Submission on behalf of Joint Creators and Copyright Owners
Class 7(a) & (b): Motion Pictures and Literary Works – Text and Data Mining

UNITED STATES COPYRIGHT OFFICE

Long Comment Regarding a Proposed Exemption Under 17 U.S.C. § 1201

[ ] Check here if multimedia evidence is being provided in connection with this comment

ITEM A. COMMENTER INFORMATION

The Motion Picture Association, Inc. ("MPA") is a trade association representing some of the world's largest producers and distributors of motion pictures and other audiovisual entertainment for viewing in theaters, on prerecorded media, over broadcast TV, cable and satellite services, and on the internet. The MPA's members are: Netflix Studios, LLC, Paramount Pictures Corporation, Sony Pictures Entertainment Inc., Universal City Studios LLC, Walt Disney Studios Motion Pictures, and Warner Bros. Entertainment Inc.

Alliance for Recorded Music ("ARM") is a nonprofit coalition comprising the many artists and record labels who together perform, create, and/or distribute nearly all of the sound recordings commercially released in the United States. Members include the American Association of Independent Music ("A2IM"), the Music Artists Coalition ("MAC"), the Recording Industry Association of America, Inc. ("RIAA"), hundreds of recording artists, the major record companies, and more than 600 independently owned U.S. music labels.

The Entertainment Software Association ("ESA") is the United States trade association serving companies that publish computer and video games for video game consoles, handheld video game devices, personal computers, and the internet. It represents nearly all of the major video game publishers and major video game platform providers in the United States.

Represented By:

J. Matthew Williams (mxw@msk.com)
Sofia Castillo (szc@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
1818 N Street, NW, 7th Floor
Washington, D.C. 20036
202-355-7904

ITEM B. PROPOSED CLASS ADDRESSED

Proposed Class 7(a) & (b): Motion Pictures and Literary Works – Text and Data Mining

ITEM C. OVERVIEW

Petitioners Authors Alliance ("AA"), the American Association of University Professors ("AAUP"), and the Library Copyright Alliance ("LCA") submitted detailed comments that provided valuable and interesting information about some of the potential research activities at

issue, including helpful statements from individual researchers. However, Petitioners propose unbounded exemptions for all motion pictures and literary works for the undefined purpose of "text and data mining" ("TDM"). The rudimentary language they propose is as follows:

> Proposed Class 7(a) Motion Pictures—Text and Data Mining: Lawfully accessed motion pictures where circumvention is undertaken in order to deploy text and data mining techniques.
>
> Proposed Class 7(b) Literary Works—Text and Data Mining: Lawfully accessed literary works distributed electronically where circumvention is undertaken in order to deploy text and data mining techniques.[1]

With respect to motion pictures, Petitioners fail to establish that all of the conduct covered by their proposed exemption is likely noninfringing or that the Section 1201(a) (1) (C) factors, as properly construed, support granting the proposed exemption. These failures stem, in part, from Petitioners' choice not to propose a carefully crafted and limited class of works and to instead propose vague language without circumscribing the scope of the covered acts of circumvention, uses or users. MPA, ARM and ESA ("Joint Creators and Copyright Owners") thus oppose the requested TDM exemptions.[2]

ITEM D. TECHNOLOGICAL PROTECTION MEASURE(S) AND METHOD(S) OF CIRCUMVENTION

The access controls at issue unnecessarily include every access control applied to motion pictures and every access control applied to literary works. The methods of circumvention at issue are any and all methods available/invented. Petitioners submit no limitations.

ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGING USES

Text and data mining has the potential to benefit copyright owners, researchers, teachers, technology companies, and other users of copyrighted works. However, it also raises complex concerns that are the subject of ongoing study and legal controversy.[3] The cases cited by

---

[1] AA *et al.*, Class 7 Long Comment at 4 (Dec. 14, 2020), https://www.copyright.gov/1201/2021/comments/Class%2007a%20and%2007b_InitialComments_Authors%20Alliance,%20American%20Association%20of%20University%20Professors,%20and%20Library%20Copyright%20Alliance.pdf ("AA 2020 Comment").

Their attempt at a definition for TDM appears to be "use of a copyrighted work in computational research." AA 2020 Comment at 4 n. 1, citing Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893, 899 n.19 (2019), https://lawreview.law.ucdavis.edu/issues/53/2/articles/files/53-2_Carroll.pdf (hereinafter "Carroll"). That phrase could describe almost anything done by any person who uses a computer.

