LOYOLA UNIVERSITY CHICAGO SCHOOL OF LAW
Philip H. Corboy Law Center
25 E. Pearson Street | Chicago, Illinois 60611

Matthew Sag
Professor of Law
Phone: 312-915-7223
Fax: 312-915-7201
Email: msag@luc.edu

June 28, 2017

Honorable Joanmariae Louise Fubbs
Chairperson, Portfolio Committee on Trade and Industry
jfubbs@parliament.gov.za

Attention Mr A Hermans
PO Box 15, Parliament,
Cape Town, 8000.
ahermans@parliament.gov.za

Comments on the Copyright Amendment Bill [B13-2017]

Dear Ms Fubbs, MP:

This letter offers comments in response to the recently released Copyright Amendment Bill (B13-2017). My comments are offered in my personal capacity as a scholar and teacher of copyright law. I am a Professor of Law at Loyola University Chicago School of Law. I am a graduated of the Australian National University and a former associate to Justice Paul Finn of the Federal Court of Australia. I have practiced law in Australia, the United Kingdom and the United States. As an attorney and as a law professor I have advised clients on the application of fair use and published several law review articles dealing with the fair use doctrine in the United States. I am currently writing a book on the fair use doctrine in the United States and my research on the copyright implications of text mining and similar processes has been highly influential in recent United States copyright cases, including the *Authors Guild v. HathiTrust* and *Authors Guild v. Google* decisions discussed below.

I commend you on the progress you have made with the Copyright Amendment Bill and I write to urge you to consider further refinements that will significantly benefit the technology sector in your country without prejudicing the interests of the authors and owners of copyrighted works.

South Africans would benefit greatly from a provision that makes is clear that the technical processes at the heart of **machine learning**, **cloud computing**, **text mining**, **plagiarism detection**, **automated detection of copyright infringement** and

constructing search engine indexes do not violate copyright law. Under current South African law, all these activities are arguably unlawful because, although they do not communicate the copyright owner's original expression to the public in any way, they all rely on copying as an intermediate technical step. Thus, it is a matter of concern that the current copyright revision bill, B13-2017) (Copyright), appears to make no provision whatsoever for important large-scale applications of new digital technology that will important to research and development in both the non-profit and for-profit sectors. As a result, the terms of the proposed revisions would leave South Africans at **a permanent and crippling disadvantage** compared to residents of the United States, Israel, South Korea and other countries that have adopted, or are considering adopting a so-called "fair use" approach to copyright limitations and exceptions, as well as other countries that may take a narrower approach to immunizing information technology innovators from liability.

The need for further reform

In the era of the printing press that gave birth to modern copyright law, making a copy of a work was a distinct activity with a well-settled meaning. Every new instantiation of a work in a physical copy made that book available to a new consumer or a new group of consumers. The exclusive right to make and sell copies made sense in this context; it created an economic system whereby copyright owners had a clear and distinct tolling point for remuneration. However, in the digital age, a large and growing number of technologies rely on intermediate copies that have no independent economic significance and do not communicate the author's original expression to the public. The copying at the heart of these technologies is "non-expressive use" (sometimes also referred to "non-consumptive use"). Specifically, the term "non-expressive use" refers to the making of intermediate copies of copyrighted works as part of an analytical process that does not communicate the work's original expression to any human end user.[1]

Technologies such as machine learning, text mining, plagiarism detection, and Internet search engines rely on the ability to "read" thousands (sometimes millions) of works and abstract metadata from those works. The metadata itself is fundamentally different to the original expression contained within the primary works: it is fact not expression; it is not similar to the primary works; and it is not authored by the creators of the primary works.[2] Every copyright system around the world recognizes that such metadata is unprotectable facts and ideas *about* the works in question, as opposed to copyrightable expression. However, there are still many countries where copyright law effectively makes it illegal to generate this kind of metadata using a computer for the simple reason that when computers "read" they also copy.

These technologies have enormous potential to advance human progress without prejudicing the interests of authors or copyright owners. To illustrate, researchers using text

[1] See Matthew Sag, Copyright and Copy-Reliant Technology, 103 Northwestern University Law Review 1607–1682 (2009); Matthew Sag, Orphan Works as Grist for the Data Mill, 27 Berkeley Technology Law Journal 1503 – 1550 (2012); Matthew Jockers, Matthew Sag & Jason Schultz, Digital Archives: Don't Let Copyright Block Data Mining, 490 Nature 29-30 (October 4, 2012).

