/

January 9, 2024

ISED Citizen Services Centre
Innovation, Science and Economic Development Canada
C.D. Howe Building
235 Queen Street
Ottawa, ON K1A 0H5
Canada

Re: Consultation on Copyright in the Age of Generative Artificial Intelligence

Dear Sir or Madam,

BSA | The Software Alliance ("BSA") welcomes this opportunity to provide comments to Innovation, Science and Economic Development Canada ("ISED") in response to its Consultation on Copyright in the Age of Generative Artificial Intelligence.[1]

BSA is the leading advocate for the global software industry before governments and in the international marketplace.[2] BSA members have a significant presence in Canada, employing thousands of workers across the country. BSA has long supported strong and effective copyright protection around the world. AI can benefit all industry sectors – including content creators – consistent with copyright law.

Given the critical importance of AI to long-term Canadian economic growth and innovation, it is important for the ISED to monitor legal developments to ensure that Canada's legal and policy environment continues to foster the development of AI. This critical advancement in technology does not need to be at the expense of artists and rightsholders.

AI encompasses a broad set of technologies and use cases. Generally speaking, AI systems learn from computational analysis of large amounts of data that can be used to train a model. The model will then make predictions when presented with a query that includes new data. Some generative AI models can take this a step further and generate new data, such as text or a picture or even new code. Regardless of what form of AI is being used, respect for copyrighted works – as well as personal data privacy and trade secret information – should be a cornerstone of responsible AI development.

We provide short answers to the ISED's three categories of questions below.

First, as regards ISED's questions about Text and Data Mining (Section 2.1), text and data mining, also referred to as computational analysis, typically involves turning data into tokens that are looking for statistical correlations with other tokenized data. Some of that data may be part of a copyrighted work, but the use normally has nothing to do with the expressive content of a work. A book may be used to learn language skills, which are then used for improved database management practices. In general, we believe that a reproduction of a lawfully accessed work for purposes of computational analysis should qualify as fair dealing under Canadian law. For greater certainty, we recommend that Canada adopt an express exception in the Canadian Copyright Act to cover copying of a lawfully accessed work for the purpose of "computational analysis."

Relatedly, we understand that there have been voluntary, industry conversations around automated tools to indicate that the rights-owner does not want a website used for training purposes, similar to the current "do not crawl" tools. We encourage further conversations to determine whether a consensus standard is possible.

Second, as regards ISED's questions about the authorship and ownership of works generated using AI systems (Section 2.2), the analytical touchstone should be whether the exercise of a sufficient degree of human skill and judgment was involved in the creation of an original work, regardless of what instrument or technology was used to aid its expression. This is consistent with established practice where software tools are used by artists to create works that are regularly deemed copyrightable, even though their creation was aided by technology. Generative AI should bolster human creativity, just as other software applications have long been an important tool of artists and storytellers. Generative AI is used, for example, in word processing for authors and photo enhancements for visual artists; it is used to create special effects in audio-visual works and arranging music for sound recordings. When generative AI is used to enhance human creativity, the resulting work should be protected by copyright so long as it meets the requirements for originality already well-established in Canadian jurisprudence.

Third, as regards ISED's questions about infringement and liability regarding AI (Section 2.3), the Canadian Copyright Act should be sufficiently flexible to determine whether copyright infringement liability exists. Stated differently, whether the output is infringing is typically not related to the technology used. Where an output is infringing, there should clearly be liability regardless of whether AI is used in producing the infringing copy. This is also consistent with established law which focuses on whether a work infringes and is less concerned with the particular method used to create the work.

I. Discussion

In Section 2.1, ISED asks about how data is used to train AI systems in the context of text and data mining – also known as computational analysis. As discussed in more detail below, there are different approaches to AI training, but this is the core of each: turning bits of data into tokens and finding statistical correlations from the tokens. The bits of data may be part of a licensed data set, they may be freely available online, or they may be part of a combination of different sets.