[2] The Joint Creators and Copyright Owners focus on motion pictures in these comments given that the Association of American Publishers is filing separate comments to address issues concerning literary works. We do not support either proposed exemption.

[3] *See, e.g.*, U.S. PATENT AND TRADEMARK OFFICE, PUBLIC VIEWS ON ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY POLICY 25 (2020), https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf ("Some mass digitization scenarios may be a fair use, whereas others may be infringements"); *id.* ("Although mass digitization for purposes of machine learning (ML) 'ingestion' processes —and large-scale ingestion of already-digitized works—has not yet been tested by the courts, some rights holders argue that AI trainers should be

Petitioners – *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014), and *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) ("*Google Books*") – involved digitizing hardcopies of literary works owned by libraries without engaging in circumvention and/or acquiring unauthorized access. They also involved specific institutions and companies, specific security measures for reproduced works, specific methods of copying, specific existing and potential licensing markets, and specific and applied models for the limited availability/dissemination of reproduced works (such as the limitations imposed by HathiTrust on access and Google's limited use of "snippets"). Nevertheless, even with all of the specifics presented to the courts, those cases, heard in only one circuit, clearly "test[ed] the boundaries of fair use."[4]

Fair use questions must be considered on a case-by-case basis,[5] and Petitioners' unrestrained proposal lacks any of the specifics necessary to establish the exemption would cover only noninfringing uses of motion pictures. Indeed, existing precedent calls into question whether the approach taken to literary works in the context of the specific facts presented in *HathiTrust* and *Google Books* makes sense when considering motion pictures, and works like sound recordings incorporated therein, even if the copying involved is assumed to be moderately transformative.[6]

Questions left unanswered by the Petitioners' comments so impact the fair use and Section 1201(a) analyses that it is difficult to conduct a factor-by-factor discussion under Section 107 or Section 1201(a)(1)(C). The burden is on Petitioners to provide clarity as to what they are proposing so that the Copyright Office may properly conduct those analyses.[7] In this instance, Petitioners have failed to do so.[8]

A non-exhaustive list of problems and open questions resulting from the unlimited scope of the proposed exemption language includes:

---

required to compensate the authors and rights holders whose copyrighted works their machines are ingesting as a simple matter of doing business.").

[4] *Google Books*, 804 F.3d at 206.

[5] *Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 US 539, 549 (1985).

[6] *See generally Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169 (2d Cir. 2018) (reproducing and distributing/performing motion pictures for research purposes was not fair use). Petitioners argue that the identified uses are "highly transformative." Setting aside that such a position is contrary to the articulation of the transformative use standard in *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578-585 (1994), a standard that may be further developed when the Supreme Court decides *Google LLC v. Oracle Am., Inc*, No. 18-956 (U.S.) (argued Oct. 7, 2020), the Copyright Office must still consider other aspects of the fair use analysis and the Section 1201 factors. Importantly, the Copyright Office must consider the long term availability of works for use and the effect of circumvention of technological measures on the market for or value of copyrighted works.

[7] SECTION 1201 RULEMAKING: SEVENTH TRIENNIAL PROCEEDING TO DETERMINE EXEMPTIONS TO THE PROHIBITION ON CIRCUMVENTION: RECOMMENDATION OF THE ACTING REGISTER OF COPYRIGHTS 12-13 (2018), https://cdn.loc.gov/copyright/1201/2018/2018_Section_1201_Acting_Registers_Recommendation.pdf ("2018 Rec.").

[8] *See* Exemptions To Permit Circumvention of Access Controls on Copyrighted Works: Notice of Proposed Rulemaking, 85 Fed. Reg. 65293, 65302 (Oct. 15, 2020) https://www.govinfo.gov/content/pkg/FR-2020-10-15/pdf/2020-22893.pdf ("NPRM") ("Proponents of exemptions should present their complete affirmative case for an exemption during the initial round of public comment, including all legal and evidentiary support for the proposal.").

- The proposed exemption is not limited to circumvention by owners of copies. It could even cover acquiring permanent, decrypted copies of motion pictures that the institutions or individuals engaged in circumvention never purchased.[9] This appears to be true even where motion pictures are available for purchase on discs, but exemption beneficiaries elect to instead pay lower prices for rental downloads or subscription-based streams to acquire copies for their databases. This deflates Petitioners' claim in the comments that an exemption will increase purchases of authorized copies of motion pictures,[10] or avoid undermining lawful business models, at least as the exemption is currently formulated.

