# Written submissions on aspects of the Copyright Amendment Bill [B13B-2017]

*Submissions requested on aspects of the Copyright Amendment Bill [B13B-2017] and the Performers' Protection Amendment Bill [B24B-2016] as requested by the Standing Committee on Finance, Economic Opportunities and Tourism (the Committee) of the Western Cape Provincial Parliament will consider the Copyright Amendment Bill [B 13D–2017]*

To:  The Standing Committee on Finance, Economic Opportunities and Tourism (the Committee) of the Western Cape Provincial Parliament

**Via Email: zadams@wcpp.gov.za**

For further information please contact:
Dr Andrew Rens
Senior Research Fellow, Research ICT Africa
arens@researchictafrica.net

Samantha Msipa,
Research Fellow, Research ICT Africa
smsipa@researchictafrica.net

Hanani Hlomani
Research Fellow, Research ICT Africa
hhlomani@researchictafrica.net

**About Research ICT Africa**

Research ICT Africa (RIA) is an African  digital policy, regulation and governance think tank based in Cape Town. It conducts research on digital economy and society that facilitates evidence-based and informed policymaking for improved access, use and application of information and communication technologies (ICTs) for social development and economic growth. RIA seeks to support policy and governance that will reduce the uneven distribution of opportunities and harms associated with the intensifying processes of digitalisation and datafication. Through active participation in international, continental and national processes of digital governance RIA provides evidence-based alternative strategies in the areas of intellectual property, internet governance, data governance, cybersecurity, algorithmic governance, innovation that will produce more equitable and just outcomes. Understanding the  digital economy, and how it can be the basis for innovation and entrepreneurship that serves the needs and challenges of marginalised communities – including women, youth, children, the elderly, and people in rural areas, for example – is an integral part of RIA's work.

**Acknowledgements and declaration**

This submission was drafted and prepared by Dr. Andrew Rens, Ms Samantha Msipa and Mr Hanani Hlomani with valuable inputs, contributions and revisions from other RIA staff members. Any errors or omissions remain the authors' own.

A similar submission was made to the Select Committee on Trade and Industry, Economic Development, Small Business Development, Tourism, Employment of the Labour National Council of Provinces. However we have had an opportunity to elaborate certain important points in this submission.

The research informing this submission is made possible by a grant from Canada's International Development Research Centre (IDRC) to develop AI policy capacity in Africa

**Availability**

We confirm our availability  to make oral representations based on this submission.

## Overview

Research ICT Africa (RIA) welcomes this opportunity to submit written comments on proposed changes to the Copyright Amendment Bill [B13B-2017].

We make this public interest submission to help ensure that the intensifying global processes of AI and datafication can be harnessed to contribute to the growth of innovation and economic opportunity in the Western Cape to reduce poverty, unemployment and inequality and ensuring that the benefits of advanced technologies and opportunities to innovate, improve lives and livelihoods by deploying them are more evenly spread.

## Innovation Friendly Copyright

In this submission RIA focuses on issues raised by the rapid technological advancements in computerised techniques that are popularly called Artificial Intelligence or AI. RIA's past submissions on the Bill have emphasised that an open ended, flexible, future proof provision is required to permit innovative uses that cannot be foreseen. A flexible provision remains essential to innovation friendly copyright not only in respect of AI but for multiple aspects of information and technology.

In addition to the specific comments made below the team at RIA are ready to offer our expertise on any aspect of the bills that affects the digital economy and information and communications technologies.

# Specific comments

## Artificial Intelligence, Data and Copyright

Artificial Intelligence is an emerging general purpose technology. AI has many beneficial uses, for example AI technologies have proven to be powerful research tools. One example is that an AI model can accurately predict the structure of proteins which will accelerate medical and biological research. Data sets are used to train algorithms that in turn produce outputs, these may include anything from weather predictions, to identifications of location of cancer to movie recommendations. Some systems produce outputs that resemble human creativity including text and images.

