

1 ROM BAR-NISSIM (SBN: 293356)
Rom@HeahBarNissim.com
2 **HEAH BAR-NISSIM LLP**
1801 Century Park East, Suite 2400
3 Los Angeles, CA 90067
Telephone: 310.432.2836

4 JARRETT LEE ELLZEY (*pro hac vice application forthcoming*)
Jellzey@EKSM.com
5 TOM KHERKHER (*pro hac vice application forthcoming*)
TKherkher@EKSM.com
6 LEIGH S. MONTGOMERY (*pro hac vice application forthcoming*)
LMontgomery@EKSM.com
7 NATISCHA VOLPE (*pro hac vice application forthcoming*)
NVolpe@EKSM.com
8 **ELLZEY KHERKHER SANFORD MONTGOMERY LLP**
4200 Montrose Street, Suite 200
9 Houston, TX 77006

10 Attorneys for Plaintiffs
11 TED ENTERTAINMENT, INC.
MATT FISHER
12 GOLFHOLICS, INC.

13
14 **UNITED STATES DISTRICT COURT**
15 **NORTHERN DISTRICT OF CALIFORNIA**
16

17
18 TED ENTERTAINMENT, INC., MATT
FISHER, and GOLFHOLICS, INC., each
19 individually and on behalf of all others
similarly situated,

20 Plaintiffs,

21 v.

22
23 BYTEDANCE INC.,

24 Defendant
25
26
27
28

Case No.: 5:25-cv-10933

CLASS ACTION COMPLAINT

DEMAND FOR JURY TRIAL

1 Plaintiffs Ted Entertainment, Inc. (“TEI”), Matt Fisher, and Golfholics, Inc. (collectively,
2 where appropriate “Plaintiffs”), each individually and on behalf of all others similarly situated, by
3 and through their undersigned counsel, file this Complaint against Defendant ByteDance Inc.
4 (“Defendant”), and alleges as follows:

5 **NATURE OF THE ACTION**

6 1. This lawsuit arises from Defendant unlawfully circumventing technological
7 measures to access and scrape millions of copyrighted videos from the online video viewing
8 platform, YouTube, in order to feed, train, improve, and commercialize Defendant’s large-scale
9 generative artificial intelligence (“AI”) model, “MagicVideo.”

10 2. MagicVideo is an artificial intelligence system designed by Defendant to generate
11 video output from text and image input. This “text-to-video” and “image-to-video” system is
12 intended to allow consumers to generate videos using text and image prompts.

13 3. The research from Make-A-Video has paved the way for more advanced tools. The
14 technology Defendant has and continues to develop is now a core feature within the products
15 Defendant offers to its users. Users can leverage this capability through features like “Effects,”
16 which use AI to enhance user content, or turning text or a static image into a video within
17 Defendant’s social media site.

18 4. YouTube allows the public to view audiovisual works only through controlled
19 streaming and never provides access to the underlying video files. Defendant intentionally bypassed
20 those restrictions by deploying scraping tools designed to evade YouTube’s access protections.
21 Defendant used a video-downloading program combined with virtual machines that rotated IP
22 addresses to avoid detection and blocking. Defendant used Plaintiffs’ and Class Members’
23 intellectual property for its own commercial gain. In doing so, Defendant has violated YouTube’s
24 Terms of Service, which were intended to protect Plaintiffs and others similarly situated.

25 5. By scraping and downloading those files in order to build MagicVideo, Defendant
26 deliberately circumvented YouTube’s access controls.

27 6. Plaintiffs and the Class Members whom Plaintiffs seeks to represent are content
28 creators who upload their audiovisual content to YouTube. In doing so, the content creators are

1 authorizing and instructing YouTube to provide protection to the video content through YouTube's
2 anti-circumvention software and Terms of Service. In fact, YouTube's anti-circumvention software
3 and protective Terms of Service are a driving factor behind content creators' decision to upload
4 their video content to YouTube.

5 7. Rather than seek permission or pay a fair price for the audiovisual content hosted on
6 YouTube, Defendant harvested content creators' protected and copyrighted videos at scale without
7 consent or compensation.

8 8. Defendant's actions were not only unlawful, but an unconscionable attack on the
9 community of content creators whose content is used to fuel the multi-trillion-dollar generative AI
10 industry without any compensation.

11 9. Content creators such as Plaintiffs and the Class Members will never be able to claw
12 back the intellectual property unlawfully copied and used by Defendant to train its generative AI.
13 Once AI ingests content, that content is stored in its neural network, and not capable of deletion or
14 retraction. Defendant's actions constitute abuse and exploitation of content creators' work for
15 Defendant's profit.

