar set of incentives for defendants not to infringe. However, if the concern is primarily that copyists are creating cheap competing models in the marketplace, disgorgement of those discounted profits may not be sufficient to fully compensate the plaintiff. In this regard, the challenges of calculating damages may be similar to those posed for copyright, with a solution being to offer the same range of damages calculations that are available in copyright.

J. Extraterritoriality

In an interconnected world, and in particular in the easily transferrable world of software and AI, extraterritoriality can be vitally important to international actors, or even those who wish to prevent abuse of their rights abroad. Yet just because international protection may be desired, it does not mean that international protection should occur automatically.

From a national perspective, a *sui generis* AI right might implicate foreign actors in a number of ways. First, it might target those actors who *acquire* the information/data/protected subject matter within the U.S. (or whichever country has granted the right). For example, if an actor takes actions within the U.S. by which they acquire information protected by the *sui generis* AI right, that might provide for jurisdiction of U.S. courts.

Alternatively, acquisition of the information might not create jurisdiction but rather *use* within the U.S. In this example, information might be acquired abroad, but if it is protected in the U.S. then any use there would be the basis of jurisdiction.

Both of these situations are viable within the trade secret context. In *Micron Technology, Inc. v. United Microelectronics Corp.*,[73] the Northern District of California interpreted 18 U.S.C. § 1836(b),[74] as requiring only *one* element to have been performed in the U.S. in order for a claim to arise. In that case, the plaintiff alleged only that the trade secret was acquired within the U.S. but then disclosed and used in China (after the date of the DTSA).[75] Similarly, in *Luminati Newtorks Ltd. v. BIScience*, the Eastern District of Texas allowed a claim of trade secret misappropriation to proceed where only *use* within the U.S. was alleged.[76] But this leaves open another possibility: information that is acquired and used abroad. Enforcing the right in this context could become a very complicated issue in international law. Whether or not the two nations have existing agreements, whether an international treaty applies, and other considerations would likely bring enforcement or jurisdiction questions that are beyond the scope of this paper.

V. CONCLUSION

As AI continues to infiltrate daily life, it is important to recognize the efforts required to make AI a reality. While certain aspects of AI may be protectable by traditional IP protection rights, such as patents, copyrights, and trade secrets, other aspects of AI are not afforded such protection.

The creation of a *sui generis* right for trained AI models may be a useful vehicle in affording protection to trained AI models which require a significant effort to create.

VI. ABOUT THE INTELLECTUAL PROPERTY OWNERS ASSOCIATION

IPO is an international trade association representing companies and individuals in all industries and fields of technology who own, or are interested in, intellectual property rights. IPO's membership includes 175 companies and close to 12,000 individuals who are involved in the association either through their companies or as inventor, author, law firm, or attorney members. IPO advocates for effective and affordable IP ownership rights and provides a wide array of services to members, including supporting member interests relating to legislative and international issues; analyzing current intellectual property issues; information and educational

---

[73] 2019 WL 1959487 (N.D. Cal., May 2, 2019).

[74] Which defines the offense of misappropriation of trade secrets as 1) acquisition of the trade secret and 2) disclosure or use of the trade secret.

[75] 2019 WL 1959487 at *11.

[76] 2019 WL 2084426 (E.D. Tex., 2019).

services; and disseminating information to the general public on the importance of intellectual property rights.

A. About the Artificial Intelligence and Emerging Technologies Committee

The Artificial Intelligence and Emerging Technologies Committee focuses on identifying, monitoring, and studying intellectual property issues raised by artificial intelligence and other emerging technologies ("AI & ET"). AI & ET create new capabilities and allow new applications that may challenge the limits of existing IP law. The Committee reports on these issues to the Board of Directors and to the IPO membership at large with the objective of being a thought leader concerning the intersection of IP law and these rapidly evolving fields. The Committee also focuses on educating IPO members about the potential impacts of these issues on their IP strategies. In appropriate circumstances, the Committee develops and recommends positions to the Board of Directors and reflects the IPO viewpoint in responses to requests for consultation from U.S. and foreign offices. Given the potentially significant impact of AI & ET on all types of IP law, the Committee coordinates its efforts with other IPO committees as warranted in developing its recommended positions.

B. Contributors to this Paper

Yeen Tham was the lead coordinator of this paper. Other contributors include:

- Andrea Evensen
- Karen Fowler
- Dan Hoovler
- Hugh Pasika
- Guillaume Jaulerry
- Chris Mammen
- Nick Palmieri
- Ryan Phelan
- Caroline Pinkston
- Steven Schwarz
- Steve Shipe
- Nicole Spence
- Kyle Trout
- Shaun Zhang