AI also raises copyright questions. Some of these questions are about copyright and the use of  about the inputs that are used to create an AI model but others are about the  outputs generated by some AI systems. The questions about inputs are analysed under flexibilities for informational analysis, and the questions about outputs are analysed in the section on authorship and AI outputs.

What is AI? The [Alan Turing Institute](#) describes AI as "the design and study of machines that can perform tasks that would previously have required human (or biological) brainpower to accomplish" (*Data Science and AI Glossary*, 2023). The current generation of AI involves the production of algorithms through iterative processes based on data sets. The United States National Institute of Standards and Technology [NIST](#) defines algorithms as "a clearly specified mathematical process for computation, a set of rules that, if followed, will provide a prescribed result". The [Alan Turing Institute](#) defines an algorithm as a sequence of rules that a computer uses to complete a task (*Data Science and AI Glossary*, 2023). Because algorithms are mathematical rather than creative expressions they have

historically been excluded from copyright.[1]

There are several different techniques for producing algorithms using training data rather than coding the algorithm directly. These include machine learning, deep learning and reinforcement learning. Machine learning is the most popular technique at present. Machine learning uses a training data set that is filtered through a blank artifical neural net (layers of nodes ) in an iterative process to produce algorithms that can produce a perform a particular task. Machine learning is used in various ways, including computer vision - computer systems that process data from digital cameras to identify and track types of entity and even individual humans, vehicles or animals. Anothe genre of machine learning system is large language learning models that can process vast amounts of text - or in some cases to associate text with visual information.

Machine Learning requires vast data sets. This data is often not subject to copyright. For example factual weather data is not copyright, nor should it be because copyright does not extend to facts so that everyone is able to use the same facts. But for some AI systems  the 'input data' is text, images, video or computer code which are often copyright works. Machine learning systems often require the reproduction and adaptation of vast numbers of copyright works, often hundreds of thousands of works. However these works are used by the system rather than a human, and for the novel purpose of machine learning. Even when human intervention is required, for example to label images, it as part of the preparation and not for a human use such as entertainment or education.

It is impossible in practical terms for AI researchers to obtain find out who all the affected copyright holders of a set of a 1000 images are, and then contact them

---

1 This is recogised in the European Union Software Directive, Recital 11: "11For the avoidance of doubt, it has to be made clear that only the expression of a computer program is protected and that ideas and principles which underlie any element of a program, including those which underlie its interfaces, are not protected by copyright under this Directive. In accordance with this principle of copyright, to the extent that logic, algorithms and programming languages comprise ideas and principles, those ideas and principles are not protected under this Directive." (Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the Legal Protection of Computer Programs, 2009)

to obtain their consent for this entirely novel use. But the size of datasets used for many machine learning systems is very much greater consisting of hundreds of thousands or millions of items.

## Flexibilities for informational analysis

While its not clear in the 1978 Act whetehr the use of copyright works to train AI does infringe copyright its also not clear that it does not infringe copyright. In similar situations researchers and entrepreneurs with fewer resources tend to avoid the risk of being sued for copyright infringement while technology companies with more resources are willing to take that risk. If parliament wants South African researchers and entrepreneurs to engage in AI research then the 1978 Copyright Act should be clearly amended to permit most uses of copyright works for AI research.

AI research is just one category of what it termed informational analysis, a broad term, that also includes text mining and data mining. Research in the age of the Fourth Industrial Revolution and Artificial Intelligence can make progress only if computational methods such as text and data mining and AI can be carried out with regard to copyright works without fear of legal liability. Informational analysis of copyright work requires clear legal authorisation by copyright legislation. South Africa needs a flexible provision in copyright that enables informational analysis including use of copyright works in training algorithms. Section 12A in the bill is the only section of the bill that explicitly protects uses for a research purpose. Explicitly permitting research is crucial to encourage the kinds of research essential to artificial intelligence - a crucial technology of the 4th Industrial Revolution.