16 10. Most YouTube videos are not registered with the U.S. Copyright Office. That lack
17 of registration, however, does not render them valueless or leave them unprotected. Content
18 creators invest time, skill, and resources into producing their works, and they rely on YouTube's
19 technological protection measures to safeguard their files from unauthorized access. Because
20 copyright registration is not a prerequisite for protection against unlawful circumvention of access
21 controls, this claim is especially critical where Defendant's misconduct lies in breaking through
22 access barriers that prevent anyone from obtaining the underlying files in the first place

23 11. Plaintiff Ted Entertainment, Inc. is an independent content creator with over 5,800
24 original videos on YouTube, with a combined total of over 4,000,000,000 YouTube views. TEI has
25 additionally amassed a substantial following on YouTube of over 2,600,000 subscribers. TEI has
26 invested substantial time and money into bringing awareness around his content.

27 12. Plaintiff Matt Fisher is an individual and resident of the State of California. He is a
28 golf content creator who posts videos on YouTube, many of which are instructional. His channel

1 is @Mr.ShortGame on the YouTube platform. He has over 500,000 subscribers and hundreds of
2 millions of views. Fisher has invested substantial time and money into bringing awareness around
3 his content.

4 13. Plaintiff Golfholics is a corporate entity organized pursuant to the laws of the State
5 of California. It is a golf content channel who posted videos on YouTube. The channel is
6 @Golfholics on the YouTube platform. It has over 130,000 subscribers and millions of views.
7 Golfholics invested substantial time and money into bringing awareness around its content.

8 14. Like Plaintiffs, the Class Members in this case are independent content creators who
9 uploaded their video content to YouTube’s video platform. The video content of Class Members
10 was also part of the dataset utilized by Defendant to train its AI model.

11 15. An essential component of Defendant’s business model—powering AI features and
12 services with large-scale training data—includes the mass acquisition and ingestion of creators’
13 videos scraped from YouTube.

14 16. Defendant has profited and will continue to profit substantially from its infringement
15 of Plaintiffs’ and Class Members’ video content through Defendant’s training of its generative AI
16 products. Defendant’s financial and technological success would not have been possible without
17 the video content created by Plaintiffs and Class Members, which was intended for streaming on
18 YouTube.

19 17. Plaintiffs bring this class action on behalf of themselves and of a nationwide
20 class of YouTube creators whose works were scraped, ingested, and trained on without
21 authorization, seeking statutory damages, injunctive relief, restitution, and all other remedies
22 allowed by law pursuant to the Digital Millennium Copyright Act, 17 U.S.C.A. § 1201(a),
23 seeking an injunction and damages commensurate with the scope of Defendant’s massive and
24 ongoing infringement. More particularly, Defendant’s conduct violates the provisions of the
25 DMCA regarding anti-circumvention (§1201) by bypassing technological protection measures
26 that control access to and copying of YouTube videos.

27 **THE PARTIES**

28 18. Plaintiff TEI is a California based media company that owns and operates the

1 YouTube channels “h3h3 Productions” and “H3 Podcast Highlights.” Both channels appear
2 extensively in the dataset Defendant used to train its artificial intelligence model. H3H3
3 Productions has 146 videos in HD-VILA-100M. H3 Podcast Highlights has 285 videos in HD-
4 VILA-100M. TEI invested significant resources into producing and publishing this video content,
5 and every one of these videos was downloaded, copied, and ingested by Defendant without
6 authorization.

7 19. Plaintiff TEI is the creator of the videos identified in Exhibit A, all of which were
8 uploaded exclusively to YouTube by Plaintiff TEI and all of which were included in the dataset
9 used by Defendant to train its generative AI models. Plaintiff TEI derives value from the works
10 identified in Exhibit A through viewership, advertising, sponsorships, licensing, and related
11 monetization.

12 20. Plaintiff TEI and its owners – Ethan and Hila Klein – are longtime champions of the
13 rights of YouTube creators. The Kleins helped define what constitutes fair use reaction videos and
14 the good faith belief standard for DMCA counternotifications. Plaintiff TEI has championed online
15 free speech and is currently helping define what reaction content does not constitute fair use to
16 ensure YouTube content creators can enjoy the fruits of their labor.

17 21. Plaintiff Matt Fisher is an individual and resident of the State of California. He is a
18 golf content creator who posts videos on YouTube, many of which are instructional. His channel
19 is “Mr.ShortGame Golf” on the YouTube platform. He has over 500,000 subscribers and hundreds
20 of millions of views. Plaintiff Fisher has invested substantial time and money into bringing
21 awareness around his content. MrShort Game has 2 videos in the HD-VILA-100M dataset.

22 22. Plaintiff Golfholics is a corporate entity organized pursuant to the laws of the State
23 of California. It is a golf content channel who posted videos on YouTube. The channel is
24 “Golfholics” on the YouTube platform. It has over 130,000 subscribers and millions of views.
25 Plaintiff Golfholics invested substantial time and money into bringing awareness around its content.
26 Golfholics has 62 videos in the HD-VILA-100M data set.

27 23. Defendant Bytedance Inc., is a Delaware corporation with its principal place of
28 business at 1199 Coleman Avenue, San Jose, CA 95110.

