Adapt Your IP Strategy for Artificial Intelligence

By Kevin M. Pasquinelli, Esq.[1]

Almost every business is exploring how artificial intelligence and machine learning ("AI/ML") systems can reduce operational costs, increase efficiency, grow revenue, and improve customers' experiences. However, the implementation and application of AI/ML technology is drastically different than traditional software and systems. It inputs and applies magnitudes more data. It applies algorithms in unforeseen ways, and outputs expressive content, even without human involvement. This creates new, sometimes unknown, legal risks that must be addressed in any intellectual property legal strategy. This paper discusses all major elements of an IP strategy for AI/ML.

ADAPT YOUR IP STRATEGY FOR AI

- Mitigate risk by judicious use of data sources and shifting risk
- Protect AI/ML algorithms and applications through patents, copyrights, and trade secrets
- Hold learnings and operational models as trade secrets
- Protect output by careful licensing terms that cover ownership of the results
- Manage risk through information containment

According to Forbes,[2] AI/ML systems have figured prominently into both the daily operation of enterprises and the daily lives of consumers. According to a study from McKinsey 47 percent of business executives state that as of the end of 2019, they are using at least one AI/ML capability in their business processes. Twenty-one percent say their organizations have placed AI/ML in several parts of their business, and 30 percent represent that they are only piloting. Nonetheless, that means any company not using AI/ML at this time is well behind the power curve.[3] With this rapid and increasing adoption, it is essential that every business, whether a software or SaaS (software as a service) supplier, services vendor, or in-house IT department, make AI/ML a core competency. Realizing that AI/ML is core to business

[1] Kevin M. Pasquinelli, is an intellectual property attorney at Robins Kaplan, LLP with more than 20 years of experience working in the technology industry, before becoming an attorney. His professional background incorporates a wide range of technologies and products, including digital hardware, video servers, medical devices, system software, software development practices, telecommunications, enterprise software, customer relationship management software, and network management software. He has represented clients ranging from start-ups to the Fortune 500 companies in numerous patent, trade secret, copyright, licensing, and trademark disputes. Before becoming an attorney, Kevin served in a variety of technology and management roles at a multinational information technology company and took on similar roles in a number of Silicon Valley startups, taking one public. Kevin can be reached at kpasquinelli@RobinsKaplan.com, 650.784.4025.

[2] https://www.forbes.com/sites/gilpress/2018/12/15/ai-in-2019-according-to-recent-surveys-and-analysts-predictions/#43b085ca14c3.

[3] https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain.

success means that the value in AI/ML systems must be captured, maintained, and protected, lest your competition gain an advantage.

Despite this rapid adoption rate, many organizations have not adequately adopted their IP strategy for AI/ML. AI/ML has moved well ahead of the law, and voices are becoming increasingly elevated for additional clarity from Congress and the courts on what can and should be protected. This article assesses the intellectual and data property rights that can be applied to capture, maintain, and protect your ownership and intellectual property in an AI/ML system. As context, this paper first discusses: (1) why big data used with AI/ML systems creates tremendous opportunity, (2) some examples of AI/M systems, and (3) the basic architecture of an AI/ML system. Second, the paper then addresses the IP risks AI/ML systems face when importing and exporting information. Third, it discusses how to use patents, trade secrets, and careful licensing to protect the value added in the AI/ML system. The paper summarizes these matters by proposing "containment" models for managing risk and concludes with a list of recommendations for adapting an IP strategy for AI/ML.

I. Say Hello to Big Data

AI/ML has existed for decades. There was a "buzz" about AI/ML in the 1980s, which fluttered and fell to Earth when the promise didn't meet the hype. Efforts at that time focused on the computing algorithms used to solve a problem and simplifying the decision logic by using rule-based systems. While that helped make fundamental logic decisions (if/then/else) available to IT departments, the systems were unable to learn, whether human aided or not. This focus on making programming easier continued through the 1990s.

Then, in 2001, a simple yet seminal paper was published by Michele Banko and Eric Brill of Microsoft Corp.[4] Banko and Brill discovered that very different machine learning algorithms, even simple ones known for decades, performed equally as well as more recent sophisticated algorithms when using orders of magnitude more pieces of data (*e.g.,* billions of labeled data versus millions of labeled data). Banko and Brill were researching natural language disambiguation (*i.e.,* how to select confusable words such as {principle, principal}, {then, than}, {to, two, too}, and {weather, whether} from one another).[5] As can be

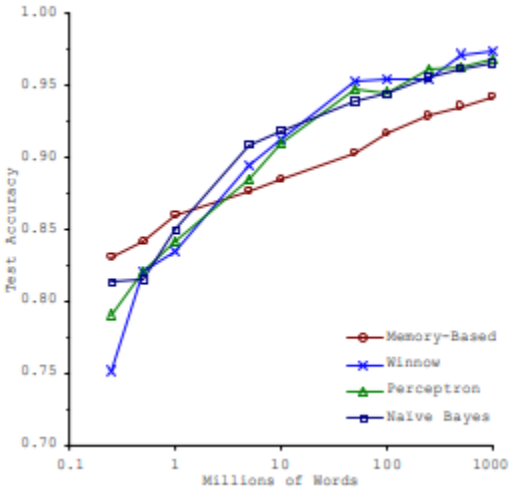

Figure 1: Learning Curves for Confusion Set Disambiguation

[4] https://www.aclweb.org/anthology/P01-1005. Scaling to Very Very Large Corpora for Natural Language Disambiguation.

[5] Probably to add such functionality to the very software application this paper is typed into, Microsoft Word®.

seen in the graph, the test accuracy of successfully disambiguating words rose from about 80 percent accuracy, when applying a few hundred thousand words, to about 97 percent accuracy, when applying a billion words. The graph also clearly demonstrates that the algorithms performed approximately equal in effectiveness. The study humbly concluded, "these results suggest that we may want to reconsider the tradeoff between spending time and money on algorithm development versus spending it on corpus development."[6] Thus, the era of big data was born.[7] Despite its birth in 2001, the proposition that data matters more than algorithms for solving complex problems did not begin to take hold until 2009,[8] when Halevy, Narvig, and Pereira of Google, Inc. published *The Unreasonable Effectiveness of Data*.[9] The Google scientists had access to an even larger corpus, up to a trillion words.[10] Given this even larger body of data, they concluded that "semantic relationships can be automatically learned from the statistics of search queries and the corresponding results or from the accumulated evidence of Web-based text patterns and formatted tables" without needing any manually annotated data.[11] Realizing that the entire internet could now be the corpus for machine learning rocketed the application of machine learning, not to mention the stock of Google. Thus, it became readily apparent that big data was not just for scientists anymore — everyone could apply these concepts given available algorithms, the right data, and the right *amount* of data.

II. Overview of AI/ML Applications and System Architecture

AI is human intelligence exhibited by computing machines. ML is an approach of AI where computing machines learn from experience to find patterns in a set of data, without the reprogramming of its software. The applications of AI/ML are bounded only by insufficiency of data. Some common and well-known applications of AI/ML include:

- Object recognition (*e.g.,* for an intelligent oven to determine what food was inserted in the oven cavity, or automated cars recognizing humans, other cars, and walls)[12]
- Speech recognition and detection (such as Amazon's Alexa)[13]
- Prediction (*e.g.,* answer a question such as, "Is this picture a work of Haring?")

