# Limitations of AI: the case against singularity

Víctor Santiago González

Keywords: *Artificial Intelligence, Singularity, Manifold Hypothesis, Machine Learning, Adversarial examples, Neural networks, Limitations*

**Abstract**

Using the excuse of Ray Kurzweil's *singularity* predictions, this essay assesses some of the limitations of current machine learning developments. Concerns about the actual process of developing a machine learning system and a few theoretical limitations grounded in recent research are exposed in order to argue against Kurzweil's predictions.

## 1 The singularity about to come

The *singularity* is a term first coined by Vernon Vinge and popularized, later on, by Ray Kurzweil in a series of books published by early 2000's. [9, 11]

It refers to a point in technologial progress (particularly in the field of Artificial Intelligence) where **machine's capabilities will overcome all human intelligence combined.**

Vinge's article hesitates to asses singularity's feasibility (He isn't 100% sure but "If it can happen, It will", he concedes), Kurzweil stance is significatively more bold: *he actually made predictions.*[1]

Regarding artificial intelligence, the steps that will pave the road to the singularity are[15]:

2009 Speech-to-speech automated translation will be available in cell phones.
2017 Computers will be ubiquitous, smaller, integrated in our clothes and some of them sel-organized.
2017 Full inmersive virtual reality will be available.
2018 10TB of memory storage (roughly the human brain capacity) will cost less than $1000.
2020 A computer is expected to pass Türing's test.
2023 $10^{16}$ calculations per second will be possible in a cheap machine.
2029 Computers will have achieved human-level intelligence.
2025 Military-grade UAV's will be 100% autonomous.
2045 **Singularity**. Artificial Intelligences will became the smartest and most skilled creatures in earth.

Kurzweil's confidence is built on top of "*The law of accelerated returns*", [10], which states that:

> *"Rate of progress of an evolutionary process increases exponentially [and] technology is such another evolutionary progress"*

Reallity seems to follow Kurzweil's predictions: it is possible to buy a 10TB Seagate Barracuda for $\sim 400€$, Virtual (or augmented) reality systems have become popular in the last few years, Machine's translation has seen spectacular[2] advances...

And inspite of, in practice, every exponential growth is exponential until is not, Microprocesors industry have conformed Moore's Law (an equivalent formulation of the same principle) for decades.

[1] It is relevant noticing that Mr. Kurzweil, Google's director of engineering, is a reputated futurist. Thus, we can see the predictions listed above as a curated summary on Artificial Intelligence's *state of the buzz*

[2] It's not only Google Translate new arquitecture, it worth to mention Waverly Labs's earphones[12], a new gadget that promises live and wearable machine translation.

But most of machine learning's whispering are built on top of modern neural network capabilities. After the publication of backpropagation [13] algorithm, and the rise of cloud computing, neural network applications have grown rapidly. Nowadays, finding impressive examples of neural networks adopting nearly-human behaviours is susrprisingly easy:

- They can *dream*.[3]
- They describe pictures.[1]
- They can code a website by their own.[7]
- They are playing video games.[8]

**Figure 1:** *A neural network dreaming*

The temptation to attribute human behaviour to AI systems is stronger today than it ever was. Therefore it looks that we should agree Mr Kurzweil: the singularity is near.

Is it?

## 2 Something is rotten in *the state of the art*

The problem of antrhopomorphizing machine learning applications is that, often, those statements come from marketing departments or reporters who've ever trained a model. They haven't realized (or deliberately omit) an obvious flaw that all the algorithms and techniques share in practice: **they don't train, they are trained by humans.**

Actually, there is an incredibly ammount of human work in every machine learning project that comes from human labour, including:

- **Data preparation:** every model has his own requirements about data format.
- **Data cleansing:** most algorithms assume that their training phase will be conducted over a perfectly curated data set: *zip codes that actually refer geographical information, missing data to be imputed...*
- **Context awareness:** That's specially important in the case of neural networks. Way before running the first epoch, the researcher or data scientist, must figure out how to express problem's logic in a language that the network can deal with. Sometimes this represent a straightforward requirement, sometimes is nearly impossible.
- **Disambiguation, cultural biases:** Machine learning inherits [4] all sort of biases included in the training dataset. It's human responsability to prune them.

Event thought their inherent difficulty and prevalence, the aforementioned arguments could be considered *low-level* tasks in the context of artificial intelligence. Not a real concern for the *singularity* concept.

Specially if we assume Kurzweil's full thesis, i.e., that the singularity will reach by a mixed human and artificial intelligence. What he called *human 2.0* would be a human body empowered with the benefits of the kind of *learning* we associate with current machine learning or deep learning systems.

On the one hand, this scenario would overcome all the problems mentioned until now: the human brain will still be there to manage them. But, in the other hand, all what w we did adopting the *human 2.0* argument is changing the perspective. Now the questions look different.

**How reliable this brand new learning is?**

Sadly, in practice, even slightly departures from the nature of training data could turn the finest tuned models into completely absurd predictors.

Recent research on "adversarial examples"[14] has shown how adding noise (inperceptible to human eye) to well classified images can substantially modify the output.