JURISDICTION AND VENUE

1
2 24. This is a civil action seeking damages and injunctive relief for infringement under
3 the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205. As such, this Court has subject
4 matter jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and 1338(a), based on federal
5 question jurisdiction.

6 25. The Court has supplemental jurisdiction over related state-law claims, if any, under
7 28 U.S.C. § 1367.

8 26. Personal jurisdiction is proper because Defendant’s principal place of business is in
9 San Jose, California, and because Defendant transacts business nationwide, purposefully avails
10 itself of this forum, and a substantial part of the events or omissions giving rise to these claims
11 occurred or were directed here.

12 27. Venue lies in this judicial district pursuant to 28 U.S.C. § 1391(b)(1) and/or 1400(a)
13 because Defendant resides in, is found in, or may be sued in this District, and a substantial part of
14 the events giving rise to the claims occurred here.

FACTUAL BACKGROUND

15
16 ***A. Content Creators, YouTube and Technological Protection Measures***

17 28. As noted above, Plaintiffs and the Class Members are independent content creators
18 of audiovisual content.

19 29. Plaintiff TEI is the creator of the videos listed in Exhibit A.

20 30. Plaintiff Matt Fisher is the creator of the videos listed in Exhibit B.

21 31. Plaintiff Golfaholics, Inc. is the creator of the videos in Exhibit C.

22 32. Plaintiffs and the Class Members create video content and upload their video content
23 onto YouTube’s video sharing platform.

24 33. YouTube’s “users”—the individuals who view the digital content available on
25 YouTube—can watch and listen to music videos for free on YouTube’s ad-supported service, but
26 YouTube does not give users access to or allow downloading of the digital files for the content
27 viewed by the user.

28 34. YouTube deploys technological protection measures (“TPMs”) designed to control

1 access to the underlying video files and prevent direct downloading outside permitted channels
2 (e.g., streaming-only delivery, application programming interface (“API”) usage limits, and access
3 controls).

4 35. YouTube’s Terms of Service expressly prohibit scraping, unauthorized
5 downloading, bulk extraction, or other forms of data mining of audiovisual content except through
6 expressly permitted features or licensed APIs.¹ These contractual restrictions operate together with
7 TPMs to prevent unlicensed access to creators’ videos.

8 36. According to YouTube’s Terms of Service, Content creators such as Plaintiffs who
9 upload content onto YouTube grant license to YouTube for certain uses as well as to other users of
10 YouTube to access content through YouTube’s services; however, the license makes clear that it
11 “does not grant any rights or permissions for a user to make use of [the] Content independent of
12 the Service.”²

13 37. This language confirms that end users are not given access to the file itself, only the
14 ability to view (i.e., stream) through YouTube’s controlled environment. YouTube’s Terms of
15 Service reflect YouTube’s intent to restrict access to the digital files underlying the videos that
16 YouTube’s users are allowed to stream through YouTube’s platform.

17 38. Streaming through YouTube and downloading permanent copies provide the user
18 with different value propositions—watching and listening for free but seeing ads, versus possessing
19 a permanent digital copy.

20 39. Accordingly, scraping or bulk downloading is not merely copying material already
21 provided; it is an act of unauthorized access to data files that YouTube affirmatively withholds
22 from public download.

23 40. In order to enforce its prohibitions on downloading content, YouTube does not
24 provide a downloading option that is readily available to users.

25 41. Although YouTube does offer downloading options to customers who subscribe and

26 ¹ “Terms of Service,” YouTube, <https://www.youtube.com/t/terms> (last viewed December 10,
27 2025) (prohibiting “access[ing] the Service using any automated means (such as robots, botnets or
scrapers) except (a) in the case of public search engines, in accordance with YouTube’s robots.txt
file; or (b) with YouTube’s prior written permission”).

28 ² *Id.*

1 pay for YouTube’s “Premium” plan, YouTube also employs TPMs to prevent access for Premium
2 customers. Specifically, YouTube prohibits all downloading audiovisual content except to the
3 YouTube app. Further, the “download” option only makes audiovisual content available for offline
4 streaming—it does not provide the Premium customer with access to the audiovisual files.
5 Accordingly, the audiovisual files cannot be transferred to any other device, but remain only for
6 streaming on the app. Finally, the files are only available for offline streaming for a limited amount
7 of time, at which point they will become available only for online streaming once again.

8 42. For users who do not have a “Premium” plan, the “download” option on YouTube’s
9 player page is non-functional.

10 43. In order to further enforce its prohibitions on downloading content, YouTube uses
11 technological processes and tools to detect and block unauthorized downloading. For example,
12 YouTube occasionally updates APIs, which operate to interfere with downloaders utilizing the non-
13 updated API to extract files. As another example, YouTube monitors downloading activity and may
14 block IP addresses that make too many download attempts in a specified time period.

15 44. YouTube’s refusal to provide users with a readily available downloading option,
16 along with the processes and tools YouTube employs operate to restrict users’ ability to access
17 audiovisual material, are part of the TPMs that YouTube has employed to prohibit access to
18 audiovisual files.