Figure 2: Style Transfer

[6] Scaling to Very Very Large Corpora for Natural Language at 3.
[7] The study also found that test accuracy can be substantially increased using active learning on smaller sets of data.
[8] This was for a variety of reasons including lack of computing power, lack of storage, and limited network speeds.
[9] https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35179.pdf.
[10] Id.
[11] Id. at 1.
[12] https://www.digitaltrends.com/home/june-oven-artificial-intelligence/.
[13] https://developer.amazon.com/alexa-skills-kit/conversational-ai.

- Language translation (*e.g.,* Google Neural Machine Translation)[14]
- Picture restoration (*e.g.,* filling in lost portions of an old picture)
- Spam identification
- Spell-checking[15]
- Customer relationship management ("CRM") applications
- Style transfer of artwork: Two images are "merged" using the first as sample content (*e.g.,* your picture or landscape photograph) and the style from the second (*e.g.,* "style it like Haring").[16] *See* Fig. *2.*

ML has become so accessible it is even a matter of a whimsical smartphone application in HBO's Silicon Valley, which answers the question, "Is it a hot dog?"[17] *See* Fig. *3.*

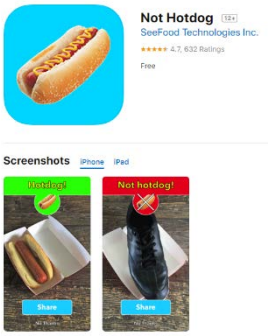

Figure 3: Is it a hot dog?

The fundamental elements of an AI/ML system are shown in Fig. 4 below, which include at least the following: (1) the data inputs, which must include the suspected signals and features that contain the necessary information to solve the proffered question; (2) the leaning algorithm; (3) the deployed operating model; and (4) the desired output. Special note should be taken of the dashed lines, because when information crosses these lines, intellectual property rights may be affected.

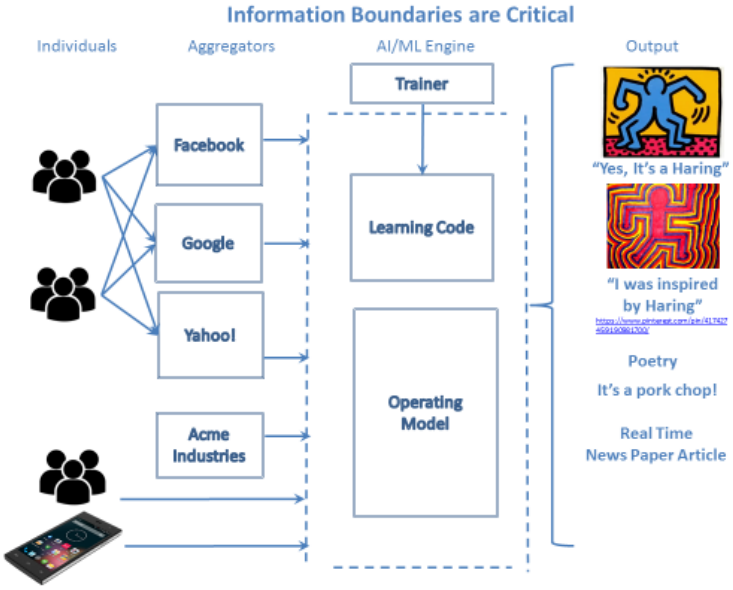

Figure 4: AI/ML Architecture

AI/ML is an iterative process upon which a chosen algorithm (also known as the learning function) improves by learning. *See* Fig. 5, below. An ML algorithm has the key

[14] https://en.wikipedia.org/wiki/Google_Neural_Machine_Translation
[15] https://cloud.google.com/blog/products/g-suite/everyday-ai-beyond-spell-check-how-google-docs-is-smart-enough-to-correct-grammar.
[16] *See* Deep Art Effects at https://www.deeparteffects.com/.
[17] This application, titled *Not Hotdog*, is available on the Apple App Store and offered by SeeFood Technologies, Inc. https://itunes.apple.com/app/not-hotdog/id1212457521.

goal of making a correct prediction, for example answering the question “Is it a Haring?” or “Is it spam?”

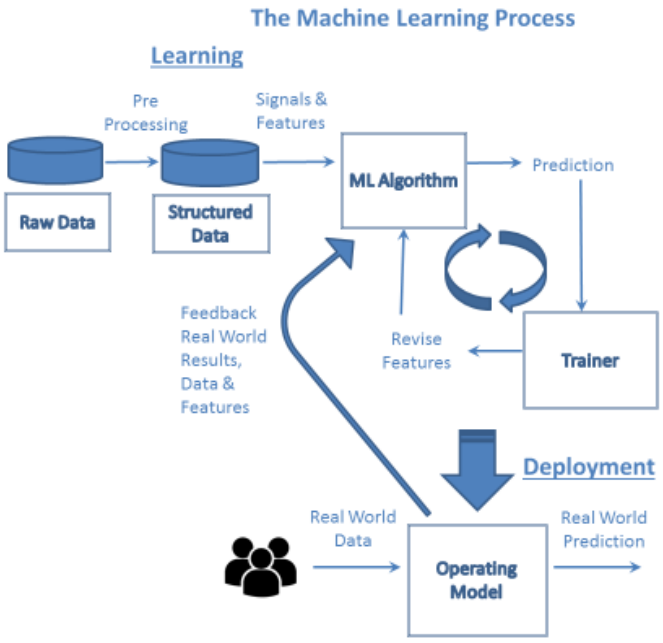

Figure 5: AI/ML Learning Process

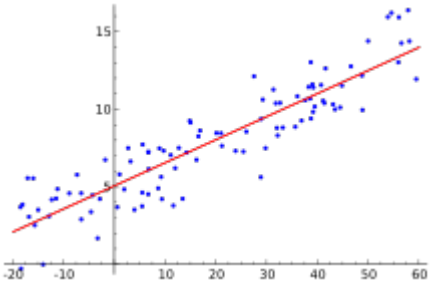

Figure 6: Linear Regression

There are many different, well-known AI/ML algorithms (*see right*), which can be applied to a given problem. Simply put, an AI/ML algorithm applies a target function (*f*), given an input set of data (x), to compute an output (y), *i.e.,* y = *f* (x). Predictive modeling is where the algorithm learns to better predict (y) over time given a greater amount of data (x). A few of the most well-known algorithms are (1) linear regression (fitting data to line),[18] (2) logistic regression (deciding if data are on one side or another of an “S” curve),[19] and (3) classification and regression trees (applying a series of binary decisions). Graphical examples of each are shown in Figs. 6–8. The number of input signals can be very high.

[18] https://en.wikipedia.org/wiki/Linear_regression.
[19] https://en.wikipedia.org/wiki/Logistic_regression.