[2] *Id.* See also Matthew Jockers, Matthew Sag & Jason Schultz, Brief of Digital Humanities and Law Scholars in Support of Defendants-Appellees and Affirmance in *Authors Guild v. Google* (13-4829) (July 10, 2014).

mining do not copy works to read them individually; they copy them by the thousands to generate abstract metadata about entire collections of works. This is true whether the works are blog posts, library books, or webpages. As the Second Circuit explained in the recent case of *Authors Guild v. Google, Inc.*,

> Google's "ngrams" research tool draws on the Google Library Project corpus to furnish statistical information to Internet users about the frequency of word and phrase usage over centuries. This tool permits users to discern fluctuations of interest in a particular subject over time and space by showing increases and decreases in the frequency of reference and usage in different periods and different linguistic regions. It also allows researchers to comb over the tens of millions of books Google has scanned in order to examine word frequencies, syntactic patterns, and thematic markers and to derive information on how nomenclature, linguistic usage, and literary style have changed over time.[3]

Beyond identify patterns in vast libraries of literature, text mining has enabled researchers to identify new treatments for diseases by observing correlations in scientific papers that were not apparent to any single researcher. Text mining is vital for machine learning, automatic translation, and developing the language models the power dictation software.

Allowing non-expressive use is consistent with the goals copyright. Copyright law is not an end unto itself, it was established to promote human progress by motivating and rewarding the creation of new and original expression. Thus, the law distinguishes between facts and idea (unprotectable) and expression (protectable). Thus, a work is only regarded as having been copied when a substantial part of its original expression has been reproduced. If the purpose of copyright is to protect original expression, it stands to reason that non-expressive use should not infringe copyright.

Options for reform

There is no need for South Africa to burden its dynamic and entrepreneurial technology sector with copyright laws that slow down innovation without doing anything to promote the goals of copyright law. The most obvious pathways to reform would be to expand the current closed list of fair dealing purposes in the current law, or to draft a specific statutory exception relating to non-expressive use.

(1) An Open Fair Dealing Provision

The current fair dealing provisions may be too confined to allow for socially beneficial non-expressive uses like those discussed above. The non-expressive use of copyrighted works does not fit neatly into the categories of fair dealing. The production and use of metadata is not exactly criticism and it is not always research or scholarship. Computational analysis may be vital to news reporting on current events, but as an intermediate step it does not clearly amount to news reporting as such.

[3] See Authors Guild v. Google, Inc., 804 F.3d 202, 209 (2d Cir. 2015) (internal citations and quotations omitted)

South African copyright law could make suitable provision for non-expressive use by simply by adding the words "**such as**" to the introductory language in the new proposed general exception in Section 12 of the Act, so that it reads:

> "In addition to uses specifically authorised, a fair dealing or use with respect to a work or performance for purposes such as the following does not infringe copyright in that work: . . ."

An open fair dealing provision would many other benefits as well. In the United States, the fair use doctrine is not confined to the patterns of fair use that could be foreseen when our copyright law was drafted in the late 1960s and early 1970s. This additional flexibility has allowed United States courts—and more recently, courts in Israel, Korea and other nations—to recognize the fairness of a variety of uses that do not substitute for or interfere with the copyright owner's ability to communicate her work to the public and are for purposes broadly compatible with the goals of copyright. As recent technologies such as Internet search engines and text mining illustrate, the fair purposes of tomorrow are not necessarily within our power to imagine today.

In the United States, courts have recognized that it is fair use for both commercial and non-commercial actors to make intermediate copies of copyrighted works as part of an analytical process that does not communicate the work's original expression to any human end user—the relevant decisions are discussed in the next section. Courts in South Africa would likely reach the same conclusion, but they may of course tailor their rulings to specific local conditions.

(2) Specific Exception

As an alternative to an open fair dealing provision, or as a clarification thereof, South Africa's copyright could be amended with a provision along the following lines:

> Any act of reproduction, adaptation or transmission of a work that is merely an intermediate technological step in the production of metadata that does not itself embody and is not capable of communicating a copyright owner's original expression, does not infringe the exclusive rights of the author or the copyright owner of that work under this Act.

United States Case Law on Fair Use and Non-Expressive Use

In the United States, the fair use doctrine is not confined to the patterns of fair use that could be foreseen when our copyright law was drafted in the late 1960s and early 1970s. This additional flexibility has allowed United States courts to recognize the fairness of a variety of uses that do not substitute for or interfere with the copyright owner's ability to communicate her work to the public and are for purposes broadly compatible with the goals of copyright. The fair use doctrine in the United States allows both commercial and non-commercial actors to make intermediate copies of copyrighted works as part of an analytical process that does not communicate the work's original expression to any human end user.

