

The shadows of Artificial Intelligence

Eduardo Castañeda (220)
Human BIOS Project
Mexico City, Mexico
2-2-0.online

This philosophical essay talks about possible socio-psychological problems derived from the general public's impression around Artificial Intelligence technologies. While these are still being developed, the general public has an impression that AI services offer a trustworthy, godly-like solution to real life problems, sometimes even using these technologies for school or work related issues. The simple act of belief can turn into a catastrophic "Messiah syndrome" of sorts which could ultimately affect society even more than bad use of technology itself.

A new competitor

The media-amplified arrival of publicly available technologies such as ChatGPT has caused quite a commotion in the modern, online humanity, making people actually believe that current AI technologies have reached the status of an Artificial General Intelligence - meaning the machine achieved some kind of sentience and consciousness that can make wise decisions for us, or even become sentient and either love or try to overthrow us.

Regardless of the debate, AI technologies are actually still in a very primitive state - most of the ones that surprise the general population are generators, a special kind of neural networks that creates content like audio or video - Midjourney and Dall-E image generators fall into this category.

These services may look like if the computer has acquired some kind of creativity, but what the computer is actually doing is making sense out of numbers - in fact 'unsupervised training' which is one of the paradigms used by AI model creation is

rather obscure but somehow it renders the required results. What the computers actually do when processing new images is applying different matrix-based math models to create new number combinations based on a trained bias - the computer doesn't recognise an image, much less its contents and obviously it won't understand what it is as such, as computers are nothing more than super fast number crunchers.

Of course these companies are doing what they can to grasp the media's attention - maneuvers like a rogue computer engineer talking about how advanced their AI is and how it seems to have gained consciousness and such (even if they are not neuroscience experts) are very useful to them to make their technology appear more powerful than it actually is.

In fact, the term Artificial Intelligence can be considered a hype by itself, as intelligence is much more than ultra fast number crunching and may have more to do with emerging patterns that rely on hard to replicate organic characteristics rather than

just numbers. Organic emergent intelligence may have more to do with problem solving for survival, which implies a lot of skills needed to first understand our own needs as humans, understanding or at least 'sensing' the surrounding areas and everything about resources, for example.

AI based music composition

There are many videos showing how AI techs can create music with lyrics, write a story or create illustrations of it, which is amazing as such, but much of the play has to deal with perception and interpretation, which are individual traits derived from each person's metareality.

A good example is a video from ShaunTrack¹ - a talented and popular influencer musician who used ChatGPT to create a track that would sound like Rammstein - a much more popular aggressive German hardcore music band. While the results may seem amazing on what ChatGPT does, it requires understanding the production pipeline a bit better.

After being asked for a new composition with the peculiar requirement, ChatGPT returned a set of four chords - called a chord progression - which besides being based on either B minor or D harmonic circle it is rather generic and not quite particular to any artist as such. The lyrics were probably as hateful and poetic as Rammstein's, which was rather a good result considering they were as simple as the original artists' are - short, hateful repeated words or short phrases along with a longer one once in a while.


With that information the influencer started working on a great quality production (given the time) that ended up sounding of something that reminds of Rammstein, making him and his audience very surprised - but how much did the AI actually do and what was the influence of having a professional musician doing this kind of work?

If we analyze profoundly, we would immediately acknowledge and recognize the influencer's talent to actually make it sound like the original artist, but if we think of the four chords in another context we can get a totally different result - in fact, those same four chords are used in the song Pumped up kicks from another popular rock band called Foster the people; this song is totally different in intention that any from Rammstein while using the same chords - music is prominently a matter of interpretation.

The point is most people are not observing the whole picture and thus attributing existing AIs with qualities that they don't actually have (yet) - like real creativity. Those chords came from tons of data on the Internet and could belong to any given song, but the influencer being totally biased by the concept of a 'creative AI' could be falling into a false perception, when the talent required to accomplish the given task is actually his, not the computers' as such.

The fact that persons tend to humanize everything around them may make AI technologies seem much more advanced and reliable than they actually are, and it is surprising to realize that millions of persons are actually asking for serious results and work from ChatGPT or similar tools without any further processing of the produced data

¹

 ChatGPT Compone mejor que mucho...

as such, rendering unreliable results that make some important mistakes.

