Can a Machine be Intelligent?: Analysing Notions in AI Discourse

BULLO KANO

Are machines intelligent? Perhaps this question is already late. It may be seen to be redundant when the term "Artificial Intelligence" or "AI" has already gone rampant in modern discussions. Further, AI fields of research may not care if the debates in philosophy are able to finally conclude if machines can think, be conscious, or be intelligent. Philosophical discussion in turn often takes on some or the other notion of *intelligence* and declare machines to be either intelligent or not. From AI cars to AI in mobile phones, the application of AI technology already, it seems, remain immune to what philosophical nitpickings have got to say on the subject. Such indifference is perhaps, not altogether unjustified. AI as an academic discipline largely has to do with computational modelling of various human capabilities; all these capabilities perhaps marked as intelligent or constituting being intelligent. As such, the discipline may correctly be considered as mostly consisting of research in techniques involved in computation such as the AI programming languages, artificial neural networks, machine learning, all this again to reach at least human level abilities if not to surpass them. In light of the success of AI research, theoretically as well as in its application, it is easy to forget its roots in philosophy.

Perhaps, much depends on how AI is defined. A controversial (yet fitting, in my opinion) definition provided by Margaret Boden is that AI is seen as "the science of intelligence in general"[1]. Its goal then is to arrive at systematic theories that provide for explanations for the various mental or psychological capacities. It then encompasses the entire range of possible minds. As such, AI cases still remains very central in philosophy, especially philosophy of mind. AI gives for interesting cases on which to test philosophical concepts and theses. For instance, human-like android, Commander Data, from the star trek series is taken up by Ned Block in posing the question behind the *harder problem of consciousness* – "Why should physically different creatures overlap phenomenally in one way rather than another or not at all?"[2]. Block's contention is that as of *now* we have no conception of consciousness that allows us to answer such questions or to test for consciousness (including human consciousness). Further, he says, "But we should not be happy with the idea that there is a science of human phenomenality, another of machine phenomenality, and so on."[3] So following Boden and Block, one might as well decree that AI is relevant to discussions on consciousness in general or the *science of consciousness in general*. Anaylses of notions in purview of AI cases then not only becomes important in expanding the old horizons of those notions but also in providing for new horizons to explore. The aim here then is to analyse the notions of *understanding* and *intelligence* with regards to AI and provide for an answer to the initial question of – Are Machines intelligent? However, the take away lessons will also include testing the boundaries of those notions irrespective of AI.

1. The Two Approaches

Can a machine think? The question was first proposed by Alan M. Turing in 1950[4]. In present day debates, the question may be more specifically applied to computers – can computers think or act intelligently? A variant of this question may be stated as – is Artificial Intelligence (AI) possible?

Journal of Comparative Literature and Aesthetics Vol. 46, No. 2, Summer 2023 [90-97]

Philosophical debates centred around these questions make use of two approaches. First way is to analyse, discuss and assert something about the two players typically involved in AI discussions – the human and the machine. With regards to human, often the assertions will be of the type that they are nothing more than a kind of machine, a physical system. Therefore, positions[5] with such assertions conclude that consciousness, awareness, intentionality, intelligence and all that mental stuff have their basis on a physical system of sort. And since they are so, it is very much plausible and possible to create a physical system (different from the physical system that a human is) that can have or do all that the human physical system can. One such system is the physical symbol system as proposed by Newell and Simon[6]. For them, symbols lie at the root of intelligent action. So being a mere physical system is not enough. To be intelligent, one essential requirement, as the duo claimed, is the ability to store and manipulate symbols. So, what follows from such claims is that humans' capacity for intelligent actions are fundamentally a capacity for symbol manipulation. Humans are themselves a physical symbol system. The point to be noted here is the *downplay* of humans and all their capacities to a physical machine, especially the downplay of intelligence to symbol manipulation. Correspondingly, opposition to the possibility of machines being capable of human capacities will oppose any such downplay of humans. A disclaimer here is that the paper does not necessarily take sides with any of the positions for its purpose nor is it the purpose here to argue for any such posits. Also, while the notion of downplay may include reductionist thesis, especially reduction to the physical, it is not limited to such theses.