19 45. Content creators, including Plaintiffs and Class Members, rely on the TPMs
20 contained in YouTube’s Terms of Service in deciding to upload their audiovisual content to
21 YouTube. Content creators, including Plaintiffs and Class Members, expect that their works will
22 not be copied at scale without consent and that any authorized uses will retain CMI so their
23 identities and rights are preserved.

24 ***B. Defendant Improperly Obtained Millions of YouTube Videos to Train Its***
25 ***Foundational AI Model***

26 46. As noted above, Defendant is the creator, owner and operator of the AI video
27 generator, MagicVideo.

28 47. In order to feed, train, improve, and commercialize MagicVideo, Defendant requires

1 significant amounts of data. Among the digital data required to train MagicVideo are datasets that
2 include millions of videos, including pre-existing data sets such as “HD-VILA-100M” and
3 “Webvid-10M,” as well as “self-collected 7M video-text samples.”³

4 48. MagicVideo was not a research tool or isolated project; it is Defendant’s generative
5 AI video model that specifically envisions creators and artists as its target market.

6 49. Because MagicVideo is intended to target consumers in an already competitive field
7 of AI text-to-video and image-to-video models, Defendant had an overwhelming incentive to
8 acquire training data on an unprecedented scale. Rather than negotiate for lawful licenses,
9 Defendant broke through YouTube’s access protections to obtain the massive dataset necessary to
10 fuel MagicVideo and, by extension, Defendant’s success in the field of AI text-to-video and image-
11 to-video models.

12 50. Defendant has invested in its MagicVideo product and has also fundamentally
13 collaborated on improving its generative capabilities. Particular to the current case, Defendant has
14 sought to develop MagicVideo so it is capable, among other things, of turning language and image
15 prompts into audiovisual content. In other words, Defendant has sought to, and has had some
16 success with, creating an AI product that accepts language instruction or images and produces a
17 video based on that instruction or image.

18 51. In order to adequately train its generative AI model, Defendant requires vast
19 amounts of data—the more data, the better the AI product.

20 52. Defendant obtained datasets from a variety of sources, including academic
21 repositories, research compilations, and other large-scale video collections created by universities,
22 corporations, and independent researchers. These datasets were treated by Defendant as raw
23 material for MagicVideo, even when the datasets were expressly licensed for academic or non-
24 commercial use and prohibited commercial exploitation, redistribution, or any use that would
25 involve downloading the underlying copyrighted works

26 53. In order to acquire high-quality text-to-video and image-to-video generation,

27 ³ Daquan Zhou, Weimin Wang, Hanshu Yan, Weiwei Lv, Yizhe Zhu, Jiashi Feng, ByteDance Inc.;
28 “MagicVideo: Efficient Video Generation With Latent Diffusion Models” (hereinafter referred to
as “ByteDance Research Paper”), p. 6, May 11, 2023, available at <https://arxiv.org/abs/2211.11018>
(last viewed December 7, 2025).

1 Defendant, directly and through agents, contractors, and affiliates, intentionally accessed large
2 volumes of YouTube videos by scraping and/or using tools and workflows that bypass or evade
3 YouTube’s TPMs and usage restrictions, and then reproduced those videos to assemble training
4 corpora for Defendant’s AI models and services.

5 54. “Scraping” content involves downloading audio and video files from a website, in
6 this case YouTube, using an automation of some type.

7 55. When Defendant scraped audio and video files from YouTube, Defendant did not
8 simply download those files onto the YouTube app for offline streaming, as envisioned by
9 YouTube’s Premium plan. Instead, Defendant improperly accessed the actual audio and video files
10 and downloaded those files into Defendant’s own system, where Defendant had control over the
11 files and where Defendant could store them indefinitely. These actions are inconsistent with
12 YouTube’s Premium plan.

13 56. To fuel its scraping activities, Defendant obtained the HD-VILA-100M dataset.

14 57. The HD-VILA-100M does not contain the actual videos or clips but instead provides
15 pointers such as video identifiers and timestamp boundaries for approximately 100 million clips
16 drawn from 3,098,462 unique YouTube videos.

17 58. Upon information and belief, the HD-VILA-100M dataset was obtained through
18 GitHub—a publicly available online code-sharing platform.

19 59. Upon information and belief, the license agreement for the HD-VILA-100M dataset
20 restricts the dataset to “academic use only,” such that any use for commercial purposes would be
21 unauthorized.

22 60. In addition, upon information and belief, the GitHub license agreement prohibits
23 distribution, reproduction, modification, or exploitation of the dataset content without the
24 permission of the copyright owner, thus recognizing that copyright ownership resides with the
25 original creators of the source material, including YouTube content creators such as Plaintiffs and
26 the Class Members.

27 61. Defendant was aware that the HD-VILA-100M dataset was for research purposes
28 only but used the dataset for commercial development (i.e., “training”) of its generative AI model.

