

UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK  
-----x  
VIACOM INTERNATIONAL INC.,  
et al.,

Plaintiffs, 07 Civ. 2103 (LLS)

v.

YOUTUBE INC., YOUTUBE LLC,  
and GOOGLE INC.,

Defendants.

-----x  
THE FOOTBALL ASSOCIATION PREMIER  
LEAGUE LIMITED, et al., on  
behalf of themselves and all  
others similarly situated,

**OPINION AND ORDER**

Plaintiffs,

v.

07 Civ. 3582 (LLS)

YOUTUBE INC., YOUTUBE LLC,  
and GOOGLE INC.,

Defendants.

-----x

Plaintiffs in these related lawsuits (the "Viacom  
action" and the "Premier League class action") claim to own  
the copyrights in specified television programs, motion  
pictures, music recordings, and other entertainment  
programs. They allege violations of the Copyright Act of  
1976 (17 U.S.C. § 101 et seq.) by defendants YouTube<sup>1</sup> and  
Google Inc., who own and operate the video-sharing website

---

<sup>1</sup> Defendants YouTube Inc. and YouTube LLC are both referred to as "YouTube."

known as "YouTube.com". Plaintiffs claim, as set forth in Viacom's First Amended Complaint ¶¶ 30-31, that:

Defendants encourage individuals to upload videos to the YouTube site, where YouTube makes them available for immediate viewing by members of the public free of charge. Although YouTube touts itself as a service for sharing home videos, the well-known reality of YouTube's business is far different. YouTube has filled its library with entire episodes and movies and significant segments of popular copyrighted programming from Plaintiffs and other copyright owners, that neither YouTube nor the users who submit the works are licensed to use in this manner. Because YouTube users contribute pirated copyrighted works to YouTube by the thousands, including those owned by Plaintiffs, the videos "deliver[ed]" by YouTube include a vast unauthorized collection of Plaintiffs' copyrighted audiovisual works. YouTube's use of this content directly competes with uses that Plaintiffs have authorized and for which Plaintiffs receive valuable compensation.

. . . .

When a user uploads a video, YouTube copies the video in its own software format, adds it to its own servers, and makes it available for viewing on its own website. A user who wants to view a video goes to the YouTube site . . . enters search terms into a search and indexing function provided by YouTube for this purpose on its site, and receives a list of thumbnails of videos in the YouTube library matching those terms . . . and the user can select and view a video from the list of matches by clicking on the thumbnail created and supplied by YouTube for this purpose. YouTube then publicly performs the chosen video by sending streaming video content from YouTube's servers to the user's computer, where it can be viewed by the user. Simultaneously, a copy of the chosen video is downloaded from the YouTube website to the user's computer. . . . Thus, the YouTube conduct that forms the basis of this

Complaint is not simply providing storage space, conduits, or other facilities to users who create their own websites with infringing materials. To the contrary, YouTube itself commits the infringing duplication, distribution, public performance, and public display of Plaintiffs' copyrighted works, and that infringement occurs on YouTube's own website, which is operated and controlled by Defendants, not users.

(Viacom's brackets).

Plaintiffs allege that those are infringements which YouTube and Google induced and for which they are directly, vicariously or contributorily subject to damages of at least \$1 billion (in the Viacom action), and injunctions barring such conduct in the future.

Among other defenses, YouTube and Google claim the protection afforded by the Digital Millennium Copyright Act of 1998 ("DMCA") (17 U.S.C. §§ 512(c)-(d), (i)-(j)), which among other things limits the terms of injunctions, and bars copyright-damage awards, against an online service provider who: (1) performs a qualified storage or search function for internet users; (2) lacks actual or imputed knowledge of the infringing activity; (3) receives no financial benefit directly from such activity in a case where he has the right and ability to control it; (4) acts promptly to remove or disable access to the material when his designated agent is notified that it is infringing; (5) adopts, reasonably implements and publicizes a policy

of terminating repeat infringers; and (6) accommodates and does not interfere with standard technical measures used by copyright owners to identify or protect copyrighted works.