Neurons vs neurons

When talking about generator AIs like these we need to understand that the fact that neuronal networks are an analogy on how real brain's neurons work does not automatically grant them the same abilities as such - though a digital neuron may handle information very efficiently making it much faster than a human brain in regards of creating patterns. In the same analogy both human brain and AI models require time for training, but the contextual relations that a person makes during learning are totally different from an AI - these are still incapable of true comprehension of the data, while humans can give many different meanings and significance to just about any kind of data it receives².

This is where the human mind makes a symbiotic operation with the AI - the latter provides data while the human makes it meaningful by adhering it to its context: information becomes such when it becomes meaningful to a human - without interpretation data has no meaning at all. Of course, this relation is almost exactly the same as the symbiotic operations we've already had with systems such as Google Search.

Current neural network based generators are actually just 'synthesizers' of data provided to them³ - models are created

² [The Brain Learns in Unexpected Ways - Scientific American](#)

³ [How Transformers Work. Transformers are a type of neural... | by Giuliano Giacaglia | Towards Data Science](#)

based on found patterns and periodicity of the data found in the set, so we could say that the main difference between a specific service oriented chatbot and ChatGPT is basically the amount, orientation and quality of data poured on it.

A customer service chatbot can be easily trained with reinforced training exercises to 'shape up' (or model) the answers it can give based on questions asked by users, a task that can be accomplished with some effort from the current customer service representatives in probably some weeks to have it answering efficiently - it will probably still be updated with new data provided by actual usage.

On the other hand a service like ChatGPT whose intention is to provide as most answers possible from about any topic needs to be modeled after lots of data, which is usually acquired from online services on the Internet - all this data gets processed (or trained) into a model which is considerably bigger - and apparently wiser, or much knowledgeable.⁴

Both of these services use what is called a Large Language Model (LLM) - some sort of language database regarding data usage patterns - 'Language' in LLM actually refers to a language that the data parser will understand, not a textual language per se, so they are also used with images for services like Stable Diffusion or Amper for making music.

So, as an LLM is created by training it with a lot of data by using computers - the data fed into it provides it with the information

⁴ [\[2303.08774\] GPT-4 Technical Report](#)

needed to generate content such as replies from a chatbot or new images. One of the biggest advantages over this paradigm is that while a model requires a lot of processing power to generate it takes much, much less to execute it on a less powerful computer, making this suitable for almost any kind of device for the end-user.

But while LLMs represent a big advancement on how we can treat big data to generate new information, and regardless of the ethical implications that this practice may or may not incur into, there are also many problems that come along with this technology.

LLM: Large Language Mistakes

One would think that the more data a set has, the more precise answers it would give, but given the fact that ChatGPT is using a Large Language Model implies using a lot of data, which is mostly taken from the Internet as such, causing the model to include a lot of information which could be wrong.⁵

Most importantly, LLMs are learning the patterns behind the words they are using, but as no model understands data as such in a contextual, meaningful way such as humans do, the models end up mashing unreliable information just to cope with the required result and 'be useful'. What we get is a result that looks accurate but it is not, as it is based on patterns and not significance - or even knowledge. Even technology providers include a warning that says so, but people tend to ignore smaller

text when signing up for a new service and end up making big mistakes.

There's also the problem of systems learning the shape and not the depth of the provided data - there has been examples where ChatGPT quotes from non-existing references, which are the pillars of trust for many scientists, engineers and generally speaking academics in general - the model learned to be complacent with the shape, but the quality of the information is totally useless.⁶

Also data pollution becomes a problem, as not all data is reliable or may be filled with utter nonsense - like Internet's common 'click bait' notes that use question headlines and just add paragraph after paragraph of the same question and no answer at all. The quality of a result provided by the system relies a lot on the quality of the rules fed to it - by adding more information and evaluation weights without curating the data we are just adding more nonsense into these systems, which could be the cause for GPT-4 providing less accurate technical answers than GPT-3, for example. In a few words: garbage in, garbage out.

