

File: c:/ddc/AIQV/AIAssessment2.docx

Date: 2022 Dec; 2023 Feb, Apr, Jun, Aug, Dec

(C) Dennis de Champeaux/ OntoOO

AI Assessment
Dennis de Champeaux
SiliconValley
ddc2@ontooo.com

Abstract

AI has been around now for over 60 years. AI remains an opaque topic especially after Gogle's neural network technology took the AI banner by suggesting a crucial breakthrough with 'deep learning' and OpenAI with 'chat functionality'. Comparing the contributions of the genome to us sapiens against robotic agents and cyberspace agents reveal significant differences. These impact what AI applications are realistic. The singularity notion woke up two decades ago and triggers in the Silicon Valley again predictions about the fate of humanity. These need to be confronted and we provide arguments to neutralize these positions. Doable AI applications given the state of art (which need elaboration) are sketched. Appendices have problematic example exchanges with the 'chat'-systems and about advanced and speculative topics.

Introduction

AI has been around now for over 60 years. Someone working recently on an AI agent aimed that it would be a pleasant member at the dinner table. This paper explores, among others, the feasibility of such an entity.

AI remains an opaque topic especially after Google's neural network technology took the AI banner by suggesting a crucial breakthrough with 'deep learning'. However it is not known what would be the difference between that and 'shallow learning'. Moreover learning is an 'umbrella term' under which reside several versions of learning, each with its own type of deliverable. More recently 'chat-functionality' is claimed to be AI.

Let's try a plausible definition of AI not relying on a list of AI's sub disciplines.

AI is about the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

Required is both problem-solving and some variants of learning. This definition is good enough for a first round investigation of AI from numerous, different perspectives. We discuss obstacles for AI: missing major advantages in the sapiens genome, gaps in the AI toolbox, frictions about employment, and the inability to specify intrinsic goals for AI agents. We provide also arguments against those who believe that AI will surpass humanity. Subsequently, we will articulate an alternative, realistic role that AI can play.

Why?

Why would we spend resources on AI – as we have been doing already for 60 years? Curiosity is a fine answer. Sapiens use a brain, which has been under development for a billion of years and hence it is not obvious that brain intelligence can be replicated on a manmade machine. We still don't know – partially because it is also ambiguous what the scope is of brain intelligence – what we would have to replicate; more about that below. Another answer would be the creation of a new class of tools. We have been

making tools for millennia, intelligence is in short supply and thus tools applicable where intelligence is needed would be worthwhile.

What machine

Our definition refers to a ‘machine’, which has different interpretations. The machine is embodied or not; the former referring to (mobile) robotic devices, the latter to cyberspace (mobile) entities. In both cases, a computer (or several) is typically the ‘brain’ equivalent of human intelligence. Experiments are underway to use machines with (additional) neural network clusters in hardware [Loihi2].

Cognition (not) embodied

Human cognition is embodied. It has ‘system’ support for social connections [EmbodiedCognition]:

Primary inter-subjectivity is the pre-theoretical, non-conceptual, embodied understanding of others that underlies and supports the higher-level cognitive skills involved in mentalizing. It is “the innate or early developing capacity to interact with others manifested at the level of perceptual experience – we see or more generally perceive in the other person’s bodily movements, facial gesture, eye direction, and so on, what they intend and what they feel”.

Primary inter-subjectivity is present from birth, but it continues to serve as the basis for our social cognition in adulthood.

AI embodied or not lacks this infrastructure and thereby it is way behind in connecting with people; just consider when contacting a call-center: ‘Please notice that our options have changed’. For now, and beyond, AI agents are autistic beyond the meaning of autism.

Human moral cognition (the distinction between right and wrong) is considered grounded in embodiment also:

Moral sentimentalism is the view that our emotions and desires are, in some way, fundamental to morality, moral knowledge, and moral judgments. A particular version of moral sentimentalism holds that emotions, moral attitudes, and moral judgments are generated by our “gut reactions,” and any moral reasoning that occurs is typically post-hoc rationalization of those gut reactions (...). Embodied moral cognition takes inspiration from this kind of moral sentimentalism. It holds that many of our moral judgments stem from our embodied, affective states rather than abstract reasoning.

AI agents lack the emotional equipment that plays a key role in our moral judgments; hence its morality is at most very shallow if there is any at all.

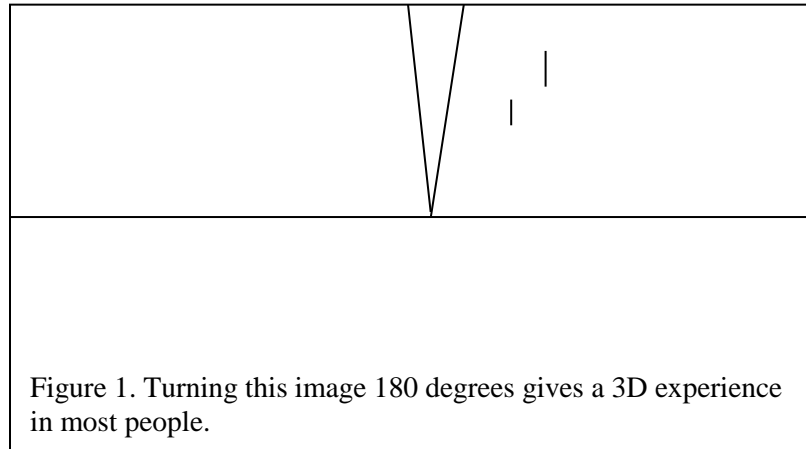
The item [EmbodiedCognition] omits the ‘detail’ that the different genders have different bodies with quite different functionalities entailing very different cognitions. Grounding of these differences in AI agents looks virtually impossible.