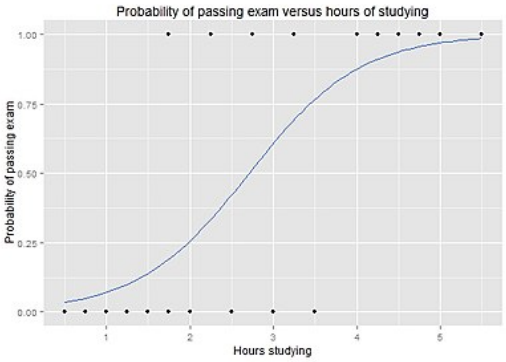

Figure 7: Logistic Regression

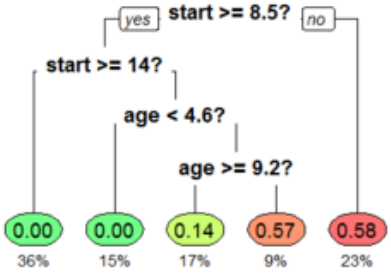

Figure 8: Regression Trees

As noted in the introductory section, AI/ML algorithms learn best given a very, very large amount of data. The data are applied to the AI/ML algorithm, and it learns to answer the desired question. As the algorithm is exposed to more and more data, the algorithm improves its predictive performance. The algorithms can learn through (1) supervised training (where the data are labeled with input and the desired output), (2) unsupervised training (where the AI/ML algorithm receives unlabeled data and is tasked with the goal of discovering structure in the input data), and (3) reinforcement learning (where the AI/ML algorithm trains itself through trial and error). The AI/ML algorithm learns with the aid of a trainer (the trainer may be a real person, data tabs, or another software program). Different AI/ML algorithms operate best on different types of problems. The data sets and learnings are iteratively applied to the AI/ML algorithm. Once the system is able to predict the outcome with a desired accuracy, the model is deployed from the learning environment to the production environment. Then, real-world data are applied. So, to determine whether a work is a Haring, the operational model would accept input of a picture and analyze the picture on a chosen set of signals. Assuming use of a logistic regression algorithm, if the resulting data point was above the S curve it would answer “Yes,” if below the S curve it would answer “No.”

Even after deployment of the ML system, the production system continues to record the real-world findings, which are fed back into the learning environment for a subsequent deployment of a more accurate operational model.

III. Legal Protections are Lacking for Information Inputs

Fair use in copyrighted works has been judicially affirmed for transformative uses such as image search engines and Google Books. No decision has been made concerning the unauthorized use of copyrighted input to AI/ML applications.

Fair use of copyrighted works is undecided

AI/ML systems, as discussed above, input large data sets that copy the data, pre-process it, clean it, tag it, and format it for training. The number of copies, whether literal or derived, can be enormous. If copyrights exist on the incoming data, there is significant risk of infringement, unless the fair-use doctrine applies. Fair use is codified in 17 U.S.C. § 107 and is evaluated according to a four-factor standard: (1) the purpose and character of use, (2) the nature of the copyrighted work, (3) the amount and

substantiality of the portion used, and (4) the effect of the use upon the potential market for or value of the copyrighted work. In the context of software and internet applications, the most significant consideration is whether or not a use is non-expressive and sufficiently "transformative."

The seminal case on non-expressive fair use is *Kelly v. Arriba*,[20] followed a few years later in *Perfect 10 v. Amazon*.[21] Defendants in both cases operated search engines that crawled the internet for images, reduced the image resolution, and displayed the resulting "thumbnails" in search results, despite copyrights existing in plaintiffs' images. Both courts held that the activities constituted transformative fair use. In *Kelly*, the court explained that "thumbnails do not stifle artistic creativity because they are not used for illustrative or artistic purposes and therefore do not supplant the need for the originals."[22] Pushing the envelope of the transformative-use doctrine further was *Authors Guild v. Google, Inc.*[23] Similar to *Kelly and Perfect 10*, the court held that the unauthorized reproduction of copyrighted texts was a transformative fair use because Google Books provided information "about" books. It did not provide the books expression. The court held that Google's views, displaying text surrounding the search terms, were sufficiently transformative, because the text beyond the search term was contextual. As such, it did not threaten the author's copyright interest.[24]

Whether fair use protects use of copyrighted input to an AI/ML system has yet to be addressed in any U.S. court. The question is important, because substantial liability could exist for AI/ML systems that import works verbatim, create intermediate derivative works, and then export expressive, or non-expressive, works. For example, prediction applications, such as "Is this picture a work of Haring?" need to learn from an existing corpus of Haring artwork. If the unlicensed use is not a fair use then a massive copyright infringement will occur when the AI/ML system searches the Internet for works of Haring. Without such a corpus, the AI/ML system will be inaccurate. Alternatively, the burden of licensing all of Haring's works as input to the AI/ML system may be exorbitant, prohibiting the application from being commercially viable. A second example is language translation where an enormous corpus of information is essential to success. It is unlikely that an independent entity could create or license a sufficiently large corpus to cover all the nuances of language translation. Given this state of the law, careful review of inbound data licensing contracts is essential.

"Hidden" copyrights exist in schema and compilations

Another source of input risk is "hidden" copyrights located in vendors' database schemas and metadata compilations. It is common for enterprises to license CRM and enterprise resource management ("ERM") software. While using these systems, enterprises store their customer and manufacturing information in the databases of

[20] *Kelly v. Ariba Soft Corp*., 336 F.3d 811 (9th Cir. 1992).
[21] *Perfect 10, Inc. v. Amazon.com, Inc*., 508 F.3d 1146 (9th Cir. 2007).
[22] *Kelly*, 336 F.3d at 819-320.
[23] *Authors Guild v. Google Inc.,* 804 F.3d 202 (2d Cir. 2015).
[24] Id. at 218.

these vendors. However, while the enterprise may own the data, it does not necessarily have rights in the schema or compilation of the information. As such, extracting information from these systems and making verbatim copies of the schema or compilation may constitute copyright infringement.

Madison River Management Co. v. Bus. Mgmt. Software Corp. is highly instructive.[25] In *Madison*, the Madison River Management Co. ("MRM"), a rural telephone company, purchased telecommunications service software from Business Management Software Corporation ("BMS"), called the Ticket Control System ("TCS"). The TCS "automatically manage[d] the supply of service to and problems with . . . telephone networks."[26] During normal use the BMS software stored MRM's data in its TCS database. The parties disputed certain contractual and licensing matters, and the case went to court. BMS accused MRM of copyright infringement for extracting information from the TCS database. The court held:

> "The TCS program utilizes Madison's raw data that is saved in an Oracle relational database by imposing on the raw data a new structure and metadata enhancements. It is the raw data subjected to the TCS structure, processes, triggers, program modules, and stored procedures which then becomes the TCS database. Thus, the TCS database is covered by . . . [BMS's] copyright over its TCS software."[27]

In other words, the database schema and any structured compilation of metadata is copyright protectable.

The court went on to conclude that the copying of the TCS database was not necessarily a fair use, due to issues of fact.[28] The court outlined ways to avoid such an infringement in the future, for example: (1) use licensed tools to extract the raw data and place it in a separate electronic file, or (2) use the underlying database to extract the raw data (here an Oracle database) and place it in a separate electronic file. The key to avoiding infringement is to not use the schema designed by TCS and to place the raw data in a separate location, formatted differently, for future access and use.[29]

Consider again, the prediction application, "Is this picture a work of Haring?" Even if the input of Haring's works were free of encumbrances, the format in which they are stored in the AI/ML or other system may be protected by a third party vendor. Therefore, ingesting data directly from vendors' databases or copying the data verbatim risks copyright infringement and tainting subsequent use in the AI/ML system. Given the

---

[25] *Madison River Mgmt. Co. v. Bus. Mgmt. Software Corp*., 387 F. Supp. 2d 521 (M.D.N.C. 2005).
[26] Id. at 528 ("The TCS suite includes 'TCS Control,' the core event manager, and five additional products: (1) 'TCS Provide' for service provisioning; (2) 'TCS Resolve' for trouble management; (3) 'TCS Force' for workforce management; (4) 'TCS Satisfy' for customer care; and (5) 'TCS Defend' for fraud detection.").
[27] Id. at 535.
[28] Id.
[29] Id.

lack of fair use and lurking copyrights within the enterprise itself, it is essential that both data and software license agreements are carefully reviewed.

