EXHIBIT FIVE

From: Keqian Li[/O=THEFACEBOOK/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=4611E44D0752455DBE72E6A1B1AF1A44-[redacted]

Sent: Fri 7/12/2024 8:41:44 PM (UTC)

To: Tony Cao[[redacted]@meta.com]; Keqian Li[[redacted]@meta.com]; Junpu Wang[[redacted]@meta.com]; William Held[[redacted]@meta.com]; Jacob Xu[[redacted]@meta.com]; Frank Zhang[[redacted]@meta.com]; Todor Mihaylov[[redacted]@meta.com]; Chandler Zuo[[redacted]@meta.com]; David Esiobu[[redacted]@meta.com]; Jian[[redacted]@meta.com]; Ellen Tan[[redacted]@meta.com]; Xuchao Jia[[redacted]@meta.com]; Sean Bell[[redacted]@meta.com]; Hanwen Zha[[redacted]@meta.com]; Xiaolan Wang[[redacted]@meta.com]; Kenneth Heafield[[redacted]@meta.com]

Subject: Message summary [{"otherUserFbId":null,"threadFbId":7654952347915810}]

Attachment: 449974340_443211872019299_2526685641170031849_n.png

Hanwen Zha (7/12/2024 00:00:47 PDT):
>[redacted] uhaul table only have 14 days retention. Now how can we select a list of rows' html by joining [redacted] table with [redacted] and ds ? (For ablate parsing changes, line-dedup etc.)
>
>Do we need to / Is it doable to uhaul a lot of our table to [redacted] in order to do the join?
>
>I try to uhaul but hit ACL warning, want to know what's best practice for this? @silent

Sean Bell (7/12/2024 00:07:58 PDT):
>Honestly I think we need to find a way to do everything in just one namespace. @silent

Sean Bell (7/12/2024 00:08:35 PDT):
>We have no more storage. We will request some but the idea of splitting everything in 2, and uhauling everything doesn't make sense

Kenneth Heafield (7/12/2024 06:25:21 PDT):
>Delia, last week on having one namespace "I am considering, but that is also something that will take bw and time but maybe we should start next week"

Kenneth Heafield (7/12/2024 12:31:06 PDT):
>Thoughts on $7 million for 5800 academic books from Emerald Publishing?
https://www.emerald.com/insight/browse/publications (They didn't want to sell the journals due to diffuse copyright ownership)

Sean Bell (7/12/2024 12:33:47 PDT):
>how many tokens?

Sean Bell (7/12/2024 12:34:12 PDT):
>feels expensive for likely ~1B tokens

Kenneth Heafield (7/12/2024 12:34:40 PDT):
>Don't worry, we already asked

Sean Bell (7/12/2024 12:39:04 PDT):
>Are we filtering this out of our datasets now?

Sean Bell (7/12/2024 12:39:11 PDT):
>I assume yes, and/or we don't already have it

Sean Bell (7/12/2024 12:39:28 PDT):
>we should keep a list of all of these conversations so that we can make a blocklist/filter-list against our PDFs/Books dataset

Kenneth Heafield (7/12/2024 12:43:13 PDT):
>In theory we shouldn't be starting the conversation without checking what we have already. In practice the metadata is pretty messy. And also OCLC sued Anna's Archive over said metadata...

Kenneth Heafield (7/12/2024 12:44:14 PDT):
>The #1 publisher is empty string.

Xiaolan Wang (7/12/2024 13:19:21 PDT):
>That's quite expensive. 5800 books will be roughly 1B tokens. Fuzzy match shows that we may have some of their collection?

Xiaolan Wang (7/12/2024 13:19:23 PDT):

shared: 449974340_443211872019299_2526685641170031849_n.png

Sean Bell (7/12/2024 13:24:59 PDT):
>We should check with legal **Redacted**

Xiaolan Wang (7/12/2024 13:30:55 PDT):
>Can we at least crawl their catalog so that we can dedup against what we have? Or once we started the conversation, we cannot crawl anything from them?

Xiaolan Wang (7/12/2024 13:41:44 PDT):
>Checked the categories (P1478035457). Mainly on Social science, economics & management, and very few STEM related.