With regards to machines, it is said that they are physical systems. Though different from the ones that humans are, nonetheless, still a physical system with the capacity for symbol manipulation and anything else that the human physical system is capable of. Here the human physical system may be capable of possessing non-physical capacities, thus, referring to them as a physical system is not a downplay. Not only that machines can be aware, understand and act intelligently, they can be so in their own subjective way. They are a species in themselves, just like any other. Such a position is held by Kevin Warwick[7]. The point to be noted here is the *up play* of machines. And correspondingly, opposition to the possibility of machines being capable of human capacities will oppose such up play of machines.

What happens in this approach is that the debate tends to focus on the entities, what they are capable of and then move on to concepts of those entities and capacities. There is a distinction between focusing on entities and focusing on concepts. For instance, let's consider the statement "Humans are mortal". One way to explain what we understand from the statement is that it points out to a set of beings out there in the world, existing, and talks about that set of beings being mortal. Another way is to see them as concepts and how the concept "human" implies attribution of the concept "mortal". Whatever is human will be mortal. Whatever is included in the concept of human or whoever the concept is applied to, will lead to the application of the concept of mortality to it. The determination of the concepts by the entities and the determination of the possibilities of the entities by the concepts might be extremely intertwined, so much so that while talking of entities, conceptual invocations may occur in the same breath and vice-versa. The AI debate as described above, in talking about the players, make claims about those players and hence focuses on the entities and what they are capable of.

But there is a second way to approach the debate – the conceptual way. In this approach one may focus on the concept of machines and see whether the concept of intelligence is applicable to it. For further demonstration of the distinction, we may take up John Searle's Chinese Room argument[8]. A computer is provided with an algorithm that enables it to simulate the understanding of Chinese. A question is given to it in Chinese, it matches the question against its database and produces appropriate answers, also in Chinese. If the answers are as good as a native Chinese speaker then according to Turing's test, this computer will have to be thought of as having a mind. The interrogators will think that they are speaking to a Chinese person. This is then taken as enough to attribute understanding

of Chinese to the computer. Searle now substitutes the computer with himself locked in a room. So, Searle is in a room and in this room are several baskets of Chinese symbols. Searle does not understand Chinese at all. But there is a rule book in English given to him for manipulating the Chinese symbols. The rules specify the manipulations of the symbol purely in formal terms. So the rule might say, "Take XYZ (Chinese symbols) from basket number one and put it next to WXY (Chinese symbols) from basket number two". Now some Chinese symbols are passed into the room and following the instructions from the rule book, Searle passes back Chinese symbols out of the room. Now again suppose that unknown to Searle, the symbols passed into the room are called 'questions' and the symbols passed out of the room are called 'answers to the questions'. By manipulating the symbols according to the rule book, Searle is said to have a conversation with the outside interrogators. He is taken to be answering Chinese questions in Chinese language. Thus, Searle is considered as being able to understand Chinese. But this is absurd as it is known that Searle does not understand Chinese. The attribution of understanding Chinese to Searle in this case is simply misplaced.

It may be observed here that the entities, the computer and Searle, are in fact incapable of understanding Chinese. But there is also a conceptual aspect to this whole argument. Searle states that mental states by their very definition have certain sorts of contents. A computer by its very definition consists in symbol manipulation. A computer program is specified only in terms of *syntax*. Syntax is just the formal structure or arrangement of symbols. If the symbols themselves are not semantical, that is, if they do not have any content, then there can be no meaning to those symbols purely based on their formal structure or arrangement which is the syntax. A computer program simply has no semantics. Therefore, it can have no understanding, no mind and surely no thinking. A point to be noted is that Searle is not referring to any set of computers existing at that period or existing now or in future. The force of his argument lies in the thrust that he denies the concept of understanding to the concept of computers. And in doing so it also holds for the entities currently existing or will come to exist in future. A conceptual contradiction will be committed if understanding is attributed to computers.

