SEPTEMBER 7, 2017 — ESSAYS

[FoR&AI] The Seven Deadly Sins of Predicting the Fu of AI

rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/

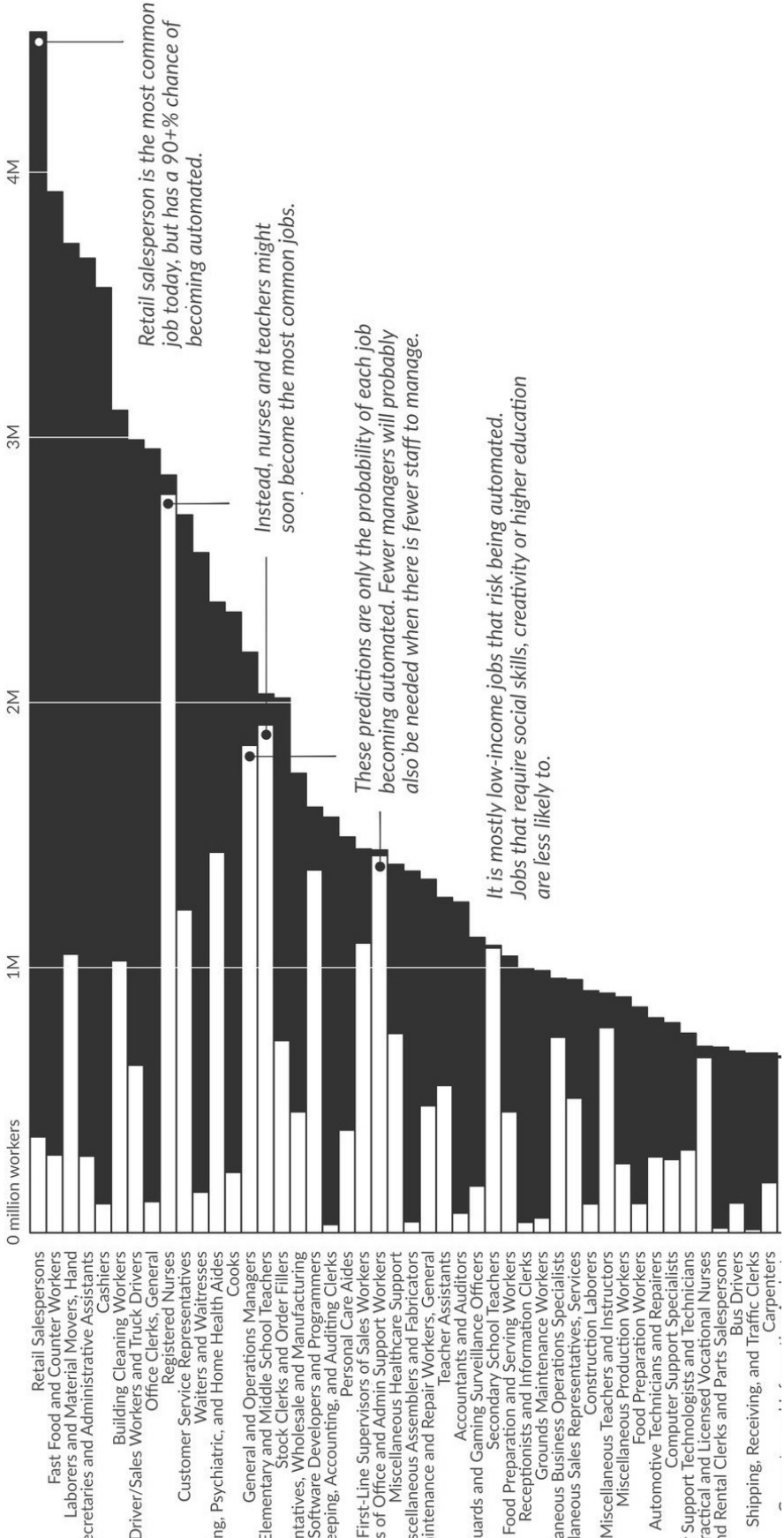
[An essay in my series on the Future of Robotics and Artificial Intelligence.]

We are surrounded by hysteria about the future of Artificial Intelligence and Rob There is hysteria about how powerful they will become how quickly, and there i hysteria about what they will do to jobs.

As I write these words on September 2nd, 2017, I note just two news stories fror last 48 hours.

Yesterday, in the New York Times, Oren Etzioni, chief executive of the Allen Insti for Artificial Intelligence, wrote an opinion piece titled How to Regulate Artificial Intelligence where he does a good job of arguing against the hysteria that Artifi Intelligence is an existential threat to humanity. He proposes rather sensible wa thinking about regulations for Artificial Intelligence deployment, rather than the chicken little "the sky is falling" calls for regulation of research and knowledge t we have seen from people who really, really, should know a little better.

Today, there is a story in Market Watch that robots will take half of today's jobs to 20 years. It even has a graphic to prove the numbers.


