JOURNAL OF LEGAL STUDIES
https://e-journal.uum.edu.my/index.php/jls

How to cite this article:
Khong, D. W. K., & Mon S. W. (2023). Artificial intelligence as a common heritage of mankind. *UUM Journal of Legal Studies*, *14* (1), *UUM Journal of Legal Studies*, *14*(1), 113-139. https://doi.org/10.32890/uumjls2023.14.1.5

ARTIFICIAL INTELLIGENCE AS A COMMON HERITAGE OF MANKIND

[1]Dennis Wye Keen Khong & [2]Su Wai Mon
[1]Centre for Law and Technology, Faculty of Law,
Multimedia University, Malaysia
[2]Faculty of Law, University of Malaya, Malaysia

[1]Corresponding author: wkkhong@mmu.edu.my

Received: 31/5/2021 Revised: 17/1/2022 Accepted: 28/1/2022 Published: 18/1/2023

ABSTRACT

Artificial intelligence technologies today employ techniques known as machine learning and deep learning, which apply datasets to a suitable mathematical or statistical technique known as an algorithm. This in turn produces a model that can be employed to predict an outcome, given a new set of data that was previously unseen by the model. The principle of common heritage of mankind, which has originated in the 19th century, promotes the concept that humanity as a whole has rights and responsibilities over territories or outer space. This study aims to advance the concept of treating the components of artificial intelligence as an intellectual common in the form of a common heritage of mankind, in order to promote the discovery and the development of more novel artificial intelligence applications for the benefit of people around the world. This work employs a mix of legal doctrinal research related to intellectual property law and conceptual

theoretical discussion. The potential application of open access and open data licensing is discussed. The history of the common heritage of mankind is covered, and the potential benefit of recognizing basic artificial intelligence components as a common heritage of mankind is explored. Finally, a novel method for implementing this idea is proposed. This work is significant in advancing a method to liberate certain artificial intelligence technologies from intellectual property rights protection, in order to promote greater experimentation and the development of artificial intelligence applications for the greater good of humanity.

Keywords: Artificial intelligence; machine learning; deep learning; intellectual property; common heritage of mankind.

INTRODUCTION

Industrial Revolution 4.0 is hailed as a new stage of the industrial revolution in human civilization, whereby the cognitive and decision-making processes of humans are automated using digital technology (Schwab, 2016). The technology underlying Industrial Revolution 4.0 is artificial intelligence, or more specifically, machine learning and deep learning technologies. The availability of cheap computing power, both using multi-core Graphics Processing Units (GPUs) and cloud computing, the amassing of big data by organizations and businesses in the last decade, and the discovery of efficient algorithms have become the catalysts for the explosion of artificial intelligence applications in recent years (Internet Society, 2017).

It is anticipated that the use of artificial intelligence technology will increase the well-being and economic condition of the global population. The patent system was undoubtedly an important contributor to the rise of the industrial revolution in the 18th century (MacLeod, 1988; Sullivan, 1989). By patenting an invention, the investors and inventors gained a time-limited monopoly to sell a product incorporating the patented invention in order to recover the costs of the invention and earn a profit, while at the same time obtaining a legal means to fend off competition from counterfeiters.

Unlike the previous industrial revolutions, Industrial Revolution 4.0 involving artificial intelligence occurs in the information age.

Information, and more particularly digital information in the form of data, is a key component. Software and data take primary positions compared to computing hardware. Hence, it is likely that the patent system does not play the same important role as in the previous industrial revolutions. In this regard, the law on copyright and trade secrets are now tasked with this responsibility.

Notwithstanding the importance of intellectual property rights in the development and diffusion of new technologies in this modern age, one has to be wary of the risk of intellectual property law becoming a barrier to wider technological adoption. Changes in the technology sphere must be matched with changes in the law (Azmi, 2020). Thus, if there is a legal means to ensure that the basic components of artificial intelligence technology remain in the public domain, knowledge may be easily diffused and new applications developed without any hindrance of intellectual property law. Therefore, in this paper, the authors have argued that one should consider treating the basic components of artificial intelligence technology as a common heritage of mankind so that the benefits of advances in artificial intelligence technology may be enjoyed by as many people as possible.

This paper starts by providing an overview of contemporary artificial intelligence technology, particularly machine learning and deep learning techniques, and its components. It then focuses on an economic theory of information, and how information is efficiently used if it is priced at zero. Next, it examines various areas of intellectual property law vis-à-vis the different components of artificial intelligence technology. It then discusses the idea of the common heritage of mankind in international law, and how the same concept may be applied to the field of artificial intelligence technology. Finally, it concludes by proposing a self-help method of implementing the concept of the common heritage of mankind to the basic components of artificial intelligence technology.

The present paper is not the first to advance the idea of treating the basic components of artificial intelligence as an intellectual common. Salameh (2017), in his Bachelor's thesis, studied the opportunities and challenges for society in treating artificial intelligence as a common. Similarly, Tzimas (2018) has briefly examined the same issues in relation to the principle of 'international law supremacy'.

In the present paper, this idea has been extended by looking at the relevant intellectual property laws which have an impact on the idea. For the purpose of this paper, the researcher have referred mainly to Malaysian law as an example of the relevant intellectual property laws.

AN OVERVIEW OF ARTIFICIAL INTELLIGENCE

The term 'artificial intelligence' was coined by the computer science pioneer John McCarthy in 1956 during the Dartmouth Summer Research Project on Artificial Intelligence (Moor, 2006). The exploratory research agenda was to investigate how to develop a thinking machine (McCarthy et al., 1955). In the early days of artificial intelligence technology, most systems employed some form of rule-based or logic-based approach.