Plaintiffs move jointly pursuant to Fed. R. Civ. P. 37 to compel YouTube and Google to produce certain electronically stored information and documents, including a critical trade secret: the computer source code which controls both the YouTube.com search function and Google's internet search tool "Google.com". YouTube and Google cross-move pursuant to Fed. R. Civ. P. 26(c) for a protective order barring disclosure of that search code, which they contend is responsible for Google's growth "from its founding in 1998 to a multi-national presence with more than 16,000 employees and a market valuation of roughly \$150 billion" (Singhal Decl. ¶¶ 3, 11), and cannot be disclosed without risking the loss of the business.

1. Search Code

The search code is the product of over a thousand person-years of work. Singhal Decl. ¶ 9. There is no dispute that its secrecy is of enormous commercial value. Someone with access to it could readily perceive its basic design principles, and cause catastrophic competitive harm to Google by sharing them with others who might create their own programs without making the same investment. Id. ¶ 12. Plaintiffs seek production of the search code to support their claim that "Defendants have purposefully designed or modified the tool to facilitate the location of infringing content." Pls.' Reply 10. However, the predicate for that proposition is that the "tool" treats infringing material differently from innocent material, and plaintiffs offer no evidence that the search function can discriminate between infringing and non-infringing videos.

YouTube and Google maintain that "no source code in existence today can distinguish between infringing and non-infringing video clips — certainly not without the active participation of rights holders" (Defs.' Cross-Mot. Reply 11), and Google engineer Amitabh Singhal declares under penalty of perjury that:

The search function employed on the YouTube website was not, in any manner, designed or modified to facilitate the location of allegedly infringing materials. The purpose of the YouTube search engine is to allow users to find videos they are looking for by entering text-based search terms. In some instances, the search service suggests search terms when there appears to be a misspelling entered by the user and attempts to distinguish between search terms with multiple meanings. Those functions are automated algorithms that run across Google's services and were not designed to make allegedly infringing video clips more prominent in search results than non-infringing video clips. Indeed, Google has never sought to increase the rank or visibility of allegedly infringing material over non-infringing material when developing its search services.

Singhal Reply Decl. ¶ 2.

Plaintiffs argue that the best way to determine whether those denials are true is to compel production and examination of the search code. Nevertheless, YouTube and Google should not be made to place this vital asset in hazard merely to allay speculation. A plausible showing that YouTube and Google's denials are false, and that the search function can and has been used to discriminate in favor of infringing content, should be required before disclosure of so valuable and vulnerable an asset is compelled.

Nor do plaintiffs offer evidence supporting their conjecture that the YouTube.com search function might be adaptable into a program which filters out infringing

videos. Plaintiffs wish to "demonstrate what Defendants have *not* done but could have" to prevent infringements, Pls.' Reply 12 (plaintiffs' italics), but there may be other ways to show that filtering technology is feasible<sup>2</sup> and reasonably could have been put in place.

Finally, the protections set forth in the stipulated confidentiality order are careful and extensive, but nevertheless not as safe as nondisclosure. There is no occasion to rely on them, without a preliminary proper showing justifying production of the search code.

Therefore, the cross-motion for a protective order is granted and the motion to compel production of the search code is denied.

---

<sup>2</sup> In the Viacom action (Housley Decl. ¶ 2):

Viacom is currently using fingerprinting technology provided by a company called Auditude in order to identify potentially infringing clips of Viacom's copyrighted works on the YouTube website. The fingerprinting technology automatically creates digital "fingerprints" of the audio track of videos currently available on the YouTube website and compares those fingerprints against a reference library of digital fingerprints of Viacom's copyrighted works. As this comparison is made, the fingerprinting technology reports fingerprint matches, which indicate that the YouTube clip potentially infringes one of Viacom's copyrighted works.

2. Video ID Code

Plaintiffs also move to compel production of another undisputed trade secret, the computer source code for the newly invented "Video ID" program. Using that program, copyright owners may furnish YouTube with video reference samples, which YouTube will use to search for and locate video clips in its library which have characteristics sufficiently matching those of the samples as to suggest infringement. That program's source code is the product of "approximately 50,000 man hours of engineering time and millions of dollars of research and development costs", and maintaining its confidentiality is essential to prevent others from creating competing programs without any equivalent investment, and to bar users who wish to post infringing content onto YouTube.com from learning ways to trick the Video ID program and thus "escape detection." Salem Decl. ¶¶ 8-12.