2. Analysis of the Notions

The strategy here is to first up play humans and downplay machines to the fullest. One may imagine a spectrum with humans on one end and the machines on the other end. Congruent and coinciding with this spectrum imagine a spectrum of a concept with humans on one end and the machines on the other end. Suppose the concept of consciousness is taken up. Then on the human end, the concept will be in its fullest form as opposed to a reduced form somewhere in between and nil in the end. The idea then is to see how the spectrum of other concepts adjusts and clicks with the human-machine spectrum.

Can a computer think? The question sort of presents two concepts that are important in this enquiry. First is the concept of intelligence. Machines can be said to be intelligent or not only if we know what intelligence is. Second is the concept of understanding. To be intelligent requires one to be able to understand. This seems to be a natural assumption among many. Thus, central to the AI debate are the two concepts of understanding and intelligence. However, another concept must be taken up before the other two are analysed. The concept of computer is the third notion central to this debate. But the paper will only briefly deal with this concept as required for the analysis of the other two concepts.

One understanding of computer is that it is mere symbol manipulation. As such computer programs are mere syntax, with no semantic content. Such a definition of computer is what Searle uses in his Chinese room argument. A slightly different understanding of computer may be had. Tim Crane roughly defines computer as a device which processes representations in a systematic way[9]. The difference of this definition with the previous one is the notion of representations. Following Crane, a general and obvious way to understand this notion is that a representation is something that

represents something. What is important is that representations are not just abstract inert symbols. As representing another thing, the notion of representation has in it this idea of being about something, that something being the object of representation. As such, it follows that – the notion of representation itself has within it this tacit reference to something, something of a kind which is a content or *content-like*. This is what distinguishes representations from mere symbols, which does not represent anything but may be used as representations. So, this definition of computers as processing representations may be understood as processing not mere abstract symbols, but processing of symbols that represent something. The symbols may be abstract in themselves, but so long as they represent, they are representations. This understanding of representation is what makes the second definition of computer not only about symbol manipulation but also something more. It is due to the power of representations themselves that the representational activity of computers can be brough about. The claim here is that a computer is more than just syntax. So, definitions that reduce computers to be essentially syntactical do not do justice to the notion of computers.

Now first, the notion of understanding is taken up. It is natural to assume the necessity of *understanding* in anything (a creature, an organism, or a system) for intelligence to be attributed to it. For instance, Searle's argument posits that the computer does not understand just as Searle in the room does not understand Chinese. He maintains, mental states involve semantic contents. The content is regarding the state of affairs in the world. If he is right then when one understands, the mental state one is therein, has a semantic content involved. There is no semantic content in the manipulation of symbols that a computer program is. So, there is no understanding involved. And hence, computers cannot be intelligent. On the other hand, Warwick posits that computers may be understood to have its own subjective understanding. It can at the least understand machine codes. Further, machines can be given its own sensory systems and made aware of their surroundings. The input from the sensors may be converted into machine codes and the machines will be aware of the surrounding. The machines now have a subjective understanding of the world around them. It seems there are semantics involved. As such they are intelligent in their own subjective way. If one were to work with the definition of computer that Searle works with, then obviously syntax is not enough for semantics, and computers does not have any semantic content. In the above discussion computers were defined as processing representations. As such they are more than syntax. However, this definition does not amount to the claim that computers have semantic contents and thereby understanding, as Warwick claims. Understanding that arises due to semantic content will surely be not possible in a computer.

But the notion of understanding is not limited to semantic content. There can be understanding that does not require semantics. The subjective character of experience of an organism is not something external in the world and thus, an understanding of it is not about something external in the world either. Nor is it about any particular experience which can serve as the object of understanding, rather the understanding is present in all subjective experience of an organism as subjective experiences for that organism. Hence, such an understanding has no semantic content. Perhaps this is the sort of understanding that a computer may have, if semantics is denied to it. May be this is what Warwick's claim amounts to when he argues for subjectivity in awareness, understanding, intelligence and so on, for machines as a species. But to assert this there is a problem. Thomas Nagel points out that members of radically different species will never know what it is like to *be* a member of another species[10]. This is because the subjective characters of experience of different species do not share a common reality. There is a pre-assumption here. The bats, the humans and most of other species are already assumed to have subjective experiences which are to be accounted for. In fact, it is the case that the science of today, does not provide a proper account of it precisely because the kind of reduction they make, walks away from the very thing they attempted to explain in the first place. Now *computationalism* as being a form of *functionalism* reduces mental states to functions which rip them off its subjective character. Further, if these functions are then taken and put (as algorithm) in a system, of which it is already known that it has no consciousness or mental states, then it is doubtful