IV. Legal Protections are Lacking for AI/ML Information Output

Consider the style transfer of artwork application where two images are "merged" using the first as context (perhaps your own picture or art work) and the style from the second (perhaps Haring) producing a merged output. Assuming that the risk of copyright infringement on the input is solved through fair use or licensing, what is the scope of IP protection in the output?

Copyright protection in AI/ML generated output are inconsistent

Figure 9: White Pages Not Subject to Protection

Under 17 U.S.C. § 102(a) copyright protection exists "in original works of authorship fixed in any tangible medium of expression . . . either directly or with the aid of a machine or device." In *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, a seminal case penned by Justice O'Connor, the Supreme Court of the United States further clarified that "to qualify for copyright protection, a work must be original to the author . . . independently created by the author" and possess at least some "minimal degree of creativity." [30] Justice O'Connor held that a white pages listing did not qualify for protection because, although compilations can be protected, there was not even the minimal degree of creativity to qualify.[31] Copyrights vest in the author of the work.[32]

After years of debate, both the Copyright Office and federal courts have afforded copyright protection to software. Therefore, entities developing AI/ML algorithms can copyright the source code.[33] Copyrights do not extend to any "idea, procedure, process, system, method or operation, concept . . . explained, illustrated, or embodied in the work."[34] In other words, the source code and any graphical user interface is copyright protectable. The functionality of the code is not. This is well-settled law. However, copyright protections for AI/ML systems become blurry from there.

Copyright ownership in the output of AI/ML systems is inconsistent

The U.S. Copyright Office has clearly stated that it will register only works created by a human being,[35] because copyright law is limited to original intellectual

[30] 499 U.S. 340, 345 (1991).
[31] Id. at 345.
[32] Joint ownership in a copyright can inure to collaborators of a work. 17 U.S.C. § 101.
[33] *See* U.S. Copyright Office, Copyright Registration of Computer Programs. At https://www.copyright.gov/circs/circ61.pdf.
[34] 17 U.S.C. 102(b).
[35] *See* U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 306.

conceptions of the author.[36] The Office will refuse to register a claim if it determines that a human being did not create the work.[37] The Office makes explicit that "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 receive registration.[38] Since registration is required to have standing to bring a copyright lawsuit, the Office's decision stands as a firewall against bringing a suit based on common law rights.[39]

Although the U.S. Copyright Act makes no explicit statutory requirement that the work's author be human, the Ninth Circuit has held that since "the Copyright Act does not expressly authorize animals to file copyright infringement suits . . . Naruto [a monkey who took selfies of himself and was represented by PETA] lacks . . . standing to sue under the Copyright Act."[40] Given this holding and the Office's decision to limit registration to humans, it is unlikely that a court would find that a fully automated AI system that generates expressive output would be considered an author. However, many, if not most AI/ML systems are focused on machine learning, where there remains a human operator, trainer, and algorithm developer. It remains undecided and unclear under what conditions an ML-generated work can be granted a copyright. Specifically, whether an AI/ML system can be a sole author, part of a jointly authored work, or part of a work for hire remains an open question.

The Ninth Circuit has acknowledged that at least one authority "suggest[s] that the copyright protection afforded a computer program may extend to the program's output if the program 'does the lion's share of the work' in creating the output and the user's role is so 'marginal' that the output reflects the program's contents."[41] This case, if widely adopted, could be the "camel's nose under the tent" that leads to wide copyright protection of output from AI/ML programs. More likely, it's an outlier. It has not been widely embraced. In the cases citing to *Torah Soft* considering the "lion's share of the work" standard, only one court denied a motion for dispositive relief (motion for

---

[36] Id.

[37] *Citing to Burrow-Giles Lithographic Co. V Sarony*, 111 U.S. 53, 58 (1884).

[38] *See* U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 313.2.

[39] Although this is not to say that copyright cannot be asserted based on the output of a software program authored by a human. *See Design Data Corp*. below.

[40] *Naruto v. Slater*, 888 F.3d 418, 420 (9th Cir. 2018).

[41] *Design Data Corp. v. Unigate Enter., Inc*., 847 F.3d 1169, 1173 (9th Cir. 2017) (*quoting Torah Soft Ltd. v. Drosnin*, 136 F.Supp.2d 276, 283 (S.D.N.Y. 2001)). In *Torah Soft*, the software at issue created a matrix in response to an end user's input of a particular term. *Torah Soft*, 136 F.Supp.2d at 283. The Southern District of New York found that the defendant — the program's user — was not the author of the copyright. Id. In its analysis, the court emphasized the end-user's role in creating the matrix:

> In addition, an end-user's role in creating a matrix is marginal. Creating a matrix is unlike the creative process used in many computer art programs, which permit an end-user to create an original work of art in an electronic medium. It is fair to say that users of such programs often supply the lion's share of the creativity to create the screen display. By contrast, an end-user of the Software merely inputs a word or phrase which the Software searches for in the Database. Thus, the Software does the lion's share of the work. In short Drosnin [the defendant] is not the author of the matrixes.

summary judgment or motion to dismiss), thereby allowing the case to move toward trial (*see Mfg. Automation & Software Sys. v. Hughes*).[42] In the most recent case citing *Torah Soft*, *Rearden v. Walt Disney Co.,*[43] the court considered whether there could be copyright protection in the output of a system of cameras closely controlled by software that "precisely captures and tracks the 3D shape and motion of a human face to sub-millimeter precision."[44] The system captures the facial expressions of human actors and the resulting output is merged with graphical images, thereby creating realistic expressions on the face of animated characters, such as the Beast in "Beauty and the Beast." The court held that capturing and formatting the images did not constitute the "lion's share" of the output such that copyright protection would attach. However, whether the output of an AI/ML system which learned from prior videos and output a unique video of its own is undecided.

