# **ML1000**

**Individual Assignment**

22nd March 2019

Salman Amin

1. **Unemployment. What happens after the end of jobs?**

Geoffrey Hinton, one of the pioneers of Deep Learning, is famous for saying that “It’s quite obvious that we should stop training radiologists[[1]](#footnote-1)” That is obviously not good news for aspiring radiologists but it is a triumph for human intelligence. Part of what makes us human is that we constantly strive to improve the quality of life. This need for constant innovation has resulted in advancements in technology in all fields over time. Advancements in medicine, communications, transportation, and automation are achievements humanity has collectively with each generation making incremental contributions.

The trucking industry is no exception. It became an industry with the proliferation of the combustion engine, which essentially automated “horse power.” While it did destroy many jobs in industries such as the horse-drawn carriages, it created many more by giving birth to new industries and improved the overall quality of human life. That is true of any advancement in technology as we have seen generational shifts from farming-based to service-based economies. Automation, therefore, is not a creator of unemployment, but another milestone in the never-ending journey of humanity’s advancement. The benefits from this shift will only benefit society and create new industries and avenues for further advancement.

1. **Inequality. How do we distribute the wealth created by machines?**

I am a strong believer in Adam Smith’s classical economic theory of the “Invisible hand.” This theory, among other things, implies that market forces are the best allocator of resources and will naturally find their equilibrium if left alone to do so. This theory also implies that we need not worry about a fair post-labor economy as market forces will allocate more resources to those who contribute the most to society through their innovations, as long as they play fair and are following the rule of law enforced by governments.

Silicon Valley entrepreneurs and tech idols such as Bill Gates and Jeff Bezos get a lot of media attention because of the wealth they have accumulated. However, it is important to note that the richest two Americans of all time, measured in today’s dollars, were John D. Rockefeller and Andrew Carnegie[[2]](#footnote-2). They were pioneers in their own right that contributed to the railroad industry and the industrial revolution. Would America as a country have generated so much wealth and become a nation of entrepreneurs if John D. Rockefeller’s and Andrew Carnegie’s wealth was taken away from them?

1. **Humanity. How do machines affect our behavior and interaction?**

Tech addiction is a serious concern. Although technology has accrued several benefits to society, there are negative side effects as well. For example, a recent study showed that social media has also been proven to cause depression and loneliness[[3]](#footnote-3). Similarly, it is quite unfortunate that the increasing intelligence of bots has resulted in better exploitation of human behaviours and emotions. Taking a holistic view of the effects of increased use of technology, I believe that they can be categorized along a spectrum of the good, the bad, and the ugly.

While there are a lot of good use cases, there are several “bad” ones as well as the case of addiction to technology and exploitation of human behaviors. Organized criminal hackers would fall in the “ugly” category. While dealing with the “ugly” requires good law enforcement, dealing with the “bad” is not as black and white. It is also unfortunate that in most cases, exploitation of human emotions and behaviors results in direct financial benefits while using these tactics to nudge society in a positive direction does not. Educating and raising awareness of these issues would be helpful but not completely eradicate these issues. Nevertheless, it is the responsibility of society as a whole, especially those who are knowledgeable in these areas, to do their part and educate others to raise a more responsible and ethical generation.

1. **Artificial Stupidity. How can we guard against mistakes?**

This is a very interesting paradox so let us consider the problem space before considering possible remedies. All systems, natural or those made by humans, have flaws. Whether we look at humans, weather, automobiles, deep learning neural networks, or a city’s transportation system. Their flaws can either be known or unknown. If the flaws are unknown, there is nothing we can really do about them. If they are known, they could either have negative consequences or positive ones. For example, having a defective gene may prevent a person from contracting another disease[[4]](#footnote-4), and hence have a positive impact. So what we really need to worry about are known flaws that cause negative impacts.

There are several ways to deal with these negative impacts but none of them would guarantee complete protection. First, we can consider checks and balances, depending on context, to ensure that systems are not being manipulated and are performing as planned. Second, we should ensure that people are using they systems appropriately and not providing biased or invalid input in the. Third, as previously unknown flaws become known, we should test their consequences and ensure that negative impact is being minimized. Finally and most importantly, we should ensure continuous learning and ongoing optimization to avoid chances of random errors and intentional manipulative input and minimize chances of unreliable output. Other than that, we can hope that one day all systems will be fool proof too.

1. **Racist Robots. How do we eliminate AI bias?**

The impact of bias in AI can be significant depending on the use case. For example, using AI to predict future criminals, or people’s sexual orientation, and worse, to determine how much prison sentence someone should receive, could be detrimental. While these are interesting applications of AI that may offer some unique insights, their application to real life may not be appropriate.

In my opinion, it is ultimately the responsibility of the researchers to remove any kind of bias in their models, and for practitioners to ensure that the models are being applied in an unbiased manner. This dilemma is quite analogous to the one faced by pharmaceuticals and doctors, especially for highly addictive drugs such as opioids. The pharmaceutical companies need to ensure that their drugs are as less addictive as possible with minimal side effects, while clearly disclosing any harmful effects for patients. The doctors, on the other hand, are responsible for ensuring that they are prescribing those drugs appropriately and monitoring the patient for side effects and potential abuse. A similar framework could be prescribed for developing and deploying AI models as well.

