What Is Artificial Intelligence?

William J. Rapaport

**Departments of Computer Science and Engineering, Philosophy, & Linguistics,
and Center for Cognitive Science
University at Buffalo, The State University of New York, Buffalo, NY 14260-2500**

rapaport@buffalo.edu ; http://www.cse.buffalo.edu/~rapaport/

Abstract

Wang (2019) claims to define AI in the sense of delimiting its research area. But he offers a definition only of 'intelligence' (not of AI). And it is only a *theory* of what intelligence is (artificial or otherwise). I offer and defend a definition of AI as computational cognition.

1. The Nature of Definitions. Forward-looking (or *prescriptive*) definitions suggest what researchers ought or want to work on. Backward-looking (or *descriptive*) definitions are based on what researchers have actually tried to do or succeeded in doing.[1] Examples abound in the history of science: Atoms were originally conceived as indivisible; we now know that they are not; electrons were originally conceived as tiny particles, a view now challenged by quantum mechanics. Reconciling such apparently incommensurable definitions or concepts is an open question in philosophy.

In the case of AI, there is an obvious candidate for the forward-looking, prescriptive definition (mentioned, but not explicitly cited, by Wang): McCarthy's definition from the Dartmouth conference that gave the field its name:

> the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. (McCarthy et al., 1955)

This definition is free from the anthropomorphism that Wang criticizes in others (although McCarthy et al. do go on to talk of solving "problems now reserved for humans").

Whether modeled on humans or not, AI has also tended to be Janus-faced, with the interaction between the naturally-occurring original and its computational model going in both directions, as in these two definitions:

1. … *artificial intelligence*, the science of making machines do things that would require intelligence if done by men. (Minsky, 1968, p. v)
2. By "artificial intelligence" I … mean the use of computer programs and programming techniques to cast light on the principles of intelligence in general and human thought in particular. (Boden, 1977, p. 5)

Here, the anthropomorphism is surely eliminable (delete "if done by men" from Minsky's, and "and human thought in particular" from Boden's). Minsky looks at naturally occurring "intelligence" and seeks to re-implement it in machines. Boden looks at computation and seeks to use it to understand "intelligence".

And, of course, there are problems (noted by Wang) raised by the "fluidity" of concepts and the difficulty (if not impossibility) of providing necessary and sufficient conditions for concepts best understood as having only family resemblances. As a consequence, one-sentence definitions such as any of those under discussion are really only acceptable for quick overviews or dictionaries. To really understand a subject, one needs at least an encyclopedia article, a textbook, or a research program (Rapaport, 2019, §3.3.3).

2. Wang's Definition. "On Defining Artificial Intelligence" offers no such definition. Ignoring 'A', Wang concentrates on 'I': "Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources" (p. 17). As definitions of intelligence go, this is not a bad one, though I find it notable that Wang pays scant attention to definitions of intelligence from the psychological literature (e.g., Gardner 1983; Sternberg 1985). Moreover, Bertram Raphael observed "that AI is a collective name for problems which we do not yet know how to solve properly by computer" (Michie, 1971, p. 101), which implies that, once we do know how to solve them, they are no longer AI (Wang, 2019, p. 11). Daniel R. Schlegel (personal communication) points out "Without the 'capacity' part of … [Wang's] definition, this would be lurking in his definition—once something is understood to the point that adaptation is no longer required, it isn't an intelligent action anymore."

What about 'A'? Wang says that he won't discuss the possible confusion with 'artificial' in the sense of "fake" (p. 3) and that "how to interpret the 'A' is not a big issue" (p. 4). I think this is a mistake. The nature of AI's "artificiality" has

[1] "The dictionary, after all, is more of a rearview mirror than a vanguard of change"—Peter Sokolowski, cited in Fortin 2019.

played an important role in philosophical discussion: The argument from biology in Searle (1980) states in essence that an AI that is A is therefore not I (Rapaport 2000b; Rapaport 2019, §19.6.2).