1 62. The HD-VILA-100M dataset is made up entirely of YouTube source material that
2 must be downloaded from YouTube to be used for AI training. This dataset utilized by Defendant
3 does not contain the videos themselves. A video dataset functions only as an index file that lists
4 pointers such as URLs, YouTube IDs, or similar location identifiers, sometimes with limited
5 metadata attached. A dataset contains no audiovisual files. To use any such dataset, a company
6 must retrieve the actual video files by downloading them directly from YouTube. Doing so requires
7 new copying of each copyrighted work and requires bypassing YouTube’s technological
8 restrictions, terms of service, and licensing limitations. The dataset merely identifies where the
9 videos are located. All copying, downloading, and circumvention is performed by the user.

10 63. The HD-VILA-100M dataset used by Defendant list thousands of specific
11 timestamps within those videos that are designated as clips. A single YouTube video can be divided
12 into many such clips, and each clip is treated as a separate training sample. As a result, the scraping
13 process requires repeated retrieval of the same underlying video for each clip. This means the same
14 copyrighted work is copied multiple times, from different starting and ending points, for the
15 purpose of extracting each designated clip.

16 64. To obtain any clip listed in the HD-VILA-100M dataset used by Defendant, the user
17 must newly copy the underlying YouTube video at the specific timestamp identified for that clip.
18 Each clip requires a separate retrieval of the source video, which results in a distinct act of copying.
19 Because YouTube provides no lawful mechanism for downloading these works, every retrieval
20 requires the user to circumvent YouTube’s technological restrictions, terms of service, and
21 licensing limits. The dataset provides only the video location and the start and end times for each
22 clip. All downloading, copying, and circumvention occurs on a clip-by-clip basis, and each instance
23 constitutes its own violation.

24 65. The HD-VILA-100M dataset is comprised entirely of YouTube videos and contains
25 3,098,462 source videos that were divided into roughly 100 million clips. Each clip represents a
26 distinct segment of the original work. The dataset was compiled by Microsoft Research Asia and
27 published in 2021 for research involving video-based AI models. It lists pointers to YouTube videos
28 and includes subtitle information extracted from YouTube’s closed captioning system.

1 66. The HD-VILA-100M dataset does not provide the video files themselves. Any user
2 of the dataset must download each referenced clip directly from YouTube. Defendant did so at
3 scale, which triggered millions of separate acts of unauthorized access and copying.

4 67. The usage license on the HD-VILA-100M dataset states it was created “solely for
5 Computational Use for non-commercial research. This restriction means that you may engage in
6 non-commercial research activities (including non-commercial research undertaken by or funded
7 via a commercial entity), but you may not use the Data or any Results in any commercial offering,
8 including as part of a product or service (or to improve any product or service) you use or provide
9 to others.”

10 68. These limitations on the dataset’s usage license is an implicit acknowledgment that
11 the HD-VILA-100M dataset is comprised of protected, copyrighted works obtained without
12 license, permission, or authorization.

13 69. Defendant was aware that the dataset was for research purposes only but used the
14 HD-VILA-100M dataset and other datasets for commercial development (i.e., “training”) of its
15 generative AI model.

16 70. The HD-VILA-100M dataset consists of location identifiers that point to millions
17 of YouTube videos or clips. The dataset does not contain the underlying audiovisual files. To use
18 the audiovisual files in training, a company must retrieve and download every referenced video
19 directly from YouTube. By targeting this dataset, Defendant initiated millions of individual
20 downloads of protected YouTube content, all without authorization, all in violation of YouTube’s
21 access restrictions, and all for the commercial purpose of building its artificial intelligence video
22 model.

23 71. Defendant used the HD-VILA-100M dataset as the foundation for its own refined
24 dataset. However, to create this reduced version, had to download every underlying YouTube
25 video, likely multiple times, listed in the original HD-VILA-100M dataset in order to extract those
26 clips.

27 72. Defendant used the HD-VILA-100M dataset to scrape audio and video content from
28 YouTube without the knowledge or consent of YouTube or the creators of the content.

1 73. Defendant also obtained its own separate audio and video datasets directly through
2 its own independent scraping of YouTube’s video sharing platform without the knowledge or
3 consent of YouTube or the creators of the content.

4 74. The scraping activities Defendant conducted on YouTube’s video sharing platform
5 required Defendant to use processes to defeat protections that YouTube put in place to prevent
6 unauthorized access to the digital files underlying the audiovisual content. The processes used by
7 Defendant to scrape audiovisual files from YouTube are not used in the ordinary course of
8 YouTube’s operations from the point of view of an ordinary consumer.

9 75. Upon information and belief, Defendant used tools and processes such as the open-
10 source YouTube video downloader “yt-dlp” combined with virtual machines that refresh IP
11 addresses to access audiovisual content from YouTube’s platform. Such tools and processes are
12 necessary for Defendant to avoid being blocked by YouTube.