Pain is a unique sensation by mammals and possibly other species. It is a crucial differentiator. Our sense of mortality is another one. AI agents don’t have a ‘grip’ on these for us crucial notions.

What cognitive function: perception

We use our senses originally for analog data, while fairly recently we use vision, hearing and touch also for symbolic data. Processing non-symbolic input is an open-ended challenge, especially for video-like perception.

If the AI agent is embodied current sensor technology does not provide, for example, the 3D/depth experience (from 2D sensors) that we (and other animals) get for ‘free’/ take for granted. Pinker [Pinker] shows a diagram with over fifty distinct brain regions devoted to visual processing to provide this 3D/depth experience. There are many heuristics involved, see Figure 1 for an example.



Whether AI will ever be able to replicate our 3D/depth experience is unclear, which has consequences for motion planning & for locomotion. It is a crucial, currently missing feature for mobile robotics.

It *appears* that AI agents can recognize symbolic input because they can provide textual representations of what is read and heard. But then the problem starts for AI agents because, in general, syntax, semantics, pragmatics, the contextual setting and the knowledge repository must cooperate to capture the *intent* of the input. Major hurdles for this task are: missing a large knowledge data base and lacking an agreed upon Language of Thought [Fodor] – an internal language, mentalese – for the representation of meanings. Such a language relies on the content of the knowledge repository, needs to interface with cognitive operations and must be able to provide input to actuators.

What cognitive function: acting

The repertoire of acting is dependent on whether an AI agent is embodied or not. If not embodied (and neither indirectly as a control agent) actions consist of communication acts for which there are different settings: dialogs, monologs, etc. Only restricted dialog settings can be currently supported because using a mental model of the dialog partner and for capturing the preceding ‘moves’ is not available.

We sapiens appear to have also a hidden gadget that maintains the envelope of free space around our body. This obviates, for example, motion planning for clapping our hands or knowing that that is impossible when we are ‘sardine-packed’ in rush hour traffic. Our embodiment with cross connected, integrated sensors provide infrastructure that supports sophisticated acting. AI agents lack this infrastructure. Still we can admire now the robots from Boston Dynamics, the self-driving cars and the many robots on assembly lines. But we should not believe that their ‘awareness’ of the World is comparable with the 3D/depth skills of animals (and us) and hence the acting of mobile AI agents is very, very constrained.

What cognitive function: (optional) deliberation

In between perception and acting we have in the sapiens case a black box known as ‘working memory’. It plays a key role in non-reflex behaviors. Intelligence operates in this black box. Hence we need to check whether the intelligence of an AI agent can match what goes on in our working memory. We have already ruled out emulating cognitive operations applied to situations where social and moral aspects play a role. This entails that those operations used by social workers, psychologists, psychiatrist, human resources, legal specialists, etc., etc. are off limits.

Reasoning and planning are operations where AI agents can actually outperform the human versions – provided that the proper data structures do not have to be invented. Humans can ‘program themselves’ for a new skill like chess, go, Sudoku, road planning, etc. and can improve performance incrementally. Instead, an AI agent needs currently reprogramming for another skill that involves reasoning and planning.

Humans switch tasks routinely, which entails that they have a repository for ‘canned’ procedures and that retrieving an appropriate procedure is ‘trivial’. Adjusting a retrieved procedure to deal with a non-standard aspect is done routinely. These are desirable features for an AI agent for sure, but it assumes access to a repository of annotated procedures that can be adjusted at run time while the annotations give pre- and post-conditions so that relevance for a task can be ascertained.

Problem solving using analogy reasoning/ case based reasoning – as done by humans – is also a desirable feature but requires knowledge repository with ‘meta’ annotations that are not available.

Natural language relies on recursion for respectively the syntactic of sentences, the semantics of meanings and the pragmatics of intentions. Recursion was made available earlier on, around 1960, in the programming language Algol60. A first-in last-out data structure can be supported by the hardware with system stacks. A brain mutation – perhaps around 40000 years ago – facilitated the current natural languages. It is not the case that thinking is done using a natural language; instead we use a kind of mentales. We must assume that this internal mentales language supports recursion also. Here a natural language example with recursion of pragmatics – from [Pinker] – that requires mentales to support recursion:

... the ability to embed one thought inside another, so that we can entertain not only the thought that Elvis lives, but the thought that the *National Enquirer* report that Elvis lives, that it is amazing that some people believe the *National Enquirer* report that Elvis lives, and so on.

The ‘implementation’ of recursion in our brains supports less depth than what is available for an AI agent. At the same time natural language processing by an AI agent is still limited due to missing common sense, know-what & know-how and pragmatics – and anyway the stance of an AI agent can be labeled metaphorically as: “What the heck *is* the goal of the human out there?”