Training data may or may not be part of a copyrighted work, since there are no formalities to copyright protection and all expressive works are protected. The training set, however, is used to enable studying of non-expressive data it contains. This enables the system to learn, for instance, how language functions or about spatial relationships. While this form of data analysis should qualify as fair dealing under Canadian law and/or should be accommodated in a statutory

exemption, there may be situations in which the copyright owner or publisher wants to prevent the data use for training purposes. There are ongoing industry conversations about how to create metatags to create a signal on an otherwise freely available website, and we hope that the Government of Canada and ISED will encourage those conversations.

A. Artificial Intelligence in a Copyright Context

AI machine learning encompasses a vast array of technologies developed or deployed for use in a variety of different industries and applications. Machine learning depends upon computational analysis of training data to identify correlations, patterns, or other metadata in order to develop a model that can make predictions or recommendations based on future data inputs.[3] More recently, generative AI has emerged, allowing for the creation of new outputs in textual, visual, or aural formats.

1. Applications of AI

We offer below a few (widely recognized) examples of insights, predictions, and other outputs derived from computational analysis in the machine learning context:

- Automated flight management and air traffic control based on computational analysis of meteorological conditions, real-time fuel consumption, aircraft operational data, nearby air traffic conditions, airport congestion, and numerous other data elements.[4]

- Identification of chemical and cellular anomalies for early diagnosis, prevention, and treatment in the fields of oncology, autoimmune disorders, and Parkinsons and Alzheimers disease.[5]

- By integrating generative AI into security operations, organizations can effectively identify and address security anomalies, as well as detect and mitigate potential threats.

- Predictive climate modeling based on computational analysis of satellite data, weather station data, topographical information, and various IoT and sensor data.[6]

- Improved carbon tracking and mitigation based on computational analysis of transportation logs, meter readings, fuel purchase records, atmospheric pollution tracking, and visual monitoring of power plants and other facilities, and other data sources.[7]

- Computational analysis to map vulnerable seaside areas to produce cyclone risk maps and guide investment plans for cyclone shelters, schools, health facilities, and other infrastructure for disaster planning and survivability.[8]

- Predictive typing and other office productivity solutions (e.g., an “auto-complete” function that suggests the letters “…cerely yours” after the typist inputs the letters, “sin”),[9] or the creation of sound effects or special effects to assist creators and film producers in developing new artworks.[10]

While the use cases are diverse, the elements of training each are very similar and discussed below.

2. The AI Development Life Cycle

The AI development life cycle typically includes the following steps:

- Project Conception: Often, AI models are designed as general-purpose tools, such as a natural language processing model, that may later be customized for multiple particular tasks. Sometimes, an AI model is developed with a specific task in mind. In that situation, the AI development team must formulate the "problem" that a system is intended to address and map the structure and target variables that the system is intended to predict. For instance, a fitness app that analyzes a consumer's heart rate to monitor for irregularities that might predict whether that person is at risk of a stroke or heart disease (i.e., the target variable).

- Raw Data Identification: Often, AI models are trained on massive amounts of unlabeled data through self-supervised learning to serve as general purpose tools, such as a natural language processing model, that may later be customized for multiple particular tasks. In other circumstances where the model is being developed for a specific task, the AI development team must identify a relevant universe of "raw data" that will subsequently be transformed and structured. Data sources are as diverse as the potential applications of machine learning AI and may include everything from machine-to-machine data (e.g., satellite transmission data) and international trade statistics to published materials, blog posts, website comments, and chat room logs. "Raw data" is frequently "messy," requiring significant work to transform the data into a usable form, as outlined below.

- Preparing the Data Set: The AI development team must modify the "raw data" so it can be semantically read by the machine and used to train the model. During this process, the team will revise, clean, and normalize the data as necessary. Data is typically semantically and structurally transformed through "tokenization," which involves breaking down a piece of text or data into smaller units (or "tokens") for purposes of computational analysis. Tokenization is necessary to improve the reliability, quality, and suitability of the data for analysis, helping to address quality challenges, such as missing values, duplicates, outliers, and inconsistent formatting across the entire data set. Training data will often ultimately appear as a sequence of tokens that stretches across the entire AI corpus.