In its current form S12A in the Copyright Amendment Bill represents not just a flexibility that permits informational analysis but sets an appropriate balance between the exclusive rights given by copyright and the flexibility needed to enable rapidly developing technologies. Innovative uses cannot by definition be predicted. Its helpful for copyright law to include detailed lists of activities that are permitted by copyright law but however long or detailed the lists innovative uses will never all be included since they are inherently unpredictable at the time

the legislation is drafted. A good example is use of copyright works for training an algorithm. The need for a legal provision for this use was not known at the time the Copyright Amendment Bill was first introduced in parliament. Therefore not only is a provision needed that permits informational analysis as its currently being practised but that is sufficiently flexible to deal technological development. The European Union has attempted to develop a stand alone exception, however this exception is regarded as unworkable (Geiger, 2021). There is  is widespread agreement that most machine learning in the United States is authorised by 17 U.S.C. §107 but that the provision is sufficiently flexible to prohibit exploitive practises (*Public Views on Artificial Intelligence and Intellectual Property*, 2020).

Section 12A of the CAB does permit informational analysis but only if it is in accord with the multi-factor balancing inquiry set out in S12A(b). S12A authorises use of copyright works for informational analysis including machine learning, provided that the machine learning in question is considered fair in view of the balancing exercise required by S12A(2). Using only the works of a particular artist in order to imitate her style would not be permitted.[2] The internal balance in section 12A enables appropriate outcomes even when applied to novel technologies.

S12A is the most appropriate way to deal with informational analysis and machine learning, since these are general purpose techniques that are constantly evolving and developing - thus an open ended approach that balances the rights and interests involved is the best way to keep pace with technological changes. In a S12A balancing inquiry some machine learning uses are not fair, for example,

While our legal analysis of Section 12A as it currently stands is that it permits uses for AI research it will encourage AI research to explicitly permit informational analsysis, which includes AI research.

---

2 That is the conclusion reached in the Opinion of the Office of Legal Counsel and Legislative Affairs of the Israel Ministry of Justice on "whether ML enterprises can make unauthorized use of copyrighted materials to train Artificial Intelligence (AI) system" 18 December 2022 available at https://www.gov.il/BlobFolder/legalinfo/machine-learning/he/machine-learning.pdf. The original opinion is in Hebrew but there is an English summary on page 3.

## Recommendation:

The following purpose be added to those explicitly authorised by the fair use clause:

In Section 12 A (a) (i) after "research" insert ",including informational analysis,"

The clause would then read:

"(i) Research, including informational analysis, private study or personal use…"

### *Authorship and AI outputs*

When the Bill was first drafted AI technology was not able to produce outputs that closely mimicked human work such as text and images. However in the meantime some AI techniques have been developed that can produce text and images that resemble those produced by humans. For example Chat GPT can produce text in response to a prompt. Another example is Stability AI that uses stable diffusion; a "technique the AI system uses to generate output images that are similar to those found in its training data".[3] These outputs are produced by models trained using very large collections of images or texts. Unlike preceding technologies the production image or text output is not directed by a human other than by the prompt supplied. Instead it is assembled by the model based on associations it has formed during training. Neither the user nor the engineers that trained the model can predict what an output will be like. Should the resulting output be treated as a copyright work?

There are a number of reasons why machine learning outputs should not be treated as copyright works. Copyright is intended to act as an incentive to human creators. Even when copyright vests in a juristic person or the State it does so only when a human creates a copyright work. The relationship of the author to the juristic person that owns the copyright as an employee or someone

---

3 Some outputs are so similar that some artists have alleged infringement. For example in Andersen v. Stability AI the plaintiff alleges that the AI generated artwork results from copies of the images used in the training dataset. However these kinds of disputes can be resolved under existing copyright law.

commissioned to produce the work is the basis for ownership vesting in the juristic person. The producers of the AI systems do not need the incentive of copyright in every one of the potential thousands or hundreds of thousands of outputs from the AI system. Millions of AI outputs are already being produced by these systems without any clear copyright incentive. Many producers make money from their AI models using a freemium model, users can have a limited number of users free, but pay for more. That model does not require copyright to exist. Granting producers of AI models copyright over millions of outputs based on the work of human artists would be to give them a windfall they don't need.