An outsider view: knowledge clusters according to Pinker

Evolution has equipped us with many tools to operate in reality: Pinker [Pinker] lists the following common sense, knowledge domain clusters:

- intuitive physics
- intuitive biology/ natural history
- intuitive engineering
- intuitive psychology
- intuitive spatial sense
- intuitive number sense
- intuitive probability sense
- intuitive economics

-- a mental data base

+ language: syntax, semantics, pragmatics

Implementing this infrastructure is a massive investment. It assumes a commitment first about an ontology that can capture any kind of knowledge – including of its own structure. The Cyc project [Cyc] is an example. It has been used in robotics projects but it has failed to be accepted as a version 0.1 for AI. That is very unfortunate because knowledge representation is a key component missing in the AI tool box since its inception.

Raw intelligence - good clock speed available in AI - does not compensate for the lack of massive amounts of common sense. Indeed, we are not the first reminding this, but we keep unfortunately forgetting it.

An outsider view: Google

Google revived from Minsky [Minsky] the idea of neural networks: “In 1951 he built the SNARC, the first neural network simulator.” Hardware improved since then and Google was able to construct many layered classifiers that could be trained by using massive amounts of data to recognize the content of images. The next step was training gaming programs by associating board configurations with value assessments. This yielded – using self play – increasing self-improvements so that world class play-levels were obtained. Protein folding has been the next success story [Protein]. These are applications that benefit from classifier technologies and thereby are indirect spinoffs of AI, but not more than that.

Google described these abilities first with ‘deep learning’ and with ‘deep mind’. Learning is a human ability with numerous different meanings – most of them not covered by training using large data sets as done by Google. The popular media ignored the nuances recklessly and the public believes now that AI will soon take over the world. At least they should have questioned what ‘deep’ has to do with ‘learning’ and with ‘mind’.

An insider view: Marcus & Davis

Marcus & Davis wrote "Rebooting AI: Building AI We Can Trust" [Marcus&Davis] in which they demolished ‘deep learning’. They recommend:

-- Robots require rich cognitive models and deep understanding

-- Representations, abstraction and generalizations, the role of theories, understand causal relations, etc. need elaboration

-- Common sense needs to be fleshed out (with proposals how to go about it)

These recommendations are actually not surprising because they were on the wish list, by and large, already in the 70s.

An insider view: Judea Pearl

Judea Pearl wrote in ‘The Book of Why’ [Pearl]:

"Deep learning has (instead) given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality."

Perhaps we can elaborate this assertion by using a design pattern for AI agents having a 3-partition with the unsurprising components: -- perception, -- deliberation/decision making, -- acting. The middle component depends crucially on ‘a model of reality’. The success of deep learning in games exploits its non-reliance on this middle component. The inability to explain the ‘why’ of its output, and the non-extendibility, etc. makes us concur with Pearl's position.

An outsider view: OpenAI and Google

ChatGPT is also not AI with the same argument by Judea Pearl. The middle component is hampered because its ‘knowledge’ is a hodgepodge of info islands that can be contradictory. It exploits that the meaning captured by natural language (the user’s input) is (generously) *correlated* with the user’s intention and hence the program can at best *guess* what the user’s intention is. If the input matches something on which it has been trained, it can dump large amounts of (often irrelevant) information.

We have encountered numerous examples where it fails dramatically when dealing with questions where expertise knowledge is necessary. Several examples are in the Appendix ‘Conversations’ with ChatGPT, among which “Why is deep learning not AI according to Judea Pearl?”. A Google search with this question as input avoids a direct answer; instead links are provided to articles that are potentially relevant for the user. Input stories/puzzles that do not match trained input show surprising good responses but also reveal major gaps in its problem solving abilities; stories with a sequence of events can yield wrong conclusions.

ChatGPT is a questionable AI agent because of its ‘knowledge’ not being integrated while its cognitive operations are quite defective. It is also autistic and has no morality. A ‘discussion’ about the topic who-pays-what US taxes and about fairness yielded a stream of statements that were mutually contradictory, but still with a pretension of (pseudo) authority.

Asking ChatGPT (and Google’s Bard) about itself – using ‘you’ in the queries – confirmed what we have claimed above:

- No, I don’t have personal experiences, emotions, or the ability to feel love.
- Emotions are complex and involve subjective experiences, consciousness, and biological processes that machines like me do not possess.
- I don’t have the ability to punish or make decisions about consequences. I am a machine learning model created by OpenAI, and I lack personal experiences, values, or intentions.

We conjecture that the ability to be interviewed in this way could not be achieved by being trained by what is available on the Web. Instead we believe that the creators of these systems wrote up documents to satisfy ‘nosy’ researchers and subsequently fed them into their system.

Whether these systems can be salvaged by not dealing with inappropriate topics is, obviously, unclear to us. ChatGPT and Bard can be misleading as acknowledged by OpenAI and Google. Releasing them, while knowing that they could ‘hallucinate’ (their term), is – in our opinion – unconscionable.

An insider view: Landgrebe & Smith

Landgrebe & Smith write “Making AI meaningful again” in 2019 before ChatGPT arrived [Landgrebe & Smith]. They provide numerous arguments that there are serious problems with capturing the nuances of natural language and argue that:

“This ... means that dNNs, whatever it is that they are doing, cannot be modeling the semantics that need to be captured in order to extract information from texts, a crucial task in natural language processing for most automation purposes. ... The failure of dNN-based approaches to compute natural language semantics is illustrated also by the recent misclassification of the United States Declaration of Independence as hate speech by the Facebook filter algorithms.”
{dNN is deep neural network.}

They provide two sentences which they consider problematic:

“The cat caught the mouse because it was slow” vs.