The following examples may help to illustrate:

(1) Software reverse engineering

Sony Comp. Entertainment, Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000) is one of many cases holding that reverse engineering computer software is fair use. In this case, the court of appeals for the Ninth Circuit held that intermediate copying of computer software was fair use because the copying was necessary to gain access to the functional (and thus uncopyrightable) elements of the software.[4] As the Ninth Circuit explained, "… the Copyright Act protects expression only, not ideas or the functional aspects of a software program . . . . Thus, the fair use doctrine preserves public access to the ideas and functional elements embedded in copyrighted computer software programs."[5]

(2) Plagiarism detection

In *A.V. ex rel. Vanderhye v. iParadigms, LLC,* 562 F.3d 630, 640 (4th Cir. 2009) the court of appeals for the Fourth Circuit held that copying student term papers for use as reference works in an automated plagiarism system was fair use. The court's finding of fair use was based in large part on the fact that "[defendant's] use of these works was completely unrelated to expressive content and was instead aimed at detecting and discouraging plagiarism."[6]

(3) Library digitization to enable text mining and research

In *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) the Second Circuit court of appeals held that the digitalization of copyrighted library books to permit full-text searching within those books was fair use. The court described the creation of a full-text searchable database of printed library books as "a quintessentially transformative use" in the sense that it lead to the creation of new insights, understandings, meaning and message without displacing the copyright owner's original expression. As the court explained, "the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn. Indeed, we can discern little or no resemblance between the original text and the results of the [HathiTrust Digital Library] full-text search."[7]

(4) Media monitoring

In *Fox News Network, LLC v. TVEyes, Inc.*, 43 F. Supp. 3d 379, 393 (S.D.N.Y. 2014) the district court for the Southern District of New York held that the core functionality of an automated media monitoring service was fair use. The TVEyes media-monitoring service recorded all content broadcast on more than 1,400 television and radio stations and

---

[4] Sony Comp. Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 607 (9th Cir. 2000)

[5] *Id.* at 603.

[6] A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 640 (4th Cir. 2009)

[7] Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 97 (2d Cir. 2014)

transformed that content into a searchable database. The court recognized that the media-monitoring service's central information location function enabling the tracking of when, where, and how certain words of interest were used in the media was fair use.[8] However, in a later decision, the court also held that some of the service's peripheral features that went beyond the production of metadata and allowed the viewing of substantial extracts of copyrighted television programing were not fair use.[9]

(5) Library digitization to enable full text searching

In *Authors Guild v. Google, Inc.,* 804 F.3d 202 (2d Cir. 2015) the Second Circuit held that library digitization by a commercial search engine company that allowed the public to search the texts of the digitally copied books and see displays of snippets of text was fair use. The decision in *Authors Guild v. Google* was consistent with the related *Authors Guild v. HathiTrust* case, however the Google book search engine was different to the HathiTrust enterprise in two respects. First, Google's search engine was commercial in the sense that it was part of a for-profit enterprise. Second, in addition to reporting metadata about books, Google displayed brief snippets of its search results. The snippets conveyed minute amounts of the original expression within the targeted works and were, according to the court, manifestly unlikely to displace any demand for the books themselves.

(6) Related cases

In addition to these examples, other courts have the creation of complete digital copies of copyrighted works was a transformative fair uses when the copies served an entirely different function from the original. In particular in *Perfect 10, Inc. v. Amazon.com, Inc.,* and *Kelly v. Arriba Soft Corp.*, the Ninth Circuit held that an Internet search engine's use of low resolution thumbnail images in a menu of search results was fair use.[10] The use was entirely different in purpose and manifestly unlikely to result in expressive substitution because of the low resolution of the images.

The Technological Advantage Fair Use has Given the United States

In a global economy where the production and analysis of data is increasingly important, it is vital for national economic competitiveness that copyright law does not needlessly impede new technologies. The United States is a world leader in various applications of text mining, starting with Internet search, but going far beyond that. This technological leadership owes a great deal to the flexible structure of United States copyright law. Unlike in closed list systems, in the United States, the fair use doctrine gives technology developers a chance to explain why their particular use of a copyrighted work is for a purpose that promotes the goals of copyright, is reasonable in light of that purpose, and is unlikely to harm the interests of copyright owners.

---

[8] Fox News Network, LLC v. TVEyes, Inc., 43 F. Supp. 3d 379, 393 (S.D.N.Y. 2014).

[9] Fox News Network, LLC v. TVEyes, Inc., 124 F. Supp. 3d 325 (S.D.N.Y. 2015).

[10] Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007); Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003)

The fair use doctrine is a real source of competitive advantage for technologists and academic researchers in the United States. Right now, there are technologies being developed and research being done in the United States that either cannot be pursued in other countries, or can only be pursued by particular people subject to various arbitrary restrictions.

Conclusion

Based on my academic research and my experience as a technology lawyer in Australia, the United Kingdom and the United States, I believe that it would be in South Africa's interest to revise its Copyright law to recognize that the non-expressive use of a copyrighted work does not infringe the rights of the work's author or copyright owner.

I am grateful for the opportunity to comment on the Bill and I would be happy to offer any assistance you might need going forward to expand upon the issues raised in this letter.

Yours sincerely,