The current forms of artificial intelligence technologies employ techniques known as machine learning and deep learning. In essence, machine learning is the application of a dataset to a suitable mathematical or statistical technique known as an algorithm, in order to produce a model which can be used to predict an outcome, given a new set of data (Lehr & Ohm, 2017). In this article, the phrase artificial intelligence refers to those technologies using machine learning and deep learning techniques.

Data, whether textual, numerical, pictorial, video or audio, first needs to be structured into a dataset before it can be used. A dataset consists of many data instances, each of which is usually coded into a row. Each row of an instance will consist of several to many hundreds of columns known as features or attributes. Some datasets may have a special column known as a target variable. The target variable provides for the outcome, given the values in the other attribute columns.

Generally, machine learning techniques are categorized into supervised learning, unsupervised learning and reinforcement learning. A supervised learning approach is used to predict an outcome based on other given data, whereas unsupervised learning does not have such a predetermined outcome and is used to cluster similar types of data instances, in order to determine which sub-group a new data instance is closest to in the dataset (James et al., 2021). Reinforcement learning

on the other hand is a more sophisticated form of supervised learning which allows an algorithm to continuously improve its own model through a feedback loop (Sutton & Barto, 2018). Reinforcement learning as a field is still very much less mature as compared to supervised learning and unsupervised learning.

In order to perform supervised learning, unsupervised learning or reinforcement learning, a suitable algorithm is needed. An algorithm is a set of computer codes which transforms data into a model. Typically, algorithms are developed based on some statistical techniques and formulated as a programming function. There are many different algorithms used in machine learning. These algorithms are usually grouped into the following three categories: classification, regression and clustering (James et al., 2021). Classification is used to predict the class of a new data instance, based on existing datasets with a target variable. Regression does the same, but instead of predicting a class, the target variable predicts a numerical value. Finally, clustering groups data instances into a fixed number of different classes, based on the characteristics of the features or attributes, and are used to identify which class a new data instance is closest to.

Quite often, a few different algorithms may be equally used to achieve the same purpose. It is the duty of the artificial intelligence expert to undertake exploratory work to identify the best algorithm for a given dataset, by comparing the performance of the models that have been generated. The datasets used in machine learning are usually randomly broken into two subsets of training data and testing data, with the training data taking a bigger portion, such as 70 percent to 90 percent (Wiley & Wiley, 2019, p. 226). The training data is applied to an algorithm to generate a model. The performance of the model is tested using the test data.

Once a model has been found, the dataset that has been used to make the model is no longer needed to make predictions. Thus, work of developing a good artificial intelligence system may lie in preparing the data, selecting the appropriate features and applying various algorithms to find the one which produces the best model for the job. For example, Schelter et al. (2018) provides a discussion on real-world challenges in developing models. Sometimes, when quick processing time is crucial, the best model may not be the one which gives the most accurate predictions, but the one which is balanced against the time used to come up with a prediction.

Deep learning extends upon the machine learning technique by using a technique called artificial neural networks. Basically, an artificial neural network can be seen as a stack of multiple layers of machine learning algorithms cascaded on each other in order to produce a more detailed analysis of the dataset. Deep learning technology is often used in image and video classification, natural language processing and speech recognition.

SOME ECONOMICS OF INFORMATION

Technology is informational in nature. As the renowned economist and Nobel laureate Kenneth J Arrow (1962) wrote, "if [the cost of transmitting a given body of information] were zero, then optimal allocation would obviously call for unlimited distribution of the information without cost". Thus, the optimal price for information is at its marginal cost, i.e. zero, whereby all users who wish to use it can do so. Due to the fact that information has the characteristics of public goods, and consumption of it is non-rivalrous, it follows that information, and technology for that matter, do not suffer from the problem typically described as the tragedy of the commons (Hardin, 1968). Instead, when information is priced at zero and is accessible to all, such as being placed at the public's disposal on the Internet, the use of information becomes efficient. Hence, the ideal scenario for access to technology is to provide free access to all.

Unfortunately, technological development is not costless. Often, businesses need to invest in research and development in order to devise useful technologies. However, due to the fact that technology is informational in nature, like all information, it bears the following twin characteristics of public goods: non-rivalry in consumption and non-excludability (Samuelson, 1954). It is the characteristic of non-excludability that gives rise to free-riding, i.e. using information without paying for it, because it is difficult to exclude non-payers. Intellectual property laws are enacted as a solution to this free-riding problem (Khong, 2019). Relevant intellectual property rights to artificial intelligence are, namely copyright (and database rights in Europe), patent and the tort of breach of confidence. Of these three, only patent rights may exhibit a blocking effect by a prior patented invention against subsequent inventions, which may deter widespread adoption of follow-up technologies due to the licensing cost (Czarnitzki et al., 2020).

INTELLECTUAL PROPERTY RIGHTS IN ARTIFICIAL INTELLIGENCE

Datasets

Traditionally, databases, or legally defined as compilations of data, are the subject matters of copyright law. Article 10.2 of the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS), as Annex 1C of the Agreement Establishing the World Trade Organization states that:

> "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself."

In the Malaysian context, the old copyright doctrine which protects 'tables or compilations' under the definition of 'literary work' in section 3 of the Copyright Act 1987 and based on a 'sweat of the brow' standard still applies. The Malaysian High Court in *Kiwi Brands (M) Sdn Bhd v Multiview Enterprises Sdn Bhd* [1998] 6 MLJ 38 (HC) cited with approval this traditional English approach:

> "In *Football League Ltd v Littlewoods Pools Ltd* [1959] Ch 637, it was held that even if the compilation of the chronological list had merely consisted of the reproduction of the clubs' fixtures lists, so that there was no element of skill and ingenuity, there was nevertheless sufficient painstaking hard work to justify a claim for copyright. As Upjohn J remarked at p 656 of the report:
>
> > But, I would add, it involves a great deal of painstaking hard work with complete accuracy as the keynote. That was all that was required . . .
>
> *Football League Ltd* merely reiterates the now entrenched principle that copyright protects compilations which may need no skill and ingenuity so long as there is

> effort in producing the copyrighted work. This principle appears to manifest itself in s 7(3) of the Copyright Act 1987 which merely requires that 'sufficient effort has been expended to make the work original in character'. Effort and hard work work hand in hand to make the copyrighted works protectable."