- Model definition: After input data has been suitably processed, the AI development team must establish the system's underlying architecture. This includes identifying the variables (i.e., "features") in the training data that the algorithm will evaluate as it looks for patterns and relationships as the basis of a rule for making future predictions. It also includes selecting the type of algorithmic model that will power the system (e.g., linear regression, logistic regression, deep neural network.)[11] Once the data is ready and the algorithm is selected, the team will train the system to produce a functional model that can make predictions about future data inputs.

- Model Validation, Testing, and Revision. After the model has been trained, it must be validated to determine if it is operating as intended and tested to demonstrate that the system's outputs fall within expected parameters and do not contain unexpected errors or unintended bias. Based on outcome of validation and testing, the model may need to be revised and refined to mitigate these risks.

B. The Use of Copyrighted Works to Train AI Models

As discussed above, computational analysis is typically applied to a large training data corpus that may comprise millions or billions of tokenized data elements. Depending on how the model is trained, data accessible over the Internet may be collected as part of the raw data set that is transformed into that corpus.

This raw data may include copyrighted works because a substantial portion of the internet is potentially subject to copyright protection, but importantly, however, not all of the material online is subject to copyright, in part because copyright protection does not extend to facts, ideas, or mathematical concepts.[12]

Computational analysis may involve two sets of reproductions that potentially implicate the Canadian Copyright Act: (1) reproductions necessary to create a corpus of "training data," and (2) temporary reproductions that are incidental to the computational process of training the AI model. In each case, the reproductions are "intermediate" in the sense that they are not visible or otherwise made available to the public. Instead, the reproductions are the necessary byproduct of a technical process that is aimed at identifying non-copyrightable information *about* the underlying corpus of works – i.e., the correlations and patterns that inform the creation of the AI model and enable it to make predictions based on future data inputs. Such intermediate, non-expressive reproductions do not compete with and have no impact on the economic interests that copyright is intended to protect.

Furthermore, computational analysis does not involve the consumption of any copyrighted works for their expressive content. Rather, such analysis involves mathematical calculations of probabilities, correlations, trends, and other patterns across the entire tokenized data set. Such analysis seeks to understand only the mathematical patterns (e.g., the relationships of specific tokens in relation to other tokens) distributed across the entire data set. These mathematical patterns are themselves not expressive content protected by copyright law.

1. Reproduction of a Work for Purposes of AI Computational Analysis Should Qualify as Fair Dealing under the Copyright Act

It is impossible to draw a generalized conclusion that *all* applications of AI involving the reproduction of copyrighted works will qualify as fair dealing. Nevertheless, in most cases, it is appropriate to treat as *fair dealing* the creation and use of AI training databases for the purposes of analyzing a large collection of individual works to identify patterns, correlations, and other metadata to develop an AI model that makes predictions about future data inputs.

Sections 29, 29.1 and 29.2 of the Copyright Act permits "fair dealing" as an exception to exclusive copyright. To qualify for this exception, the "dealing" in copyrighted content must be "fair" and for the purpose of research, private study, criticism, review, or news reporting. For the reasons stated below, the AI developer would ideally be able to rely on the fair dealing doctrine to construct an AI training database using materials to which he or she has lawful access.[13]

We offer an illustrative explanation of these points below. For example, an AI developer seeking to create a natural language processing model – such as an AI-driven predictive typing model – can rely on publicly available textual material (ranging from poetry or novels to anonymous website commentary) to create the training database. In such a scenario, the AI developer would not be reproducing this text for its expressive purpose. Rather, the reproductions would be made solely for the purpose of extracting unprotected information *about* the English or French language

– i.e., the correlations, patterns, and relationships among the 26 letters of the alphabet and the millions of English and French language words, as they appear in thousands of stock phrases, figures of speech, similes, metaphors, grammar patterns, and common linguistic formulations and expressions. Neither these letters, words, and phrases, nor the mathematical patterns among them across thousands or millions of writings, are copyright protectable subject matter.