13 76. The use of virtual machines to refresh IP addresses is necessary because YouTube
14 uses programs to monitor the activity from IP addresses for downloading activity and block those
15 IP addresses from accessing audiovisual data on the YouTube platform. Refreshing the IP addresses
16 operates to defeat YouTube’s monitoring programs.

17 77. The yt-dlp downloader is a tool used for downloading videos and audio from online
18 platforms such as YouTube. The yt-dlp downloader can be used to automatically merge separate
19 audio and video streams and download entire playlists, among other tasks, and Defendant used the
20 yt-dlp downloader for these very purposes in downloading files from YouTube.

21 78. To use yt-dlp, Defendant first had to install the program, which can be downloaded
22 from the official yt-dlp page on GitHub, along with other programs necessary to effectively
23 download files, such as programs for merging video and audio files. Once installed, yt-dlp can be
24 used to download individual files or entire playlists by adding the Uniform Resource Locator
25 (“URL”), which is essentially a web address and mechanism for retrieving a specific file, for each
26 desired video to the yt-dlp program.

27 79. Defendant used programs such as yt-dlp to improperly access and download files
28 and merge data in order to feed complete audiovisual packages to its generative AI model for

1 training purposes.

2 80. Processes such as these allowed Defendant to bypass YouTube’s player page and
3 avoid YouTube’s monitoring systems in order to scrape content from YouTube, and feed content
4 directly to its generative AI model.

5 81. Defendant did not obtain the consent of YouTube, Plaintiffs or other creators of the
6 audiovisual content to conduct its scraping activities.

7 82. On information and belief, and based on the sheer volume of video data included,
8 the dataset of videos that included YouTube videos were scraped without permission from
9 YouTube and in violation of YouTube’s Terms of Service.⁴

10 83. The scraping and acquisition processes used by Defendant were inconsistent with,
11 and in violation of, YouTube’s Terms of Service, which forbid scraping and mass downloading of
12 videos.⁵

13 84. In fact, YouTube’s CEO, Neal Mohan has stated that, “From a creator’s perspective,
14 when a creator uploads their hard work to our platform, they have certain expectations. One of
15 those expectations is that the terms of service is going to be abided by,” Mohan said. “It does not
16 allow for things like transcripts or video bits to be downloaded, and that is a clear violation of our
17 terms of service. Those are the rules of the road in terms of content on our platform.”

18 85. The automated system used by Defendant to access video content from YouTube
19 was intentionally designed and utilized to circumvent YouTube’s TPMs.

20 86. To be clear, Defendant’s generative AI model is not “watching” YouTube in order
21 to be trained. Rather, data improperly accessed and downloaded from YouTube is being uploaded
22 into Defendant’s generative AI model in order to develop and improve Defendant’s products.

23 87. Defendant knew or should have known its conduct contravened YouTube’s rules
24 and the TPMs that enforce them.

25 88. Despite its knowledge, Defendant continued its program of improperly accessing

26 ⁴ “Terms of Service,” YouTube, <https://www.youtube.com/t/terms> (last viewed December 5,
27 2025) (prohibiting “access, reproduce, download, distribute, transmit, broadcast, display, sell,
28 license, alter, modify or otherwise use any part of the Service or any Content except: (a) as
expressly authorized by the Service; or (b) with prior written permission from YouTube and, if
applicable, the respective rights holders;”).

⁵ *Id.*

1 YouTube data and using the audiovisual content it improperly accessed to train its generative AI
2 model.

3 ***C. Defendant's Actions Were At All Times Commercially Driven***

4 89. Defendant's efforts to train its generative AI products, including MagicVideo, were
5 not merely for "research" purposes, but rather primarily served Defendant's own commercial and
6 financial interests.

7 90. As noted above, MagicVideo was not a research tool or isolated project, but rather
8 is Defendant's generative AI video model that envisions creators and artists as its target market.

9 91. Defendant has numerous financial interests related to its generative AI products,
10 such as MagicVideo.

11 92. Defendant is the owner of TikTok, a social media platform that allows users to
12 network through posts that include images and video segments.

13 93. Defendant maintains a website at <https://effecthouse.tiktok.com/>, where users can
14 select from among options to enhance the images and video segments they post on Defendant's
15 social media platform. Among the options are AI editing and creation that allow users to use AI to
16 alter the content they create.

17 94. Accordingly, Defendant has an interest in improving upon its AI features in order to
18 increase its market among both content creators and viewers on Defendant's social media site.

19 95. Defendant's effort to train its generative AI product is commercially-driven.

20 96. Defendant has at all times intended to create a well-trained generative AI product
21 that Defendant can integrate into its existing technology and social media-based business structure,
22 that will give Defendant an advantage in competition against its competitors in technology and
23 social media spheres, and that Defendant can leverage for commercial purposes in order to increase
24 its own profitability.