0 million workers
1M
2M
3M
4M
Retail Salespersons
Fast Food and Counter Workers
Laborers and Material Movers, Hand
Secretaries and Administrative Assistants
Cashiers
Building Cleaning Workers
Driver/Sales Workers and Truck Drivers
Office Clerks, General
Registered Nurses
Customer Service Representatives
Waiters and Waitresses
Nursing, Psychiatric, and Home Health Aides
Cooks
General and Operations Managers
Elementary and Middle School Teachers
Stock Clerks and Order Fillers
Sales Representatives, Wholesale and Manufacturing
Software Developers and Programmers
Bookkeeping, Accounting, and Auditing Clerks
Personal Care Aides
First-Line Supervisors of Sales Workers
First-Line Supervisors of Office and Admin Support Workers
Miscellaneous Healthcare Support
Miscellaneous Assemblers and Fabricators
Maintenance and Repair Workers, General
Teacher Assistants
Accountants and Auditors
Security Guards and Gaming Surveillance Officers
Secondary School Teachers
Supervisors of Food Preparation and Serving Workers
Receptionists and Information Clerks
Grounds Maintenance Workers
Miscellaneous Business Operations Specialists
Miscellaneous Sales Representatives, Services
Construction Laborers
Miscellaneous Teachers and Instructors
Miscellaneous Production Workers
Food Preparation Workers
Automotive Technicians and Repairers
Computer Support Specialists
Health Practitioner Support Technologists and Technicians
Licensed Practical and Licensed Vocational Nurses
Counter and Rental Clerks and Parts Salespersons
Bus Drivers
Shipping, Receiving, and Traffic Clerks
Carpenters
Retail salesperson is the most common job today, but has a 90+% chance of becoming automated.
Instead, nurses and teachers might soon become the most common jobs.
These predictions are only the probability of each job becoming automated. Fewer managers will probably also be needed when there is fewer staff to manage.
It is mostly low-income jobs that risk being automated. Jobs that require social skills, creativity or higher education are less likely to.

The claims are ludicrous. [I try to maintain professional language, but sometime
For instance, it appears to say that we will go from 1 million grounds and
maintenance workers in the US to only 50,000 in 10 to 20 years, because robot
take over those jobs. How many robots are currently operational in those jobs?
How many realistic demonstrations have there been of robots working in this ar
ZERO. Similar stories apply to all the other job categories in this diagram where
suggested that there will be massive disruptions of 90%, and even as much as
in jobs that currently require physical presence at some particular job site.

Mistaken predictions lead to fear of things that are not going to happen. Why ar
people making mistakes in predictions about Artificial Intelligence and robotics,
that Oren Etzioni, I, and others, need to spend time pushing back on them?

Below I outline seven ways of thinking that lead to mistaken predictions about
robotics and Artificial Intelligence. We find instances of these ways of thinking i
many of the predictions about our AI future. I am going to first list the four such
general topic areas of predictions that I notice, along with a brief assessment of
where I think they currently stand.

A. Artificial General Intelligence. Research on AGI is an attempt to distinguish a
thinking entity from current day AI technology such as Machine Learning. Here
idea is that we will build autonomous agents that operate much like beings in th
world. This has always been my own motivation for working in robotics and AI, l
the recent successes of AI are not at all like this.

Some people think that all AI is an instance of AGI, but as the word “general” w
imply, AGI aspires to be much more general than current AI. Interpreting curren
an instance of AGI makes it seem much more advanced and all encompassing t
really is.

Modern day AGI research is not doing at all well on being either general or getti
an independent entity with an ongoing existence. It mostly seems stuck on the
issues in reasoning and common sense that AI has had problems with for at lea
years. Alternate areas such as Artificial Life, and Simulation of Adaptive Behavi
make some progress in getting full creatures in the eighties and nineties (these
areas and communities were where I spent my time during those years), but th
have stalled.

My own opinion is that of course this is possible in principle. I would never have
started working on Artificial Intelligence if I did not believe that. However perha
humans are just not smart enough to figure out how to do this–see my remarks
humility in my post on the current state of Artificial Intelligence suitable for
deployment in robotics. Even if it is possible I personally think we are far, far fu
away from understanding how to build AGI than many other pundits might say.

[Some people refer to “an AI”, as though all AI is about being an autonomous a
think that is confusing, and just as the natives of San Francisco do not refer to t
city as “Frisco”, no serious researchers in AI refer to “an AI”.]

B. The Singularity. This refers to the idea that eventually an AI based intelligent
entity, with goals and purposes, will be better at AI research than us humans ar
Then, with an unending Moore’s law mixed in making computers faster and fast
Artificial Intelligence will take off by itself, and, as in speculative physics going
through the singularity of a black hole, we have no idea what things will be like
other side.

People who “believe” in the Singularity are happy to give post-Singularity AI
incredible power, as what will happen afterwards is quite unpredictable. I put th
word believe in scare quotes as belief in the singularity can often seem like a re
belief. For some it comes with an additional benefit of being able to upload thei

minds to an intelligent computer, and so get eternal life without the inconvenie having to believe in a standard sort of supernatural God. The ever powerful technologically based AI is the new God for them. Techno religion!

Some people have very specific ideas about when the day of salvation will com followers of one particular Singularity prophet believe that it will happen in the 2029, as it has been written.

This particular error of prediction is very much driven by exponentialism, and I address that as one of the seven common mistakes that people make.

Even if there is a lot of computer power around it does not mean we are close t having programs that can do research in Artificial Intelligence, and rewrite their code to get better and better.

Here is where we are on programs that can understand computer code. We cur have no programs that can understand a one page program as well as a new st in computer science can understand such a program after just one month of tal their very first class in programming. That is a long way from AI systems being at writing AI systems than humans are.

Here is where we are on simulating brains at the neural level, the other method that Singularity worshipers often refer to. For about thirty years we have known full "wiring diagram" of the 302 neurons in the worm C. elegans, along with the connections between them. This has been incredibly useful for understanding h behavior and neurons are linked. But it has been a thirty years study with hund people involved, all trying to understand just 302 neurons. And according to the OpenWorm project trying to simulate C. elegans bottom up, they are not yet ha there. To simulate a human brain with 100 billion neurons and a vast number of connections is quite a way off. So if you are going to rely on the Singularity to u yourself to a brain simulation I would try to hold off on dying for another couple centuries.