The Copyright Office considers that a copyright in source code also covers any graphical user interface ("GUI") produced from the code and its binary form. "As a general rule, a computer program and the screen displays generated by that program are considered the same work."[45] "The U.S. Copyright Office views source code and object code as two representations of the same work."[46]

Outside the United States there is a split among countries that have, and have not, embraced copyright ownership in the output of AI/ML technology. Common law

---

[42] There are four cases citing to *Torah Soft* discussing the "lion's share standard."
(1) *See Dig. Drilling Data Sys. LLC v. Petrolink Servs*., No. 4:15-CV-02172, 2018 U.S. Dist. LEXIS 83158, at *26 (S.D. Tex. May 16, 2018) ("the Court finds that though the DataLogger database schema is subject to Digidrill's copyright, the actual data entries contained in the database are not copyrightable.").
(2) *See Design Data Corp. v. Unigate Enter., Inc*., 847 F.3d 1169, 1173 (9th Cir. 2017) (In affirming a lower court's motion for summary judgment the Ninth Circuit held: "Assuming, without deciding, that copyright protection does so extend, we nonetheless conclude that Design Data did not raise a question of material fact that the imported SDS/2-generated images and files reflected the contents of its program. Design Data did not present evidence establishing that SDS/2 "does the lion's share of the work" in creating the steel detailing files or that the user's input is "marginal." *Torah Soft*, 136 F. Supp. 2d at 283. Thus, the district court correctly rejected Design Data's argument that the SDS/2 copyright protects the images and files that UE imported and distributed.").
(3) *Mfg. Automation & Software Sys. v. Hughes*, No. 2:16-cv-08962-CAS(KSx), 2018 U.S. Dist. LEXIS 106797, at *27 n.8 (C.D. Cal. June 25, 2018). (In denying a motion for summary judgement, the Court held that "there is a triable issue of material fact as to whether the source code or user input does the "lion's share" of the work in generating the output.).
(4) *See Rearden LLC v. Walt Disney Co*., 293 F. Supp. 3d 963, 971 (N.D. Cal. 2018). (Dismissing the claim without prejudice because "Rearden has not alleged ownership of the output [consistent with producing the lion's share of the output]).
[43] *Rearden LLC v. Walt Disney Co*., 293 F. Supp. 3d 963, 971 (N.D. Cal. 2018).
[44] *Id.*
[45] The Compendium of U.S. Copyright Office Practices: Chapter 700 VX.Y § 721.10(A) ("If the applicant states "computer program" in the Author Created/New Material Included fields or in spaces 2 and 6(b), the registration will cover the copyrightable expression in the program code and any copyrightable screen displays that may be generated by that code, even if the applicant did not mention the screen displays and even if the deposit copy(ies) do not contain any screen displays.").
[46] The Compendium of U.S. Copyright Office Practices: Chapter 700 VX.Y § 721.5 *citing GCA Corp. v. Chance,* 217 U.S.P.Q. 718, 719-20 (N.D. Cal. 1982) ("[b]ecause the object code is the encryption of the copyrighted source code, the two are to be treated as one work…").

countries such as the U.K. and New Zealand have updated their copyright acts to explicitly include copyright protection for the output of computer programs.[47] The author of the computer-generated work is the person "by whom the arrangements necessary for the creation of the work are undertaken."[48] For AI/ML where a person is using, directing, and training the system, authorship rights attach. However, in a fully automated system it's less clear that any copyright would vest in the owner of the system alone. Alternatively, civil law countries, such as France, Germany, Greece, and Switzerland, reject the notion of non-human authorship.[49]

Thus, consider again the style transfer application. In the U.S., copyright cannot reside with the AI/ML machine as an author, therefore the output is arguably not protected. This outcome is inconsistent with a human being inspired by the work another. For example, if a human is inspired to style their art work after Haring, the work is copyright protectable; if an AI/ML program modifies your work in the style of Haring, it is not protected. Copyright protection in common law countries outside the U.S. provides some protection, however using protected works on the inputs may still encumber any copyright in the output. Therefore, whether in the United States or worldwide, it will be insufficient to depend on current copyright law to protect the output of your AI/ML system. Strong licensing protections on the output of your system are highly recommended. See section VII below for recommendations on protecting data output. Common law countries should be closely monitored to see how aggressively AI/ML creations are afforded copyright protections in the courts.

Ownership of data output is also inconsistent

It is well-established that there can be no valid copyright in facts.[50] So, there is generally no reason to believe that the facts output by an AI/ML system are protected, *per se* (at least in the United States), unless contract establishes otherwise. However, since facts, *per se*, are not protectable, entities that have not spent the effort to organize and publish those facts have incentive to "free ride" off others' works. Many see this as an unjust enrichment. For example, in *National Basketball Association v. Motorola*, the National Basketball Association ("NBA") asserted that Motorola and STATS (a real-time statistics program) infringed the NBA's copyright by misappropriating the data presented on its real-time data feed, named SportsTrax pager.[51] The United States Court of Appeals for the Second Circuit held that a "hot-news" claim exists in "cases where: (i) a plaintiff generates or gathers information at a cost; (ii) the information is

[47] *See* Copyright Designs and Patents Act 1988 (CDPA). UK copyright law acknowledges the possibility that works can be "computer-generated" ("generated by computer in circumstances such that there is no human author of the work").
[48] Section 9(3) CDPA.
[49] These countries heavily emphasize moral rights due the personality of the author. For example, under Art. L121-1 the author is the sole judge as to when the work may be first available to the public (Art L 121-2), can prevent modification of the work, and can prevent further reproduction and distribution (Art. L 121-4).
[50] *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 344, 111 S. Ct. 1282, 1287, 113 L. Ed. 2d 358 (1991) ("That there can be no valid copyright in facts is universally understood.").
[51] 105 F.3d 841 (2d Cir. 1997).

time-sensitive; (iii) a defendant's use of the information constitutes free riding on the plaintiff's efforts; (iv) the defendant is in direct competition with a product or service offered by the plaintiffs; and (v) the ability of other parties to free-ride on the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened."[52] However, after the passage of the 1976 Copyright Act, the hot-news doctrine suffered some reversals. The Second Circuit held that the "hot-news" tort was largely preempted by the 1976 Act in the *NBA* case.[53]

So, despite the lack of protection for data or facts, in general, there are isolated examples where the enrichment seems so manifestly unjust as to cause courts to protect the investment in publishing the facts.[54] Due to the rise and current availability of real-time AI/ML journalism applications, the "hot-news" doctrine is likely to be asserted again.

AI/ML protection under the EU Database Directive is hotly debated

In contrast to the U.S. approach to protecting data, which weakly relies on unfair competition and essentially limits the protections to trade secrets and contract, the European Union has passed Directive 96/9/EC ("the EU Database Directive" or "the Directive"). It states in part:

> 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 quantitatively, of the contents of the database.
>
> * * * * * * * * *
>
> The repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interest of the maker of the database shall not be permitted.
>
> * * * * * * * * *

[52] *National Basketball Ass'n*, 105 F.3d at 845.
[53] *Barclays Capital Inc. v. Theflyonthewall.com, Inc*., 650 F.3d 876 (2d Cir. 2011), in reversing the decision of the US District Court, the Second Circuit held that the claims by three major financial investment firms (Barclays Investment Bank, Morgan Stanley, and Merrill Lynch) against an internet subscription stock news service (theflyonthewall.com) for "Hot-news" Misappropriation under state common law doctrine could not stand because they were pre-empted by several sections of the Federal Copyright Act (17 U.S.C. § 106, 17 U.S.C. § 102, and 17 U.S.C. § 103).
[54] *See International News Service v. Associated Press*, 248 U.S. 215 (1918).