25 ***D. Defendant's Actions Caused Harm to Plaintiffs and Class Members***

26 97. At no time did Defendant seek or obtain Plaintiffs' or Class Members' authorization
27 to access and use the videos Defendant used to train its generative AI models

28 98. At no time did Defendant seek or obtain Plaintiffs' or Class Members' authorization,

1 nor did Defendant compensate Plaintiffs or the Class Members for the access, copying, ingestion,
2 and use of their YouTube videos in AI training and related workflows.

3 99. Plaintiffs and Class Members used YouTube as their platform of choice to upload
4 their video content in substantial part due to YouTube's terms and service that prohibit the type of
5 scraping, unauthorized accessing and downloading, bulk extraction, and other forms of data mining
6 of audiovisual content utilized by Defendant to obtain Plaintiffs' video content.

7 100. Defendant intentionally violated Plaintiffs' and Class Members' intent and rights to
8 their content by accessing Plaintiffs' and Class Members' video content to train Defendant's
9 generative AI model.

10 **CLASS ACTION ALLEGATIONS**

11 101. Plaintiffs bring this action on behalf of themselves and on behalf of all other persons
12 similarly situated.

13 102. Plaintiffs propose the following Class definition, subject to amendment as
14 appropriate:

15 **All persons in the United States who uploaded original videos to**
16 **YouTube and whose videos (or material portions thereof) were**
17 **included in the HD-VILA-100M dataset scraped and**
18 **downloaded by Defendant.**

19 103. Excluded from the Class are Defendant's officers and directors, and any entity in
20 which Defendant has a controlling interest; and the affiliates, legal representatives, attorneys,
21 successors, heirs, and assigns of Defendant. Excluded also from the Class are Members of the
22 judiciary to whom this case is assigned, their families and Members of their staff.

23 104. Plaintiffs reserve the right to amend or modify the class definition with greater
24 specificity or division after having an opportunity to conduct discovery. The proposed Class meets
25 the criteria for certification under Rule 23 of the Federal Rules of Civil Procedure.

26 105. Numerosity. The Members of the Class are so numerous that joinder of all of them
27 is impracticable. The exact number of Class Members is unknown to Plaintiffs now, but Plaintiffs
28 estimate that there are thousands of Class Members.

1 106. Commonality. There are questions of law and fact common to the Class, which
2 predominate over any questions affecting only individual Class Members. These common
3 questions of law and fact include, without limitation:

- 4 a. Whether Defendant used an automated tool to download or otherwise obtain video
5 content created and uploaded by Plaintiffs and Class Members onto YouTube;
- 6 b. Whether Defendant’s downloading of the video content was intended to bypass the
7 TPMs put in place by YouTube without the authorization of Plaintiffs and Class
8 Members;
- 9 c. Whether Defendant used Plaintiffs’ and Class Members’ content for purposes of
10 training its generative AI models;
- 11 d. Whether Plaintiffs and Class Members suffered damages from Defendant’s
12 misconduct;
- 13 e. Whether Plaintiffs and Class Members are entitled to damages, civil penalties,
14 punitive damages, and/or injunctive relief.

15 107. These issues are common to the Class, and their resolution would advance matter
16 and the parties’ interests therein.

17 108. Typicality. Plaintiffs’ claims are typical of those of other Class Members because
18 Plaintiffs’ video content, like that of every other Class Member, was made available on YouTube
19 and improperly altered and used by Defendant to train Defendant’s generative AI models for
20 commercial purposes. Plaintiffs’ claims are typical of those of the other Class Members because,
21 among other things, all Class Members were injured through the common misconduct of
22 Defendant. Plaintiffs are advancing the same claims and legal theories on behalf of themselves and
23 all other Class Members, and no defenses are unique to Plaintiffs. Plaintiffs’ claims and those of
24 Class Members arise from the same operative facts and are based on the same legal theories.

25 109. Adequacy of Representation. Plaintiffs will fairly and adequately represent and
26 protect the interests of the Members of the Class. Plaintiffs’ Counsel are competent and experienced
27 in litigating class actions.

28 110. Predominance. Defendant has engaged in a common course of conduct toward

1 Plaintiffs and Class Members, in that Plaintiffs' and Class Members' video content was unlawfully
2 downloaded, altered and used by Defendant in the same way. Defendant's conduct was uniform
3 across the Class: every video unlawfully accessed from YouTube was ingested into the same core
4 model that underpins Defendant's product ecosystem. This commonality further demonstrates the
5 predominance of shared legal and factual issues among Class Members. The common issues arising
6 from Defendant's conduct affecting Class Members set out above predominate over any
7 individualized issues. Adjudication of these common issues in a single action has important and
8 desirable advantages of judicial economy.

9 111. Superiority. A Class action is superior to other available methods for the fair and
10 efficient adjudication of the controversy. Class treatment of common questions of law and fact is
11 superior to multiple individual actions or piecemeal litigation. Absent a class action, most Class
12 Members would likely find that the cost of litigating their individual claims is prohibitively high
13 and would therefore have no effective remedy. The prosecution of separate actions by individual
14 Class Members would create a risk of inconsistent or varying adjudications with respect to
15 individual Class Members, which would establish incompatible standards of conduct for
16 Defendant. In contrast, the conduct of this action as a class action presents far fewer management
17 difficulties, conserves judicial resources and the parties' resources, and protects the rights of each
18 Class member.

