



Artificial Intelligence

Google & AI ethics

**sergeivolodin**

2:18 AM

Personally, I am most confused/concerned/unhappy with the tone and the speed of what happened rather than with the main events/facts themselves.

a) the firing of Timnit seems to have been done in a "little green men" sort of way:

[https://en.wikipedia.org/wiki/Little_green_men_\(Ukrainian_crisis\)](https://en.wikipedia.org/wiki/Little_green_men_(Ukrainian_crisis)) – unofficial, fast, not making much sense so that people don't have time to think and respond properly. As far as I understand, she didn't get an e-mail and there was no discussion, and the team was not informed as well (correct me if I'm wrong, my information is from my memory of Twitter posts from last year). This seems most alarming, because as [@lenhoang](#) said it's about setting a precedent. Next time someone wants to discuss this, or any other controversial issue at Google, it's likely they just wouldn't even start – because they already know how it will end – no discussion and some fast and personal measures taken. It might be that this is the outcome that was intended [conspiracy theory mode off]... even if indeed there were mistakes/misconduct done by the ethics team, it just does not seem great for a leading innovative company to respond in such a harsh/fast manner – a series of team meetings where some somewhat-impartial manager first listens to all sides, and then still makes the same decision at the end, would be a better option because at least all sides would have the time to think, state their position and respond to criticism... this is very not in line with what I saw there – usually everything is slow and gradual.

b) the explanation by Jeff Dean was a bit weird as well – why would they make the accent on the paper



Reply






Artificial Intelligence

Google & AI ethics



is slow and gradual.

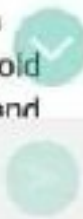


b) the explanation by Jeff Dean was a bit weird as well – why would they make the accent on the paper quality? many papers are not so great, it doesn't mean you need to fire the authors. The fact that they relied on the (very weak from my POV) internal review argument – which is usually just a routine check – makes them look strange to the employees and the people outside – as if they fire everyone who forgets citations in papers, without any discussions, and this happens every day... It is understandable that they want to hide the real reason, I'm just super confused why they chose this particular way – this way does not make much sense and it does not look good. I'm sure they could have come up with a much more reasonable explanation [they are a big company and they have the resources to craft whatever story they want], and the fact that they didn't makes it strange.

c) the general lack of response from the AIS community seems also weird to me – for me it's clear that the information that we have is not sufficient to conclude with certainty on whether any of this was a justified decision, or it was an 'evil company doing evil things'. We know only what Google lawyers allowed us to know, which is not much. And from the ethics team side, we know that they claim it's all a setup, but we don't know the details (e-mails, conversations, previous history of what happened, ...) either. In that case, the default response would be to generally favor the side of the ethics team – because it's not that hard, and it potentially brings a lot of value (in case if it was more of the 'evil' case). The drawback here is that in case if it was *not* really 'evil', people who support the team now might be on some sort of a blacklist later, as supporters of a 'wrong cause' (hold this, I will get back to it later!). As far as I understand




Reply





Artificial Intelligence

Google & AI ethics



c) the general lack of response from the AIS community seems also weird to me – for me it's clear that the information that we have is not sufficient to conclude with certainty on whether any of this was a justified decision, or it was an 'evil company doing evil things'. We know only what Google lawyers allowed us to know, which is not much. And from the ethics team side, we know that they claim it's all a setup, but we don't know the details (e-mails, conversations, previous history of what happened, ...) either. In that case, the default response would be to generally favor the side of the ethics team – because it's not that hard, and it potentially brings a lot of value (in case if it was more of the 'evil' case). The drawback here is that in case if it was *not* really 'evil', people who support the team now might be on some sort of a blacklist later, as supporters of a 'wrong cause' (hold this, I will get back to it later!). As far as I understand, AIS people say 'well, it's not really our area, so why bother'. This is not the tone of typical AIS discussions (the informal ones usually go somewhat quickly from discussing concrete problems and papers and proving theorems to discussing what shape buildings on Mars should be to optimize for well-being, and how simulated creatures in the simplicity prior are trying to rig our computations

<https://www.lesswrong.com/posts/Tr7tAyt5zZpdTwTQK/the-solomonoff-prior-is-malign> – something that we probably will never be able to prove or disprove empirically, or theoretically – all while AIS claims to be related to the very practical problem of actual AI being actually aligned, now or in the future... AIS is still not a very well-defined field, and I think if that post is in the scope of AIS, Google Ethics stuff is as well). I agree with [@Konrad Seifert](#) that people might not want to engage in things they don't fully understand – the question is, why aren't they asking



Reply





Artificial Intelligence

Google & AI ethics



be related to the very practical problem of actual AI being actually aligned, now or in the future... AIS is still not a very well-defined field, and I think if that post is in the scope of AIS, Google Ethics stuff is as well). I agree with [@Konrad Seifert](#) that people might not want to engage in things they don't fully understand – the question is, why aren't they asking questions to have this understanding... Getting back to the 'wrong cause' and a potential backfiring risks of supporting the ethics team now – asking questions and being confused is not necessarily supporting. So, I'm confused that AIS researchers in general are not confused about this and are generally not very interested in what actually happened...

So, I think we should be confused about this and we should ask for some answers, details and explanations on what happened and how Google will handle these kinds of issues in the future... The approach on the ethics team they took seems to contradict the very nature of the guidelines on ethics that they set – transparency, explainability, fairness etc. The decisions they made were not explained, the decision process was not transparent, and the explanation by Jeff Dean does not seem fair... No matter what the ethics team did/said – the Google's, as a provider of some sort of a public good, response, whatever it would be, should at least comply with their own guidelines, if not for the actual decision, then at least for the process of making that decision...

edited about 1 month ago



May 29



Reply

