

APPENDIX A

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00074729 at -730	Melanie Kambadur stated on a message chain, “I don’t think we should use pirated material. I really need to draw a line there.” The four messages that follow are redacted.	These redactions show attorney-client communications in furtherance of Meta’s decision to use copyrighted data it knew to be pirated.	October 19, 2022
Meta_Kadrey_00218170 at -170-71	Joelle Pineau responds to Eleonora Presani’s statement that “using pirated material should be beyond our ethical threshold.” Ms. Pineau then asks, “You think it’s problematic to use even for this phase?” followed by a redacted sentence. Presani then says “SciHub, ResearchGate, LibGen are basically like PirateBay or something like that, they are distributing content that is protected by copyright and they’re infringing it.”	This redaction shows attorney-client communications in furtherance of Meta’s decision to take and use copyrighted works without permission that it knew to be pirated, despite clear ethical concerns.	October 19, 2022

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00218543 at -547	This document appears to be notes from a January 2023 meeting that Mark Zuckerberg attended. It is heavily redacted, including a large section titled “Legal Escalations.” Immediately after that section the document states “[Zuckerberg] wants to move this stuff forward,” and “we need to find a way to unblock all this.”	These redactions show attorney-client communications in furtherance of Meta’s decision to use copyrighted material without permission that it knew to be pirated.	January 17, 2023
Meta_Kadrey_00171374 at -375-76	On a message thread, Sy Choudhury discusses OpenAI’s use of LibGen. An update from “the partnerships side” is redacted. On the same thread, Alex Yu asks Erin Murray, “who is the legal person that you are working with [redacted]. That is critical to our strategy.”	This redaction likely shows Meta in-house counsel’s services were utilized in furtherance of Meta’s decision to halt licensing efforts for copyrighted works and instead use pirated works.	March 31, 2023
Meta_Kadrey_00207827	In a document that appears to be Alex Boesenberg’s notes from an internal meeting, several messages are redacted within a broader section noting the challenges of licensing data and the existence datasets like LibGen which “are not fully approved for commercial settings.”	These redactions likely shows Meta in-house counsel’s services were utilized in furtherance of Meta’s decision to halt licensing efforts for copyrighted works and instead use pirated works.	April 4, 2023

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00155709 at -709-10	<p>On a message chain, Erin Murray explains that OpenAI’s model is likely trained on Smashwords and LibGen. The latter half of her message, addressed to Beau James, an in-house counsel, is redacted.</p> <p>A message from James later on the same thread is entirely redacted.</p>	This redaction shows attorney-client communications in furtherance of Meta’s decision to use copyrighted data it knew to be pirated.	April 5, 2023
Meta_Kadrey_00054419	The first message on a thread about data acquisition, which discusses “tradeoffs” of using LibGen and other “publicly available datasets” is redacted.	This redaction is likely to show Meta in-house counsel’s services were utilized in furtherance of Meta’s decision to halt licensing efforts for copyrighted works and instead use pirated works.	April 6, 2023
Meta_Kadrey_00089791 at -861.00017	Nikolay Bashlylov suggested that Meta conceal its downloading of LibGen data using a VPN (“Can we load libgen data using Meta IP ranges? Or should we use some vpn?”). All three bullet points that follow are redacted. ¹	These redactions indicate Meta’s in-house counsel was not only aware that Meta was using LibGen, but potentially also advised other Meta employees on how to conceal doing it.	April 17, 2023

¹ One “comment” on this document, Meta_Kadrey_00101591, shows some of this language unredacted and specifically alludes to the known “legal risks of using an internal tool for external datasets[.]” Another, Meta_Kadrey_00101583, shows that word “pirated” was

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00204227	Melanie Kambadur said in a chat message that “the next step is actually us starting to process libgen more thoroughly...the lawyers are [redacted].”	This redaction shows attorney-client communications in furtherance of Meta’s unlawful torrenting of pirated copyrighted works.	April 19, 2023
Meta_Kadrey_00204223 at -224	In an internal message, Nikolay Bashlykov expresses concern about using Meta IP addresses “to load through torrents pirate content,” and says, “torrenting from a corporate laptop doesn’t feel right 😞.” A response from David Esiobu is redacted.	This redaction shows attorney-client communications in furtherance of Meta’s unlawful torrenting of pirated copyrighted works.	April 21, 2023
Meta_Kadrey_00204208	In a calendar invitation from Nikolay Bashlykov, for which the agenda mentions downloading LibGen, he also asks a question addressed to “Legal,” which is then redacted.	This redaction likely shows Meta in-house counsel’s services were utilized in furtherance of Meta’s decision to use LibGen and download it via torrenting.	April 24, 2023

replaced with the word “external” by Todor Mihalov, suggesting the document originally mentioned the “legal risks of using an external tool for *pirated* datasets[.]”.