It defines a database as "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."[55] Any computer program used to create the database is not protected.[56] These rights "apply irrespective of the eligibility of that database for protection by copyright or by other rights."[57]

Initial application of the Directive seemed skewed positively in the database owner's favor. Similar to the real-time data feeds in *National Basketball Association v. Motorola*, William Hill was found to have infringed on the database rights of BHB (the governing authority of the British Racing Industry) when it forwarded real-time information accessible from the BHB racing database to its customers.[58]

The EU Database Directive is a *sui generis* right, separate and apart from copyright, and has no authorship requirement. Its owner's key threshold is to prove that there has been "substantial investment . . . in [creating] the contents of the database" and that the data is "arranged in a systematic or methodical way." However, in 2004, the Court of Justice of the European Union ("CJEU") clarified the scope of the Directive finding that it does not apply to databases that are merely the by-products of the main activity of an organization.[59] In April 2018, the European Commission ("the Commission") published a study evaluating the Database Directive, finding that "in the current context, it seems that the Database Directive does not apply to the databases generated with means of machines, sensors and other new technologies (such as the Internet of Things or artificial intelligence)."[60] The study acknowledged that "generation of these databases is closely interlinked with the creation of their content (*i.e.* data)." It continued: "However, case law indisputably excludes investment in data creation from the scope of the *sui generis* right."[61] It then concluded: "The Internet of Things, Artificial Intelligence, algorithm- and sensor-generated data, Big Data are all gaining increasing economic importance. It is nevertheless unclear how they are regulated, e.g. whether the current definition of a database embraces them, and, even more importantly, whether they should benefit from protection under the sui generis right."[62] The study

---

[55] Article 1(2) of Directive 96/9/EC.
[56] Para. 23 of the Preamble to Directive 96/9/EC.
[57] Directive 96/9/EC.
[58] *British Horseracing Board Ltd v. William Hill* (C-203/02, 9/11/2004) ("It follows that William Hill's action of taking information from the [BHB] . . .and loading it onto its own computers for the purpose of making it available on its Web site is an unlicensed act of extracting a substantial part of the BHB Database and the subsequent transmission of that data onto its Web site for access by members of the public is a reutilization. The defendant infringes BHB's rights in both ways." available at https://www.law.uh.edu/faculty/gvetter/classes/IntlIPfall2013/BritishHorseRacingBoard-v-WilliamHill_CA201320_July2005.pdf.
[59] *Fixtures Marketing Ltd v. Oy Veikkaus Ab* (C-46/02, 9/11/2004), *Fixtures Marketing Ltd v. Svenska Spel Ab* (C-338/02, 9/11/2004*) Fixtures Marketing Ltd v. OPAP* (C-444/02, 9/11/2004).

[60] Final Report of the Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases. Executive summary, Pg. ii.
[61] Id.
[62] Id. at Pg. v.

summarized the views on whether the Directive should apply to these emerging technologies and whether the Directive should be repealed. It acknowledged that views are "polarized" between database creators, who desire protection, and users, who desire unfettered access rights.[63]

Therefore, U.S. protection for publicly available data remains weak, while the EU Database Directive, which initially appeared protective of databases, is headed for continued uncertainty. A "showdown" between data producers and consumers in the EU is inevitable. Many parts of the journalism profession are being enhanced using AI/ML, including data correlation tools, context checking tools, and data visualization.[64] In some instances, even complete automation is possible. For example, using Wordsmith, a natural language AI/ML system from Automative Insight, Yahoo! Sports and the Associated Press are automatically generating personalized news stores from sports data.[65] Therefore, data output needs to be carefully protected by contract restrictions and trade secrets, wherever possible.

V. Use Patents to Protect Novel AI/ML Algorithms

Subject-matter eligibility for software patents has been under considerable debate in light of the Supreme Court's *Alice/Mayo* test.[66] Consistent application of the test, whether at the USPTO or in the Federal Courts, has been elusive. Given the inconsistencies, inventors and businesses have been uncertain which of their inventions are eligible for patent protection. In response, on December 20, 2018, the USPTO released revised subject-matter eligibility guidance[67] and soon thereafter held a conference on AI policy considerations.[68] At the conference, Director Iancu expressed his commitment that the USPTO would take a leadership role in this emerging technology.[69] The European Patent Office ("EPO") has also updated its examination guidelines to account for AI/ML-based inventions[70] following its May 2018 conference on patenting AI.[71]

---

[63] Id.
[64] *See* *https://emerj.com/ai-sector-overviews/automated-journalism-applications/.* (last accessed May 5, 2019).
[65] *See* https://emerj.com/ai-case-studies/news-organization-leverages-ai-generate-automated-narratives-big-data/ (last accessed May 5, 2019).
[66] *Alice Corp. Pty. Ltd. V. CLS Bank Int'l,* 573 U.S 208, 217-18 (2014) (citing *Mayo Collaborative Servs. V. Prometheus Labs., Inc.,* 566 U.S. 66 (2012)).
[67] "2019 Revised Patent Subject Matter Eligibility Guidance" available at https://www.govinfo.gov/content/pkg/FR-2019-01-07/pdf/2018-28282.pdf.
[68] "Artificial Intelligence: Intellectual Property Policy Considerations" USPTO (January 31, 2019). Recordings of the conference are available here: https://www.uspto.gov/about-us/events/artificial-intelligence-intellectual-property-policy-considerations.
[69] https://www.uspto.gov/about-us/news-updates/remarks-director-iancu-artificial-intelligence-intellectual-property.
[70] https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/j.htm.
[71] https://www.epo.org/learning-events/events/conferences/2018/ai2018.html.

In light of these public commitments endorsing the patentability of AI/ML systems, AI/ML patent grants have risen. USPTO allowance rates have increased from 15 percent before the guidance, to 38 percent after the guidance. The change correlates highly with training of examiners in January 2019.[72]

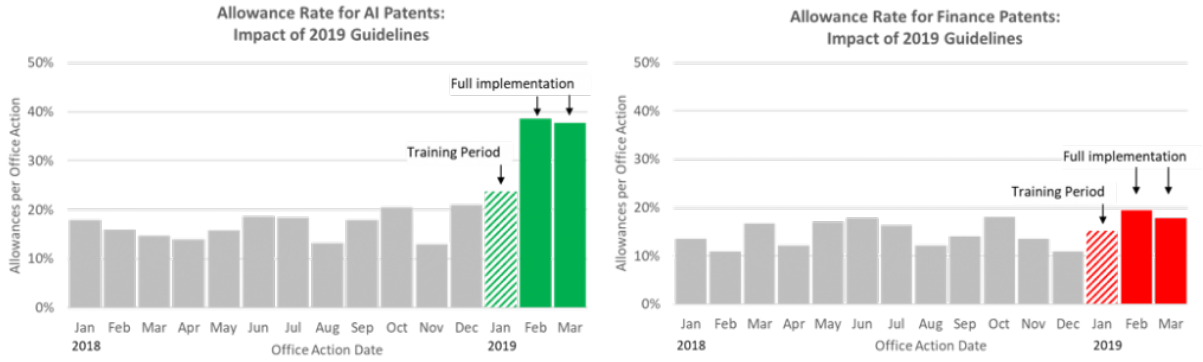

Figure 10: U.S. Allowance Rates Rising

The World Intellectual Property Organization ("WIPO") has recently released a technology trends study[73] concluding that "AI-related inventions are booming, shifting from theory to commercial application"[74] and that "machine learning is the dominant AI technique disclosed in patents."[75] AI/ML inventions are patentable at the USPTO as long as they are novel, non-obvious, and include a practical application.[76] The EPO allows patent grants for computer programs that have a "technical character."[77] The EPO gives some AI/ML guidance, clarifying that terms such as "reasoning engine" and "neural network" will be examined with great care for concern of being abstract and not of a technical character.[78] The guidance also makes clear that "use of a neural network in a heart-monitoring apparatus for the purpose of identifying irregular heartbeats makes a technical contribution. The classification of digital images, videos, audio, or speech signals based on low-level features (*e.g.,* edges or pixel attributes for images) are further typical technical applications of classification algorithms."[79] However, "classifying text documents solely in respect of their textual content is however not regarded to be *per se* a technical purpose but a linguistic one" and not patent eligible. The guidance