Notwithstanding this, an opposite position was made in an obiter dictum in *Hardial Singh Hari Singh v Daim Zainuddin & Ors* [1991] 2 CLJ (Rep) 701 (HC). In *Hardial Singh*, the court noted that:

> "[The appellant's] failure to show that his compilations were original lay in his incapacity to demonstrate that he had imposed some sort of unique pattern or order on the material he had copied which was not to be found in the Government publications. Mere listing of facts is not enough to make something a literary work, however laborious the undertaking."

Curiously, in compliance with Article 10.2 of the TRIPS Agreement, the Malaysian Copyright Act 1987 has another provision which corresponds to the TRIPS requirement. Section 8(1)(b), which was incorporated through a series of amendments to the Copyright Act reads:

> "(1) The following derivative works are protected as original works:
>
> ...
>
> (b) collections of works eligible for copyright, or compilation of mere data whether in machine readable or other form, which constitute intellectual creation by reason of the selection and arrangement of their contents."

Therefore, the paragraph has made it clear that the standard of protection for 'compilation of mere data' is that of 'intellectual creation', much like the position of the case in *Hardial Singh*, although that case did not refer to section 8(1)(b) and the 'intellectual creation' standard must be applied only to the 'selection and arrangement of [the compilation's] contents'. The discrepancy between the English 'sweat of the brow' position and the 'intellectual creation' standard is due to the fact that

the former follows a Lockean approach to property rights favored in English law and, by extension, copyright law (Moore, 1997), while the latter is influenced by a Hegelian personality theory of copyright law in Continental Europe (Hughes, 1988).

The existence of two separate provisions in the Malaysian copyright law inevitably leads to some confusion as to the correct standard to be applied (Manap, 2012). Unfortunately, no reported court decision has discussed the application of section 8(1)(b) vis-à-vis 'tables or compilations'. If one can take a cue from cases such as in *Kiwi Brands*, it would appear that the English standard of 'sweat of the brow' applies, notwithstanding the additional provision of section 8(1)(b). This is to be contrasted to the position in the European Union where a two-tiered system exists to protect databases by copyright under an 'intellectual creation' standard and a separate *sui generis* database right under a 'substantial investment' standard (Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases [1996] OJ 2 77/20).

If one applies the 'sweat of the brow' standard to datasets used in artificial intelligence, one can be fairly certain that they qualify for copyright protection under the Malaysian copyright law and other jurisdictions adopting a similar standard. The effort to collect the data, organize them into suitable structures such as in a data frame, clean up the data using data wrangling techniques, and manually labelling data would likely be sufficient to pass the bar for copyright protection under the 'sweat of the brow' standard.

On the other hand, under the 'intellectual creation' standard, it is not certain whether most artificial intelligence datasets will qualify for copyright protection. This is because the 'selection or arrangement' requirement was first formulated in the age of printed directories, whereby data has to be selected because paper was expensive and it was not practical to have content running into hundreds of thousands of pages. Henceforth, information has to be arranged in a user-friendly manner in order to be readable. In this era of Big Data, the cost of digital storage is no longer a significant barrier (Klein, 2017) and it is not necessary to pre-arrange information in a visually pleasing manner because the data is not meant to be read by a living person but is to be processed by computer programs and algorithms. Furthermore, sorting and filtering functions can be easily applied to

databases to produce different views of data on the fly. In conclusion, 'selection and arrangement' has no practical meaning in the context of electronic databases, and it is difficult to sincerely demonstrate that there is intellectual input in the 'selection and arrangement' of the content of datasets which constitutes an 'intellectual creation'.

Copyright Protection for Algorithms

Article 10.2 of the TRIPS Agreement states that:

> "Computer programs, whether in source or object code, shall be protected as literary work under the Berne Convention (1971)."

Thus, computer programs are protected accordingly as literary works under section 3 of the Malaysian Copyright Act. The term 'computer program' refers to the set of instruction codes for a computer.

The term 'computer program' in copyright law is to be contrasted with the concept of 'software' used in patent law. Article 9.2 of the TRIPS Agreement states that:

> "Copyright protection shall extend to expression and not to ideas, procedures, methods of operations or mathematical concepts as such."

The Malaysian Copyright Act's equivalent is found in section 7(2A). It is argued that the 'ideas, procedures, [and] methods of operations' behind a computer algorithm is not protected by copyright law, although it may be considered for protection under patent law. Copyright protects the actual implementation in the form of computer codes. However, the position in the United States is slightly different. Following the authority of *Computer Associates International, Inc v Altai, Inc*, 982 F 2d 693 (2d Cir. 1992), non-literal aspects of a computer program may be protected under copyright law in the United States. This approach, however, has been rejected by the English Chancery Division in *IBCOS Computers Ltd v Barclays Mercantile Highland Finance Ltd* [1994] FSR 275 as not helpful in English copyright law.

It is undeniable that algorithms implemented in computer codes are computer programs and can rightly be protected under copyright.

However, it does not follow that artificial intelligence algorithms *per se* are automatically protectable under copyright law. In the United States' copyright law, a merger doctrine excludes an expression from being considered for protection by copyright law, if "the expression and idea have merged, which has been stated to occur where there are no or few other ways of expressing a particular idea" (*Apple Computer, Inc v Franklin Computer Corporation*, 714 F 2d 1240 (3d Cir. 1983)).