“The cat caught the mouse because it was quick.”

What is the “it” in each of these sentences?

Feeding a variant into ChatGPT and asking ‘What happened?’:

“The cat didn't catch the mouse because it was slow.”

yielded the bizarre response:

“... it means that the cat was unable to catch the mouse, and the reason for this failure was not the mouse being slow. In other words, the mouse might have escaped due to other factors or the cat's inability to catch it, but the mouse's slowness was not the cause of the cat's failure to catch it.”

We have in an appendix other simple examples with unacceptable responses.

Their summary:

Taken together, these problems rule out entirely the use of machine learning algorithms *alone* to drive mission-critical AI systems—for example with capability such as driving cars or managing nuclear power stations or intensive care units in hospitals. They are too brittle and unstable against variations in the input, can easily be fooled, lack quality and precision, and fail completely for many types of language understanding or where issues of liability can arise. Even at their very best, they remain approximate, and so any success they achieve is still, in the end, based on luck.

Economic headwinds

Let's ignore for now all the problematic aspects for AI agents outlined thus far and ask the unpleasant question why AI has not delivered a blockbuster financial success after 60 years of funding. AI must operate in anomalous welfare states in which majorities of the workforce have negative economic value and depend on system charity (C. Murray's: Custodial Democracy). Maintaining employment has a high priority so that those workers can maintain their self-esteem in the fabric of the society. This entails that an AI application should not endanger employment by being too successful.

The self driving car under development for over a decade has been touted as *the* breakthrough AI application. Not thus far. Feasibility on freeways has been demonstrated decades ago. A self-driving mini-van was driven in 1995 by Dean Pomerleau and Todd Jochem of CMU on a 2850 mile trip from Pittsburgh to San Diego [Pomerleau]. However, in 2017 the International Brotherhood of Teamsters, a union representing 1.4 million members, successfully lobbied Congress to place a 10,000 pound weight-limit on pending driver-less vehicle legislation to protect their employment. A Congressional Research Service document [R45985] (2021 April) explains:

"... it is likely that widespread use of fully autonomous vehicles – with no driver attention needed – lies many years in the future."

Unprotected left turns, bad weather, no 3D/depth vision, infrastructure optimized for drivers but not for automatic vehicles etc. remain technical challenges. Geo-fencing has been used to demarcate regions where self-driving cars/ trucks cannot do harm.

There is still another daunting obstacle. Self driving vehicles would lead to self-driving taxis, which would remove the need for personal vehicles, which stand idle 97% of the time. This would have a major impact on the percentage of the transportation sector on the GDP, impacting its employment as well as the taxation revenue stream of this sector. Oregon prohibited self-pumping gas to protect employment. Self-driving taxis could be blocked similarly.

IBM ran into trouble in the healthcare sector after it won a "Jeopardy" contest in 2011 through its ability to find answers in unstructured data more effectively than standard search technology and tried immediately to apply this functionality.

Strickland elaborates in [Strickland] 'How IBM Watson over-promised and under delivered after its triumph on Jeopardy':

... encoding a human doctor's expertise in software turns out to be a very tricky proposition.

This statement echoes AI's lacking a generic knowledge representation formalism, missing a subset for a medical ontology and more important a medical discipline which over the last half century itself has not embarked on developing such an ontology:

... doctor's notes and hospital discharge summaries. That narrative text accounts for about 80 percent of a typical patient's record - and it's a stew of jargon, shorthand, and subjective statements.

AI's stagnation in natural language processing and without a medical ontology blocked commercialization of a diagnostic tool. The [Strickland] paper describes numerous other attempts by IBM in other medical sub-disciplines with their failures, based on - we believe - omissions in the AI basic tool box:

... in trying to apply Watson to cancer treatment, one of medicine's biggest challenges, IBM encountered a fundamental mismatch between the way machines learn and the way doctors work. -- "All the prior guidelines have been thrown out, based on four patients."

But Watson won't change its conclusions based on just four patients.

The [Strickland] paper is recommended for any party venturing in the healthcare system.

These are daunting technical issue. Now consider the context in which this IBM system had to operate. US Healthcare is 18% of GDP versus 10% elsewhere. The best explanation for the discrepancy is that the healthcare sector is used to give citizens jobs – irrespective of economic rational as required elsewhere in the private sector. This is possible because healthcare is funded mostly by government programs that rely on taxation. Proposing functionality that would reduce employment is not 'welcome'.

We had a similar experience with a web-based diagnosis system Health Check (1999-2005) that can be used in the range of self-help, physician mediated up to supervised by a call center. It has an ontology, a scalable design for medical knowledge and an integrated diagnosis algorithm. It generates machine readable health records that go beyond what physicians typically scribble in these records. The HC system contains: 236 locations, 102 body systems, 498 diseases (and 94 'super-class' diseases), 83 abnormal conditions, 928 symptoms, 153 external causes, and 244 medication/drugs. This HC system is available uninterrupted since 2005. While physicians routinely miss-diagnose, this system has no takers (but could still be relevant for under-served regions). [The Clinton administration advocated over two decades ago standardizing medical billing. The HC machine readable records could feed into such a system, but who is interested in improving operating efficiencies?] The advantage of machine readable records for the medical practices would materialize only after years of usage. Just *seeing* the patient with one or more follow up questions is typically good enough satisfying current practices. The medical discipline has disease taxonomies ICDX (X=8,9,10,11, ...). These taxonomies have not been extended with attributes (symptoms, etc.) to support a diagnosis process and neither with best treatments. A standard for eRecords is lacking and there is no agreement in the health industry to establish it. Increased transparency is an issue for physicians. And again, cost reduction is not welcome.