Plaintiffs claim that they need production of the Video ID source code to demonstrate what defendants "could be doing — but are not — to control infringement" with the Video ID program (Pls.' Reply 6). However, plaintiffs

can learn how the Video ID program works from use and observation of its operation (Salem Decl. ¶ 13), and examination of pending patent applications, documentation and white papers regarding Video ID (id.), all of which are available to them (see Defs.' Opp. 7). If there is a way to write a program that can identify and thus control infringing videos, plaintiffs are free to demonstrate it, with or without reference to the way the Video ID program works. But the question is what infringement detection operations are possible, not how the Video ID source code makes it operate as it does. The notion that examination of the source code might suggest how to make a better method of infringement detection is speculative. Considered against its value and secrecy, plaintiffs have not made a sufficient showing of need for its disclosure.

Therefore, the motion to compel production of the Video ID code is denied.

### 3. Removed Videos

Plaintiffs seek copies of all videos that were once available for public viewing on YouTube.com but later removed for any reason, or such subsets as plaintiffs designate (Pls.' Reply 41). Plaintiffs claim that their

direct access to the removed videos is essential to identify which (if any) infringe their alleged copyrights. Plaintiffs offer to supply the hard drives needed to receive those copies (id. 41), which defendants store on computer hard drives.

Defendants concede that "Plaintiffs should have some type of access to removed videos in order to identify alleged infringements" (Defs.' Opp. 27), but propose to make plaintiffs identify and specify the videos plaintiffs select as probable infringers by use of data such as their titles and topics and a search program (which defendants have furnished) that gives plaintiffs the capacity both to run searches against that data and to view "snapshots" taken from each removed video. That would relieve defendants of producing all of the millions of removed videos, a process which would require a total of about five person-weeks of labor without unexpected glitches, as well as the dedication of expensive computer equipment and network bandwidth. Do Decl. ¶¶ 5-7.

However, it appears that the burden of producing a program for production of all of the removed videos should be roughly equivalent to, or at least not significantly greater than, that of producing a program to create and

copy a list of specific videos selected by plaintiffs (see Davis Decl. ¶ 21).

While the total number of removed videos is intimidating (millions, according to defendants), the burden of inspection and selection, leading to the ultimate identification of individual "works-in-suit", is on the plaintiffs who say they can handle it electronically.

Under the circumstances, the motion to compel production of copies of all removed videos is granted.

#### 4. Video-Related Data from the Logging Database

Defendants' "Logging" database contains, for each instance a video is watched, the unique "login ID" of the user who watched it, the time when the user started to watch the video, the internet protocol address other devices connected to the internet use to identify the user's computer ("IP address"), and the identifier for the video. Do Sept. 12, 2007 Dep. 154:8-21 (Kohlmann Decl. Ex. B); Do Decl. ¶ 16. That database (which is stored on live computer hard drives) is the only existing record of how often each video has been viewed during various time periods. Its data can "recreate the number of views for any particular day of a video." Do Dep. 211:16-21.

Plaintiffs seek all data from the Logging database concerning each time a YouTube video has been viewed on the YouTube website or through embedding on a third-party website. Pls.' Mot. 19.

They need the data to compare the attractiveness of allegedly infringing videos with that of non-infringing videos. A markedly higher proportion of infringing-video watching may bear on plaintiffs' vicarious liability claim,<sup>3</sup> and defendants' substantial non-infringing use defense.<sup>4</sup>

Defendants argue generally that plaintiffs' request is unduly burdensome because producing the enormous amount of information in the Logging database (about 12 terabytes of data) "would be expensive and time-consuming, particularly in light of the need to examine the contents for privileged and work product material." Defs.' Opp. 22.

But defendants do not specifically refute that "There is no need to engage in a detailed privilege review of the logging database, since it simply records the numbers of

---

<sup>3</sup> See Fonovisa, Inc. v. Cherry Action, Inc., 76 F.3d 259, 263 (9th Cir. 1996) ("financial benefit prong of vicariously liability" claim may be satisfied by demonstrating that "infringing performances enhance the attractiveness of the venue to potential customers" and act as "a 'draw' for customers" from whom the venue's operator derives income).