It is not common in English copyright law and by extension, Malaysian copyright law, to recognize a merger doctrine. However, such an idea has previously been raised in an English court. In *Total Information Processing Systems Ltd v Daman Ltd* [1992] FSR 171 (ChD), it was held that:

> "... stemming from the principle that copyright does not exist in ideas but in the expression of them, is the line of authorities commencing with *Kenrick & Company v. Lawrence & Company* (1890) 25 Q.B.D. 99 that if there is only one way of expressing an idea that way is not the subject of copyright."

Furthermore, the basis for the American merger doctrine, the idea-expression dichotomy, is statutorily recognized in section 7(2A). Thus, specific algorithms may be declared as ideas and not subject to copyright protection, as Lord Hailsham in *LB (Plastics) v Swish Products* [1979] RPC 551 at 629 (HL) observed:

> "Of course, it is trite law that there is no copyright in ideas … But, of course, as the late Professor Joad used to observe, it all depends on what you mean by 'ideas'."

A similar effect to the merger doctrine can potentially be achieved by not protecting commonplace expressions (Ang, 1994), as stated by Lord Millett in *Designers Guild Ltd v Russell Williams (Textiles) Ltd* [2001] 1 WLR 2416 (HL):

> "... similarities may be disregarded because they are commonplace, unoriginal, or consist of general ideas."

In *Petraware Solutions Sdn Bhd & Anor v Readsoft Aktiebolag & Anor* [2013] MLJU 1606 (CA), the plaintiffs failed in their claim on appeal

when the judge held, following 'international approaches to copyright protection of non-literal components of computer programs', that the plaintiffs did not show that they owned the copyright to a graphical user interface, system flow of the modules and terms used in their menus. The court found generally that these are ideas not capable of being protected under copyright law.

In conclusion, although it is doubtless that computer programs are protected under copyright law, artificial intelligence algorithms may or may not be so protected. Copyright protection may be unavailable to an algorithm if the idea behind the said algorithm is well-known, and the computer codes merely implement the algorithm.

Patent Protection for Algorithms

Patent law in Malaysia is governed by the Patents Act 1983. Section 13(1)(a) excludes 'discoveries, scientific theories and mathematical methods' as the subject matters of a patent. This phrase is *in pari materia* to the same exception in the European Patent Convention (EPC) and functionally the same in the United Kingdom's Patents Act 1977. However, unlike the case of the EPC and the Patents Act 1977, there is no patentability exclusion for 'programs for computers' in the Malaysian Patents Act 1983.

To date, there is no reported court decision on the patentability of software in Malaysia. Nevertheless, as a practice, there is no reason for the Intellectual Property Corporation of Malaysia (myIPO) to reject patent applications solely because they involve a software-implemented invention. Contemporary global developments, particularly those in the United Kingdom, are influential on the development of patent laws in Malaysia. Thus, the English Court of Appeal decision in *Aerotel Ltd v Telco Holdings Ltd* [2006] EWCA Civ 1371 is instructive.

In the well-known *Aerotel Ltd* decision, it was held that the accepted approach in determining whether an invention involving software is patentable is the 'technical effect approach':

> "Ask whether the invention as defined in the claim makes a technical contribution to the known art—if no, Art. 52(2) applies. A possible clarification (at least by way of

> exclusion) of this approach is to add the rider that novel or inventive purely excluded matter does not count as a 'technical contribution'."

In general terms, the *Aerotel Ltd* decision affirmed the position that, just because an invention uses software is not a reason to reject the invention for patentability. What is necessary is that the invention demonstrates a novel technical contribution. On the other hand, if the novel invention is just a mathematical method, then it does not count as a qualifying technical contribution.

The implication of this decision to artificial intelligence algorithms is that it is likely that the algorithms *per se* would be excluded from patentability as a form of mathematical methods, but the use of an artificial intelligence algorithm in an invention is no bar to patentability.

Models

The third component of an artificial intelligence system is a model. An artificial intelligence model is a set of parameters associated to the features applied to a specific algorithm. Between the two, the parameters are the identifying characteristics of a model because a model without parameters is just an algorithm. Thus, from the point of view of copyright law, a model can be considered as a database. Applying the 'sweat of the brow' standard for copyright protection to a model may result in the database being accepted for copyright protection. However, it is not certain that the same result may be achieved by the 'intellectual creation' standard, since the effort in creating the model is not in the selection or arrangement of its content, but in calculating the values of the parameters therein. This conclusion is supported by the decision of the European Court of Justice in *Football Dataco Ltd & Ors v Yahoo! UK Ltd & Ors*, C-604/10, EU:C:2012:115 where the court held that:

> "the concepts of 'selection' and of 'arrangement' ... refer respectively to the selection and the arrangement of data, through which the author of the database gives the database its structure. By contrast, those concepts do not extend to the creation of the data contained in that database."

It is very unlikely that a model can be considered as an invention capable of patent protection. Nevertheless, once developed, the model may be a valuable part of an artificial intelligence system because it can be incorporated into software for real-world applications. Surprisingly, intellectual property protection of machine learning models is not a topic discussed in the legal literature, although the risk of machine learning and deep learning models being stolen is real (Hitaj & Mancini, 2018; Tramer et al., 2016).

Confidential Information Protection

Article 39 requires WTO member states to have laws to protect undisclosed information. In Malaysia, there is no specific statute governing the protection of undisclosed information. Instead, the English common law tort of breach of confidence is applicable. Trade secrets, which are confidential information used in business, are protected under this tort.

In *Alfa Laval (M) Sdn Bhd v Ng Ah Hai* [2009] 7 CLJ 1 (HC), the High Court of Malaya in an obiter dictum accepted the proposition that the source code of a computer program may be protected as confidential information, on condition that what is claimed to be confidential is separated from non-confidential components. Similarly, Jacob J in *IBCOS Computers Ltd v Barclays Mercantile Highland Finance Ltd* [1994] FSR 275 (ChD) held that the "source code is normally kept confidential by software houses … source code … was confidential. It follows that the plaintiffs [can] succeed, so far as breach of confidence is concerned …".