AI gaps

The AI toolbox has splintered since its inception in different sub-disciplines which hardly cross fertilize each other.

-- Automatic deduction/ theorem proving is focused on assisting correctness proofs for challenging algorithms

-- (Heuristic) search has not progressed in how to generate the insides of a 'node'

-- Problem solving has split off two decades ago with its ad hoc PDDL formalism for representing problem families and problem instances

-- Natural language processing is stuck because commonsense representation is not available, Fodor's Language of Thought is still not available, the formal pragmatics of the use of language is not available, and the formal semantics of performatives is not available

- Knowledge representation for symbolic knowledge (ontology) is not off-the-shelf available (although OpenCyc [Cyc] has been used by the robotics community)
- Knowledge representation for procedural know-how is not available; one can program a robot how to boil an egg, but one cannot tell it how to do it, because the representation of the semantics of actions using Hoare-logic is absent
- Learning was discouraged in the 70s because it presupposes how the many different kinds of learned 'stuff' get stored and be retrieved as needed, which was not available then and which is still not available.

AI limits

Limits are not popular. A level headed discussion is in J.D. Barrow 'Impossibility/ The limits of science and the science of limits' [Barrow]. AI as a sub discipline of Informatics is subject to the limits of complexity theory. Most cognitive operations have exponential or worse complexity. Hence AI is as bounded as we are and an exponential intelligence explosion is simply ludicrous.

We suspect that our elaborate (Math) theories are just the result of lucky (depth first) search operations and our ability to capture the outcomes and pass them on to next generations. There are no reasons to believe that AI, by itself, can do better.

The outcomes of our 'mysterious' curiosity lead to questions that 'somehow' we are pursuing topics based on a social reward system (which is error prone). How can AI by itself pull this off?

Goals for AI agents

Humanity has an abundance of prohibitions about what is *not* allowed. Virtually nothing is agreed upon what is allowed and what we should strive for. (This has potential fatal consequences for the state of the biosphere.) The Maslow hierarchy describes needs that culminate in the goal of self-realization, which does not have an operational interpretation [Maslow]. Given that *we sapiens* do not have agreed on transcendental goals; we are empty handed in how to provide AI agents an intrinsic sense of purpose if they would have to coexist with us. Since AI agents are not equipped with morality it is not feasible to give them goals with choices. (This has a distinct advantage: self-driving cars don't have road rage.)

AI Singularity

Kurzweil revived in 2005 an old idea that future robots will surpass humanity in "The Singularity is Near" [Kurzweil]. His arguments are based on further exponential growth of the hardware as it happened before 2005. These arguments have failed. Hardware got only marginally faster in the previous decades. Multicore machines did arrive; however parallel software does not scale up. Quicksort which is 'absurdly' easy to parallelize does not run N times faster with N cores and maxes out already for low values of N .

Few people can produce requirements, an architecture and sequential software. Parallel software is way, way harder. The notion that AI will create its own software is science fiction; creating data structures/ (mini) ontologies is a black art. Even if that would happen, computational complexity prevents an exponential explosion.

All living things are in conflict of interests with all living things resource wise. The notion that we will grant AI agent's unlimited access to resources is beyond absurd. There are already communities that reject data centers because they are energy hogs. Humanity has already in 1980 surpassed the carrying capacity of the Planet [Ecofootprint]. We need to reduce *our* (global) consumption and giving resources to AI agents needs realistic cost-benefit justifications.

Another notion that AI agents will surpass and subsequently eliminate us sapiens assumes that they have our questionable human nature. That assumption – a kind of paranoia – overestimates what we can/ will design.

In drive by fashion, we can dismiss Tegmark's "Life 3.0" [Tegmark]. It is a 'feverish' extrapolation of the singularity into the next 10000 years and billion years beyond. It fails for the arguments given above.

AI next description

We have used:

AI is about the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

This definition needs an upgrade. IQ has a bell curve distribution. Hence the weird question: what human level should AI compete with?

The notion of a bell curve suggests that IQ is like length and weight: just a linear metric. This notion could be justified when IQ is driven by the equivalent of the clock speed of a cpu: based on, say, different neuron signaling speeds.

There are more factors relevant in human intelligence. Personality theory uses a 5 dimensional space with openness as one component. Another component is antagonism/ disagreeableness. The latter has certainly a negative connotation, but we suspect that it plays a role in driving creativity/ out of the box thinking. Fresh anger can be a source for sublimation: another driver for exceptional performances. AI intelligence does not have these drivers but has as many cycles as we are willing to provide them. However, exponential challenges cannot be conquered with more cycles. Exceptional AI performance needs more.

