EXHIBIT P

Channel #chatgpt-feedback - 2023-08-05 (UTC)

August 5th 2023, 12:20 AM UTC

Adrien Ecoffet, Isaac Wolkerstorfer (Wasileski), Joshua Gross and 2 others

AE Adrien Ecoffet
Thoughts on this? https://twitter.com/davidmanheim/status/1687188005948153857 8/4/2023, 9:05 PM UTC

TS Ted Sanders
How do we know it's not leaking conversations? Are conversations not in the post-training data? 8/5/2023, 12:20 AM UTC

JG Joshua Gross
@Ted Sanders we've repro'd and traced it down to the engine level 8/5/2023, 12:30 AM UTC

TS Ted Sanders
Right, but how do we know the engines aren't reproducing conversations from post-training data? 8/5/2023, 12:31 AM UTC

IW Isaac Wolkerstorfer (Wasileski)
you mean, verbatim reproducing memorized responses from the post-training data? if there's a risk of memorizing and reproducing post-training data I would assume we wouldn't be doing that, but I am not a machine learning researcher 8/5/2023, 12:46 AM UTC

TS Ted Sanders
Yes. I saw some people identify one of these 'hallucinations' as coming from an actual comment on a website somewhere, so looked to be regurgitated from pre-training. Makes me wonder if some of these conversations are similarly regurgitated from post-training. In that sense, this would be leaking conversations but in a very time-delayed fashion. 8/5/2023, 12:48 AM UTC

1

IW Isaac Wolkerstorfer (Wasileski)
it's the same basic responses you'd get with e.g. [redacted] so, if that's a thing you can access our post-training data/real responses pretty trivially 8/5/2023, 12:49 AM UTC
I thought pre-training memorization is expected but post-training is not, but again, not my domain per se 8/5/2023, 12:50 AM UTC

JG Joshua Gross
@Ted Sanders I guess in theory possible but seems incredibly unlikely (and much harder to prove it's not happening since you'd need to do a fulltext search through all of our conversation DBs.... and some of the convos may have been deleted already) 8/5/2023, 12:51 AM UTC

it would also be *extremely* unlikely but I guess not impossible to regurgitate a conversation that was used in post-training *without any prompt,* since the way these conversations are used in post-training is to treat the model how to respond *to those conversations*

8/5/2023, 12:52 AM UTC

whereas regurgitating something from pretraining seems quite expected

8/5/2023, 12:52 AM UTC

+ 1

VB

Valerie Balcom

sometimes during streaming, my browser doesn't maintain "scroll to bottom" if a fix can't be found for this, i wouldn't mind a keyboard shortcut that would pin scrolling to the bottom during streaming

8/5/2023, 5:26 PM UTC