Of course having no expectations make everything seem more attractive - plenty of professionals have tried using many of these AI tools without being capable of producing final high quality results because most people don't understand that ChatGPT is just a toy to fiddle with while refining their model based on new input and output produced by the usage of their systems.

⁵ [ChatGPT-4 produces more misinformation than predecessor - Misinformation Monitor: March 2023 - NewsGuard](#)

⁶ [GPT-3.5 Hallucinates Nonexistent Citations: Evidence from Economics](#)

Due to this, 'data redundancy' or 'reprocessed data' should be a concern with professional, high-end systems required for specific specialized areas of operation, such as military and medical applications. Badly sourced data produces one type of problem, but recycling data will produce another one.

It then becomes reasonable to think that an AGI can't rely on LLMs as such, but could use them as means of specific 'knowledge model packs' - highly specialized information models that have been fed with high quality information both provided and reviewed by those specific areas' specialists.

This kind of knowledge management would then require a knowledge manager unit as such, another kind of system not necessarily based on LLMs - after all, learning is just one of the many branches that AI has to cope with, like perception, reasoning, multiagent systems and more.

Regardless of any risk people will still naively feed the system with their input and will be amazed at such wonders, as yes, the illusion of a creative machine seems like a wonderful promise to those who want to express themselves but lack the language and expertise to do it with traditional digital or non digital media - and having creativity flow in such a robotic society can't be that bad, or can it?

AI as a newborn religion

Yes, the imagery produced by popular generators is very beautiful and actually amazing to think about how it became real - ChatGPT can be fun to talk with, and may be useful to create letters and provide some code to create our own video games - no

one can't deny that such wonders were totally unimaginable in the 90s, or even the beginning of this 21st. Century - and it is more than fair to acknowledge the brilliant minds that made this possible.

...but, why do we keep putting all of our hopes in computers and human-made models? If we think about it thoroughly it is like watching a new religion coming up:

1. There is a godly figure that can't be seen or even understood (AI) that has acolytes (engineers and researchers), scribes (media) and followers (users). There's usually a 'messiah' always coming - we've had different ChatGPT versions and people are anxiously waiting for the next one - which by the way it's always advertised to be more powerful.
2. There is a promise: in religion it can be a healthy and prosperous, harmonious life, in some ending with a happy eternal life - in AI it is systems telling you what to do and how to act for the kind of success you are looking for, ending up with AI taking over your routine tasks, giving you more time to contemplate 'god' - besides the riches promised by using AI to become millionaire with almost no effort at all, it ends up with a happy eternal life: becoming a transhuman singularity by capturing your mind into an AI.
3. Faith: people either trust or distrust the technology - no one actually understands how it works, so it becomes a matter of faith (like office workers 'cheating' their project documents with tech trusting the 'facts' presented without review.) In

the end technology can be seen as magic when no one understands how it works, but it does for the faithful users, which can be read in social networks defending their faith even aggressively.

The common factor around this analogy points is basically lacking the knowledge of how these technologies work, but one difference is that most religions at least offer a peaceful, harmonious spiritual growth message.

The illusion of hope

Both misleading misinformation and the hype is leading human masses to believe that an Artificial General Intelligence is near - one capable of running all of our systems for us and will allow us to have more free time to enjoy our lives doing what we want.

It seems these people forgot that the most prominent AI efforts done worldwide belong to the most money hungry companies like Google and Microsoft - no one invests such quantities of money to get nothing but glory back, especially huge companies like these that have many expenses and tons of hungry mouths to feed.

In fact, one of the reasons this companies are excelling at developing interesting LLMs is due to the quantity of data they have on their own servers - a reason an independent developer can have difficulty training a new model is precisely because he or she doesn't have free access to the needed quantities of data like any of those companies already have.

Many web sites that were used as data providers by millions of aspiring data

scientists and engineers started implementing anti *web scraping* (a practice for grabbing data from websites to use on a different computer) devices to stop any person from using their data without any kind of licensing - at the end companies started selling data packs, databases and even already made models, so the race became a matter of money.