- The proposed exemption does not limit beneficiaries from circulating decrypted copies to other institutions or researchers (or even to commercial, corporate actors) who never purchased copies of them or even acquired authorized, temporary access to them. Nor does it prohibit the creation of databases that aggregate all of the decrypted copies created by all exemption beneficiaries, or that link to each other. In fact, the comments at times seem to imply that is the expected outcome.[11] Nor is the proposal even limited to circumvention for the *sole purpose* of text and data mining or to noncommercial users or uses.[12] Assuming *arguendo* that it is lawful to reproduce a motion picture for computational study (however that is defined), it requires a giant leap to conclude that one copy of each motion picture may be reproduced by one person, institution, or organization and then shared (including remotely) with the world, even if only for research purposes.[13] It is also misguided to conclude that it is justifiable to rent temporary access to a motion picture that is also available for more permanent access at a higher price, and then to decrypt the motion picture to obtain permanent access. Approving such conduct could impair the availability of works at affordable price points.

- The proposed exemption does not require security measures to be put in place to protect databases of decrypted copies of works.[14] Petitioners claim academic institutions are unlikely to have security breaches.[15] But this is unsupported. And, the proposed exemption is not limited to academic institutions and does not even require that its use not facilitate copyright infringement. Petitioners suggest that all academics are responsible and will protect the decrypted copies they create, but they include no

---

[9] *Google Books* discussed the relevance of the libraries owning copies: "Google's creation for each library of *a digital copy of that library's already owned book* in order to permit that library to make fair use through provision of digital searches is not an infringement. If the library had created its own digital copy to enable its provision of fair use digital searches, the making of the digital copy would not have been infringement." *Google Books*, 804 F.3d at 229.

[10] AA 2020 Comments at 28.

[11] *Id.* at 12.

[12] *Id.* at 25.

[13] 2018 Rec. at 236 (stressing importance of on-premises restrictions in preservation context).

[14] Even the law review article primarily relied upon by the petitioners concedes that "the user's security precautions are relevant under the first fair use factor…" Carroll, *supra* note 1, at 893. We believe it is also relevant to the fourth factor.

[15] AA 2020 Comment at 27-28 ("Higher education institutions and other research institutes are extraordinarily well equipped to help researchers secure their data.").

language to require either and even allude to the fact that many academics are already spreading infringing copies of literary works. At the same time, other academics are devising legal theories as to why academics may use these infringing copies (in some cases originating or stored in countries notorious for their disregard for American intellectual property, such as Russia) in their research without liability.[16] Indeed, the comments discuss the Sci-Hub database, which distributes infringing copies of literary works despite being enjoined.[17]

- As mentioned above, the proposed exemption beneficiaries are not limited to institutions or non-profit organizations/academic researchers.[18] Almost anyone could qualify as a "researcher" of motion pictures, and limiting the scope of beneficiaries to those involved in "computational research" does little more to narrow the class of proposed beneficiaries. A person can research a movie by watching it on a computer, for example. That conduct does not appear to be the intended target of the proposal, but Petitioners do not offer a method to achieve appropriate line-drawing.

- The proposed exemption does not prohibit beneficiaries from making works or portions thereof publicly available. Petitioners claim that researchers will not make available to the public decrypted copies of motion pictures or portions thereof aside from when used in published works of scholarship.[19] But, understanding the extent to which works or portions of works are made publicly available is a critical component of the overall scheme envisioned and is required for a proper legal analysis. Nothing in the exemption defines what may be made available, or in what fashion, and once circumvention is achieved for TDM, a question arises as to whether any of the other existing exemptions for criticism and comment remain relevant. What could result is an exemption that allows all motion pictures to be decrypted and aggregated ostensibly for research

[16] Carroll, *supra* note 1, at 898 ("This Article concludes that a researcher maintains the right to conduct computational research on the literature even when the material is copied from an infringing source. This argument has two subparts: (1) a user's good faith is irrelevant to the fair use analysis; and (2) even if good faith were relevant, a TDM researcher would be acting in good faith even when knowing that her sources are infringing because of the net social benefits of conducting TDM research."); *id.* at 952-53 (discussing infringing Sci-Hub database, that academics use as a matter of preference even when they have access to legitimate databases, how unauthorized copies were obtained (including by academics sharing log-in credentials to Sci-Hub), the issuance of an injunction, and where infringing copies are stored). *See also Elsevier Inc. v. www.Sci-Hub.org*, No. 15 CIV. 4282 RWS, 2015 WL 6657363, at *2 (S.D.N.Y. Oct. 30, 2015) ("Elsevier has made a substantial evidentiary showing, documenting the manner in which the Defendants access its ScienceDirect database of scientific literature and post copyrighted material on their own websites free of charge. According to Elsevier, the Defendants gain access to ScienceDirect by using credentials fraudulently obtained from educational institutions, including educational institutions located in the Southern District of New York, which are granted legitimate access to ScienceDirect."); *Elsevier Inc. v. Sci-Hub*, No. 15-CV-4282 (RWS), 2017 WL 3868800, at *1 (S.D.N.Y. June 21, 2017) ("This court hereby finds that Defendants are liable for willful copyright infringement under 17 U.S.C. §§ 101, *et seq*., and this Default Judgment and Permanent Injunction ('Permanent Injunction') is entered against each Defendant.").