that these functions will in this new system get back that subjective character that was ripped off of them. It may be argued that in this case the subjective character is already in the new system, perhaps because if something can be unique in its physical parameters, then it must surely be unique in its subjective parameters. But this amounts to pre-assuming consciousness in the system. So while it may be possible that the functions, now as algorithm, when put into a conscious system may regain a subjective character from the new system, it seems not at all possible that functions when put in a non-conscious system will have any such subjective character by virtue of being functions. Thus, a computer defined as processing of representations cannot be attributed the kind of understanding that has no semantic content. In fact, such an understanding has to precede the understanding that consists of semantic content. So, just fitting computers with sensors and making them aware of their surrounding world, giving them some semantic content, will not amount to understanding them.

But what about intelligence? Can a computer be intelligent? Given the natural assumption of the necessity of understanding in attributing intelligence, the answer may be an obvious denial. The concept of intelligence is next taken up. Newell and Simon described intelligence as a composite, for no single thing accounts for intelligence in all its manifestation. There simply is no intelligence principle. One requirement for intelligence according to them is the ability to store and manipulate symbols. Computers being such systems can, therefore, be intelligent. It seems here that the notion of understanding plays no role in the notion of intelligence. Warwick draws an analogy between intelligence and cake. He says, "The question is perhaps similar to asking: when baking a cake, how much of the quality of the cake is down to the original mix of ingredients how much is down to how it is cooked? ... What people have been looking for… is the recipe for intelligence"[11]. One may as well say, "While creating intelligence, how much of the quality of intelligence depends on…?" It may immediately be noted in the asking of this question the awkwardness of the term "quality of intelligence". Can there be any quality of intelligence? Can there be, for instance, a "bad" quality intelligence just as there can be a "bad" quality cake? Or if one says "intelligent action" can it be also said "bad intelligent action"? To say "bad intelligence" creates confusion. This is precisely because intelligence itself is a quality attributed to other things, namely, action and behaviours. Both "bad" and "intelligent" are adjectives. As such even if they are used together, they only affect the noun or verb they are connected to. So to say "bad intelligent action" may actually mean "bad action" and "intelligent action" where both are about the same action, and not that the quality of intelligence is bad. For if there was lacking in the action, then the quality will be of lesser intelligence and not bad intelligence.

To say human beings are intelligent is to attribute them the quality of intelligence. This attribution should not be confused with something, processes and capacities, in the human that is intelligence itself. For instance, one way to understand intelligence, as George Butterworth points out, may be by reference to the psychological processes which give rise to knowledge[11]. Richard Gregory describes intelligence as – 1) intelligence of stored knowledge termed *Potential Intelligence*, and 2) intelligence of processing for problem solving termed *Kinetic Intelligence*.[13] What may be observed in these descriptions of intelligence is that intelligence is defined as a product of processes in correlation to knowledge. Intelligence as a product is therefore taken to be an entity or perhaps a more appropriate term will be a *quantitative property*, just like energy. Also, there are those who take intelligence to be something in the mind or brain. For instance, Warwick defines intelligence as: "the variety of information-processing processes that collectively enable a being to autonomously pursue its survival"[15]. His definition of intelligence takes specific processes as constituting intelligence. Searle, who correlates intelligence with consciousness, brain processes and so on, may be taken as providing a similar kind of understanding of intelligence. In all these definitions, it may be observed that intelligence is taken to be an entity or a composite or both together. In case of Warwick it seems to be a complex entity. In case of definitions where intelligence is an abstract correlated with knowledge and hence a product, it is taken to be a composite. My claim is therefore two-fold. Firstly, intelligence is not an entity and secondly, it is not a composite. Together, as a conjunction, my assertion is that intelligence is not a composite entity.