Just in case I have not made my own position on the Singularity clear, I refer yo my comments in a regularly scheduled look at the event by the magazine IEEE Spectrum. Here is the the 2008 version, and in particular a chart of where the p stand and what they say. Here is the 2017 version, and in particular a set of bo where the players stand and what they say. And yes, I do admit to being a little snarky in 2017...

C. Misaligned Values. The third case is that the Artificial Intelligence based mac get really good at execution of tasks, so much so that they are super human at getting things done in a complex world. And they do not share human values a leads to all sorts of problems.

I think there could be versions of this that are true–if I have recently bought an ticket to some city, suddenly all the web pages I browse that rely on advertisem for revenue start displaying ads for airline tickets to the same city. This is clearl dumb, but I don't think it is a sign of super capable intelligence, rather it is a ca poorly designed evaluation functions in the algorithms that place advertisemen

But here is a quote from one of the proponents of this view (I will let him remai anonymous, as an act of generosity):

> *The well-known example of paper clips is a case in point: if the machine's goal is maximizing the number of paper clips, it may invent incredible technologies as it sets about converting all available mass in the reachabl universe into paper clips; but its decisions are still just plain dumb.*

Well, no. We would never get to a situation in any version of the real world whe such a program could exist. One smart enough that it would be able to invent w

subvert human society to achieve goals set for it by humans, without understar the ways in which it was causing problems for those same humans. Thinking th technology might evolve this way is just plain dumb (nice turn of phrase...), and on making multiple errors among the seven that I discuss below.

This same author repeatedly (including in the piece from which I took this quote also at the big *International Joint Conference on Artificial Intelligence (IJCAI)* tha held just a couple of weeks ago in Melbourne, Australia) argues that we need research to come up with ways to mathematically prove that Artificial Intelligen systems have their goals aligned with humans.

I think this case **C** comes from researchers seeing an intellectually interesting research problem, and then throwing their well known voices promoting it as ar urgent research question. Then AI hangers-on take it, run with it, and turn it int existential problem for mankind.

By the way, I think mathematical provability is a vain hope. With multi-year larg team efforts we can not prove that a 1,000 line program can not be breached b external hackers, so we certainly won't be able to prove very much at all about AI systems. The good news is that us humans were able to successfully co-exist and even use for our own purposes, horses, themselves autonomous agents wit going existences, desires, and super-human physical strength, for thousands of And we had not a single theorem about horses. Still don't!

D. Really evil horrible nasty human-destroying Artificially Intelligent entities. Th category is like case **C**, but here the supposed Artificial Intelligence powered machines will take an active dislike to humans and decide to destroy them and them out of the way.

This has been a popular fantasy in Hollywood since at least the late 1960's with movies like *2001: A Space Odyssey* (1968, but set in 2001), where the machine wreaked havoc was confined to a single space ship, and *Colossus: The Forbin P* (1970, and set in those times) where the havoc was at a planetary scale. The th has continued over the years, and more recently with *I, Robot* (2004, set in 203 where the evil AI computer VIKI takes over the world through the instrument of new NS-5 humanoid robots. [By the way, that movie continues the bizarre conv from other science fiction movies that large complex machines are built with sp that have multi hundred feet heights around them so that there can be great pl peril for the human heroes as they fight the good fight against the machines go bad...]

This is even wronger than case **C**. I think it must make people feel tingly thinkin about these terrible, terrible dangers...

In this blog, I am not going to address the issue of military killer robots–this ofte confused in the press with issue **D** above, and worse it often gets mashed toget people busy fear mongering about issue **D**. They are very separate issues. Furthermore I think that many of the arguments about such military robots are misguided. But it is a very different issue and will have to wait for another blog

Now, the seven mistakes I think people are making. All seven of them influence assessments about timescales for and likelihood of each of scenarios **A**, **B**, **C**, ar coming about. But some are more important I believe in the mis-estimations th others. I have labeled in the section headers for each of these seven errors whe think they do the most damage. The first one does some damage everywhere!

1. [A,B,C,D] OVER AND UNDER ESTIMATING

Roy Amara was a futurist and the co-founder and President of the *Institute For Future* in Palo Alto, home of Stanford University, countless venture capitalists, a intellectual heart of Silicon Valley. He is best known for his adage, now referred Amara's law:

> *We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.*

There is actually a lot wrapped up in these 21 words which can easily fit into a t and allow room for attribution. An optimist can read it one way, and a pessimist read it another. It should make the optimist somewhat pessimistic, and the pess somewhat optimistic, for a while at least, before each reverting to their norm.

A great example[1] of the two sides of Amara's law that we have seen unfold ove last thirty years concerns the US Global Positioning System. Starting in 1978 a constellation of 24 satellites (30 including spares) were placed in orbit. A groun station that can see 4 of them at once can compute the the latitude, longitude, height above a version of *sea level*. An operations center at Schriever Air Force in Colorado constantly monitors the precise orbits of the satellites and the accu of their onboard atomic clocks and uploads minor and continuous adjustments t them. If those updates were to stop GPS would fail to have you on the correct r you drive around town after only a week or two, and would have you in the wro town after a couple of months.

The goal of GPS was to allow precise placement of bombs by the US military. Th was the expectation for it. The first operational use in that regard was in 1991 d Desert Storm, and it was promising. But during the nineties there was still much distrust of GPS as it was not delivering on its early promise, and it was not until early 2000's that its utility was generally accepted in the US military. It had a ha time delivering on its early expectations and the whole program was nearly can again and again.

