

Why (so many) humans would fail the Turing Test

The mythical Turing Test has for almost 70 years been seen as the litmus test to determine whether machines could think as humans, but how valid is it really? What is it that the test actually measures, and what if humans fail it, does that mean that they are not capable of reasoning? These questions come to the forefront in the development of AdBots designed to be indistinguishable from humans.

The Turing Test was developed by the English mathematician Alan Turing in 1950 as a mean to assess whether machines had the ability to display humanlike intelligent behaviour to the point that they no longer could be distinguished from humans.

Turing proposed that the test should be designed so that a human evaluator would judge text-only conversations between a human and a machine from a natural language perspective, being aware that one of the participants was indeed a machine. Typically, the tester tries to snare the assumed machine with plethora of cognitively related questions, which could be of an arithmetic nature, i.e. *what is 12×12 ?* that a human should be able to answer, blended with inhumanly difficult questions that a human should fail, as well as, questions of a more philosophical nature, such as, *if you close your eyes, does that mean that the car in front of you no longer exist?* (Whilst the answer to this question might seem obvious to you readers, in fact most people, due to a common psychological phenomena, often in practice do chose to overlook the obvious, pretending it is not there, or the reverse assuming things that are not, in accordance with *The Emperor's New Cloth Syndrome*.)

The machine would pass the test, if the evaluator was not able to identify it as a machine through this battery of various questions aiming at seeking out the characteristics of human reasoning. Given that level of natural language is a screening criteria, the decisive qualifying quality is more about the machine's ability to mimic how a human would communicate and reason rather than giving correct answers to all queries.¹

And whilst the Turing Test has been criticised from philosophical reasons, it remains a standard test to determine whether AI applications can assume humanlike reasoning, however the definition of human reasoning remains elusive, *what is it really?* Then there also is the much less discussed other side of the coin, what about humans failing the test? Would that mean that they are bereft of human reasoning? This has been something of a blind spot in the exclusively one-sided Turing Test, but carries a significant impact as the assumption has always been that whilst machines could be caught out as just being machines, humans would always pass the test and never assumed to be machines. If the test can identify an actual human lacking the ability to reason like a human, well, how can we then know for sure that a machine is not reasoning like a human?

What 'type' of humans, if any, can one then expect to fail the Turing Test, and more interestingly, why would that be?

To start with, it is important to have an understanding of the variation of mankind's cognitive capabilities as it closely relates to the *quality* of human reasoning. If using IQ as a rough proxy measure of cognitive capability, the mean global IQ is estimated to be about 88.² Assuming a normal distribution, half of the world's population are above 88, and the other half are below it. Having an IQ below 90 however comes with noted consequences, in antiquated psychological publications it is classified as *dullness*, characteristics including;

- Generally not suitable for higher education, and:

- They remain above the threshold for normal independent functioning and can perform explicit routinised hands-on tasks without supervision as long as there are no moments of choice and it is always clear what has to be done.

If having an IQ below 80, previously classified as *borderline deficiency*, the limitations become even more profound;

- Limited trainability and have difficulties with everyday demands, and therefore require assistance from family or social workers to manage their lives;
- Will struggle with even low-level education, and;
- Are generally unemployable unless for simple tasks requiring supervision.

And an IQ under 70 marks the boundary for where investigations of mental retardation usually commences.^{3,4}

Whilst these kinds of descriptions rarely are accepted today due to the sensitivities in talking about cognitive levels, various euphemisms are instead routinely deployed to tone down the previous bluntness but however vague the contemporary wordings are, the gist of the previous descriptions still apply, as they remain painfully factual. So, regardless of how unflattering and uncompromising the characteristics of the below IQ 90 group are, they highlight difficulties with comprehending features such as simple arithmetic and more abstract philosophical ponderings. In other words, they would struggle to answer the aforementioned types of questions, with the risk that they by the Turing Test evaluator be declared lacking the capacity to reason like humans, *and this might apply for half of the world's population.*

A side point is the estimate on how many of these individuals that will be considered employable in a highly automated knowledge-based economy, where machines' level of reasoning starts to supersede theirs, let alone having considerably higher work capacity. For a truly gloomy scenario of the future, please see *Idiots Breed Idiots, why men no longer are created equal.*⁵

But there is also another cognitive cluster that might fail the Turing Test, and they are at the other end of the IQ distribution. Individuals with a noted high IQ but with autistic (or similar) traits that always takes things a bit too literal and tend to respond in a more robotic rather than in a human fashion. This as there appears to be something lacking that is above and beyond pure logic, perhaps it is not having the capability to comprehend the emotional status of their counterparts, not being able to fathom what is conveyed *when reading between the lines*. This is also the group that is likely to occasionally be able to answer inhumanly difficult questions.