The reification of intelligence as an entity is a mistake. It follows from what has been discussed above with regards to the cake analogy by Warwick. This fact is quite commonly understood among psychologists, who in fact often uses tests to investigate intelligence by treating it as an entity. W. B. Dockrell writes: "The different concepts of intelligence held by the participants in the symposium minimize the danger of accepting any one point of view about intelligence as correct. There remains the danger of unconsciously reifying the concept of intelligence and treating it as though it were an entity and not merely 'a convenient manner of speech'"[15]. Intelligence is a quality that is attributed to doings and doers, that is, actions and actors of those actions. Another concept that applies to actions is that of "good" or "goodness". A such my claim is that the analysis of the concept of intelligence may be carried out just like the analysis of the concept of good as given by G. E. Moore. Moore takes good/goodness to be a simple notion and hence, not definable[16]. By not definable he means, good cannot be further decomposed into constituents. He says that propositions about "the good" are all synthetic and never analytic. "The good" is the thing which has good/goodness as its quality. So in the statement, "Pleasure is the only good", the "only good" is the thing that has as its quality goodness and pleasure is asserted to be that thing. The statement, according to Moore, is not *analytic*. It does not give the meaning of the word "good". *Analytic* statements are those where the relation between the subject and predicate is drawn based upon the meanings of the constituent terms. *Synthetic* statements are those whose truth depends upon the facts about the world, which is known through experience. The relation between the constituent terms is not based on their meanings, but on facts or state of affairs in the world. So, synthetic statements will be of the kind – "Pleasure is good", "Happiness is good" and so on. Moore also explains that simple terms are "simply something which you think of or perceive, and to anyone who cannot think of or perceive them, you can never, by definition, make their nature known."

A similar case can be made for intelligence. The reason why many feel "intelligence" is a complex, multifaceted thing or entity is precisely because they feel that its meaning can be understood in terms of some constituent parts, such as communicating, thinking, processing, being conscious and so on. But then it may be realised that statements such as "He is intelligent", like statements such as "Pleasure is good" are all synthetic. Also intelligence is such a term which, to anyone who cannot think of it or perceive it, can never be made to know its nature, simply by definition. Imagine for instance that one has to explain an AI computer what intelligence is. How will one do it? It seems almost impossible to define to do so. Given that computers follow strict rules, telling it "This is intelligent" while pointing at a game of chess or rules of logic will only lead the computer to equate the game of chess or rules of logic to the term "intelligent". It is often said by AI supporters that a computer, which has machine learning or deep learning capabilities, can extract features from cases which are marked as intelligent and then have a kind of learning about intelligent acts. The question then is how will it know what the features are being marked as? Suppose in the game of chess, it may learn which moves to make when faced with a given situation. How will it know those features are intelligent? Such marking just give rise to synthetic relations and it does nothing to explain what intelligence is to the computer. Perhaps, explaining qualities to the computer might be too much to ask. It may also be asked: "Do Chimpanzees ever think that humans are more intelligent than them?" or "Do they ever wonder if monkeys can be as intelligent as them?" Thus, I posit that intelligence is a simple notion and cannot be defined. Also, it follows then that "intelligence is a simple quality". This then contradicts any understanding of the concept which takes it as a complex entity. It is more sensible to attribute that simple quality of intelligence to individual actions and processes, rather than try and define intelligence in terms of those actions and processes.

3. Answering the Question

Now, can computers be intelligent? If this question amounts to asking if computers can be intelligent and thereby can computers understand, then the answer according to what I have argued is that

they cannot have understanding and thereby cannot be intelligent. But the question has another sense in it. Can the concept of intelligence be attributed to the concept of computers? So far it is a quality that can be attributed to actions, computers can be attributed intelligence if what they do are considered as actions. An implication of such attribution will be that the concept of intelligence is no longer as strongly connected to the concept of understanding. A computer might not understand at all and yet be considered intelligent.