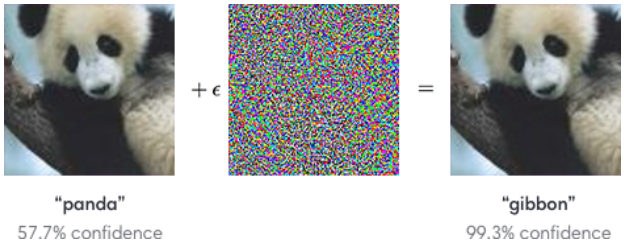


**Figure 2:** *Adding noise ruin the predictions*

The system starts, suddenly, saying that panda bears are gibbons; [2] or refuses to recognize image's contents depending on the process followed to capture them.

**Deep learners, learn. But they don't understand.**

**You want me to become a cyborg, right? Does it worth? Will it add something fundamentally valuable to my current intelectual toolbox?**

Regarding this question (which, basically, inquires on what are the theoretical limitations of machine learning) we should consider John Lauchbury's "*manifold hypothesis*" [5]

In short, this hypothesis states a generalization to a high-dimensional space of the well-known linear separability problem. Where a simple perceptron isn't able to *learn* a non-linearly separable group of clusters in a given dataset; the extension to multidimensional spaces considers that clusters adopt the form of manifolds and entangled manifolds that cannot be separated produce datasets that cannot be learned.

This geometric stance links the problem of deciding when culstering (or "classification" or "learning") can be performed over a general dataset with the mathematical field of topological manifolds, particularly with the question of entangled manifolds separability.[6]

Both questions lack from a general answer. But the adoption of this *geometrical* view fills one important gap in some machine learning (deep or not) setups: the explainability of the models after training.

Lookgin through the topology glass, neural networks have only one job: to perform continuous transformations between spaces.


**Figure 3:** *Artificially generated caption: toothbrush = bat*

And humans are overwhelmingly good at it. That's why we'd never confuse a bat with a toothbrush.

## 3 A world of quiet machines

At this point, our assesment of machine intelligence is quite negative: they aren't capable to perform the most basic tasks they need to start working, they can't contextualize their scope. They aren't understanding the problems they claimed to had learnt. And they aren't using any magical reasoning, but basic geometric transformations.

Still, someone could argue (maybe Mr Kurzweil himself) that some systems are outperforming humans in some areas: Didn't Deep Blue win Kasparov? Didn't IBM's Watson win Jeopardy?

No, they didn't.

None of these machines decided by its own to train his skills in order to participate in the contest. Humans did.

And this is a big *no*. Because, in my opinion, what truly defines intelligence is

> the ability to come up with questions, not the ability to provide answers.

Chances are that the post-singularity world (if any) will be populated by millions of super-skilled machines quietly waiting for instructions. Doing nothing.

## References

[1] Jason Brownlee. *8 inspirational applications of deep learning*. 2016. URL: https://machinelearningmastery.com/inspirational-applications-deep-learning/ (visited on 2016).

[2] Ian J. G. and Christian S. Jonathon S. "Explaining and Harnessing Adversarial Examples". In: *Arxiv.org* (2015). URL: https://arxiv.org/abs/1412.6572.

[3] Deep dream generator. *Deep dream generator*. 2017. URL: https://deepdreamgenerator.com/ (visited on 2017).

[4] The Guardian. *Google says sorry for racist auto-tag in photo app*. 2015. URL: https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app (visited on 2015).

[5] DARPA John Launchbury. *A DARPA perspective on artificial intelligence*. 2017. URL: https://www.darpa.mil/attachments/AIFull.pdf (visited on 2017).

[6] DARPA John Launchbury. *Neural Networks, Manifolds, and Topology*. 2014. URL: http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/ (visited on 2014).

[7] Andrej Karpathy. *Software 2.0*. 2017. URL: https://medium.com/@karpathy/software-2-0-a64152b37c35 (visited on 2017).

[8] Slava Korolev. *Neural Network to play a snake game*. 2017. URL: https://towardsdatascience.com/today-im-going-to-talk-about-a-small-practical-example-of-using-neural-networks-training-one-to-6b2cbd6efdb3 (visited on 2017).

[9] Ray Kurzweil. *The age of spirtitual machines. When computers exceed human intelligence*. Viking Press, 1999.

[10] Ray Kurzweil. *The law of accelerating returns*. 2001. URL: http://www.kurzweilai.net/the-law-of-accelerating-returns (visited on 2001).

[11] Ray Kurzweil. *The singularity is near. When humans trascend biology*. Viking, 2005.

[12] Waverly labs. *Waverly labs pilot earphones*. 2018. URL: https://www.waverlylabs.com/pilot-translation-kit/ (visited on 2018).

[13] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536. URL: doi:10.1038/323533a0.

[14] Christian S. et al. "Intriguing properties of neural networks". In: *Arxiv.org* (2014). URL: https://arxiv.org/abs/1312.6199.

[15] Computer Worldk. *Interview with futurist Ray Kurzweil*. 2007. URL: https://www.computerworld.com/article/2477417/smartphones/the-kurzweil-interview--continued--portable-computing--virtual-reality--immortality--and.html (visited on 2007).