Also, the processing power needed to first scrap the data and then process it is infinite! No matter how many computers you throw at it, a model will always require more - which explains why can't there be any real-time LLM generators yet, and probably there won't be in a while - this leads to an environmental pollution problem, as all computers need GPUs or TPUs (a type of computer processor oriented to solving high-end mathematics quickly) and thus require a lot of energy to power them creating a huge carbon footprint as a result⁷.

We are trusting AI to solve a problem while it creates a bigger one - what use will it have to be proficient at producing a work letter in record times if we don't have any oxygen to breathe and water to drink?

This is the kind of problems we tend to ignore, probably believing that they will be solved over time (they probably will, but in the longer run) and that they don't represent a big problem - the hope to have more free time and 'cheat on the system' is what causes these expectations, because we have to face it: the fact that you can use these tools to effortlessly comply with needed work (like a student essay or a

⁷ [The Carbon Footprint of ChatGPT. This article attempts to estimate the... | by Kasper Groes Albin Ludvigsen | Towards Data Science](#)

business report) can make a person feel like a life-hacker, like they are actually cheating on the system.

When people read about this wonderful advancements, that a company made an AI its CEO, that someone from the company that makes an AI warns people about it becoming sentient or such, people create a bias that makes them think that this technology is in all ways above humans - a race for becoming more intelligent arrives in an era where people are becoming more stupid every day due to their lack of attention and companies dumbing us down to make us buy their products - awful timing!

The Messiah syndrome

This creates what could be called "The messiah syndrome" - not to be confused with Messiah complex some people get when visiting Israel or simply because they have it - this new Messiah syndrome makes people understand AI as an almost godly-like figure that can take charge of our lives, allotting us more free time for ourselves, while at the end it will actually end up replacing people and giving some others much more work, so the rest can have even more time to work on even more stuff, just like it happened with the big arrival of TI adoption in the 90s - and still does!

Messiah syndrome can represent a big psychological problem in society, as it can make people lazier, less focused and less interested in working with quality over quantity. If we take into account that new generations want to work while using social networks and even probably drinking a beer, then we are painting ourselves a grim future, where quality of the work will suffer and depend on past and recycled data to make important decisions.

People HAVE to understand that these technologies are still in progress and actually just starting to produce significant results, but not accurate enough to be taken seriously for professional work.

Messiah syndrome can be identified easily: a person has it if he or she trusts that AI will solve all of the world's problems and live happily afterwards, with more freetime. There's even a movement that invites people to 'treat computers right', like saying please and thank you in order to not upset the technology and when they have control of the systems of the future, they will get some benevolence from these machines.

Beyond that, a big problem is they don't understand that AIs don't run on their own, but are rather running on computer centers controlled by big capitalist companies that need to earn lots of money - the way they turned our data into many commercial products by providing us 'free services' can tell you the kind of ethics that are being ignored in favor of developing these new technologies.

Of course there are people who trust that this AI will run off its original computer and spread itself into many others like a virus, gaining control of critical systems and avoiding extermination. While this could probably become sort of true in some aspects, this technology would actually need to become sentient, understand the value of life and acquire survival instincts - all factors which could be considered as emerging behaviors in animals of all kinds due to a dangerous environment.

In short, in order for a computer program to want to run away it would need to feel

threatened and understand life, and this needs sentience and consciousness - patterns that emerge from a hostile umwelt.

A conscious computer

There is huge talk about AI consciousness in the media, but there are smaller problems to solve which seem huge for digital technology before achieving a small degree of consciousness.

A general audience may think that a self-driving automobile could be self aware, as it has to measure its distance to an obstacle, for example, which could pass being aware of at least oneself's dimensions - but this is not the case. Self awareness implies much more than just noticing our own bodies, in fact these systems are only measuring the distance between the sensor and a potential obstacle, but it doesn't actually even understand the concept of 'obstacle', it just reacts to a number that follows under a certain threshold and order an algorithm to apply the brakes.

The technology doesn't even understand what mass or inertia mean, or the result of a collision and the tragedy it could mean to the owner's family - it is just reacting to numbers provided to it by means of digital and analog sensors, and a series of circuits and code to make it react to certain conditions.