19 112. Defendant has acted on grounds that apply generally to the Class as a whole, so that
20 class certification, injunctive relief, and corresponding declaratory relief are appropriate on a Class-
21 wide basis.

22 113. Ascertainability. Finally, all members of the proposed Class are readily
23 ascertainable. Defendant used a dataset that contain the full lists of URLs and video identifiers for
24 every YouTube video incorporated into the training pipeline. Those identifiers allow the parties to
25 determine exactly which videos were used and to match each video to its creator through
26 YouTube's public channel and authorship information. Because the dataset provides a complete
27 map of the videos Defendant downloaded, the identities of the affected creators are identifiable
28 through straightforward reference to the URLs and corresponding channel data.

CLAIMS FOR RELIEF

FIRST CAUSE OF ACTION

Violation of the DMCA (Anti-Circumvention) 17 U.S.C. § 1201(a)

114. Plaintiffs repeat, reallege, and incorporate the allegations contained in the previous paragraphs as if fully set forth herein.

115. YouTube’s Terms of Service prohibit scraping, unauthorized downloading, bulk extraction, or other forms of data mining of audiovisual content. These restrictions operate together with TPMs to prevent unlicensed access to the underlying video files, not simply reproduction of content.

116. Plaintiffs and Class Members uploaded their original video content to YouTube’s video sharing platform due in substantial part to YouTube’s Terms of Service and TPMs prohibiting bulk downloading of creators’ videos.

117. Content creators, including Plaintiffs and Class Members, expect that their works will not be accessed at the file level and copied at scale without consent.

118. YouTube’s Terms of Service and associated access controls constitute “effective technological measures” within the meaning of 17 U.S.C. §1201.

119. Defendant used automated tools for the sole purpose of circumventing YouTube’s access barriers and extracting files never made available to the public. In doing so, Defendant improperly obtained millions of videos from YouTube’s platform.

120. This distinction is critical: viewing a YouTube video through YouTube’s platform does not provide access to the underlying file. Defendant’s circumvention tools broke through that access barrier, triggering liability under §1201.

121. Defendant accessed, downloaded, stored and utilized those videos to assemble training corpora for Defendant’s AI models and services.

122. Each act of circumvention constitutes a separate violation of §1201. Plaintiffs and the Class Members are entitled to statutory damages, injunctive relief, impoundment, and attorneys’ fees and costs under 17 U.S.C. §1203.

123. Each of Defendant’s acts of infringement is a willful violation as Defendant

1 specifically utilized automated tools designed to evade YouTube’s TPMs.

2 **PRAYER FOR RELIEF**

3 WHEREFORE, Plaintiffs respectfully request a judgment in their favor and against
4 Defendant as follows:

- 5 A. For a declaration that Defendant has willfully circumvented the copyright
6 protection systems of YouTube intended to protect Plaintiffs’ and the Class
7 Members’s audiovisual content.
- 8 B. For statutory damages (up to the maximum allowed by law per violation),
9 injunctive relief, and attorneys’ fees and costs under 17 U.S.C. §1203;
- 10 C. For such equitable relief under Title 17, Title 28, and/or the Court’s inherent
11 authority as is necessary to prevent or restrain infringement of Plaintiffs’ and
12 the Class Members’ copyright-protected content, including a preliminary
13 and permanent injunction requiring that Defendant and its officers, agents,
14 servants, employees, attorneys, directors, successors, assigns, licensees, and
15 all others in active concert or participation with any of them, cease
16 infringing, or causing, aiding, enabling, facilitating, encouraging,
17 promoting, inducing, or materially contributing to or participating in the
18 infringement of any of Plaintiffs’ or the Class Members’ exclusive rights
19 under federal law, including without limitation in the content identified in
20 Exhibit A;
- 21 D. For an award of Plaintiffs’ costs and disbursements in this action, including
22 reasonable attorneys’ fees, pursuant to 17 U.S.C. § 505;
- 23 E. For an award of pre-judgment and post-judgment interest, to the fullest
24 extent available, on any monetary award made part of the judgment against
25 Defendant; and
- 26 F. For such other and further relief as the Court may deem just and proper.

27 **JURY DEMAND**

28 Plaintiffs demand a trial by jury on all claims for which trial by jury is proper.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Dated: December 23, 2025

HEAH BAR-NISSIM LLP

By /s/ ROM BAR-NISSIM
ROM BAR-NISSIM

**ELLZEY KHERKHER SANFORD
MONTGOMERY LLP**

JARRET LEE ELLZEY
TOM KHERKHER
LEIGH S. MONTGOMERY
NATISCHA VOLPE

Attorneys for Plaintiffs