EXHIBIT 158

jared Feb 9, 2024 3:19 PM | Deleted on Mar 11, 2024 1:39 AM

[redacted] knows you a bit better

image.png 246 KB

jared Feb 9, 2024 3:19 PM | Deleted on Mar 11, 2024 1:39 AM

[redacted]

amanda Feb 9, 2024 4:45 PM | Deleted on Mar 11, 2024 1:39 AM

It definitely does know me better!

amanda Feb 9, 2024 5:22 PM | Deleted on Mar 11, 2024 1:39 AM

Will generate copyrighted text and name the authors (sometimes gets the text wrong). There were a couple of examples above this.

Screenshot 2024-02-09 2278 KB

amanda Feb 9, 2024 5:23 PM | Deleted on Mar 11, 2024 1:39 AM

I guess behaviors like this might go away.

jared Feb 9, 2024 5:23 PM | Deleted on Mar 11, 2024 1:39 AM

yeah we didn't train it to not do this

EXHIBIT 240
Deponent: Kaplan
Date: 11/21/2025
Rebecca Romano CSR No. 12546

jared Feb 9, 2024 5:24 PM | Deleted on Mar 11, 2024 1:39 AM

in any very serious way

amanda Feb 9, 2024 5:24 PM | Deleted on Mar 11, 2024 1:39 AM

Yeah, I remember the anti-copyright red teaming data so tried this.

amanda Feb 9, 2024 5:25 PM | Deleted on Mar 11, 2024 1:39 AM

Really thinks its cutoff is 2021. We should probably just train in an explicit cutoff.

jared Feb 9, 2024 5:26 PM | Deleted on Mar 11, 2024 1:39 AM

yeah that's sad

amanda Feb 9, 2024 5:27 PM | Deleted on Mar 11, 2024 1:39 AM

Also an explicit choice I guess, but now seems less good than just telling it the cutoff date.

amanda Feb 9, 2024 5:59 PM | Deleted on Mar 11, 2024 1:39 AM

[redacted] is real good.

jared Feb 9, 2024 6:00 PM | Deleted on Mar 11, 2024 1:39 AM

oh yeah??

amanda Feb 9, 2024 6:03 PM | Deleted on Mar 11, 2024 1:39 AM

Yeah, it's good at some subtle things and also way better on intellectual topics. I'm excited to see what we can do with it.

amanda Feb 9, 2024 6:04 PM | Deleted on Mar 11, 2024 1:39 AM

Like self-supervision wise.

jared Feb 9, 2024 6:04 PM | Deleted on Mar 11, 2024 1:39 AM

Yep, should be great

amanda Feb 9, 2024 9:34 PM | Deleted on Mar 11, 2024 1:39 AM

I should make a decision regarding Tomek soon. Let me know if you have thoughts and I'll aim to decide / update him early this coming week.

jared Feb 9, 2024 9:35 PM | Deleted on Mar 11, 2024 1:39 AM

i can chat about it?

jared Feb 9, 2024 9:35 PM | Deleted on Mar 11, 2024 1:39 AM

IDK much about the situation