Now, I am not the first to point out this issue, far from it, as others have noted the difference between, at least some, intelligent behaviour and human behaviour, and the Turing Test values human behaviour ahead of intelligence in order for a machine to pass it. Pointing out humans as machines in evaluations is referred to as *the confederate effect*.^{6,7} But beyond observing and labelling this effect, there have been few, if any, studies of the practical consequences.

Human communication is miscommunication

If the Turing Test is likely to fail large segments of humanity, what it is that it then actually is testing? Implicitly, it seems to expect a minimum quality of human reasoning, which might be in excess of what the average < IQ 90 person could muster, however it must also capture the ability to grasp emotional insights and ambiguities, both which appear to relax certain deductive properties.

But what many, including Turing Testers, tend to forget is that human communication is entangled with miscommunication, and much of our communication and reasoning is devoted to the attempts, often futile, to overcome it. This miscommunication is due to humans need to be able to handle and respond to conversations that contains;

- Ambiguity;
- Incomplete information;
- Incorrect Information, and;
- Multiple points of view, including opinions & hypotheses

The recognition of the above utterances, including any emotions they might convey, even covertly (*the understatement* often seen as the definition of Englishness being a case in point), require considerable cognitive enterprise and energy. But unfortunately, the complexities of reasoning do not end here, this as the human mind consists of a conscious and an unconscious part with separate logic structures and these absorb reality in diverging chunks, with the former, truncated through narratives and norms, and the latter able to amass broader perceptions of reality. These are held together and controlled through a governing mechanism. They interact in accordance to a protocol which often manifests in decision-making that can be perceived as seemingly irrational but it is far from it, rather it follows a diverging schema aligned to attain goal maximation. This brings an element of irrationality in human reasoning that is difficult for a machine to replicate.

So, to correctly interpret a counterpart's opinions and questions, it calls for an ability to understand his *specific* narrative, whether that be of a cultural, political, or religious nature, or usually a combination thereof. As narratives curtail reality into a *social reality* where reasoning is often confined to the information that exists within its boundaries, ignoring a diverging actual reality through dogmatic tenets. This means that even when arriving at conclusions following the correct deductive steps, for a person operating on a differing narrative it is seen as an aberration, even bordering to self-deception.

In essence, to develop machines exhibiting human like behaviour, they must be capable of replicating and standardising all of the above features, including unconscious contents that influence our language and actions, into a fully computational model. A difficult but not impossible feat, far from it.

Why is the question of what human reasoning actually is so important to us?

Well, I work with a startup that are developing *AdBots*, where one of the defining design features is their indistinguishable dispositions from online manifests of human reasoning. In that sense, they must be devoid of any *mechanistic* thought- and language patterns that might out them as bots. In attempting to harmonise artificially induced thought patterns with human thought patterns, come the insight that the modeling of human reasoning needs to mimic the imperfections that is what makes it precisely human. Thus, understanding what the Turing Test is trying to measure, and ultimately how humans appraise social media interactions, therefore serve as quality markers in the calibration and testing of bots.



Niklas Hageback the CEO of The Virtual Mind, a Sweden based startup which is developing AdBots. He is also a published author, including a number of titles in psychology, and is currently working on a book on how to develop machine generated creativity.

Footnotes & References

1. Turing, Alan. "Computing Machinery and Intelligence" *Mind - A Quarterly Review of Philosophy and Psychology*, Vol. LIX. No. 236. October, 1950.
2. Lynn, Richard and Vanhanen, Tatu. "*IQ and the Wealth of Nations*" Westport, CT: Praeger, 2002.
3. Kaufman, Alan S. "*IQ Testing 101*" New York: Springer Publishing, 2009. p. 110 ctd.
4. Terman, Lewis M. "*The Measurement of Intelligence: An Explanation of and a Complete Guide to the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*" Riverside Textbooks in Education. Ellwood P. Cubberley (ed) (Boston: Houghton Mifflin, 1916). p. 79.
5. Hageback, Niklas. "*Idiots Breed Idiots, why men no longer are created equal*" Sweden: Logik Förlag, 2018.
6. Saygin, A. P. & Cicekli, I. "Pragmatics in human-computer conversation" *Journal of Pragmatics*, 34 (3), 2002. p. 227 – 25.
7. Shah, Huma & Henry, Odette. "*The Confederate Effect in Human-Machine Textual Interaction*" 2005.