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00207830	In a document that appears to be Alex Boesenberg's notes from an internal meeting, after a bullet point saying licensing efforts have stopped, the following two bullet points discuss data acquisition priorities are partially redacted.	These redactions likely show Meta in-house counsel's services were utilized in furtherance of Meta's decision to halt licensing efforts to obtain copyrighted works and instead use pirated works.	April 28, 2023
Meta_Kadrey_00207833	There are two lines redacted on a document that appears to be notes from an internal meeting. The document references "Fair use datasets" such as Books3 and LibGen.	These redactions show in-house counsel advised employees that using pirated datasets was acceptable because it determined that using pirated datasets was fair use.	May 9, 2023
Meta_Kadrey_00209134	In this document, Sergey Edunov explains "if we license one single book, we won't be able to lean into fair use strategy. So we will have to drop all of the libgen and books datasets[.]"	This document suggests Meta in-house counsel advised Meta to stop its efforts to license copyrighted works and instead utilize pirated works exclusively.	August 23, 2023
Meta_Kadrey_0089791 at - 856	Nikolay Bashlykov writes, "using torrents would entail 'seeding' the files – i.e., sharing the content outside, this could be legally not OK" while implying Meta would direct download instead.	As of September 5, 2023, Meta knew torrenting/seeding LibGen and related sites was legally problematic.	September 5, 2023

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00211699 at -699-703	<p>This document contains admissions that Meta knew that LibGen was pirated (i.e., illegal) and expresses concern over what will happen if regulators learn that Meta is training Llama on pirated copyrighted data.</p> <p>The “Legal Risk” section is entirely redacted.</p>	<p>This redaction shows attorney-client communications and the utilization of in-house counsel in furtherance of Meta’s decision to use copyrighted material without permission that it knew to be pirated.</p>	December 13, 2023
Meta_Kadrey_00209314 at -314-22	<p>In this email chain with the subject line “FW: [A/C Priv] LibGen Approval for OneLLM,” a Meta employee states that she wants to “flag an issue that’s going to be significantly gnarly.” Then the next seven pages are entirely redacted.</p>	<p>This redaction shows attorney-client communications in furtherance of Meta’s decision to use copyrighted data it knew to be pirated.</p>	January 4, 2024
Meta_Kadrey_00211852	<p>Jelmer Van Der Linde admits, between redacted messages, that Bashlykov said he torrented the scimag data and downloaded the rest of LibGen.</p>	<p>These redactions show attorney-client communications in furtherance of Meta’s unlawful torrenting of pirated copyrighted data.</p>	January 5, 2024
Meta_Kadrey_00204235	<p>After one redacted message, Jelmer Van Der Linde says “code seems to suggest you at least tried torrenting[.]”</p>	<p>This redaction shows attorney-client communications in furtherance of Meta’s unlawful torrenting of pirated copyrighted data.</p>	January 5, 2024

EVIDENCE	RELEVANT COMMUNICATION	SIGNIFICANCE	DATE OF COMMUNICATION
Meta_Kadrey_00108336	This document shows Meta employees deciding not to use “FB [Facebook] infra[structure]” for its “data downloading” from pirated databases in order to “avoid[] risk of tracing back the seeder/downloader [] from FB servers.”	This document shows that Meta’s unlawful activity is ongoing.	April 2, 2024
Meta_Kadrey_00107954	This document shows that Meta tormented at least 81.7 terabytes of data across multiple shadow libraries through the site Anna’s Archive, including at least 35.7 terabytes of data from Z-Library and LibGen.	This document shows the volume of data that Meta tormented and that the crime is ongoing.	April 11, 2024