Given the uncertain nature of copyright and patent protections for algorithms and models, confidential information protection may by far be the most effective way to protect the different components of artificial intelligence. The necessary conditions for confidentiality information protection according to Megarry J in *Coco v A N Clark (Engineers) Ltd* [1968] FSR 415 (ChD) are:

> "First, the information itself, in the words of Lord Greene, M.R. in the *Saltman* case ... must 'have the necessary quality of confidence about it'. Secondly, that information must have been imparted in circumstances importing an obligation of confidence. Thirdly, there

> must be an unauthorised use of that information to the detriment of the party communicating it."

Through the use of cloud computing technology, it is now possible to offer artificial intelligence services to clients without disclosing the underlying algorithms and models (Gill et al., 2019). This has the advantage of keeping the algorithms and models secret, by transforming hardware, software and data into an Artificial Intelligence as a Service (AIaaS). As long as the source codes and datasets to the underlying deployed artificial intelligence technology are not made available to the public, it is possible that confidential information protection can be co-opted to give an additional layer of legal protection to these technological components.

Open Access and Open Source Movements

Despite the potential availability of intellectual property protection to artificial intelligence technologies, such as data, algorithms and models, not all creators are interested in keeping these technologies proprietary. Instead, these publicly spirited creators may want to leave their artificial intelligence technologies in the public domain so that others can employ them to create useful applications.

Unfortunately, copyright statutes are written to only afford copyright protection to works, and leave scant attention to the possibility of dedicating one's work to the public domain (Johnson, 2008). Hence, there is a legal risk involved in simply taking and using codes and datasets from the Internet without prior verification of whether the authors have given consent for the use of their works. Hence, in the absence of a legal mechanism to place a copyrighted work in the public domain, the practical alternative is to publicly license the codes and datasets under an open source license (Paton & Kobayashi, 2019). Correspondingly, an open science and open data movement attempts to encourage the release of scientific data to the public in order to accelerate scientific research (Benchoufi & de Fresnoye, 2020).

An open source license is one of the many forms of copyright license which permits licensees to use protected computer programs under certain conditions. Open source licenses range from permissive licenses with minimal conditions, to restrictive licenses such as the GNU Public License that require a derivative computer program or

codes to be re-licensed under the same license as the original computer program. Generally, regardless of whether the open source licenses are permissive or restrictive, they all have a few features which are similar, namely source codes of the computer programs must be made available, and users can both use the source codes for modification and learning, as well as to create compiled object codes.

Open data takes the concept of open source for computer source codes and applies it to datasets. It is possible to apply most open source licenses such as the GNU General Public License, and open access licenses, such as the Creative Commons licenses to datasets, although specific open data licenses are also readily available. For example, the Open Knowledge Foundation provides a set of Open Data Commons licenses dedicated to covering rights on the use and sharing of datasets. National organizations responsible for collecting and compiling data can also play an important and positive role in overcoming barriers to the adoption of artificial intelligence technology. In particular, they can make the data they have collected available free of charge to the public under an open data license.

Risk of Intellectual Enclosure

As can be seen from the foregoing discussions, the components of artificial intelligence are likely to be subjected to intellectual property rights protection such as copyright, patents and confidential information. Strong and broad intellectual property rights protection lead to barriers to entry by competitors (Heger & Zaby, 2017), and this in turn deters widespread adoption. High transaction costs in obtaining licenses may deter widespread adoption of technology and useful information (Gordon, 2002).

Given the prevalence of the global intellectual property regime that potentially covers several artificial intelligence technologies, open source and open access licensing appears to be a feasible short-term solution to quickly build up intellectual commons in artificial intelligence, despite its reliance on the consent of all contributors. Furthermore, there remain variations among domestic laws on intellectual property, such that their treatments to open source and open access licensing are not consistent. In the long term, it is proposed in this paper that there should be an alternative method to designate components of artificial intelligence technology as a common heritage of mankind.

COMMON HERITAGE OF MANKIND

Origin of 'Common Heritage of Mankind'

The common heritage of mankind principle can be traced back to the 19th century, when geologists discovered polymetallic nodules on the deep seabed beyond the territorial sea in the Arctic Ocean off Siberia (Barkenbus, 1979). This led to the need for a new legal regime to govern access to this deep seabed when nations wanted to commercially mine the polymetallic nodules using new technology (Guntrip, 2003). In August 1967, Ambassador Arvid Pardo of Malta first proposed in his speech at the General Assembly of the United Nations that "[t]he seabed and the ocean floor are a common heritage of mankind and should be used and exploited for peaceful purposes and for the exclusive benefit of mankind as a whole" (Mirzaee, 2017).

Initially, it was difficult to ascertain whether the common heritage of the mankind principle bears a status of legal standard or merely the theory, philosophical and political concept in international law. As such, there have been many debates among developed and developing countries on the interpretations of the principle (Guntrip, 2003). The developing countries endorsed the principle to establish a more equitable distribution of resources and income between developed and developing states. On the other hand, the developed states, being technologically more advanced in deep seabed mining, rejected the principle for its lack of legal meaning and claimed that deep seabed resources could not be considered as the common resources of the global community (Guntrip, 2003).

In order to further develop the common heritage of mankind principle, the General Assembly passed a series of resolutions relating to deep seabed explorations. The most important resolutions are the Moratorium Resolution 1969 which restricted the exploration and exploitation of deep seabed resources, and the Declaration of Principles 1970 which declared deep seabeds as a common heritage of mankind.

