United States District Court
Northern District of California
Δ DEFENDANT Δ
Case No. 4:24-CV-04722-YGR
Case Title Musk et al. v. Altman et al.
Exhibit No. DX-0758
Date Entered
Mark Busby, Clerk
By:_____, Deputy Clerk

From: Shivon Zilis <Shivon@spacex.com>
To: Elon Musk <erm@spacex.com>
Cc: Sam Teller <steller@spacex.com>
Subject: OpenAI / Possible AI scenarios
Date: Tue, 13 Feb 2018 02:17:35 -0800
Importance: Normal
Attachments: Possible_Scenarios.docx

Hi Elon,

Realize the AGI stuff is a bear to think about. When I make myself think it through it's honestly with the heaviest of hearts, so I can only imagine how difficult it is for you.

You've said you aren't sure if there is or isn't a worthwhile experiment to run for an AGI counterbalance at this point. I realize how short you are on time – so I suspended disbelief for a moment to jot down a set of possible paths forward (attached).

Please know that they are not pitches, but thought starters. Some are a bit bonkers but, as you've said, it's kind of Hail Mary time so why not at least think them through. Figured even seeing some of these scenarios and saying "hell no" doesn't hurt to help you figure out what does / doesn't feel right to you.

Feel free to disregard if this kind of thing isn't helpful, or if you don't feel like thinking about AI. The point was merely to try to take some of the weight off of your shoulders for getting to a decision.

Shivon

Possible Scenarios

1. **TeslaAI: Ace-Centric Strategy**
    - *What:* Do this entirely at Tesla, where E hires 1-2 best in class "ace" researcher cornerstones to shore up AI talent across both research and engineering.
    - *Why:* E may want to make a full court press and do it quickly. He likely needs at least one cornerstone he can trust to do that.
    - *Pros:* Already has Andrej, who isn't the cornerstone but a big piece. E could be architect vs. having to be in the weeds. This would also allow someone to be 100% time on leading it.
    - *Cons:* List of people is **very** short. Jeff Dean (Google) and Oriol Vinyals (DeepMind) are top of mind. If you could get Shane Legg (DeepMind) it would be a coup. This strategy also leaves the vision in the hands of the ace, in large part. That could be ok if it aligns well, but unsure if it will since a lot of these guys are really enjoying living the rockstar life vs. worrying about AGI.
    - *1st step:* Meet with the handful of researchers that could potentially run this. Figure out what would make them join and what their fastest path to safe AGI plan would be. (May make sense to do the latter regardless of plan – would be good to reconnect with folks like Dean)

2. **TeslaAI: Project-Centric Strategy**
    - *What:* E defines 1-2 projects that he thinks aren't being done that would lead to a faster path to safe AGI than what DM is up to. No small task but I think it's possible.
    - *Why:* Would be a good opportunity for have E think deeply about what the best technical plan and experiments are to create a true counterbalance. Would be great for him to think this through because he will either come up with something insanely amazing or conclude there is nothing worth doing at this point. In either case at least there is some semblance of resolution.
    - *Pros:* If E does come up with something, it will almost certainly be better than Demis' or OpenAI's current plans. This doesn't necessitate an ace and he can truly lead vision, which would be awesome.
    - *Cons:* Would take a lot of E's time so may have to spin up post Model 3? He'd also have to be sufficiently inspired by the project to do this version because it would be too emotionally hard to execute otherwise. Hiring is near impossible right now but I think what he comes up with could likely be done with a high ratio of engineers to researchers, which is in his favor.
    - *1st step:* **please take this with 11 grains of salt** E figures out what subset of projects he thinks are directionally correct. Perhaps mildly amusing, but my best guess for what this looks like him creating some version of the simulation that he thinks we are in. Naïve take, but just feels like smart interplay between agents and rich/well-constructed environments will be the safest of the fast paths to AGI -- and no one is doing anything particularly intelligent on the environment

front. This also plays to his strengths around gaming and envisioning complex systems. Everyone else seems more focused on building algorithms for existing static environments, would be fascinating to have the environment evolve as well. It could be a massive training accelerant, and potentially allow for richer moral and behavioral training which may lead to safer breeds of AGI than Ilya and Demis' alternatives.

3. **TeslaAI: Research -> Applications Strategy (1st Race vs. 2nd Race theory)**
    - *What:* Convince a bunch of the multi-agent / generalization type researchers to co-locate with engineers that run experiments on how to make their research useful in the real world. Try to shore up the folks doing multi-agent work with massive salaries and equity upside.
    - *Why:* Belief: at some point the suite of algorithms in research world (that are currently useless in the real world) will generalize enough to become massively useful and eclipse current applied AI techniques. If this belief is true – the probable path to getting to AGI first would be winning the race to building and transitioning research algorithms to the real world, since they'd be worth trillions pretty quickly (this is the 1st race). These are well sub AGI, but arguably the person who gets there first will have a lead and resources to win the race to true AGI (2nd race). E already has an advantage on seeing trajectories before anyone else and applying technology cleverly in the real world, so if this belief is true he has an edge to help compensate for the late start.
    - *Pros:* Fits in nicely with for-profit company structure. Aligns upside. Would slow Demis because you'd have to pull from his most valuable talent group.
    - *Cons:* If you even believe the assumptions above, this probably doesn't happen for 2-5 years. Could be hard to argue for at public company as a result, when there appears to be so much other low hanging AI fruit that a layman investor would want one to focus on. Also, could ignite an arms race if done too publicly.
    - *1st step:* Figure out what areas of AGI trajectory research are likely to transition to the real world and eclipse current methods.