[17] AA 2020 Comment at 3-4 (describing Sci-Hub as containing copies of "unlawfully liberated texts.").

[18] *Id.* at 26.

[19] *Id.* at 17. There is no explanation as to how authors could extract content from TDM databases for use in published works, which begs the question whether such databases would be adequately secure.

purposes and then used for any other purpose deemed lawful by users of the copies. That would, in practice, create an impermissible exemption for motion pictures for all (purported) lawful uses that could swallow other exemptions.

- The comments do not address whether licenses have been requested or proposals made to allow for authorized decryption within specified parameters. For purposes of the fourth fair use factor, there is a potential market for copyright owners of motion pictures to enter. The Copyright Office has questioned whether unlicensed uses could qualify as fair, or whether they should be licensed.[20] Petitioners do not address whether they have been denied licenses or whether they would be willing to work with copyright owners. As discussed in our comments on pending *Proposed Class 1: Audiovisual Works – Criticism and Comment*, which we incorporate by reference, copyright owners of motion pictures already license other educational uses, such as remote streaming, and could potentially license the uses at issue.

- The proposal does not preclude circumvention where viable alternatives exist. Petitioners fail to articulate with specifics why text and data mining programs cannot be pursued without circumvention. They do not explain why software engineers cannot design programs to extract information about dialogue, imagery, or other aspects of motion pictures using encrypted discs/downloads/streams. Some of their statements imply they are asking for an exemption for convenience, not necessity. They also claim, in a conclusory manner, that all researchers need the highest quality copies.[21] Even if one concedes that it would be valuable to use computer power to study images used in films directed by David Lynch,[22] why is the highest quality required to do so? Can a computer not identify "average shot length, proportions of shot types (close-up vs. long shot), and the color palette used over the course of a movie"[23] without Blu-ray or Ultra HD quality? The Petitioners also admit they can create copies using screen capture. The failure of screen capture to preserve bookmarks/chapters within discs/copies, which Petitioners point to as a problem, appears to present an issue of convenience, not necessity. And, when Petitioners object to the investment of time purportedly required by screen capture, they do not explain why it is necessary for "a human operator [to be] present for the duration of the screen capture for each movie[.]"

---

[20] U.S COPYRIGHT OFFICE, LEGAL ISSUES IN MASS DIGITIZATION: A PRELIMINARY ANALYSIS AND DISCUSSION DOCUMENT 30 (2011), https://www.copyright.gov/docs/massdigitization/USCOMassDigitization_October2011.pdf, ("This is an area that might benefit from further discussions among stakeholder groups who may be best suited to tailor collective licensing solutions for mass digitization to the evolving digital marketplace."); U.S COPYRIGHT OFFICE, ORPHAN WORKS AND MASS DIGITIZATION 76 (2015), https://www.copyright.gov/orphan/reports/orphan-works2015.pdf ("The Office is not persuaded that fair use has achieved the predictability and stability that [libraries and other user groups] ascribe to it" and concluded that "as a means of providing a coherent and reliable set of standards to govern the broad variety of digitization activities throughout the marketplace, fair use appears ill-suited.").

[21] AA 2020 Comments at 18-19.

[22] *Id.* at 2 (statement of David Bamman).

[23] *Id.*

We look forward to addressing these and other matters with Petitioners and with Copyright Office staff during the public hearings.

ITEM F: DOCUMENTARY EVIDENCE

We have included hyperlinks to webpages/documents within the body of this document. We are not submitting any other documentary evidence.

Respectfully submitted:

/s/ J. Matthew Williams
J. Matthew Williams (mxw@msk.com)
Sofia Castillo (szc@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
1818 N Street, NW, 7th Floor
Washington, D.C. 20036
202-355-7904