The common heritage of mankind principle was later incorporated into the United Nations Convention on the Law of the Sea 1982 (UNCLOS). Part XI of the UNCLOS embodies provisions governing the deep seabed via Article 1(1), which defined the 'Area' as 'the

seabed, ocean floor and subsoil thereof, beyond the limits of national jurisdiction'. Article 136 proclaims that 'the Area and its resources are the common heritage of mankind'. In addition, Article 140 states that 'activities in the Area shall be carried out for the benefit of mankind as a whole'.

Unfortunately, the incorporation of Part XI in UNCLOS was not supported by the developed nations. The most significant opponent was the United States of America, which claimed that Part XI would deter future development of deep seabed mining activities.

Although the UNCLOS claims that the deep seabed or the 'Area' is a common heritage of mankind, it does not provide a concrete meaning of the common heritage of mankind principle. Nevertheless, four elements of the common heritage of mankind principle under the corpus of international law, particularly relating to deep seabed activities may be identified in the UNCLOS as follows:

(1) The prohibition on the acquisition of the deep seabed which confirms that no state can exercise sovereignty or control over the deep seabed (Art 137);
(2) The deep seabed must be used only for peaceful purposes (Art 141);
(3) There must be 'equitable sharing of benefits' gained from deep seabed mining (Art 160); and
(4) The conservation of natural resources and marine environment which requires the International Seabed Authority (ISA) to appropriate rules and regulations for such purposes (Art 145).

The Usage of '**Common Heritage of Mankind**' in Other Fields of International Law

Apart from the deep seabed beyond national jurisdictions, a concept similar to the common heritage of mankind has also been used in other areas of international law, such as the outer space law and the international environmental law.

For example, the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including Moon and Other Celestial Bodies 1967 mandates that '[t]he exploration and use of outer space, including the moon and other celestial bodies,

shall be carried out for the benefit and in the interests of all countries, irrespective of their degree of economic or scientific development, and shall be the province of all mankind' (Outer Space Treaty, Article 1).

In the area of international environmental law, the Rio Declaration on Environment and Development 1992 (Rio Declaration) and the Declaration of the United Nations Conference on the Human Environment 1973 (Stockholm Declaration) reaffirmed the common heritage of mankind principle in developing the principles of both Declarations. The preamble of the Rio Declaration urges the global community to recognize 'the integral and interdependent nature of the Earth, our home.' Principle 2 of the Rio Declaration claims that states are responsible, in accordance with the Charter of the United Nations and the principles of international law, to exploit their own resources and to ensure such activities within their jurisdiction or control do not cause damage to the environment of other states or of areas beyond the limits of national jurisdiction.

One such example would be the effects of transboundary haze pollution. Based on the principles in the Rio Declaration, the common heritage of mankind becomes the legal standard for imposing responsibilities and conferring rights to individual nation states in order to achieve the common goal of global sustainable development. Similarly, the Stockholm Declaration proclaims that the protection and improvement of the human environment affects the well-being of humanity as a whole, and it is the duty of all the governments of the world to protect the environment. Therefore, both Declarations promote the idea of a common heritage of mankind in achieving sustainable development goals as an inclusive effort, by sharing burdens and enjoying the common benefits from the environment.

Defining '**Common Heritage of Mankind**'

When one examines the historical background of the common heritage of mankind concept, it appears that there was no explicit meaning given to the phrase 'common heritage of mankind'. This is one of the reasons which has led to the controversy on whether the common heritage of mankind concept has any legal effect among the developing and developed states.

In order to give the phrase a literal meaning, the word 'common' can be defined as 'a thing shared in respect of title, use or enjoyment, without

apportionment or division into individual parts' (Arnold, 1975). The word 'heritage' suggests property or interest which are reserved to a person by birth or something handed down from one's ancestors. In defining 'mankind', it is necessary to make a distinction between mankind and man. Mankind refers to the human race as a whole, whereas man refers to an individual man and woman. Since mankind is not yet unified under a single world government, and therefore, the collective entity of mankind is represented by the various nations of the world. It thus, follows that the 'exercise of rights to the common heritage of mankind pertains to nations, representing mankind, and not individuals (Arnold, 1975).

Owolabi (2013) summarized the core elements of the common heritage of mankind principles as follows:

> (1) 'No state or person can own common heritage spaces or resources (the principle of non-appropriation). They can be used but not owned and when common heritage of mankind applies to areas and resources within national jurisdiction, exercise of sovereignty is subject to certain responsibilities to protect the common good;
> (2) Common heritage of mankind shall be reserved for peaceful purposes (preventing military uses);
> (3) Equitable sharing of benefits associated with the exploitation of the resources in question, paying particular attention to the interests and needs of developing states in accordance with a system of cooperative management for the benefit of all humankind; and
> (4) Common heritage of mankind shall be transmitted to future generations in substantially unimpaired condition.'

ARTIFICIAL INTELLIGENCE AS A COMMON HERITAGE OF MANKIND

Using the concept of 'common heritage of mankind', which was first developed in relation to deep seabed mining for the benefit of both

developed and developing nations, the present paper is proposing that a similar concept may be applied to core artificial intelligence technology, so that its benefits may be enjoyed by humanity as a whole. If one imagines artificial intelligence technology as consisting of both core knowledge and applied knowledge, the core knowledge is the building blocks for developing more advanced artificial intelligence applications.

Part of these core components are the algorithms that are already in use today. Nevertheless, algorithms without datasets are not very useful. In line with the Open Science and Open Data movements, more data should be made available to the public. For a start, data that has been acquired and developed through public funding should be open-accessed, since the cost of creation has already been borne by the public.

Creating well-developed models available to the public is also a desirable exercise. Training an algorithm using a dataset which contains a large number of features with an even larger number of rows may require considerable computational power. Thus, it would be more efficient if the model is developed and shared by others, without having to retrain it. The computation requirement for training a model becomes even more demanding when it is trained using deep learning algorithms. If pre-trained models could be shared and regarded as belonging to the common heritage of mankind, then more artificial intelligence applications could be used or developed.