[72] https://www.ipwatchdog.com/2019/03/26/ai-patents-make-comeback-finance-patents-still-struggling/id=107680/.
[73] WIPO, Technology Trends 2019 "Artificial Intelligence" (January 31, 2019) available at https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.
[74] Id.
[75] Id.
[76] "2019 Revised Patent Subject Matter Eligibility Guidance" available at https://www.govinfo.gov/content/pkg/FR-2019-01-07/pdf/2018-28282.pdf. At 18.
[77] EPO Guidelines of Examination, Part G, Chap., II 3.6. https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_6.htm.
[78] EPO Guidelines of Examination, Part G, Chap., II 3.3.1 Artificial Intelligence and Machine Learning. https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_3_1.htm.
[79] Id.

acknowledges that processes for training and classifying data may contribute to the technical character of the invention.[80]

Therefore, patent protection for AI/ML systems remains a critical element of any IP strategy where the inventor has created novel algorithms with practical applications, or novel steps for training and classifying data. Simple applications of known algorithms are probably not patentable. AI/ML patents are most important to AI/ML technology vendors, less so for their customers who use the technology. This proposition is confirmed by reviewing who is obtaining AI/ML patents. WIPO reports that IBM has the largest portfolio of AI patent applications, with 8,290 inventions, followed by Microsoft, with 5,930. IBM and Microsoft are followed by Toshiba, Samsung, and NEC.[81] And other non-computing entities have focused on obtaining protection in their respective field. For example, Baidu ranks high in deep learning, Toyota is prominent in transportation, and Siemens in life sciences.[82]

Determining which AI/ML inventions to patent should be based on a number of factors, including the following: (1) whether the invention is revolutionary or incremental, (2) size of the market for the invention, (3) whether the invention can be designed around, and (4) whether the infringement is easily ascertainable in competitors' products. For users and customers of AI/ML technology, patents are probably less valuable. For these entities the value of AI/ML technology lies more in the learnings and in the resulting operational model. As such, trade secrets become more important.

VI. Trade Secrets are a Critical AI/ML IP Protection

Given the unpredictability of copyright ownership and the uncertainty of enforcing AI/ML patents, renewed attention should be given to protecting your AI/ML through trade secrets. Trade secrets are well-positioned to protect technologies that are not easily reverse-engineered, experience rapid change, and require significant development expenditures. Therefore, trade secret protection applies well to AI/ML systems.

A trade secret consists of information that provides a competitive advantage because it is not known to others, and for which reasonable safeguards are maintained to protect its secrecy.[83] The Defend Trade Secrets Act ("DTSA") protects trade secrets nationwide. Trade secret protection is very flexible in its application to AI/ML systems. For example, trade secrets can protect "know-how." Know-how need not be novel, as in patents. Know-how can include the following:

---

[80] Id. ("where a classification method serves a technical purpose, the steps of generating the training set and training the classifier may also contribute to the technical character of the invention if they support achieving that technical purpose").
[81] https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.
[82] Id.
[83] 18 U.S.C. § 1836 et seq.

- The selection of data sources, the pre-processing of those data sources, including reviewing, culling, and anonymizing, take considerable effort. The details of these efforts constitute protectable know-how.
- The methods used to train the model, the algorithms used, and, perhaps most important, the resulting operational model and related learnings of the AI/ML system can comprise know-how.
- Specific implementations of novel algorithms are patent protectable, with the limitations discussed above. Even if the algorithms are not novel, the selection, application, and use of known algorithms to a particular problem is protectable as a trade secret.

Therefore, trade secret protection can be as deep and as broad as desired.

Of course, the primary goal of the AI/ML system is to solve a business or technological problem. The solution is captured in the operational model. The operational model characterizes the cumulative and consolidated efforts of developers applying algorithms and relevant information at any point in time to the problem. The operational model is not well-protected by patents, because it may constantly change and evolve. Trade secrets automatically cover these changes and adaptations without additional investment in the protection.

Unlike patents and copyrights, trade secrets do not grant their owner a monopoly in the subject matter. Trade secrets only protect against misappropriation by another, such as an employee taking the trade secrets to a competitor or a third party hacking into the AI/ML system. As such, trade secrets best protect know-how that cannot be easily reverse-engineered. Patents, copyrights, and trade secrets can protect the AI/ML system in complementary ways. For example, patents can cover any novel algorithm that has been invented, copyrights can protect the source code of the algorithm, and trade secrets can cover the remaining learnings and models. The coverage can be overlapping and need not be exclusive to one another.

Further, trade secrets are generally respected in many countries. Trade secrets are protected by article 39 of the 1995 Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement. The agreement protects “undisclosed information,” which is described in terms similar to trade secrets. Despite TRIPS, many emerging countries, such as China, India, and Brazil, fail to zealously enforce trade secret protections. Of course, there are many ways to protect the sharing of trade secrets with employees or competitors in these countries.

VII. Use Licensing to Close Intellectual Property Exposures

To the degree that copyright, patents, or trade secrets offer insufficient protection, which will be common, careful licensing of the AI/ML technology is warranted. Licensing of AI/ML software will vary widely depending on the particular business model and AI/ML technology model. Particular tension may exist between AI/ML technology suppliers and their customers. The number of AI/ML technology models are as numerous as business models.

Here are some general guidelines that differ from standard software licensing terms but must be considered when licensing AI/ML technology.

1. **IP Ownership:** Ownership of the learnings and operational model must be explicitly specified, including any derivative works, further use, licensing, or sublicensing.

2. **Scope of Use:** Scope and Term of License of use of the AI/ML system must be explicitly specified. It may be wise to limit rights in the AI/ML system to specific usage, access, or application, as opposed to a grant of perpetual use for any application. If third-party data is being used, specify if the data can merge with other data or whether the data must be quarantined. Any license agreement should discuss extraction of the AI/ML system from a customer environment and what residual rights survive with the customer or the AI/ML vendor.

3. **Data Licensing:** Similar to IP ownership in general, data ownership and use must be explicitly specified, including the treatment and usage of the original data, derived data, and retention of both the original and derived data.

4. **Responsible AI Licenses ("RAIL"):** Many are concerned that AI/ML systems will be used for unethical applications. As such, some responsible-use advocates have introduced terms for licensing source code related to AI/ML.[84] AI/ML software licensed according to RAIL prohibits or partially limits use of the software for (1) surveillance used to detect identities, (2) computer-generated media without disclosure that AI/ML generated the work, (3) predictions for health care insurance, and (4) prediction that someone will behave in criminal activities.[85] Similarly, the Electronic Frontier Foundation ("EFF") poses related questions, asking what should be the proper restriction and transparency of AI/ML systems when applied to surveillance, criminal sentencing, and profiling.[86] Application, privacy, and transparency-related licensing concerns of these types are likely to continue proliferating until some consensus or legal framework is reached.[87]

VIII. Manage Risk Through Containment

Another way to manage risk is to contain the AI/ML system within a specific environment. For example, rather than offer an AI/ML system to the public through a cloud-based system, the system can be licensed and restricted for use only within an enterprise or for use only within a client device (*e.g.,* a smartphone).