What has happened here is that the human-machine spectrum for the concept of *intelligence* does not coincide with the human-machine spectrum for the concept of *understanding*. The spectrum of the concept of intelligence with humans on one end and machines on the other end is not congruent to the spectrum of the concept of understanding with humans on one end and machines on the other end. This juxtaposition is different from the alignment of spectrums of those concepts according to the common or natural way of relating understanding and intelligence. For according to the natural assumption, that understanding is necessary for attributing intelligence, the intelligence concept spectrum will superimpose on the understanding concept spectrum. The spectrums, tuned as ranging from humans to machines, will be the same for both the concepts. Being intelligent then depends on being able to understand. However, the analysis here shows that such superimposition of the concept spectrums for understanding and intelligence are not correct, or at least not necessary.

Jawaharlal Nehru University, New Delhi

Notes

[1] Boden, *The Philosophy of Artificial Intelligence*, 1990, p. 1.
[2] Block, "The Harder Problem of Consciousness", *The Journal of* Philosophy, August 2002, Vol. 99, No. 8 (Aug., 2002), p. 423.
[3] Ibid., p. 408.
[4] Turing, "Computing Machinery and Intelligence", *Mind 49*, 1950.
[5] One such position is that of physicalism, the thesis that the physical exhausts the nature of the world, including the mental. However, such assertions may not be limited to physicalism. For instance, one may assert physicalism about awareness and intentionality while taking consciousness (the phenomenal) to be non-physical. Such a position is held by David Chalmers. See Chalmers, 1995, 1996.
[6] Newell and Simon, "Computer Science as Empirical Enquiry: Symbols and search", reprinted in *The Philosophy of Artificial Intelligence*, 1990.
[7] Warwick, *Artificial Intelligence: The Basics*, 2012.
[8] Searle, *Minds, Brains and Science*, 1984.
[9] Crane, *The Mechanical Mind*, 2003.
[10] Nagel, "What is it like to Be a Bat?", *The Philosophical Review*, vol. 83 no.4, 1974.
[11] Warwick, 2012, p. 24.
[12] Butterworth, "Infant Intelligence", *What is Intelligence?*, 1994.
[13] Gregory, "Seeing Intelligence", *What is Intelligence?*, 1994.
[14] Warwick, 2012, p. 17.
[15] Dockrell, *On Intelligence: The Toronto Symposium On Intelligence, 1969*, 1970, p. 5.
[16] Moore, *Principia Ethica*, 1903.

Works Cited

Boden, Margaret A. *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden. Walton Street, Oxford: Oxford University Press, 1990.

Block, Ned. "The Harder Problem of Consciousness." *The Journal of Philosophy*, vol. 99, no. 8, 2002, pp. 391–425. *JSTOR*, https://doi.org/10.2307/3655621. Accessed 15 Feb. 2023.

Chalmers, D. J. (1995). Facing Upto The Problem Of Consciousness. David Chalmers. Retrieved February 17, 2022, from http://consc.net

Chalmers, D. J. (1996). The conscious mind : in search of a fundamental theory. Oxford University Press, USA.

Crane, Tim. *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations*, first published in 1995, second edition. London: Routledge, 2003.

Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903

On Intelligence: The Toronto Symposium on Intelligence, 1969, edited by W. B. Dockrell. London: Methuen & Co Ltd, 1970.

Searle, John. *Minds, Brains And Science*. Cambridge, Massachusetts: Harvard University Press, 1984.

The Philosophy of Artificial Intelligence, ed. Margaret A. Boden. Walton Street, Oxford: Oxford University Press, 1990

Turing, Alan M. "Computing Machinery and Intelligence". *Mind 49*, 1950, pp. 433-60.

Warwick, Kevin. *Artificial Intelligence: The Basics*. Abingdon, Oxon: Routledge, 2012.

What is Intelligence? ed. Jean Khalfa. Cambridge: Cambridge University Press, 1994.