4. **OpenAI Becomes a B Corp Subsidiary of Tesla**
    - *What:* Convince OpenAI to be a B corp subsidiary. I have no idea if this is possible but… brainstorming ftw.
    - *Why:* They will be shielded from normal Tesla pressures and would have the flexibility to fundraise externally specifically for mission or get internal funding.
    - *Pros:* Keep the band together. Would alleviate some of their concerns around building AGI at a public company.
    - *Cons:* No clue if they are still willing to work with them or if the structure is feasible.
    - *1st step:* Figure out of structure is possible then float by OpenAI.

5. **Altman as anchor for TeslaAI**
    - *What:* Do TeslaAI properly, but have Altman run it.

- *Why:* He is the only person I know other than Demis, Greg, and Ilya who wants to spend a significant chunk of his life on safe AGI specifically and is capable of a large-scale experiment. If everything is left as is, he will be CEO of OpenAI anyhow, so his fate is likely to run a big AI experiment in either case.
- *Pros:* Altman will try to aim big. He, of his own admission, has some skill gaps but the will is there and he is always willing to learn. E would be perfect person to guide him. As much as he and E don't have perfect trust, he is one of the few people I know who actually pushes and tells E hard truths regularly. If OpenAI doesn't have Altman they may just decide TeslaAI is the better option too.
- *Cons:* Trust continues to be an issue. Is there a way to solve that? Altman is also not a researcher / practitioner in the field so would have to be paired. Andrej a good start but he'd need more too.
- *1st step:* Develop a compelling plan, both in terms of ideas and funding, and discuss it with Altman. Alternatively, hear out Altman's plan and see if it's a fit to do at Tesla.

6. **If you can't beat 'em join 'em (or, really, influence them)**
    - *What:* Find a way to get Demis. Seriously. Maybe he comes to Tesla somehow or DeepMind is spun out. Perhaps convince him that tech backlash is about to hit Google hard (I hope it is... Google is terrifying) and that he's better off working with E.
    - *Why:* Demis really does fanboy hard and I don't think he's immoral... just amoral. If he hung around E perhaps it would force him to think about humanity more.
    - *Pros:* You definitely either slow or alter Deepmind's trajectory for good. E has not tried very hard to influence Demis of late and it would be good if he did, whether in this construct or otherwise.
    - *Cons:* Obviously can't be a counterbalance when you join forces. Probably only the right move if E decides he's too far off to be a counterbalance. (All of that said, there is an argument to be made that slowing and altering Demis' trajectory is the best outcome in the world, regardless)
    - *1st step:* Suspend disbelief for a hot second and REALLY think about it, lol.

7. **Funding and advisory**
    - *What:* E does not build something himself, nor aligns time with OpenAI, but funds the suite of things he thinks are most likely to have a positive impact. He will learn a bunch and retain the opportunity to do something when the time is right.
    - *Why:* If he can't find a way to positively frame this for himself, doing the AI thing is probably terrible for the world over all. It's so depressing if one doesn't believe they are building an amazing thing that will benefit the world so there is a risk of becoming less effective at Tesla and SpaceX and we need to leave the planet!!! He will also retain a ton of goodwill in the ecosystem for if he ever wants to do something himself.

- *Pros:* Can still be friends with OpenAI, who are likely to get 100's of millions unless they completely screw up. Can be Switzerland.
- *Cons:* E is really the only person who can be a counterbalance to Demis so that would be out. This is certainly conceding on some level as well.
- *1st step:* Hope we don't get to this one, but just acknowledging it may be the right answer under certain conditions.

8. **Stick with OpenAI**
    - *What:* Altman becomes CEO. They do an equity offering instead of a private ICO.
    - *Why:* They still hope for a world where this can happen and would welcome it. Would take a lot of the weight off E's mind but he'd have to believe they are on the right trajectory.
    - *Pros:* They already have kickass researchers. They are upskilling on the rest.
    - *Cons:* He doesn't do his own thing. Do we think OpenAI is on a counterbalance trajectory? Even if they don't do ICO there still may be reputational flak for for-profit equity thing.
    - *1st step:* Decide if OpenAI could conceivably be a counterbalance if Altman is CEO. Decide if he's comfortable with idea of Altman being CEO in general.

9. **OpenAI / TeslaAI concrete partnership and division of labor**
    - *What:* Tesla explicitly collaborates on hardware, gets help on factory/robotic automation from OpenAI. A bunch of employees are 50/50 but no one is fully poached. Would need to find a way to align some upside – would likely be in the 50/50 arrangement.
    - *Why:* Tesla can do hardware. But it will be better with a tight collaboration. OpenAI will really struggle to do hardware by themselves. World is clearly better if we share on that front.
    - *Pros:* Stay friends. You get two experiments for the forseeable future.
    - *Cons:* Falls apart if E does his own lab, so should only be done if no line of sight to that. If OpenAI acquires Cerebras, the hardware collaboration may be a moot point.
    - *1st step:* Discuss with OpenAI