<sup>4</sup> See Sony Corp. of America v. Universal Studios, Inc., 464 U.S. 417, 442 (1983) (barring secondary liability based on imputed intent to cause infringements from the design or distribution of a product "capable of substantial noninfringing uses"); Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 933-34 (2005) (declining "to add a more quantified description" of how much non-infringing use qualifies as substantial under Sony).

views for each video uploaded to the YouTube website, and the videos watched by each user" (Pls.' Reply 45). While the Logging database is large, all of its contents can be copied onto a few "over-the-shelf" four-terabyte hard drives (Davis Decl. ¶ 22). Plaintiffs' need for the data outweighs the unquantified and unsubstantiated cost of producing that information.

Defendants argue that the data should not be disclosed because of the users' privacy concerns, saying that "Plaintiffs would likely be able to determine the viewing and video uploading habits of YouTube's users based on the user's login ID and the user's IP address" (Do Decl. ¶ 16).

But defendants cite no authority barring them from disclosing such information in civil discovery proceedings,<sup>5</sup> and their privacy concerns are speculative. Defendants do not refute that the "login ID is an anonymous pseudonym that users create for themselves when they sign up with

---

<sup>5</sup> The statute defendants point to, 18 U.S.C. § 2710 (titled "Wrongful disclosure of video tape rental or sale records"), prohibits video tape service providers from disclosing information on the specific video materials subscribers request or obtain, and in the case they cite, In re Grand Jury Subpoena to Amazon.com, 246 F.R.D. 570, 572-73 (W.D.Wis. 2007) (the "subpoena is troubling because it permits the government to peek into the reading habits of specific individuals without their prior knowledge or permission"), the court on First Amendment grounds did not require an internet book retailer to disclose the identities of customers who purchased used books from the grand jury's target, a used book seller under investigation for tax evasion and wire and mail fraud in connection with his sale of used books through the retailer's website.

YouTube" which without more "cannot identify specific individuals" (Pls.' Reply 44), and Google has elsewhere stated:

We . . . are strong supporters of the idea that data protection laws should apply to any data that could identify you. The reality is though that in most cases, an IP address without additional information cannot.

Google Software Engineer Alma Whitten, Are IP addresses personal?, GOOGLE PUBLIC POLICY BLOG (Feb. 22, 2008), <http://googlepublicpolicy.blogspot.com/2008/02/are-ip-addresses-personal.html> (Wilkins Decl. Ex. M).

Therefore, the motion to compel production of all data from the Logging database concerning each time a YouTube video has been viewed on the YouTube website or through embedding on a third-party website is granted.

5. Video-Related Data from the User and Mono Databases

Defendants' "User" and "Mono" databases contain information about each video available in YouTube's collection, including its user-supplied title and keywords, public comments from others about it, whether it has been flagged as inappropriate by others (for copyright infringement or for other improprieties such as obscenity) and the reason it was flagged, whether an administrative

~~third party private videos it receives during the process of resolving copyright infringement claims against such videos (see Wilkens Decl. Ex. T, ¶ 4). The record shows that defendants do not disclose to content owners any private videos processed for potentially infringing the owners' copyrights unless defendants receive the express consent of the users who designated the videos as private (Salem Sur-Reply Decl. ¶¶ 1-5), and that the clause plaintiffs rely upon merely requires content owners to maintain the confidentiality of such consensually divulged private videos (id.).~~

#### CONCLUSION

For the reasons set forth above:

- (1) The cross-motion for a protective order barring disclosure of the source code for the YouTube.com search function is granted, and the motion to compel production of that search code is denied;
- (2) The motion to compel production of the source code for the Video ID program is denied;
- (3) The motion to compel production of all removed videos is granted;
- (4) The motion to compel production of all data from the Logging database concerning each time a YouTube video has been viewed on the YouTube website or through embedding on a third-party website is granted;

(5) The motion to compel production of those data fields which defendants have agreed to produce for works-in-suit, for all videos that have been posted to the YouTube website is denied;


(6) The motion to compel production of the schema for the Google Advertising database is denied;

(7) The motion to compel production of the schema for the Google Video Content database is granted; and

(8) The motion to compel production of the private videos and data related to them is denied at this time except to the extent it seeks production of specified non-content data about such videos.

So ordered.

Dated: July 1, 2008  
New York, New York

  
\_\_\_\_\_  
Louis L. Stanton  
U.S.D.J.