The Internet is a highly beneficial technology that can be used to deliver such resources. Since the marginal cost of distributing data and information on the Internet is extremely low, to the point of approaching zero, it would be desirable to make data and information free in order to achieve optimal use.

Rather than relying on the central planning of the government to recognize artificial intelligence as a common heritage of mankind, perhaps now is the time to democratize the liberation efforts of artificial intelligence technology. One effort that can perhaps be made by world governments is to establish an international treaty to facilitate the designation of information and knowledge by their creators as the ‘common heritage of mankind’. This will overcome the shortcomings in the existing intellectual property regime which hinders the dedication of knowledge to the public domain.

Currently, many components such as datasets and algorithms have been made available to the public using open access and open source licensing. Programming languages such as Python, R, Julia and others are open-sourced, and many free machine learning libraries are also readily available. Although having a 'common heritage of mankind' branding will give significant prominence to the technologies that have been offered, using an open source and open access license is nevertheless, a second-best mechanism. Furthermore, having an Internet repository of all designated components such as datasets, algorithms in various languages, and pre-trained models, will make it easy for learners and developers to access and use these resources for economic development and the common good.

IMPLEMENTATION

As a long term solution to recognizing fundamental artificial intelligence technologies and components as within the ambit of the common heritage of mankind principle, a legislative approach is desired. This can be achieved through signing a multilateral treaty to be sponsored by an international organization such as the World Intellectual Property Organization. However, such efforts usually take a long time and require many rounds of meetings and negotiations among member states. This is especially so when the treaty proposed is not to further the interests of right-holders.

The Marrakesh Treaty to Facilitate Access to Published Works by Visually Impaired Persons and Persons with Print Disabilities, or commonly known just as the Marrakesh Treaty is a good example. It was first proposed in 2013 in Marrakesh, Morocco (Vleugels, 2020). The initial effort for such a treaty started with the formation of the joint WIPO and UNESCO Working Group on Access by the Visually and Auditory Handicapped to Material Reproducing Works Protected by Copyright in 1981 (https://www.sutori.com/story/libraries-and-marrakesh-a-history-of-engagement--3W5Tf1QAfab4aBeHVaDezRi3). From the stage of proposal to ratification and implementation, the Marrakesh Treaty met with various opposition from the publishing industry. It came into force in 2016, only after twenty countries had ratified or accessioned to the treaty. Drawing the lessons from the Marrakesh Treaty episode, Land (2018) has raised the concern that its arduous ratification journey demonstrates that the prospect for other similar copyright exceptions does not look promising.

A short term solution would have to rely on open source and open access licensing. Perhaps taking a leaf from the idea of a treaty illustrated above, a multi-faceted intellectual property rights license can be crafted to cover all the various types of artificial intelligence components. This approach is not unprecedented. For example, the Creative Commons Public Licenses 4.0 cover both rights under copyright and the European *sui generis* database right (see https://creativecommons.org/about/cclicenses/). Similarly, the GNU General Public License version 3 includes a section on the licensor refraining from enforcing one's patent rights relating to the copyright work licensed under the General Public License (see https://www.gnu.org/licenses/gpl-3.0.html).

Although a multi-faceted license approach appears to be a possible short-term solution, it is not without its limitations. Despite widespread adoption around the world, the validity of open source and open access licenses have not been tested in most national courts. Some intellectual property statutes require licenses to be in the form of contracts, and the country's contract law may require additional elements such as the elements of consideration. Indeed, section 41 of the Malaysian Patents Act 1987 speaks of 'license contract' and that these license contracts must be 'in writing signed by or on behalf of the contracting parties'. Furthermore, following English contract law, Malaysian contract law requires an offeree to provide a consideration to the offeror (section 26, Contracts Act 1950). Thus, a public license for patents is likely unenforceable because public licenses do not require a licensee to provide any consideration, such as the payment of a fee or royalty, to the licensor. Furthermore, licensees are anonymous to licensors, so a signed agreement by both parties are unavailable in most instances. Bearing this in mind, an international treaty approach requiring amendment to existing intellectual property regimes may be the only feasible approach in making artificial intelligence technology a form of the common heritage of mankind.

CONCLUSION

There is currently no simple way to place various components of machine learning and deep learning technologies into the public domain. Depending on the laws of specific countries, some components may or may not be protected under copyright law, patent

law, or trade secret law. Open source and open access licensing may to some extent, achieve that aim, but the effectiveness of such an approach is questionable in some circumstances. Furthermore, such licenses may not be enforceable under the intellectual property laws of some countries.

The idea of having a common heritage of mankind designation to artificial intelligence technology does not preclude the possibility of protecting the implementation of artificial intelligence technology for real-world applications through the copyright and patent system. Economic development is premised on the ability of inventors and entrepreneurs being able to offer solutions to real-world problems in return for appropriate remuneration. Indeed, many of the current richest persons in the world have substantial investments and products using artificial intelligence in some form or another.

On the other hand, by designating core components of artificial intelligence as a common heritage of mankind, it would be easier for students and developers to adopt and use them without fear of negative legal repercussions. More users worldwide will become familiar with these technologies and with this ability to share freely, potentially more novel applications will be discovered and developed for the benefit of the general worldwide population. The ideals of the Industrial Revolution 4.0 will be actualized much faster.

ACKNOWLEDGMENT

An early draft version of this paper has been presented at the Machine Lawyering Conference 2021: Human Sovereignty and Machine Efficiency in the Law, organized by the Chinese University of Hong Kong, 14–16 January 2021.

REFERENCES

Ang, S. (1994). The idea-expression dichotomy and merger doctrine in the copyright laws of the US and the UK. *International Journal of Law and Information Technology*, *2*(2), 111–111.

