EXHIBIT FOUR

Message

From: Andrew Poulton [[redacted]@meta.com]
Sent: 2/15/2023 7:30:06 PM
To: Melanie Kambadur [[redacted]@meta.com]; Andrew Poulton [[redacted]@meta.com]; David Esiobu [[redacted]@meta.com]; Hugo Touvron [[redacted]@meta.com]; Armand Joulin [[redacted]@meta.com]; Jacob Xu [[redacted]@meta.com]; Moya Chen [[redacted]@meta.com]; Peter Albert [[redacted]@meta.com]; Marie-Anne Lachaux [[redacted]@meta.com]; Guillaume Lample [[redacted]@meta.com]; Gautier Izacard [[redacted]@meta.com]
Subject: Message summary [{"otherUserFbId":null,"threadFbId":6750324401662966}]

Melanie Kambadur (2/15/2023 08:35:46 PST):
>quick question for this group: do we know if we have any poetry, short story, or other creative writing data in our corpus?

Melanie Kambadur (2/15/2023 08:36:01 PST):
>if not, might be interesting to acquire some

Guillaume Lample (2/15/2023 08:36:41 PST):
>beyond what is in common crawl, not specifically. there were some in CC-stories but this corpus was a bit noisy

Melanie Kambadur (2/15/2023 08:38:33 PST):
>would it be good to get some more do you think if we want to generate poetry?

Guillaume Lample (2/15/2023 08:39:03 PST):
>yes we should. for this we should look at datasets on Kaggle. it's full of pretty good datasets. sometimes quite large

Guillaume Lample (2/15/2023 08:39:15 PST):
>there is one with a few GBs of poesy

Guillaume Lample (2/15/2023 08:39:30 PST):
>and there are 100k+ datasets as far as I remember

Guillaume Lample (2/15/2023 08:39:38 PST):
>but we have to get a SRT approval for each one of them :D

Melanie Kambadur (2/15/2023 08:44:49 PST):
>yes but good news is we can group this into batches

Melanie Kambadur (2/15/2023 08:45:12 PST):
>e.g. we can do the stack and a poetry dataset together if we know which ones

Hugo Touvron (2/15/2023 08:48:16 PST):
>https://www.kaggle.com/datasets/tgdivy/poetry-foundation-poems poems com from https://www.poetryfoundation.org

Peter Albert (2/15/2023 09:01:13 PST):
>How does this work? Can we download and preprocess datasets first, and do the SRTs before training?

Melanie Kambadur (2/15/2023 09:02:37 PST):
>technically - no; especially if they are going onto RSC

Melanie Kambadur (2/15/2023 09:02:52 PST):
>practically: as long as the license is not super questionable

Melanie Kambadur (2/15/2023 09:03:08 PST):
>I think we can bend the rules and always delete later if we get a 'no'

Melanie Kambadur (2/15/2023 10:05:49 PST):
>Another set of questions came up on this legal risks doc: https://docs.google.com/document/d/[redacted]4e_ij3Lc/edit#

Redacted

Redacted

David Esiobu (2/15/2023 11:24:16 PST):

Redacted

Melanie Kambadur (2/15/2023 11:29:42 PST):
>If you can do it without disrupting your other work, Redacted maybe @Moya can help pitch in if not?

Melanie Kambadur (2/15/2023 11:30:06 PST):
>(context is Redacted