Copyright has historically extended only to human activities, and efforts to extend copyright to outputs from non-humans have been rejected in several jurisdictions. The United States Copyright Office refused to register an image generated by an AI, finding that "human authorship is a prerequisite to copyright protection" (*Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071)*, 2022). Extending copyright to AI outputs would be a far reaching extension of copyright. If AI outputs are awarded copyright status now it will become close to impossible to undo since the grant of exclusive rights will give rise to an entrenched interest which will resist withdrawal of copyright. Extending copyright to AI outputs shouldn't be permitted until the costs and benefits are clear,

If AI outputs are treated as copyright works the result will be a very unequal competition between human creators and AI models that are able to produce outputs at a much greater speed and scale than is possible for any human authors. The result will be devaluing of human creativity.

Parliament has an opportunity to deal with this issue in the Copyright Act. If it does not do so then it leaves courts without the guidance of democratically elected representatives. If copyrihg law is left unclear then copyright may be extended to AI works through incremental changes in applying the law, first byb extending copyright to outputs thave have some human authorship and gradually shifting to works where there is insignificant human input. Parliament should clarify that only human creativeity is under copyright.

Attempting to define AI or AI outputs in order to explicitly exclude them from copyright creates a risk that as technology changes that the definitions will become outdated. Instead the CAB should makes clear that copyright attaches only to human creativity.

## Recommendations

Insert an sub-section in S2A that states:

2A (3) (a) Copyright extends only to the products of a natural person's skill, effort and creativity.
(b) In any dispute concerning copyright or authors rights the author or their successor shall bear the onus of proving that a disputed work or aspect of a work is the product of the skill, effort and creativity of a natural person or persons.
(c) Prompting, selection, filtering and sequencing the products of an automated system or systems is not sufficient to meet the requirements of 2A(3)(a).

Amend the definition of author in Section 1 by adding 'natural' in front of 'person' in each of (a), (b), (c), (d), (h) and (i).

### Algorithms and Copyright

As explained above algorithms have historically not been subject to copyright because they don't express an individual's creativity. As mathematical functions they can be incorporated in software in multiple ways. However this is not explicitly set out in South Africa's copyright legislation. This could lead to assertions of copyright over algorithms. That would inhibit innovation in which multiple innovators each implement an algorithm in a different way. There is an additional reason to clearly exclude algorithms from copyright. AI algorithms may be biased or produce other harms that require a mandatory intervene including 'monitoring, audit and validation input" (Gaffley et al., 2022). Permitting copyright claims over algorithms, either delibrately or through lack of clarity will inhibit responses to algorithmic bias by enabling the producers of algorithms to claim exclusive control.

The Copyright Amendment Bill already contains a provision to clarify the scope of copyright subject matter. Section 2A(1) reads: "Copyright protection subsists in expressions and not—
(a) in ideas, procedures, methods of operation or mathematical concepts".

Algorithms implicitly fit this description, however it would be simple and advantageous to include algorithms explicitly.

## Recommendation

Insert the word 'algorithm' in S2A(1) sub-section after 'operation' and before 'or mathematical concepts so that the subsection states:

"Copyright protection subsists in expressions and not—
(a) in ideas, procedures, methods of operation, algorithms or mathematical concepts"

*References:*

*Data science and AI glossary*. (2023). The Alan Turing Institute. https://www.turing.ac.uk/news/data-science-and-ai-glossary
Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, EP, CONSIL, 111 OJ L (2009). http://data.europa.eu/eli/dir/2009/24/oj/eng
Gaffley, M., Adams, R., & Shyllon, O. (2022). *Artificial Intelligence. African Insight. A Research Summary of the Ethical and Human Rights Implications of AI in Africa*. HSRC & Meta AI.
Geiger, C. (2021). *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive* (No. 66; PIJIP/TLS Research Paper Series 2021). Program on Information Justice and Intellectual Property. https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1069&context=research
*Public Views on Artificial Intelligence and Intellectual Property* (p. 45). (2020). https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-05.pdf

*Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071)*. (2022). Copyright Review Board, United States Copyright Office. https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf