

**“Our technology, our machines, is part of our humanity. We created them to extend ourselves, and that is what is unique about human beings” - Ray Kurzweil.**

The quote above was written by futurist Ray Kurzweil in his book *The Singularity is Near*. In it, he postulates that in the next fifty years, we will experience “a singularity:” a time in which computational knowledge will usurp humanity, and a superintelligent computer will design itself to become the final invention humanity will create. Perhaps to soothe his troubled soul, Kurzweil writes that computers are an extension of ourselves and that this singularity will benefit humanity just as a generation of super-geniuses would. But I believe, at least in the modern day, that the limb of technology is gangrenous. Left untreated, we will succumb to the infection of ignorance plaguing this cold appendage. Our doctors, the scientists meant to decide the pursuits of algorithms, stand idle, salting the wound while the cure lies in the hands of the ignored Philosophers. No single instance has relayed this injustice stronger than the recent “Facebook Scandal,” where engineers not only cut the wound but stood idle as infection continued to spread.

In mid-October, Frances Haugen testified before a senate subcommittee, providing documents supporting her case against a company that had chosen years of intentional ignorance. Surrounding her arguments was a core theme of Facebook’s continued “profits over people” mindset. Essentially, the loss function Facebook (among other companies) posed was simple, general, and inane: keep people on our site as long as possible. The “genius technicians” of Facebook, in all of their sagacity on the human brain and mechanical wit, decided to pose one of the broadest, most profit maximizing, and frankly most damaging functions to a machine with influence over millions of people. This artificially intelligent algorithm, to no surprise of my own, found the creative solution of politically radicalizing the masses and damaging the mental health of those with eating disorders and tendencies for self-harm and suicide for the sake of a little more screen time and a little more profit. Facebook also showed a resounding apathy towards violence in countries outside of the United States: while 90% of Facebook’s audience is outside of the US, 90% of the safety budget is allotted inside of the US. To quote Haugen, “in places like Ethiopia, it's literally fanning ethnic violence.” And even more astounding than the posing of the function was the response to it: to do nothing at all. This response suggests a general theme of Facebook willingly choosing ignorance and apathy in the face of a decision.

While my opinion on the matters of Facebook has perhaps been made clear by my wording, I will address the counterclaims Facebook and their technicians have made in the face of their obvious failings. The first of which, the strongest of their claims, is that the documents used by Haugen were stolen. Herein we must consider the moral and just argument that claims made through unjust means must therefore be unjust. Stealing confidential information, even for the purpose of justice, violates Kant’s categorical

imperative, suggesting that stealing this information is immoral. But is stealing for the sake of justice just? We reach our first crossroads that I believe is solved with a combination of Mill's flexibility in Justice and Nietzsche's view on punishment. Haugen was acting on a conviction of personal justice that overpowered the justice of stealing outlined through the law. Mill (and I) would argue this act demonstrates true justice. Nietzsche would also argue that Haugen held Facebook accountable and punished them to dissuade them from future unjust actions. I believe this also demonstrates justice.

The second of the popular counterarguments against Haugen is based on questioning her credibility as a computer scientist. To quote one engineer "She didn't know how basic Stacks worked." Now disregarding that many data scientists in the modern age who focus on machine learning (like Haugen) may never need to look at a basic stack in their career, this claim underlines a serious threat to the Philosophy of morality and justice as it intersects with technology. Those working on technical models like the ones used by Facebook are more concerned with technical feats than with any consideration of humanity. Whenever a claim of some moral weight is posed against many of today's computer scientists, they attack the problem from a technical lens, seeing such philosophical dilemmas as an affront to technical innovation. What they fail to realize is the accountability they hold as moral agents in our society. In the rest of this essay, I shall outline the responsibilities of our technicians, the role Philosophers play in their pursuits, and techniques we may use to balance humanity with technical innovation.

I am not the first Philosopher to criticize the works of engineers as unjust and immoral. But many humanitarians overlook some of the failings of modern-day algorithms, failings which I hope to clarify here. The first of which is that computer scientists have a duty to "program in" failsafes and stoppers on algorithms, which should be created in a "veil of ignorance" of Rawlsian proportions. Unfortunately, though, the issues in combining the chaotic world of moral dilemmas with the logical world of Artificial Intelligence have resulted in the dreadfully unanimous acceptance of Utilitarianism as a placeholder for the vastness of all ethics and justice has to offer. In its defense, Utilitarianism has an elegance unmatched by even the simplicity of the Categorical Imperative or the blindness of the "veil of ignorance" to the point where even a machine can understand its laws at face value. It requires no access to human empathy, a problem that I believe will plague machines far after the singularity (the point at which Artificial Intelligence can design itself). The trouble with Utilitarianism is that its simplicity is also its downfall. One should be wary of any Philosophy which treats people like numbers and is characterized by crippling apathy. Ignoring the fact that it often permits killing, I found a more nuanced critique of the school of thought. Imagine we could image the brain in such a way that we could measure one's "capacity for happiness." Let us assume that Plato's argument holds that those who are treated justly and behave justly are happier. Would one with a greater capacity for this happiness be more deserving of just

treatment in society, since they improve the net happiness of society to a greater degree? Under Utilitarianism, one would be forced to say “yes.” Indeed, under an extreme form, a response to the loss function “improve the net happiness of society to the greatest extent” could be to “get rid of” those suffering from depression, chronic pain, or even a tendency for migraines! Hopefully one might see now why Utilitarianism in the presence of a loss function can be so dangerous. Unfortunately, I can provide no solutions as to how to improve the morality of machines. I would love to attempt such problems, but any moral philosophy past Utilitarianism requires a deep understanding of humanity. As the human brain is much better at this way of thinking as of late, perhaps a better solution is to control what we can: our loss functions.

I believe our attitude towards loss functions carries itself with some unseen vicissitude. We must take more seriously the Philosophers and theoretical scientists who push for change in the way we prompt machines. I read the most fascinating piece about a year ago, which posed a prompt that struck me so deeply, I might argue that it is the solution to our troubles with Artificial Intelligence if it were to be taken seriously by researchers in the field. The prompt is known as “The Midas Problem,” modeled on the Greek myth of King Midas, who was punished for his foolish wishes to greater powers. The problem goes: if we can decide on a universal desire for humanity without the possibility of misinterpretation, we may then design a superintelligence to maximize this desire without the need for a human level of empathy or morality. Essentially, empathy and morality are “baked into” the proposal.

I would be a fool to attempt to solve this problem on my own. My very Philosophy prohibits me from making such broad sweeping generalizations about people and their desires. My only attempt at a solution is to assemble people from all walks of life. Perhaps --and I even hazard a “most likely”--, we will never find such a desire. Indeed, the relativistic nature of desire seems to prohibit any commonalities among all people. In this case, I would suggest computer scientists pursue specialized problems for sub-communities of humanity. No grand gesture should be done in relativistic chaos. But I have yet to guide computer scientists to a future where they will think more broadly about the implications of their algorithms and the accountability they hold in a just society. This thinking begins as most thinking does: in education.

I see two routes we may take in the development of Computer Science education that may liberate us from a dire future. The first is the promotion of a humanitarian education for students in the “STEM” field. Some institutions of higher education somewhat promote this goal, but most do so from artificial angles. Imagine walking along a concrete path for your entire life. Perhaps as you move along, you develop some strategies for which to progress faster such as making a bike. Then, 18 years later, the path suddenly has some roots and gravel. Your bike wheels bump and hop over the new terrain, and you find yourself falling and wishing you could progress as fast as you once were. One might view the roots

as detriments, hoping to return to the “basic stacks” of the concrete that once promoted ease of progression. But now imagine those roots have been there since you started your travel. Walking on and adapting to this new surface would be slower, but undoubtedly you would become stronger, plastic, and more curious as to the ins and outs of the surface on which you reside. I understand we have an interdisciplinary elementary school education system, but I hope we may develop a love for the roots that extends beyond adaptation and into appreciation. A core disadvantage in such a society would be the slowing of certain types of learning. I understand the desires of modern-day computer scientists (mostly because I have felt them too): the slowing of scientific progress just doesn’t inherently feel good. While I have already questioned these desires prior to this paragraph, I support and respect the feelings of my peers, and so I shall outline a middle ground to conclude.

We may promote the actions of interdisciplinary communication among specialized communities. The advantages to such a society are similar to the advantages found in a multi-tool. Perhaps one community develops wheels to move on cement, another invents spikes to climb trees, and yet another builds large tracks to move over sand or snow. With communication between members of these communities **without the presence of a power imbalance**, we may develop “spike-wheel-track” shoes that do a decent job of dealing with everything. Importantly, a multi-tool is not in and of itself a tool. Just as I would not replace all the spoons and forks in my house with sporks, I would not suggest multi-tools in place of fully developed tools in matters of purely technical or humanitarian pursuits. Overall, I consider this purpose a weaker option, but it will at least lead to some middle ground among the isolated communities.

Ironically, Facebook has attempted to connect isolated communities since its conception. They now plan to mitigate their controversy through rebranding to “Meta,” and building a “Metaverse,” a concept adopted from the dystopian science fiction novel *Snow Crash* by Neal Stevenson. In the novel, the metaverse facilitates social interaction, serves as an escape from the increasingly dystopian reality, and exacerbates the economic divide between those who can and cannot pay for the service. Yet Facebook pushes forward, for reasons which I believe are thoroughly unjust. For one, facilitating social interaction places an inordinate amount of power in the hands of Facebook (or Meta, or whatever I’m supposed to call them now), power which this company has never handled well in the past (ie. fueling genocide and data mining). As of yet, Facebook has no plan in place to protect the new types of data that will be available for collection through the Metaverse. This data includes conversations, facial scanning, movement habits, and aesthetic tendencies, among others.

Even worse than Facebook’s lust for power and data are their apathy, and I believe the Metaverse will serve as an icy monument to this. As I have mentioned, Facebook facilitates text interactions with a lack of censorship and even an encouragement of violence, leading to mass murders, suicides, and riots. If

the programmers of this company are unable to empathize with the power of the written word on a person, how can we expect this company to understand the impact of a full conversation with retained anonymity? Some optimism is held in vain for the Metaverse that I simply cannot share. I rather prescribe to Nietzsche's *schadenfreude* in matters involving the internet and anonymous pretenses. I remember watching a video of a "drawing room" in the virtual reality program AltspaceVR. Altspace has similar intentions to the Metaverse: to create a platform to facilitate virtual socialization. A certain individual noticed that the erasers in this drawing room could erase ANY drawing in this room and not just his own. And so he recorded himself running around the room, erasing everyone's drawings to the dismay of all. And I'm ashamed to admit this, but I laughed at this video before feeling guilty for finding humor in it. The combination of interactivity and anonymity breeds a perfect storm of human apathy. And now we must question: what sort of apathy does it take to facilitate the apathy of others? That menace with the eraser turned that video into a series. Twenty videos last I could find. It feels to me that this is useless violence, a true evil.

And the facilitation of such evil is, I believe, at the crux of Facebook's chosen ignorance. These oversights, in both Facebook's past and present, are in some ways microcosms of the failings of our attitude towards science. We could claim, as we often do, that the bureaucrats of Facebook are responsible for the dehumanization of the user base and the constant "money-grubbing" at the expense of lives. Indeed, such a characterization is comfortable, because it's rather easy to hate a businessman and much more difficult to criticize a scientist. But if we truly hope to provide for a better life in our scientific pursuits, we must hold scientists accountable for foresight on the basis of humanity, a practice that I believe can be learned through the means of education provided above. I don't mean to end this essay on a note of optimism. On the contrary, I am terrified by the crippling apathy of those meant to build our future. But I also believe bitter optimism is the only way to progress if we wish to avoid future failings by companies such as Facebook. As Zuckerberg builds a "utopian metaverse" in true supervillainous fashion, I can only hope in vain that an advising, worldly, educated Philosopher remains in the room with him, cooling the feverous heads of those wounded by Facebook's dastardly infection.