1. **Security. How do we keep AI safe from adversaries?**

This problem can also be considered with an analogy. It is quite similar to preventing nuclear weapons getting in the hands of rogue or evil regimes. And if these regimes are somehow able to acquire these, the task is to minimize the possibility that it is ever used, or putting in defence mechanisms to minimize the damage if they are every used. While the use of these technologies by the wrong actors can be extremely harmful, the benefits they accrue to the possessor and the power it bestows upon its owners attracts more bad actors to try and get it. Organize cybercrime is already a reality and there is no reason to assume that an arms race over AI will not ensue anytime soon.

This is also an area where the forces of good must prevail by forming alliances and collaborating to put in policies and defence mechanisms to prevent harmful attacks from occurring. It will probably never be possible to completely eliminate the possibility of these technologies getting into the wrong hands, but it can be made extremely difficult through robust monitoring, counter-intelligence algorithms, and strict enforcement. The battle between the forces of good and evil will never end.

1. **Evil genies. How do we protect against unintended consequences?**

First of all, I believe that humans should always maintain their superiority by ensuring that they understand the input and the output of these complex algorithms. Not doing that means that we have set ourselves up for failure anyways. Second, thorough testing and evaluation of these models should always be maintained to ensure that they are not producing unintended consequences.

Hypothetically speaking, if there is still an evil genie despite all of these control measures, we should also establish an AI police force that would monitor and act against rogue, manipulative, or simply stupid algorithms. These police algos would be trained and tested to spot and predict the end goal of these potentially harmful algorithms and either stop them before causing any harm, or eliminate them after they have committed the harmful act to minimize further damage. With a good cop on our side, we need not fear the evil genie.

1. **Singularity. How do we stay in control of a complex intelligent system?**

Singularity may have been a concept of fiction in the past but we are moving closer and closer to it with advancements in AI and computational power. Academics are actively researching Artificial General Intelligence (AGI) and achieving artificial superintelligence does not seem to be out of the realm of possibilities. On top of that, pioneers such as the late Stephen Hawking and Elon Musk have warned that “AI poses an existential threat to humanity.[[5]](#footnote-5)”

I am strong proponent of human intelligence and believe that we will continue to dominate all life forms. If superintelligence did become a possibility, then rational human agents should also be smart enough to develop counter-intelligence bots that would limit the power of any machines that evolve towards a path that could overpower their human counterparts. For every new innovation in missile technology, there is also similar innovation in anti-missile defense technology. We should apply the same approach to AI, especially as its military applications grow. In other words, we should have a defence against the terminator before it is deployed to production.

1. **Robot rights. How do we define the humane treatment of AI?**

This is a loaded dilemma with religious, and hence, political aspects to be considered. I would personally take a very conservative approach and reject any consideration of humane treatment of AI, or any machine for that matter. First of all, machines cannot be considered natural entities and so they should not have similar rights and protections as living beings, let alone humans. They should not even be considered in the same group as plants and trees since machines are not living things. This would remain true even if machines are one day able to achieve consciousness. This would also be consistent with the views of many religions due to the absence of a soul.

Second, the only ethical dimension that would apply here would be to ensure fair and legitimate use of AI algorithms. For example, AI algos should not be manipulated or mistreated than what their true intended purpose is. Considering any protection beyond that would be unfair to humanity and would ensue a never-ending debates. For example, should they ever be allowed to vote? How about run for elected office?Before we get there, I think there are a lot of problems we can solve that will stop the inhumane treatment of humans first. What about world hunger?

1. **Post Privacy era. How do we define and protect against privacy in the age of machine learning?**

As the examples demonstrate, privacy is already a fallacy as we don’t even know where our data resides, how it is being used, or how it can be used. The regulatory bodies are always behind the curve and lobbying efforts by the big firms block any attempts by civil rights groups to get the government to introduce laws that protect the general public. We definitely need better solutions and creative business models that can defend us against invasion of our private data.

The GDPR regulation in Europe is an effort in the right direction, but its not sufficient. We need more services that could act like a shield against illicit breaches of privacy, sort of like an antivirus does against malware. The data privacy tools introduced by Microsoft in Office 365b is an example[[6]](#footnote-6). These services could also have intelligence built into them to manipulate or misdirect algorithms that violate our privacy. Unfortunately, these services are not a reality as there is no incentive structure for them to deployed. It would require a combination of education and awareness to create demand for such services, government support to incentivize their use, and private companies to innovate, create, and market these services.

1. <https://www.economist.com/leaders/2018/06/07/ai-radiology-and-the-future-of-work> [↑](#footnote-ref-1)
2. <https://en.wikipedia.org/wiki/List_of_richest_Americans_in_history> [↑](#footnote-ref-2)
3. <https://www.sciencedaily.com/releases/2018/11/181108164316.htm> [↑](#footnote-ref-3)
4. <https://www.utoronto.ca/news/bad-genes-not-always-bad-news> [↑](#footnote-ref-4)
5. <https://en.wikipedia.org/wiki/Open_Letter_on_Artificial_Intelligence> [↑](#footnote-ref-5)
6. <https://products.office.com/en-ca/business/office-365-trust-center-privacy> [↑](#footnote-ref-6)