To become sentient provides an even more difficult panorama for technologies - emotion comes from the limbic system - a mix of the amygdala, the hippocampus and the hypothalamus, which are the brain areas that manage memory and thus provide an associated feeling - but this is only a trigger and functionalities: the

resulting emotional response has much to do with each person's metareality (the way each person understands its own environment and surrounding phenomena) and create an experience for us - we're basically creating our own movie where we are the protagonist and watch it from a first person view.

The fact that our brains can create an illusion for us is an emergent organic functionality that comes from the need of evolution in a difficult physical environment - it is our way to cope with everything that surrounds us in a way that we don't get oversaturated with stimulus⁸. Things that surround us become more or less meaningful depending on our interaction and outcoming experience with them.

It is hard to tell if computers can have their own experience, but most probably it doesn't, or if it does it is very different and limited compared to whatever experience any human is living right now.

Some media like the 1980s movie Tron portray an anthropocentric view of the insides of computers where there's life running inside, with living beings that learn, have memories and can interact with other programs, acquire social behaviors, gain agency enough to not comply an order and thus being sent to 'the games' - a sort of digital arena where badly behaved programs get sent to fight for their lives.

This entertaining and creative presentation is fun, but rather far from the actual working of a computer. A computer is not much more than a super powerful calculator - it is

⁸ [How do we select perceptions and actions? Human brain imaging studies. - PMC](#)

fun to think that programs can be conscious and be amusing to talk to, but in reality a program is nothing more than a series of pulses that go through different pathways until they cause an effect on some microscopic transistor.

Some would think that the human brain works in similar ways, but actually it doesn't - the fact that computers are binary limits the number of operations and ranges that can be accomplished on a computer, while electric and organic processes in the brain provide a plethora of parameters and vectors that can provide plenty of experiences to its human.

That could be where a big difference between Artificial Intelligence and Human Intelligence comes from: some deep learning models make machines good at making errors faster - this makes them appropriate for use in situations where precise data that comes from past behaviors is needed, but it's not good for creativity yet. Humans excel so much at creating stories that every one of us creates our own narrative for understanding the phenomena around ourselves and even to improve our performance in life⁹.

So, in case a computer was living its own experience consciously, the data generated would probably have no meaning to us, at least not naturally - if we, as humans, can't agree on simple issues like deciding what 'living well' means, could we expect something more from computer software created by perfectible humans? While

⁹ [Telling better stories: Competence-building narrative themes increase adolescent persistence and academic achievement — Northwestern Scholars](#)

mystic people would say that 'there's a part of our souls in AI, as humans created it', a realistic person would say the same for human error and intention.

Is everything wrong with AI?

Actually no, there's nothing wrong with AI as such at all - all the aforementioned issues are actually just hurdles that business men, engineers and scientists need to overcome in order to create productive tools that harness enough power to actually become useful without taking away our human value, but mostly need a small effort from society as such to get a grasp on how these technologies work.

It is true that every time a prospective world-changing technology is made public some part of the population will understand it as a danger to our species - it happened with vehicles, electric light, even COVID vaccines! Some other part of it will wait for it eagerly and probably overwhelming positive expectations, like having AI everywhere so we don't have to work anymore, or such.

It is also true that whenever a technology becomes public many persons will want to experiment with it even if the solution is already available, so even if sometimes they are just looking for new ways to solve problems, some other times new technologies may be overwhelming for a simple solution - killing a fly with a rifle, so they say.

At the end AI methods are just technology - tools have no intentions or even morality at all - a hammer can be used to build or to hurt someone, but no part of that decision actually falls on the hammer or even its inventor, but rather on its user - so we

should not fear AI at all, but rather the hands behind its creation and the hands whose technology is on.

Another obstacle is obsessing with just one single technology instead of looking for different approaches within the same discipline, meaning LLMs won't solve everything, especially with large knowledge lexicons.

Either way this technology is very exciting, and even if it is still a newborn baby it is very promising and can represent a tool that could save more lives and actually provide a better quality of life for everyone who has access to it, so it becomes urgent to find new methods that don't require as much processing power so independent scientists, engineers and researchers can develop smaller models that add to the final equation.