Today GPS is in the long term, and the ways it is used were unimagined when it first placed in orbit. My Series 2 Apple Watch uses GPS while I am out running t record my location accurately enough to see which side of the street I ran along tiny size and tiny price of the receiver would have been incomprehensible to th GPS engineers. GPS is now used for so many things that the designers never considered. It synchronizes physics experiments across the globe and is now ar intimate component of synchronizing the US electrical grid and keeping it runni and it even allows the high frequency traders who really control the stock mark mostly not fall into disastrous timing errors. It is used by all our airplanes, large small to navigate, it is used to track people out of jail on parole, and it determir which seed variant will be planted in which part of many fields across the globe tracks our fleets of trucks and reports on driver performance, and the bouncing signals on the ground are used to determine how much moisture there is in the ground, and so determine irrigation schedules.

GPS started out with one goal but it was a hard slog to get it working as well as originally expected. Now it has seeped into so many aspects of our lives that w would not just be lost if it went away, but we would be cold, hungry, and quite possibly dead.

We see a similar pattern with other technologies over the last thirty years. A big promise up front, disappointment, and then slowly growing confidence, beyond the original expectations were aimed. This is true of the blockchain (Bitcoin was first application), sequencing individual human genomes, solar power, wind pow and even home delivery of groceries.

Perhaps the most blatant example is that of computation itself. When the first commercial computers were deployed in the 1950's there was widespread fear they would take over all jobs (see the movie Desk Set from 1957). But for the n years computers were something that had little direct impact on people's lives even in 1987 there were hardly any microprocessors in consumer devices. That all changed in the second wave over the subsequent 30 years and now we all w around with our bodies adorned with computers, our cars full of them, and they all over our houses.

To see how the long term influence of computers has consistently been underestimated one need just go back and look at portrayals of them in old scie fiction movies or TV shows about the future. The three hundred year hence spa ship computer in the 1966 Star Trek (TOS) was laughable just thirty years later, alone three centuries later. And in Star Trek The Next Generation, and Star Trek Space Nine, whose production spanned 1986 to 1999, large files still needed to carried by hand around the far future space ship or space station as they could sent over the network (like an AOL network of the time). And the databases ava for people to search were completely anemic with their future interfaces which pre-Web in design.

Most technologies are overestimated in the short term. They are the shiny new Artificial Intelligence has the distinction of having been the shiny new thing and overestimated again and again, in the 1960's, in the 1980's, and I believe again (Some of the marketing messages from large companies on their AI offerings ar delusional, and may have very bad blowback for them in the not too distant fut

Not all technologies get underestimated in the long term, but that is most likely case for AI. The question is how long is the long term. The next six errors that I about help explain how the timing for the long term is being grossly underestim for the future of AI.

2. [B,C,D] IMAGINING MAGIC

When I was a teenager, Arthur C. Clarke was one of the "big three" science ficti writers along with Robert Heinlein and Isaac Asimov. But Clarke was more than science fiction writer. He was also an inventor, a science writer, and a futurist.

In February 1945 he wrote a letter[2] to *Wireless World* about the idea of geostati satellites for research, and in October of that year he published a paper[3] outlini how they could be used to provide world-wide radio coverage. In 1948 he wrote short story *The Sentinel* which provided the kernel idea for Stanley Kubrick's ep movie *2001: A Space Odyssey*, with Clarke authoring a book of the same name film was being made, explaining much that had left the movie audience somew lost.

In the period from 1962 to 1973 Clarke formulated three adages, which have co be known as Clarke's three laws (he said that Newton only had three, so three w enough for him too):

1. *When a distinguished but elderly scientist states that something is possibl is almost certainly right. When he states that something is impossible, he probably wrong.*
2. *The only way of discovering the limits of the possible is to venture a little past them into the impossible.*
3. *Any sufficiently advanced technology is indistinguishable from magic.*

Personally I should probably be wary of the second sentence in his first law, as much more conservative than some others about how quickly AI will be ascend But for now I want to expound on Clarke's third law.

Imagine we had a time machine (powerful magic in itself…) and we could trans Issac Newton from the late 17th century to Trinity College Chapel in Cambridge University. That chapel was already 100 years old when he was there so perhap would not be too much of an immediate shock to find himself in it, not realizing current date.

Now show Newton an Apple. Pull out an iPhone from your pocket, and turn it on that the screen is glowing and full of icons and hand it to him. The person who revealed how white light is made from components of different colored light by pulling apart sunlight with a prism and then putting it back together again woul doubt be surprised at such a small object producing such vivid colors in the dar of the chapel. Now play a movie of an English country scene, perhaps with som animals with which he would be familiar–nothing indicating the future in the co Then play some church music with which he would be familiar. And then show h web page with the 500 plus pages of his personally annotated copy of his masterpiece Principia, teaching him how to use the pinch gestures to zoom in o details.

Could Newton begin to explain how this small device did all that? Although he invented calculus and explained both optics and gravity, Newton was never abl sort out chemistry and alchemy. So I think he would be flummoxed, and unable come up with even the barest coherent outline of what this device was. It woul no different to him than an embodiment of the occult–something which was of g interest to him when he was alive. For him it would be indistinguishable from m And remember, Newton was a really smart dude.

If something is magic it is hard to know the limitations it has. Suppose we furth show Newton how it can illuminate the dark, how it can take photos and movies record sound, how it can be used as a magnifying glass, and as a mirror. Then w show him how it can be used to carry out arithmetical computations at incredib speed and to many decimal places. And we show it counting his steps has he ca it.