Furthermore, the ultimate use of the computational analysis applied to the data set is often far removed from the intended use of the original content. Auto-complete functionality in predictive typing software comprises a new creation in the form of software code that is distinct both from the entire AI training corpus and from any single work within that corpus. Software code that can suggest the endings of commonly used phrases is many steps removed from any copyrighted works found in the underlying raw data. Finally, such functionality simply does not compete with any copyrighted works in any manner that copyright is intended to protect.

2. Canada Should Amend the Copyright Act to Adopt an Express Exception Covering Copying of a Lawfully Accessed Work for the Purpose of Computational Analysis

As discussed above, reproduction of a lawfully accessed work for purposes of computational analysis should qualify as fair dealing under Canadian law.

However, for greater certainty, BSA recommends that Canada adopt an express exception in the Canadian Copyright Act to cover copying of a lawfully accessed work for the purpose of "computational analysis." Consistent with international best practice, ISED should ensure that any such exception is technologically neutral, sufficiently flexible so as to be future-proof, and agnostic as to purpose and user. With that in mind, the exception should extend to:

- Commercial and non-commercial uses;
- All works and other subject matter; all copyright-relevant acts, including retention of data for purposes of verification and validation of results; and
- The provision of AI services (i.e., permitting service providers to perform computational analysis on behalf of end-users).

Importantly, such an exception would be consistent with Canada's international obligations. The TRIPs Agreement and Berne Convention require Member States to ensure that exceptions to copyright are confined to "certain special cases which do not conflict with the normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the rights holder."

An exception to facilitate computational analysis of lawfully accessed works is consistent with each of these requirements. It would meet the "certain special cases" requirement insofar as it is clearly and narrowly tailored to advance a significant public interest in the development of AI. It will not conflict with the normal exploitation of copyrighted works because reproductions made during a computational analysis process are not visible to humans and cannot substitute for or displace markets for the original works.

And, finally, a computational analysis exception will not prejudice the legitimate interests of a rights holder because the value derived from such processes is unrelated to the expressive content that copyright is intended to protect. Indeed, the very purpose of performing computational analysis is identify non-copyrightable information – such as the relationships and correlations between large corpuses of works – that can be used to train AI models. The value derived from

such activity lies not in the factual information that is gleaned from any single source, but rather in the discovery of entirely new forms of knowledge that emerge from the identification of patterns and correlations that exist between large bodies of disparate sets of data.

3. Any Other Reproductions Will Qualify as a Temporary Reproduction for a Technological Process under Section 30.71 of the Copyright Act

In addition to the reproductions that may be needed to create an AI training database, reproductions may also be made when the training data undergoes the computational analysis that occurs during the machine learning process. The Copyright Board of Canada interpreted this Section 30.71 as "intended to capture copies that happen automatically, or without the direct control of the user", and that are automatically deleted once the technological process is completed.

These fleeting reproductions are an inevitable byproduct of working with digital media, which must be loaded into the random access memory (RAM) of a computing device to be accessed, analyzed, manipulated or even deleted. In some circumstances, the creation of RAM copies may constitute a reproduction that is cognizable under the Canadian Copyright Act. However, the fleeting nature of the RAM copies involved in the machine learning process are unlikely to be considered "copies" that implicate the reproduction right.

4. Voluntary Measures and Consensus Solutions

There have been calls to create mechanisms that restrict access to works online for any purpose, including training an AI system. It should be remembered that automated "do not crawl" tools can prevent access to works online for any purpose, including training an AI system. We support further voluntary conversations between creators and AI developers and deployers to arrive at effective, consensus technical mechanisms.