A 2nd component is the amount of relevant knowledge that can be recruited. A person's knowledge is unique, based on the history, what knowledge has been accumulated, how many abstractions have been created and the number of knowledge clusters that have been extended and created. A person has numerous, different goals and experts have been trained with applying and/or extending a knowledge cluster. An AI agent is different by typically being customized for specific functionality. This can provide for exceptional performance in a specific domain indeed. In contrast, people are able to extend their abilities by creating new (mental) operations and get a 'feel' for the domain, which gets captured in heuristics.

By now it should be clear that the definition for AI:

... simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

is misleading. It promises capabilities that cannot be achieved due to the consequences of non-human embodiment: lacking emotions, lacking morality, lacking social mindreading, lacking integrated sensors that have been enhanced over a billion years to enrich incoming data streams, lacking genome provided bootstrapping learning abilities, etc.

What remains is:

AI is about the creation of dedicated tools that simulate human intelligence in machines that 'think' like humans and mimic their actions.

There are also constraints on these tools. For example, they should currently not cause massive employment destruction.

This leads to the question what tools have we developed already and what can we expect? We have stationary industrial robotics. Non-stationary robots are hard to find because – as discussed – they don't have 3D/depth vision. Self driving vehicles will require restrictions, adjustments of the infrastructures, further enhancements, etc. before they will be accepted – assuming that the induced destruction of employment in the transportation sector will not lead to a rejection. Cyberspace AI agents are somewhat hiding. We do have maps that provide travel and distance estimates as well as tools that guide how to go from A to B. There is no equivalent of an enhanced 'yellow page' service that would provide a programmatic interface to websites. This would allow scheduling a trip involving multiple modalities

(car/ plane/ ferry/ train). This was discussed two decade ago but there was no follow up. Since cyberspace AI agents can leverage browser technology for perception and acting we expect more applications forthcoming like apps that provide matches for demand and supply services. Power tools to support research through the generation of conjectures would be quite desirable. As discussed interactions involving social-‘mind-reading’ appear – for now – not feasible. Corporations will have their own in-house versions.

Hopefully it is clear by now that the spin-off applications generated by Google neural networks that are customized by massive data streams and chat-agents based on Large Language Models are not AI applications according to our portrayal.

The balance of this study considers topics that are speculative, but may still shed a different light on AI. Appendices discuss AI Consciousness, Hypothetical developments for AI, and AI Creativity.

Conclusion

We have provided a battery of arguments that AI is currently massively misunderstood. Our analysis is an antidote to the ‘gospel’ spread by Google in the previous decade. We have demolished the Singularity. Unrealistic expectations have been stripped and AI has been repositioned as ‘just’ yielding tools with narrowly prescribed functionality. We foresee limited opportunities for mobile robotic AI due to their lack of our 3D/depth vision. Cyberspace AI applications are also hampered by lacking social mindreading abilities; they are beyond autism. In addition we have outlined numerous gaps in the AI tool box since its inception. These gaps restrict the kind of application tools we can develop.

The balance of the paper has in appendices speculations on debatable topics: AI consciousness, launching AI agents in space and AI creativity. There is also an appendix with a more recent test of OpenAI’s ChatGPT and Google’s Bard on a simple, four line story; both fail to generate the correct answer. They admitted in ‘interviews’ that they have no emotions, values, intentions, consciousness,

This result – and numerous others – support our conjecture that these systems are brittle, cannot be trusted and don’t live up to the hypes of this decade. Also, General Artificial Intelligence is not coming soon – if ever.

Acknowledgement

Years of zoom sessions with Frank Crossman made this piece possible.

References

[Barrow] Barrow, J.D. ‘Impossibility/ The limits of science and the science of limits’, Oxford University Press, 1998.

[Carruthers] Carruthers, P., “The Centered Mind”, Oxford University Press, 2015.

[Chalmers] Chalmers, D.J., “The Conscious Mind”, Oxford University Press, 1996.

[Cyc] <https://en.wikipedia.org/wiki/Cyc>

[deChampeaux] Champeaux, D. de, “Faster Linear Unification Algorithm”,

Journal of Automated Reasoning volume 66, pp 845–860, 2022.