What else might Newton conjecture that the device in front of him could do? Wo he conjecture that he could use it to talk to people anywhere in the world immediately from right there in the chapel? Prisms work forever. Would he conj that the iPhone would work forever just as it is, neglecting to understand that it needed to be recharged (and recall that we nabbed him from a time 100 years the birth of Michael Faraday, so the concept of electricity was not quite around) can be a source of light without fire could it perhaps also transmute lead to gol

This is a problem we all have with imagined future technology. If it is far enough from the technology we have and understand today, then we do not know its limitations. It becomes indistinguishable from magic.

When a technology passes that magic line anything one says about it is no long falsifiable, because it is magic.

This is a problem I regularly encounter when trying to debate with people about whether we should fear just plain AGI, let alone cases **C** or **D** from above. I am t that I do not understand how powerful it will be. That is not an argument. We ha idea whether it can even exist. All the evidence that I see says that we have no idea yet how to build one. So its properties are completely unknown, so rhetoric quickly becomes magical and super powerful. Without limit.

Nothing in the Universe is without limit. Not even magical future AI.

Watch out for arguments about future technology which is magical. It can never refuted. It is a faith-based argument, not a scientific argument.

3. [A,B,C] PERFORMANCE VERSUS COMPETENCE

One of the social skills that we all develop is an ability to estimate the capabilit individual people with whom we interact. It is true that sometimes "out of tribe" issues tend to overwhelm and confuse our estimates, and such is the root of th perfidy of racism, sexism, classism, etc. In general, however, we use cues from person performs some particular task to estimate how well they might perform different task. We are able to generalize from observing performance at one tas guess at competence over a much bigger set of tasks. We understand intuitivel to generalize from the performance level of the person to their competence in r areas.

When in a foreign city we ask a stranger on the street for directions and they re the language we spoke to them with confidence and with directions that seem t make sense, we think it worth pushing our luck and asking them about what is local system for paying when you want to take a bus somewhere in that city.

If our teenage child is able to configure their new game machine to talk to the household wifi we suspect that if sufficiently motivated they will be able to help get our new tablet computer on to the same network.

If we notice that someone is able to drive a manual transmission car, we will be pretty confident that they will be able to drive one with an automatic transmiss too. Though if the person is North American we might not expect it to work for t converse case.

If we ask an employee in a large hardware store where to find a particular item, home electrical fitting say, that we are looking for and they send us to an aisle garden tools, we will probably not go back and ask that very same person wher find a particular bathroom fixture. We will estimate that not only do they not kn where the electrical fittings are, but that they really do not know the layout of g within the store, and we will look for a different person to ask with our second question.

Now consider a case that is closer to some performances we see for some of to AI systems.

Suppose a person tells us that a particular photo is of people playing Frisbee in park, then we naturally assume that they can answer questions like "what is th shape of a Frisbee?", "roughly how far can a person throw a Frisbee?", "can a p eat a Frisbee?", "roughly how many people play Frisbee at once?", "can a 3 mo person play Frisbee?", "is today's weather suitable for playing Frisbee?"; in con we would not expect a person from another culture who says they have no idea is happening in the picture to be able to answer all those questions. Today's ima labelling systems that routinely give correct labels, like "people playing Frisbee park" to online photos, have no chance of answering those questions. Besides t fact that all they can do is label more images and can not answer questions at a they have no idea what a person is, that parks are usually outside, that people ages, that weather is anything more than how it makes a photo look, etc., etc.

This does not mean that these systems are useless however. They are of great to search engine companies. Simply labelling images well lets the search engin bridge the gap from search for words to searching for images. Note too that sea engines usually provide multiple answers to any query and let the person using

engine review the top few and decide which ones are actually relevant. Search companies strive to get the performance of their systems to get the best possib answer as one of the top five or so. But they rely on the cognitive abilities of th human user so that they do not have to get the best answer first, every time. If only gave one answer, whether to a search for “great hotels in Paris”, or at an e commerce site only gave one image selection for a “funky neck tie”, they would be as useful as they are.

Here is what goes wrong. People hear that some robot or some AI system has performed some task. They then take the generalization from that performance general competence that a person performing that same task could be expecte have. And they apply that generalization to the robot or AI system.

Today’s robots and AI systems are incredibly narrow in what they can do. Huma generalizations just do not apply. People who do make these generalizations ge things very, very wrong.

4. [A,B] SUITCASE WORDS

I spoke briefly about suitcase words (Marvin Minsky’s term[4]) in my post explain how machine learning works. There I was discussing how the word learning can so many different types of learning when applied to humans. And as I said there surely there are different mechanisms that humans use for different sorts of lea Learning to use chopsticks is a very different experience from learning the tune new song. And learning to write code is a very different experience from learnir your way around a particular city.

When people hear that Machine Learning is making great strides and they think a machine learning in some new domain, they tend to use as a mental model th in which a person would learn that new domain. However, Machine Learning is brittle, and it requires lots of human preparation by researchers or engineers, s purpose coding for processing input data, special purpose sets of training data, custom learning structure for each new problem domain. Today’s Machine Learn by computers is not at all the sponge like learning that humans engage in, mak rapid progress in a new domain without having to be surgically altered or purpo built.