C. The Copyrightability of Material Generated by an AI System

Section 2.2 asks about copyrightability. For both policy and legal reasons, copyright protection should be available for works generated by a human using AI. As long as sufficient human skill and judgment was involved in the creation of the work, the tool used in aiding the expression should not matter.

Inquiries regarding copyrightability will turn on a close examination of the degree of human skill and judgment. Artificial intelligence is increasingly being used as a tool in all categories of creative works. A decision to limit copyrightability when AI is used would significantly chill adoption of AI solutions. In turn, this would limit creative expression and impact Canadian AI development.

Works that emerge as outputs of AI systems and meet the human skill and judgment requirement should continue to be eligible for copyright protection. In most cases, AI systems will function as tools used by human authors to execute upon their creative vision. For instance, photographers will use AI-enabled tools to automate the tedious process of editing their images,[14] architects will use AI to augment their designs to enhance their energy efficiency,[15] and filmmakers will use AI to ensure that the movement of their animated characters appear more life-like.[16] In each of these cases, the creative contribution and exercise of skill and judgment by the human user makes it easy to conclude that the output would be copyrightable.

The use of generative AI should not change the analysis. Certainly, there will be cases at either extreme, in which it is either clear that there was a lack of or trivial amount of human involvement

or, on the other hand, generative AI was not part of the creative expression. We encourage ISED, in the first instance, to take a broad view of copyrightability. The Canadian Copyright Act provides courts with sufficient flexibility to closely examine these considerations when needed.

D. Potential Liability for Infringing Works Generated Using AI Systems

Section 2.3 relates to potentially infringing works generated using AI systems. BSA offers the following observations.

1. Current Copyright Law Protects Copyright Holders from Infringement, Including in Cases Arising from AI Generated Content

Copyright holders should have full and effective remedies when their rights are infringed. This principle applies equally to outputs generated using AI systems and outputs generated in other ways. Whenever such infringement is found, it is critical to fully compensate artists and creators for any damages caused.

In our view, existing copyright law should prove adequate to address questions of infringement. In most AI use cases, such as those described above, the output of an AI system will not implicate copyright at all. However, AI, like other technologies, could be used to create infringing material. Generally speaking, in those situations, infringement liability would be premised upon an assessment of the degree of similarity with the copyrighted work. Many users of generative AI will be small businesses experimenting with new tools, and enforcement decisions should take account of this. Liability would also arise in other appropriate cases, such as those involving derivative works.

Plaintiffs may also seek to bring infringement actions against providers of AI-related services. Here too, existing copyright doctrines should prove adequate in evaluating and apportioning liability. If a plaintiff demonstrates that a direct infringement has occurred, courts will evaluate whether the service provider should be deemed contributorily and/or vicariously liable for its users' conduct. In the context of contributory liability claims, the staple article of commerce doctrine will ensure that providers of AI services with substantial non-infringing uses are not liable for their users' infringing activity unless there is evidence that the service was made available with the object of promoting its use to infringe copyright.

2. AI Models Trained on Sufficiently Large Data Sets Are Less likely to Produce Infringing Outputs

While it is important not to conflate training data with the output of an AI system, it is worth noting that the more data available for training, the less likely the system will produce a copy or derivative of an input (in the absence of a user's intent to infringe). In any well-designed AI model trained on a sufficiently large data set, computational analysis should never (or only in the rarest of circumstances) produce outputs that are "substantially similar," let alone "virtually identical," to any specific copyrighted work. As Professor Matthew Sag of Emory University has explained,

> At the moment, memorization [i.e., the accidental replication of a particular work from a training set] is an edge case. For the most part, the link between the training data and the output of generative AI is attenuated by a process of decomposition, abstraction, and remix. Generally, pseudo-expression generated by large language models does not infringe copyright because these models "learn" latent features and associations

> within the training data, they do not memorize snippets of original expression from individual works.[17]

Additionally, some AI developers and deployers are already taking steps to engage with artists and creators on how to support their work in a changing digital environment and taking steps to limit misuse of an AI system, such as limiting what prompts can be used.