<https://doi.org/10.1007/s10817-022-09635-1>

[Ecofootprint] <https://www.footprintnetwork.org/our-work/ecological-footprint/>

[EmbodiedCognition] <https://plato.stanford.edu/entries/embodied-cognition>

- [Fodor] Language of Thought Hypothesis, https://en.wikipedia.org/wiki/Language_of_thought_hypothesis
- [Hacker] Hacker, P.M.S., “Is there anything it like to be a bat?”, *Philosophy*, volume 77, #3000, pp 157-174, Cambridge Oxford Press, 2002 Apr.
- [Harnad] Harnad, S., The Symbol Grounding Problem., *Physica D* 42: 335-346, 1990
<https://arxiv.org/html/cs/9906002>
- [Kurzweil] Kurzweil, R. , “The Singularity is Near”, Viking 2004.
- [Landgrebe & Smith] Landgrebe, J. & B Smith, “Making AI meaningful again”, *Synthese* **198**, 2061–2081 (2021). <https://doi.org/10.1007/s11229-019-02192-y>
- [Loihi2] <https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html>
- [Marcus&Davis] Marcus, G. & E. Davis, “Rebooting AI: Building Artificial Intelligence We Can Trust”, Knopf Doubleday Publishing Group, 2019.
- [Maslow] https://en.wikipedia.org/wiki/Abraham_Maslow
- [Minsky] <https://web.media.mit.edu/~minsky/>
- [Pearl] Pearl, J. , "The Book of Why", Basic Books, 2018.
- [Pinker97] Pinker, S., “How the Mind Works”, Norton & Company, 1997.
- [Pinker] Pinker, S., “The Blank Slate/ The Modern Denial of Human Nature”, Penguin Books, 2002.
- [Pinker12] Pinker, S., “The Better Angels of our Nature”, Penguin Books, 2012.
- [Pomerleau] <https://apnews.com/article/pittsburgh-plays-7eaf98e853534958a422ec71e100c4c0>
- [Protein] <https://www.nature.com/articles/d41586-021-03499-y>
- [R45985] Congressional Research Service, "Issues in Autonomous Vehicle Testing and Deployment".
<https://fas.org/sgp/crs/misc/R45985.pdf>
- [Side-effects] “Side Effects: Impacts on the 21st Century”, 2021,
<https://www.amazon.com/-/es/Dennis-Champeaux-ebook/dp/B09MHJ5W48>
- [Strickland] <https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care>
- [Tegmark] Tegmark, M., “Life 3.0” , Vintage, 2017.

Appendix Three ‘conversations’ with ChatGPT

Input starts with ‘**’ and comments use ‘<...>’.

** Why is deep learning not AI according to Judea Pearl?

Judea Pearl is a computer scientist and a pioneer in the field of artificial intelligence (AI) and causality, and he has argued that deep learning, which is a subset of machine learning, is not a form of AI. According to Pearl, the main reason deep learning is not AI is because it lacks the ability to reason and understand cause-and-effect relationships.

<That is NOT his argument. A model of the world is missing that plays a role between perception and acting.>

Pearl argues that true AI must be able to understand and reason about the causal relationships that exist in the world. He believes that deep learning, which is based on neural networks, is a powerful tool for pattern recognition and prediction, but it is not capable of understanding the underlying causal relationships that generate the data it is trained on. He believes that in order to achieve true AI, we need to develop methods that can infer causality from data, such as the causal inference methods he has proposed.

<Keeps repeating wrong argument.>

Pearl's view is that deep learning is a powerful tool for pattern recognition and prediction but it does not provide the kind of understanding that we associate with true AI, which is the ability to reason about cause-and-effect relationships.

<Way too verbose & repeating again.>

It's worth noting that Pearl's view is not shared by all researchers working in the field of AI, and there are other researchers who argue that deep learning is a form of AI and that it has the potential to learn and understand causal relationships.

<An incredible last paragraph. It imitates mentioning anonymous scientists, researchers, experts, advocacy groups, etc. as done in the mainstream press. >

** Generate an in-place sorting algorithm with $N \log N$ complexity.

"Quicksort" Algorithm:

Pick a pivot element from the array. Reorder the elements such that all elements less than the pivot come before it, and all elements greater than the pivot come after it. Recursively repeat the same process for the sub-arrays on either side of the pivot. The pivot element is now in its final position and will not need to be moved again. Repeat the process until the entire array is sorted.

Quicksort has a time complexity of $O(N \log N)$ on average and $O(N^2)$ in the worst case, but the average case is much more likely.

<This content is likely taken from the Quicksort wiki, which is out of date.>

** You fail again.

I apologize for my previous response not meeting your expectations. Can you please specify what you are looking for so I can assist you better?

** I specified worst case complexity $N \log N$, not average.

Internal server error

There was an error generating a response

<A better answer would have been: I don't know. The correct answer is actually missing in the Quicksort wiki: apply the Musser defense as described by David Musser in his 1997 paper.>

Stephen Wolfram wrote in his remembrance of Doug Lenat:

But with a word problem like "If Mary has 7 marbles and 3 fall down a drain, how many does she now have?" we didn't stand a chance. Because to solve this requires commonsense knowledge of the world, which isn't what Wolfram|Alpha is about. But it is what CYC is supposed to be about.

We tried something similar with ChatGPT:

** John bought a prime number of oranges.

He ate one.

He made juice using a prime number of oranges.

He had one left.

How many oranges did he buy?

ChatGPT; If John bought a prime number of oranges, and he ate one, then the number of oranges he had before making juice is also a prime number, let's call it "p" ...

<Obviously the reasoning derailed already and its subsequent reasoning concluded erroneously – with conviction "Therefore, there is no solution to this problem." There are actually many and possibly an infinite number when the conjecture holds that there are an unlimited number of prime pairs two apart.>

Appendix 'Conversations' with ChatGPT and Bard

An example input given to OpenAI ChatGPT and Google Bard. Input starts with ‘**’.

** John left A at noon and arrived in B three hours later. Peter left an hour later than John starting in C. C is further away from B than A, and the distance between A and C is half the distance between A and B. John and Peter arrived at the same time in B.

What is the speed ratio of John and Peter?

A solution: Peter’s travel time is 2 hours. If L is the distance between A and B then the distance between C and B is $L*3/2$. The speed of John VJ is $L/3$; the speed of Peter VP is $(L*3/2)/2$. Hence the speed ratio $VP/VJ = (L*3/4) / (L/3) = 9/4$.

ChatGPT concluded that the ratio is 2. Its reasoning derails already with “Since they arrived at the same time, their travel times are equal ...”