Likewise when people hear that computers can now beat the world chess cham (in 1997) or the world Go champion (in 2016) they tend to think that it is “playi the game just like a human would. Of course in reality those programs had no i what a game actually was (again, see my post on machine learning), nor that th are playing. And as pointed out in this article in The Atlantic during the recent G challenge the human player, Lee Sedol, was supported by 12 ounces of coffee, whereas the AI program, AlphaGo, was running on a whole bevy of machines as distributed application, and was supported by a team of more than 100 scientis

When a human plays a game a small change in rules does not throw them off–a player can adapt. Not so for AlphaGo or Deep Blue, the program that beat Garry Kasparov back in 1997.

Suitcase words lead people astray in understanding how well machines are doi tasks that people can do. AI researchers, on the other hand, and worse their institutional press offices, are eager to claim progress in their research in being instance of what a suitcase word applies to for humans. The important phrase h “an instance”. No matter how careful the researchers are, and unfortunately no them are so very careful, as soon as word of the research result gets to the pre office and then out into the unwashed press, that detail soon gets lost. Headline

trumpet the suitcase word, and mis-set the general understanding of where AI i how close it is to accomplishing more.

And, we haven't even gotten to saying many of Minsky's suitcase words about A systems; *consciousness*, *experience*, or *thinking*. For us humans it is hard to thi about playing chess without being conscious, or having the experience of playi thinking about a move. So far, none of our AI systems have risen to an even elementary level where one of the many ways in which we use those words abc humans apply. When we do, and I tend to think that we will, get to a point wher will start using some of those words about particular AI systems, the press, and people, will over generalize again.

Even with a very narrow single aspect demonstration of one slice of these word afraid people will over generalize and think that machines are on the very door of human-like capabilities in these aspects of being intelligent.

Words matter, but whenever we use a word to describe something about an AI system, where that can also be applied to humans, we find people overestimati what it means. So far most words that apply to humans when used for machine only a microscopically narrow conceit of what the word means when applied to humans.

Here are some of the verbs that have been applied to machines, and for which machines are totally unlike humans in their capabilities:

> *anticipate, beat, classify, describe, estimate, explain, hallucinate, hear, im intend, learn, model, plan, play, recognize, read, reason, reflect, see, understand, walk, write*

For all these words there have been research papers describing a narrow sliver rich meanings that these words imply when applied to humans. Unfortunately t of these words suggests that there is much more there there than is there.

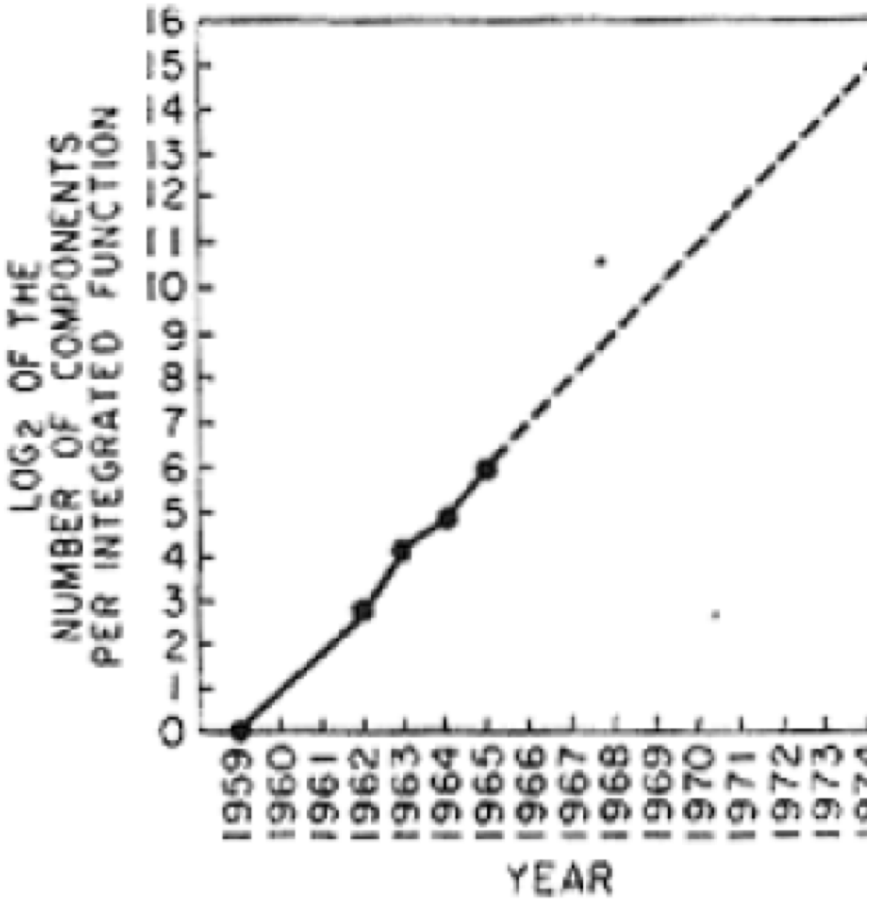
This leads people to misinterpret and then overestimate the capabilities of toda Artificial Intelligence.

5. [A,B,B,B,...] EXPONENTIALS

Many people are suffering from a severe case of "exponentialism".

Everyone has some idea about Moore's Law, at least as much to sort of know th computers get better and better on a clockwork like schedule.

What Gordon Moore actually said was that the number of components that coul on a microchip would double every year. I published a blog post in February abc this and how it is finally coming to an end after a solid fifty year run. Moore had his predictions in 1965 with only four data points using this graph:



He only extrapolated for 10 years, but instead it has lasted 50 years, although t time constant for doubling has gradually lengthened from one year to over two and now it is finally coming to an end.

Double the components on a chip has lead to computers that keep getting twic fast. And it has lead to memory chips that have four times as much memory ev two years. It has also led to digital cameras which have had more and more resolution, and LCD screens with exponentially more pixels.