II. Conclusion

BSA thanks the ISED for the opportunity to provide comments in response to the ISED's consultation. As discussed above, the non-consumptive computational analysis of copyrightable content in AI training data sets should not be treated as an infringement of copyright under Canadian law. While we believe that the computational analysis of copyrightable content in AI training data sets is not a copyright infringement under Canadian law, to the extent that Canadian law is unsettled in this regard, we urge the development of a specific statutory exemption to the use of copyrighted material within AI training sets. At the same time, BSA also supports multi-stakeholder efforts relating to AI training processes as well as efforts to minimize the risk of infringement. And to the extent that infringement occurs, BSA strongly supports fully protecting content creators.

Thank you for the opportunity to share our views on these important issues.

Sincerely,

Joseph Whitlock

Joseph Whitlock
Director, Policy
BSA | The Software Alliance

---

[1] Innovation, Science and Economic Development Canada, Consultation paper: Consultation on Copyright in the Age of Generative Artificial Intelligence (2023), at: https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/consultation-paper-consultation-copyright-age-generative-artificial-intelligence

[2] BSA's members include: Adobe, Alteryx, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, Hubspot, IBM, Informatica, Kyndryl, MathWorks, Microsoft, Okta, Oracle, Palo Alto Networks, Prokon, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, TriNet, Twilio, Workday, Zendesk, and Zoom Video Communications, Inc.

[3] *See* BSA | The Software Alliance, *Confronting Bias: BSA's Framework to Build Trust in AI* (2022), at: https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai; *See also,* National Institute of Standards and Technology, *NIST Risk Management Framework* (Jan. 2023), at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (describing an AI system "as an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments.")

[4] *See e.g.,* M. Durgut, *Artificial Intelligence and Air Traffic Control*, Aviationfile.com website (Jan. 2023), at: https://www.aviationfile.com/artificial-intelligence-and-air-traffic-control/#:~:text=One%20of%20the%20primary%20applications%20is%20to%20help,make%20informed%20decisions%20on%20routing%20and%20scheduling%20flights; Degas et al., *A Survey on Artificial Intelligence and explainable AI in Air Traffic Management*, 12 Applied Sciences 1295 (2022), at: https://www.mdpi.com/2076-3417/12/3/1295; Hanneke Weitering, *How Artificial Intelligence is Transforming Aviation*, Futureflight.aero website

(2023), https://www.futureflight.aero/news-article/2023-07-13/beyond-automation-how-artificial-intelligence-transforming-aviation

[5] *See e.g.,* Hunter et al., *The Role of Artificial Intelligence in Early Cancer Diagnosis*, 14(6) Cancers 1524 (2022), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8946688/; Stafford et al., *A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases*, 3 NPJ - Digital Medicine 30 (2020), at: https://www.nature.com/articles/s41746-020-0229-3; Diogo et al., *Early diagnosis of Alzheimer's disease using machine learning*, 14 Alzheimers Research and Theory 107 (2022), at: https://alzres.biomedcentral.com/articles/10.1186/s13195-022-01047-y; Ahsan et al., *Machine-Learning-Based Disease Diagnosis: A Comprehensive Review*, 10(3) Healthcare Basel 541 (2022), at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950225/; Luchini et at., *Artificial Intelligence in Oncology*, 126 British J. of Cancer 1 (2022), at: https://www.nature.com/articles/s41416-021-01633-1; Yumar et al., *Artificial intelligence in Disease Diagnosis*, 14(7) J. Ambient Intell Humaniz Comput. 8459 (2023), at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8754556/