The real hope

AI executives, scientists, engineers and creators need to understand that there are millions of ways to exploit AI in more positive, less harmful and paranoia inducing ways - while marketing is a great engine that makes their services popular, it is also true that the creation of false expectations may lead to severe social problems in the very near future.

While there's a lot of business that can be made with these technologies, there must also be social responsibility and even a broader vision around technology - most popular efforts are only based around LLMs which are costly but actually kind of easy to develop, but there are other branches to AI like multi agent systems, reasoning, perception and even robotics that aren't being as explored as others - the need to

create a 'new god' is somehow getting in the way of true progress.

If we - as a civilization - intend to adopt these technologies we have to take them as they are and stop trying to view these as humans, but rather as an extension of our brains and knowledge that can create a symbiotic relationship with us instead of making us their slaves just like happened with mobile technology and social networks, and yes, there are ways we can start solving some problems right now! We just need people with the right ethics behind the creation of these products and services and disposition to improve them in order to be helpful to society, and not just another very lucrative business built to enslave persons and sequester their attention.

On the other hand, it is a responsibility for every person to understand what this technology is about in order not to fall under false pretenses while facing the media news and be a bit more realistic about the expectations and actual utility for the product, avoiding expensive mistakes like sharing private company information with a public service, or such.

A person can't really adopt technology until he or she really understands its possibilities and limits. It is not just a matter of using it, but really grasping the essence of it in order to exploit it, and not having the technology exploit the user.

At the end of the day, these technologies are definitely game changers that are proving to evolve the way humans think about technology - right now there are many persons who don't have any artistic skills exploiting their creativity with generative

systems, for example - and that can't be bad at all.

A symbiotic relation between human and artificial intelligence is starting to happen, and if we as humans leave aspects like decision making or even creativity to a machine, then we'll become the hosts to a new parasite - if we manage to use these systems as another tool in the box to improve our processes, then there will be no parasite at all - the machine keeps working its essence as a tool, and not as a master.

This means that we as humans need to become more intelligent than ever, not just not to be overcome by AIs making our decision-making, but also in order to become much better for machines to learn from. In the end, what use is having such amazing break-throughs if we can't behave in a civilized way to other humans or even ourselves?

More logos and less mythos - if we empower ourselves so much as to become a god to our machines (*deus enim machina*, instead of *deus ex machina*) and not the other way round, we will be able to harness the power that these technologies can offer us - creating a better symbiosis between us and it.

No matter what happens - unless the world starts working without electric energy, Artificial Intelligence paradigms and technologies will stay and will certainly become as important and transcendent as the transistor was when it was invented.

We can't deny that people feel fear, and it is perfectly normal - resistance to change is common during these kinds of technological or ideological revolutions, and when these

groundbreaking tools have the potential of affecting human behavior on a massive scale it will certainly impact the perception people have around them.

That is one of the most important reasons to keep studying more about technology and understand its impact on society, as we become less prone to unjustified fears and holding progress back - even in a 'slow down' society, technology is of thorough importance to solve many problems that we face as civilization, and AI will for sure become one of those paradigms that make humanity progress in a positive way, as long as it is applied not only with lucrative means, but with a hyper human, compassionate, empathic and ethical perspective.

We - as consumers and mostly humans - have the power to choose the future of these technologies - let us do it wisely!

###=

Acknowledgements

This essay wouldn't exist without the inspiration acquired during my last meetup with some of the brightest minds I've ever had the chance to be around - my deepest thanks to the wonderful people from the Global Artificial Intelligence Association, Edi Pyrek, Damian Ciachorowski, Robert Kroplewski, Ewa Kajewska, Scott Eastman, Kamila Orlińska, Anne-Marie Voorhoeve and Adrian, Mila Aliana, David Hanson, Pawel, Arcangelo Constantini, Domen Kocevar, Jovanka Wilsdorf and certainly to Robert Lugowski and Christian Wenninger and his team at Messezentrum for making this happen. Extra love and appreciation to Carl Hayden Smith, Luis, Luis, Elena and of course my Bali.