The reason Moore's law worked is that it applied to a digital abstraction of true false. Was there an electrical charge or voltage there or not? And the answer to yes/no question is the same when the number of electrons is halved, and halve again, and halved again. The answer remains consistent through all those halvi until we get down to so few electrons that quantum effects start to dominate, a that is where we are now with our silicon based chip technology.

Moore's law, and exponential laws like Moore's law can fail for three different re

a. It gets down to a physical limit where the process of halving/doubling no lo works.
b. The market demand pull gets saturated so there is no longer an economic for the law to continue.
c. It may not have been an exponential process in the first place.

When people are suffering from *exponentialism* they may gloss over any of the three reasons and think that the exponentials that they use to justify an argum are going to continue apace.

Moore's Law is now faltering under case (a), but it has been the presence of Mo Law for fifty years that has powered the relentless innovation of the technology industry and the rise of Silicon Valley, and Venture Capital, and the ride of the g

to be amongst the richest people in the world, that has led too many people to that everything in technology, including AI, is exponential.

It is well understood that many cases of exponential processes are really part o "S-curve", where at some point the hyper growth flattens out. Exponential grow the number of users of a social platform such as Facebook or Twitter must turn S-curve eventually as there are only a finite number of humans alive to be new and so exponential growth can not continue forever. This is an example of case above.

But there is more to this. Sometimes just the demand from individual users can like an exponential pull for a while, but then it gets saturated.

Back in the first part of this century when I was running a very large laboratory M.I.T. (CSAIL) and needed to help raise research money for over 90 different res groups, I tried to show sponsors how things were continuing to change very rap through the memory increase on iPods. Unlike Gordon Moore I had **five** data poi The data was how much storage one got for one's music in an iPod for about $4 noted the dates of new models and for five years in a row, somewhere in the Ju September time frame a new model would appear. Here are the data:

Year	GigaBytes
2003	10
2004	20
2005	40
2006	80
2007	160

The data came out perfectly (Gregor Mendel would have been proud...) as an exponential. Then I would extrapolate a few years out and ask what we would c all that memory in our pockets.

Extrapolating through to today we would expect a $400 iPod to have 160,000 GigaBytes of memory (or 160 TeraBytes). But the top of the line iPhone of today (which costs more than $400) only has 256 GigaBytes of memory, less than dou the 2007 iPod, while the top of the line iPod (touch) has only 128 GigaBytes whi years later is a decrease from the 2007 model.

This particular exponential collapsed very suddenly once the amount of memor to the point where it was big enough to hold any reasonable person's complete library, in their hand. Exponentials can stop when the customers stop demandi

Moving on, we have seen a sudden increase in performance of AI systems due t success of deep learning, a form of Machine Learning. Many people seem to thi that means that we will continue to have increases in AI performance of equal multiplying effect on a regular basis. But the deep learning success was thirty y in the making, and no one was able to predict it, nor saw it coming. It was an is event.

That does not mean that there will not be more isolated events, where backwat AI research suddenly fuel a rapid step increase in performance of many AI applications. But there is no "law" that says how often they will happen. There i physical process, like halving the mass of material as in Moore's Law, fueling th process of AI innovation. This is an example of case (c) from above.

So when you see exponential arguments as justification for what will happen wi remember that not all so called exponentials are really exponentials in the first

and those that are can collapse suddenly when a physical limit is hit, or there is more economic impact to continue them.

6. [C,D] HOLLYWOOD SCENARIOS

The plot for many Hollywood science fiction movies is that the world is just as it today, except for one new twist. Certainly that is true for movies about aliens a on Earth. Everything is going along as usual, but then one day the aliens unexpectedly show up.

That sort of single change to the world makes logical sense for aliens but what for a new technology? In real life lots of new technologies are always happening the same time, more or less.

Sometimes there is a rational, within Hollywood reality, explanation for why the singular disruption of the fabric of humanity's technological universe. The Term movies, for instance, had the super technology come from the future via time t so there was no need to have a build up to the super robot played by Arnold Schwarzenegger.

But in other movies it can seem a little silly.

In Bicentennial Man, Richard Martin, played by Sam Neill, sits down to breakfas waited upon by a walking talking humanoid robot, played by Robin Williams. He up a newspaper to read over breakfast. A newspaper! Printed on paper. Not a ta computer, not a podcast coming from an Amazon Echo like device, not a direct connection to the Internet.

In Blade Runner, as Tim Harford recently pointed out, detective Rick Deckard, p by Harrison Ford, wants to contact the robot Rachael, played by Sean Young. In plot Rachael is essentially indistinguishable from a human being. How does Dec connect to her? With a pay phone. With coins that you feed in to it. A technolog many of the readers of this blog may never had seen. (By the way, in that same Harford remarks: "Forecasting the future of technology has always been an entertaining but fruitless game." A sensible insight.)

So there are two examples of Hollywood movies where the writers, directors, ar producers, imagine a humanoid robot, able to see, hear, converse, and act in th world as a human–pretty much an AGI (Artificial General Intelligence). Never mi the marvelous materials and mechanisms involved. But those creative people l the imagination, or will, to consider how else the world may have changed as th amazing package of technology has been developed.

It turns out that many AI researchers and AI pundits, especially those pessimist indulge in predictions **C** and **D**, are similarly imagination challenged.

Apart from the time scale for many **C** and **D** predictions being wrong, they ignor fact that if we are able to eventually build such smart devices the world will hav changed significantly from where we are. We will not suddenly be surprised by existence of such super intelligences. They will evolve technologically over time our world will be different, populated by many other intelligences, and we will h lots of experience already.