Bard replied with:

John traveled for 3 hours, so the distance between A and B is 3x.

John traveled half as fast as Peter, so Peter's speed is 2x.

Peter traveled for 2 hours, so the distance between C and B is 4x.

The speed ratio of John and Peter is 2:1.

So the answer is 2.

Bard derails with “John traveled half as fast as Peter ...” for which there is no evidence.

Appendix Creativity in Math

Math has numerous ‘tricks’ to foster creativity. One of them is taking a well-developed theory: select a subset of the axioms, add some trusted theorems and reprove theorems from the original theory.

Topology is an example. Another trick has been starting with a minimal trusted set of axioms and developing a theory that explains, among others, numbers. When stuck another axiom is added. Set theory is an example. It was necessary to add an infinity axiom without which Math could not be grounded.

Statistics emerges from probability theory, which in turn emerges from measure theory. Groups, rings, fields and analysis is another chain of Math theories.

Examples of progress/creativity in CS/ Informatics:

- The search algorithm begat the A* algorithm
- The A* algorithm begat the bi-directional A* algorithm
- The depth first search algorithm begat the iterative deepening algorithm
- The depth first search algorithm begat the bi-directional depth first search algorithm (submitted)
- The exponential unification algorithm begat linear unification algorithms [deChampeaux]

These are examples of analogy reasoning at the meta level.

Appendix AI Consciousness

Consciousness is a large topic. This section provides only a testable conjecture. We limit ourselves to access consciousness and self-consciousness and dismiss/ignore the sentience version [Chalmers] following [Pinker97] and [Hacker]. We reject the Tegmark’s definition of consciousness = ‘subjective experience’, because every experience is subjective.

Consider a system with the following features:

- It is directly or indirectly embodied through sensors with the World
- It has an internal, recursive meaning representation language as postulated by [Fodor] containing “self” and “I”, or their equivalent

- It has an internal, sufficiently rich knowledge representation, which includes the self as an entity where the content is grounded in Harnad's sense [Harnad]
- It has the functional equivalent of a working memory as understood by Psychology, Neuroscience and Neuro-Philosophy [Carruthers]
- It has a special, short-term episodic memory in which recent events about, among others, the self are being recorded
- It has a thread of control that deals with the processing of here-and-now events, deliberations and actions
- It has at least a 2nd, meta thread of control that monitors the self using, among others, the short term episodic memory

We claim that an AI agent equipped with these features will have access and self-consciousness, although the 'quality' of the consciousness is a 'watered down' version of sapiens consciousness. Such an agent can be queried about its status with: 'How are you doing?' We suspect that sapiens self-consciousness itself varies greatly.

Appendix AI Hypothetical

We must give credit to Tegmark [Tegmark] for exploring launching AI agents in the galaxy and beyond, say, because self-consciousness is a precious asset and we may/should have the responsibility to export it. Immediately the question arises whether this can be done without exporting (all) the negatives that evolutionary psychology has identified in our species in the previous century (while their findings are still ongoing, [Pinker12]). The track record of the previous millennia with dark periods does not foster optimism that we can avoid numerous, current collapse scenarios [Side-effects]. Our ability to self-destruct has still no solution. We may have to wait until we reach a new steady state after the dust of the upcoming turmoil has settled. Unresolved – as discussed above – will be the question what transcendental goals to equip these future AI agents. Just 'exploration' doesn't cut it. Expanding knowledge may be OK, but how to share new knowledge universe-wide is ultimately not doable. We have unresolved ethics regarding interactions like: AI agent – AI agent, AI agent – an encountered other life form, AI agent – an AI tribe/ community/ ... What to do with potential new knowledge requires (meta) ethics, which we can't agree on currently. We need way more solid insights about ethics/ transcendental goals before we can launch anything.

Appendix AI Creativity

Intelligence and creativity have overlap but they are not the same. Intelligence uses acquired knowledge and applies it; creativity generates something new or using existing resources in a new way. Creativity is characterized by the ability to perceive the World in new ways, to find hidden patterns, to make connections between seemingly unrelated phenomena, and to generate solutions. Hence creativity involves two processes: thinking and then producing. We can distinguish creativity with a small 'c' and with a large 'C'. In case of the former, the newness pertains to an individual's self-assessment. In case of the latter, there is the additional need of recognition of newness and its value by the 'forum': public performances, published books, published scientific/ engineering articles, etc.

AI creativity depends on who is doing the assessment. There have been cases where an AI system was (somewhat) creative because we sapiens were pleasantly surprised by an unexpected, novel outcome.

It is a different situation when a community of AI agents must decide using *their* knowledge and value system whether something new deserves the label of a Creation.

Perceiving the World in a new way is non-trivial for us sapiens. It relies on extensive familiarity with a particular domain, on certain personality traits, on 'enough intelligence', on the maturity of the domain, and there is the acid test whether there is usefulness. Literature helps us identifying new, emerging

(mental/ social) pathologies. Science attempts to help us resolve incomprehensible tests outcomes. Engineering attempts to compensate for ongoing self-inflicted collective damages.

Analogy reasoning has helped your author with new solutions – it seems. Whether sapiens creativity can be enhanced by team efforts is unclear. See the appendix for Math creativity.

The prerequisites for sapiens creativity are daunting. How to equip an AI agent/ an AI tribe with creativity is quite unclear.