Cloud Model: IP risk is probably highest in a cloud-based model. In a cloud model, inputs can arrive from any producer and be produced to any user. This model

---

[84] Responsible AI Licenses https://www.licenses.ai/ (for source code and end-user license agreements).
[85] Id.
[86] *See* https://www.eff.org/issues/ai.
[87] *See https://www.eff.org/deeplinks/2019/03/openais-recent-announcement-what-went-wrong-and-how-it-could-be-better* where EFF questions OpenAI's release of its language model as insufficient.

has the benefit of learning from a wide variety of sources which yields better learning. However, if the user at least partially controls the input, such as in a style transfer system, then the potential for legal encumbrances on the input side is significant. Correspondingly, the resulting output will be encumbered. Such a risk level may not be acceptable. One solution to managing this risk is to contain the data and system within an environment.

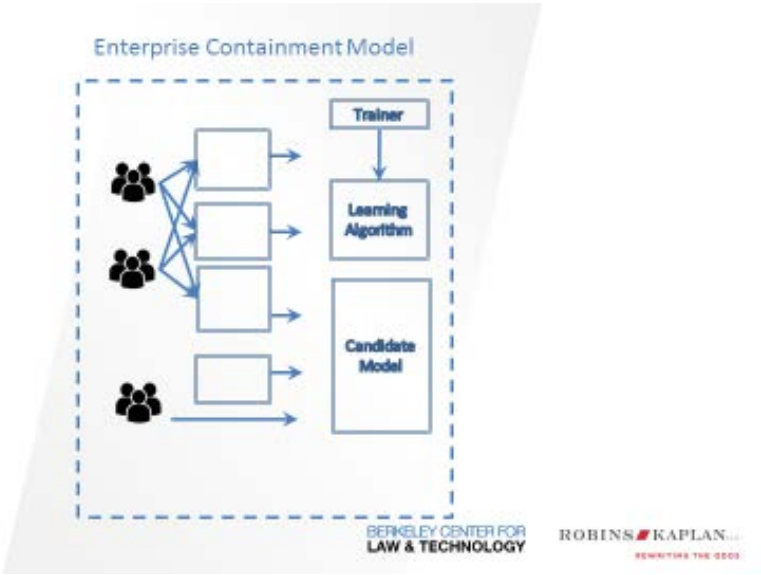

Figure 11: Enterprise Containment Model

Enterprise Model: For example, in an enterprise model the AI/ML system is contained within the enterprise, as shown in Fig. 11. The enterprise controls the data inputs and uses the output for its own business decisions. The enterprise builds its own AI/ML system or licenses one from a supplier. In such a model, the enterprise is able to manage its risks by keeping the majority, if not all, of its data inputs and outputs within the enterprise. However, while limiting the input data to a narrow set of sources may limit legal risk, it may also narrow the quality of learning by the AI/ML system. Negotiation between the AI/ML supplier and the enterprise concerning ownership rights in the learnings and operational model is essential.

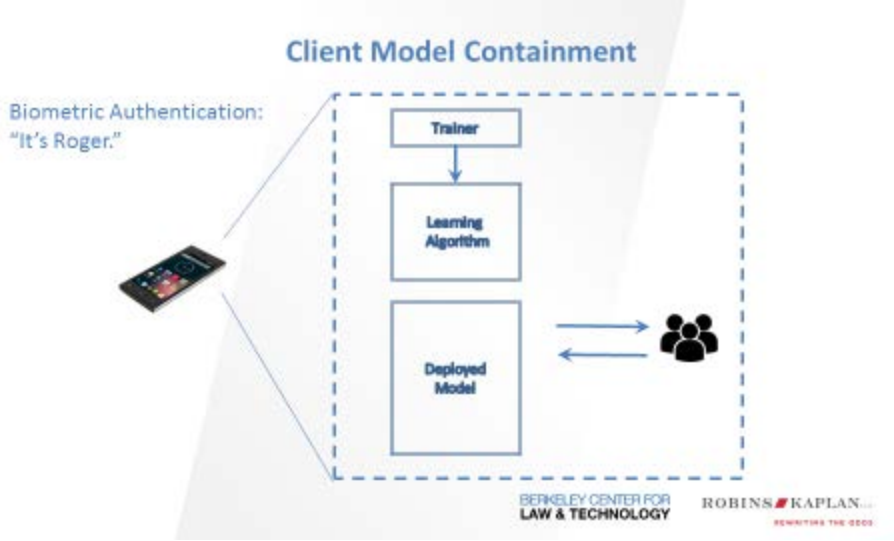

Figure 12: Client Containment Model

Consumer Model: In a consumer model, the AI/ML system is contained within the consumer device, as shown in Fig. 12. The data inputs and outputs are driven, or even selected, by the consumer. For example, biometric authentication based on a local AI/ML system can learn based on a user's walk, touch, or gait. Similarly, Google has disclosed work on a lightweight, machine learning architecture that allows Android users to generate predictive text without having to copy the data to cloud servers.[88] Just as in the enterprise model, the consumer model helps contain the risk of input from outside sources. However, also just like the enterprise model, the restriction of data may limit the AI/ML system's ability to learn.

Thus, containment is a wise approach to managing risk, if it can be done without too heavily affecting the quality of learning.

IX. Four Recommendations for Your AI/ML IP Strategy

1. **Mitigate Input Risk:** Contain the AI/Ml system locally, if possible. For example, contain information within the enterprise or within a client device. Choose data sources carefully, and assess terms of the data license. If possible, create a plan for unwinding data from the system in case it is tainted. Manage data risk through containment of information, if architecturally feasible.

2. **Protect Your Novel AI/ML Algorithms:** Patenting an AI/M algorithm is particularly attractive if the invention is revolutionary (not just an incremental improvement), the applicable market is significant (or the key target of your business plan), there is no known workaround, and no reverse engineering is necessary to identify its use in the market. Copyright the source code of the

[88] Sujith Ravi, *On-Device Machine Intelligence,* Google Research Blog (Feb 9, 2017), https://ai.googleblog.com/2017/02/on-device-machine-intelligence.html (last accessed April 29, 2019).

algorithm, and maintain all non-patentable technology as a trade secret. If you decide to use open-source AI software, carefully consider its restrictions on use and sharing.

3. **Protect Your Operational Model:** Maintain all learnings and the resulting operation model as a trade secret, as this information is not protected by copyright or patent. Publicize the effectiveness of the model, without publicizing details of the model itself. The desire to protect learnings, operational models, and improvements to existing models may create tension between technology suppliers and users. Be very careful in assessing when to release this information; it may contain the very value added by your organization.

4. **Protect Your Information Output by Careful Licensing:** All license agreements must address any limits or restrictions of use of the system and data. The agreement must also specify ownership in resulting work or derivative works and the data associated therewith. If extraction of the system is likely, specify what residual rights survive with the customer of the system and which with the technology or data supplier.