For instance, in the case of **D** (evil super intelligences who want to get rid of us) before we see such machines arising there will be the somewhat less intelligent belligerent machines. Before that there will be the really grumpy machines. Bef that the quite annoying machines. And before them the arrogant unpleasant machines.

We will change our world along the way, adjusting both the environment for ne technologies and the new technologies themselves. I am not saying there may challenges. I am saying that they will not be as suddenly unexpected as many p think. Free running imagination about shock situations are not helpful–they will be right, or even close.

"Hollywood scenarios" are a great rhetorical device for arguments, but they usu do not have any connection to future reality.

7. [B,C,D] SPEED OF DEPLOYMENT

As the world has turned to software the deployment frequency of new versions become very high in some industries. New features for platforms like Facebook deployed almost hourly. For many new features, as long as they have passed integration testing, there is very little economic downside if a problem shows up the field and the version needs to be pulled back–often I find that features I use such platforms suddenly fail to work for an hour or so (this morning it was the p down menu for Facebook notifications that was failing) and I assume these are deployment fails. For revenue producing components, like advertisement placer more care is taken and changes may happen only on the scale of weeks.

This is a tempo that Silicon Valley and Web software developers have gotten us It works because the marginal cost of newly deploying code is very very close t

Hardware on the other hand has significant marginal cost to deploy. We know th from our own lives. Many of the cars we are buying today, which are not self dr and mostly are not software enabled, will likely still be on the road in the year 2 This puts an inherent limit on how soon all our cars will be self driving. If we bui new home today, we can expect that it might be around for over 100 years. The building I live in was built in 1904 and it is not nearly the oldest building in my neighborhood.

Capital costs keep physical hardware around for a long time, even when there a high tech aspects to it, and even when it has an existential mission to play.

The US Air Force still flies the B-52H variant of the B-52 bomber. This version wa introduced in 1961, making it 56 years old. The last one was built in 1963, a me years ago. Currently these planes are expected to keep flying until at least 204 perhaps longer–there is talk of extending their life out to 100 years. (cf. The Millennium Falcon!)

The US land-based Intercontinental Ballistic Missile (ICBM) force is all Minutema variants, introduced in 1970. There are 450 of them. The launch system relies o eight inch floppy disk drives, and some of the digital communication for the lau procedure is carried out over analog wired phone lines.

I regularly see decades old equipment in factories around the world. I even see running Windows 3.0 in factories–a software version released in 1990. The think that "if it ain't broke, don't fix it". Those PCs and their software have been runn the same application doing the same task reliably for over two decades.

The principal control mechanisms in factories, including brand new ones in the Europe, Japan, Korea, and China, is based on Programmable Logic Controllers, o PLCs. These were introduced in 1968 to replace electromechanical relays. The " is still the principal abstraction unit used today, and the way PLCs are programm as though they were a network of 24 volt electromechanical relays. Still. Some direct wires have been replaced by Ethernet cables. They emulate older networ (themselves a big step up) based on the RS485 eight bit serial character protoc

which themselves carry information emulating 24 volt DC current switching. An Ethernet cables are not part of an open network, but instead individual cables a point to point physically embodying the control flow in these brand new ancient automation controllers. When you want to change information flow, or control fl most factories around the world it takes weeks of consultants figuring out what there, designing new reconfigurations, and then teams of tradespeople rewiring reconfiguring hardware. One of the major manufacturers of this equipment rece told me that they aim for three software upgrades every twenty years.

In principle it could be done differently. In practice it is not. And I am not talking just in technological backwaters. I just this minute looked on a jobs list and eve today, this very day, Tesla is trying to hire full time PLC technicians at their Frer factory. Electromagnetic relay emulation to automate the production of the mos software advanced automobile that exists.

A lot of AI researchers and pundits imagine that the world is already digital, and simply introducing new AI systems will immediately trickle down to operational changes in the field, in the supply chain, on the factory floor, in the design of products.

Nothing could be further from the truth.

The impedance to reconfiguration in automation is shockingly mind-blowingly impervious to flexibility.

You can not give away a good idea in this field. It is really slow to change. The example of the AI system making paper clips deciding to co-opt all sorts of resc to manufacture more and more paper clips at the cost of other human needs is indeed a nutty fantasy. There will be people in the loop worrying about physical for decades to come.

Almost all innovations in Robotics and AI take far, far, longer to get to be really deployed than people in the field and outside the field imagine. Self driving cars an example. Suddenly everyone is aware of them and thinks they will soon be deployed. But it takes longer than imagined. It takes decades, not years. And if think that is pessimistic you should be aware that it is already provably three de from first on road demonstrations and we have no deployment. In 1987 Ernst Dickmanns and his team at the Bundeswehr University in Munich had their autonomous van drive at 90 kilometers per hour (56mph) for 20 kilometers (12 on a public freeway. In July 1995 the first no hands on steering wheel, no feet o pedals, minivan from CMU's team lead by Chuck Thorpe and Takeo Kanade drov coast to coast across the United States on public roads. Google/Waymo has bee working on self driving cars for eight years and there is still no path identified fo large scale deployment. It could well be four or five or six decades from 1987 b we have real deployment of self driving cars.

New ideas in robotics and AI take a long long time to become real and deployed

EPILOG

When you see pundits warn about the forthcoming wonders or terrors of robotic Artificial Intelligence I recommend carefully evaluating their arguments against seven pitfalls. In my experience one can always find two or three or four of thes problems with their arguments.

Predicting the future is really hard, especially ahead of time.