Arnold, R. P. (1975). The common heritage of mankind as a legal concept. *The International Lawyer*, *9*(1), 153–158.

Arrow, K. J. (1962). Economic welfare and allocation of resources for inventions. In C. on E. G. of the S. S. R. C. Universities-National Bureau Committee for Economic Research (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press. http://www.nber.org/books/univ62-1

Azmi, I. M. A. G. (2020). Challenges for legal education in the era of IR 4.0. *UUM Journal of Legal Studies*, *11*(2), 27–51.

Barkenbus, J. N. (1979). *Deep seabed resources: Politics and technology*. Free Press.

Benchoufi, M., & de Fresnoye, O. (2020). Open science and open data: Accelerating scientific research. In B. Nordlinger, C. Villani, & D. Rus (Eds.), *Healthcare and Artificial Intelligence* (pp. 215–221). Springer International Publishing.

Czarnitzki, D., Hussinger, K., & Leten, B. (2020). How valuable are patent blocking strategies? *Review of Industrial Organization*, *56*(3), 409–434.

de Bossey, C. (2005). *Report of the working group on internet governance*. http://www.wgig.org/docs/WGIGREPORT.pdf

Drake, W. J. (Ed.). (2016). *The working group on Internet governance: 10th anniversary reflections*. Association for Progressive Communications.

Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., Tuli, S., Smirnova, D., Singh, M., Jain, U., Pervaiz, H., Sehgal, B., Kaila, S. S., Misra, S., Aslanpour, M. S., Mehta, H., Stankovski, V., & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, *8*, 100118.

Gordon, W. J. (2002). Excuse and justification in the law of fair use: Transaction costs have always been part of the story. *Journal of the Copyright Society of the USA*, *50*, 149–198.

Guntrip, E. (2003). The common heritage of mankind: An adequate regime for managing the deep seabed? *Melbourne Journal of International Law*, *4*(2), 376–405.

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248.

Heger, D., & Zaby, A. K. (2017). Patent breadth as effective barrier to market entry. *Economics of Innovation and New Technology*, *27*(2), 174–188.

Hitaj, D., & Mancini, L. V. (2018). *Have you stolen my model? Evasion attacks against deep neural network watermarking techniques*. https://arxiv.org/abs/1809.00615

Hughes, J. (1988). The philosophy of intellectual property. *Georgetown Law Journal*, *77*(2), 287–366.

Internet Society. (2017). *Artificial intelligence and machine learning: Policy paper*. https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning with Applications in R (2nd ed.). Springer.

Johnson, P. (2008). Dedicating copyright to the public domain. *Modern Law Review*, *71*(4), 587–610.

Khong, D. W. K. (2019). Intellectual property: Economic justification. In A. Marciano & G. B. Ramello (Eds.), *Encyclopedia of Law and Economics* (pp. 1178–1185). Springer New York.

Klein, A. (2017, July 11). The cost of hard drives over time. Backblaze Blog. https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/

Land, M. K. (2018). The Marrakesh Treaty as bottom up lawmaking: Supporting local human rights action on IP policies. *UC Irvine Law Review*, *8*(3), 513–553.

Lehr, D., & Ohm, P. (2017). Playing with the data: What legal scholars should learn about machine learning. *UC Davis Law Review*, *51*(2), 653–718.

MacLeod, C. (1988). *Inventing the Industrial Revolution: The English patent system, 1660–1800*. Cambridge University Press. https://doi.org/10.1017/CBO9780511522673

Manap, N. A. (2012). Intellectual creation in database: A superfluous test? *Current Law Journal (Legal Network Series)*, *1*, lxxi–lxxi.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf

Mirzaee, S. (2017). The conceptual foundations of the common heritage of mankind. *Евразийский Юридический Журнал*, *10(113)*, 50–54.

Moor, J. (2006). The Dartmouth College artificial intelligence conference: The next fifty years. *AI Magazine*, *27*(4), 87–91.

Moore, A. D. (1997). A Lockean theory of intellectual property. *Hamline Law Review*, *21*(1), 65–108.

Owolabi, K. M. W. (2013). The principle of the common heritage of mankind. *Nnamdi Azikiwe University Journal of International Law and Jurisprudence*, *4*(3), 51–56.

Paton, C., & Kobayashi, S. (2019). An open science approach to artificial intelligence in healthcare: A contribution from the International Medical Informatics Association Open Source Working Group. *Yearbook of Medical Informatics*, *28*(1), 47–51.

Salameh, A. (2017). *Artificial intelligence as a commons: Opportunities and challenges for society*. Hochschule Hannover.

Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, *36*(4), 387–389.

Schelter, S., Biessmann, F., Januschowski, T., Salinas, D., Seufert, S., & Szarvas, G. (2018). On challenges in machine learning model management. *Bulletin of the Technical Committee on Data Engineering*, *41*(4), 5–15.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

Segura-Serrano, A. (2006). Internet regulation and the role of international law. *Max Planck Yearbook of United Nations Law Online*, *10*(1), 191–272.

Sullivan, R. J. (1989). England's "Age of invention": The acceleration of patents and patentable invention during the industrial revolution. *Explorations in Economic History*, *26*(4), 424–452. https://doi.org/10.1016/0014-4983(89)90017-X

Surden, H. (2014). Machine learning and law. Washington Law Review, 89, 87–115.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Tramer, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. *Proceedings of the 25th Usenix Security Symposium*, 601–618.

Tzimas, T. (2018). *Artificial intelligence as global commons and the 'international law supremacy' principle*, 83–88.

Vleugels, C. (2020). The Marrakesh Treaty. *Journal of Disability Policy Studies*. https://doi.org/10.1177/1044207320937419

Wiley, M., & Wiley, J. F. (2019). Advanced R statistical programming and data models: Analysis, machine learning, and visualization. Apress.