[6] Schneider et al., *Harnessing AI and computing to advance climate modelling and prediction*, 13 Nature Climate Change 887 (2023), at: https://www.nature.com/articles/s41558-023-01769-3; World Economic Forum, *The role of machine learning in helping to save the planet* (2021), at: https://www.weforum.org/agenda/2021/08/how-is-machine-learning-helping-us-to-create-more-sophisticated-climate-change-models/; Kaak et al., *Aligning artificial intelligence with climate change mitigation*, 12 Nature Climate Change 518 (2022), at https://www.nature.com/articles/s41558-022-01377-7; Xin et al., *Artificial Intelligence for Climate Change Risk Prediction, Adaptation, & Mitigation*, Ecological Processes (2021), at: https://www.springeropen.com/collections/AICC; Chantry et al., *Opportunities and challenges for machine learning in weather and climate modelling*, 379 Phil. Trans. R. Soc. 83 (2020), at: https://doi.org/10.1098/rsta.2020.0083 (2020).

[7] *See e.g.,* Global Data Alliance, *Cross-Border Data Transfers & Environmental Sustainability* (2023) (internal citations omitted), at: https://globaldataalliance.org/wp-content/uploads/2023/04/04192023gdacbdtsustainability.pdf

[8] *See id.*

[9] *See e.g.,* S. Ashraf, *Desmystifying Autocomplete,* Towards Data Science (2020), at: https://towardsdatascience.com/index-48563e4c1572; A. Wickramarachchi, *Machine Learning: Word Embedding and Predicting,* Towards Data Science (2020), at: https://towardsdatascience.com/machine-learning-word-embedding-and-predicting-e603254e4d7b

[10] *See e.g.,* D. Nelson, *AI Researchers Design Program To Generate Sound Effects For Movies and Other Media*, Unite AI Website (2022), https://www.unite.ai/ai-researchers-design-program-to-generate-sound-effects-for-movies-and-other-media/

[11] *See* BSA AI Bias Framework, *infra* note 3.

[12] See "Article 9.2 of the TRIPS Agreement provides that copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such. In other words, copyright protection does not cover any information or ideas contained in a work; it only protects the original way in which such information or ideas have been expressed. Thus everyone is free to use the information contained in a work, including for the purpose of creating new works. For example, the idea behind a detective novel is not protected as such, but an unauthorized reproduction of that novel that is an expression of the idea is prohibited. This principle, commonly referred to as the 'idea/expression dichotomy', has always been present in copyright doctrine, although it had not been explicitly set out in the provisions of the Berne Convention. Article 9.2 of the TRIPS Agreement is therefore the first explicit confirmation of the principle in multilateral IP law." WTO, Overview of the Agreement on Trade-Related Intellectual Property Rights (2023), at: https://www.wto.org/english/tratop_e/trips_e/ta_docs_e/modules2_e.pdf

[13] Applying the fair dealing factors to AI training:

1. The purpose and character of the dealing is highly tangential insomuch as the computational analysis is unrelated to the expressive content of any copyrighted work;
2. The amount of the dealing is *de minimis* insomuch as it involves computational analysis of non-copyrightable mathematical patterns and correlations in a larger corpus;
3. The nature of the work and the effect of the dealing on it are *de minimis* or immaterial, given the non-expressive use of non-copyrightable content across the corpus;
4. There are no feasible alternatives given that a substantial portion – perhaps the vast majority – of the diffuse body of content found across the Internet is potentially subject to copyright, yet there is often no reasonable means to identify authors, negotiate terms of remuneration, or so forth. Nevertheless, for those authors who wish to affirmatively remove their works from AI training sets, the discussions referenced in Section B.4 below should continue.

[14] https://theblog.adobe.com/adobes-general-counsel-makes-the-case-for-ai/.

[15] https://www.autodesk.com/redshift/machine-learning-in-architecture/.

[16] https://theblog.adobe.com/state-of-ai-in-animation/.

[17] *Matthew Sag, Copyright Safety for Generative AI*, 61 Hous. L. Rev. 101 (Aug. 10, 2023), at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593