Wang suggests that his definition of intelligence "corresponds to a *working condition* and a *coping strategy* that are both different from those of computation" (p. 17). If so, then what does AI's artificiality consist in? Yet he suggests that AI both will and will not be algorithmic:

> ... an intelligent system defined in this way cannot always solve problems by following problem-specific algorithms .... On the other hand, a computer system eventually runs according to algorithms. The solution of this dilemma is to combine algorithm-specified steps to handle each problem-instance in a *case-by-case* manner .... (p. 20)

He seems to think that if AI is computational, then there must be a *single* algorithm that does it all (or that is "intelligent"). He agrees that this is not possible; but whoever said that it was?

He also puts a lot of weight on the view that "A program is traditionally designed to do something in a predetermined *correct* way ..." But AI researchers from the very beginning have relied on "heuristics", not in the sense of vague "rules of thumb" or fallible suggestions of how to do something, but in a very precise *algorithmic* sense:

> A *heuristic for problem p* can be defined as an *algorithm* for some problem p', where the solution to p' is "good enough" as a solution to p (Rapaport, 1998, p. 406). Being "good enough" is, of course, a subjective notion; Oommen and Rueda (2005, p. 1) call the "good enough" solution "a *sub-optimal* solution that, hopefully, is arbitrarily close to the *optimal*." (Rapaport 2017, p. 15; Rapaport 2019, §3.15.2.3; see also Romanycia and Pelletier 1985; Chow 2015)

Thus understood, an AI heuristic is a "predetermined correct way" to do something that is (arbitrarily) *close to* what minds do. It is related to Simon's notion of bounded rationality; so (given Wang's remarks in §4.1), Wang should be sympathetic to it.

As for his comment that

> traditional computer systems should be taken as unintelligent, as they are designed according to principles that are fundamentally different from what we call intelligence. From a theoretical point of view, AI should not be considered as the same as computer science, or a part of it. (p. 16)

one should consider the fact that Turing Machines themselves were conceived along the lines of McCarthy's and Minsky's methodology: Analyze how humans solve a certain problem, and then devise an algorithm that does the same thing in the same way (Rapaport, 2017, p. 12).

3. My Definition. AI is a branch of computer science (CS), which is the scientific study of what problems can be solved, what tasks can be accomplished, and what features of the world can be understood computationally (i.e., using the language of Turing Machines), and then to provide algorithms to show how this can be done efficiently, practically, physically, and ethically (Rapaport 2017, p. 16; Rapaport 2019, §3.15). Given that CS's primary question is "What is computable?", I take the focus of AI to be on whether *cognition* is computable.

I agree with Wang that both 'A' and 'I' are not the best terms, so I replace 'A' by 'computational' and 'I' by 'cognition': Computational cognition (which we can continue to abbreviate as 'AI') is the branch of CS that tries to understand the nature of cognition (human or otherwise) computationally. By 'cognition', I include such mental states and processes as belief, consciousness, emotion, language, learning, memory, perception, planning, problem solving, reasoning, representation (including categories, concepts, and mental imagery), sensation, thought, etc. AI's primary question is "How much of cognition is computable?"; its working assumption is that *all* of cognition is computable (echoing McCarthy's original definition); and its main open research question is "Are *aspects of cognition that are not yet known to be computable* computable?" If they are, does that mean that computers can "think" (i.e., produce cognitive behavior)? If there are *non*-computable aspects of cognition, *why* are they non-computable? An answer to this question should take the form of a logical argument such as the one that shows that the Halting Problem is non-computable. It should not be of the form: "All computational methods tried so far have failed to produce this aspect of cognition". After all, there might be a new kind of method that has not yet been tried.

Wang's definition of intelligence is a proposal about *how to go about finding* computational solutions to cognitive abilities. Do any of those solutions also need to be solutions to the problem of how *living* entities cognize? *Pace* Boden, not necessarily, for at least two reasons. First, a process is comput*able* iff there is an algorithm (or perhaps multiple

interacting algorithms) that is input-output equivalent to the process. There is no requirement that natural entities that exhibit a comput*able* behavior must themselves do it computa*tionally* (Rapaport, 1998, 2012, 2018). Second, as Shapiro (1992)[2] has urged, there are 3 distinct goals of AI: (1) AI as advanced computer science or engineering extends the frontiers of what we know how to program and to do this *by whatever means will do the job*, not necessarily as humans do it. (2) AI as computational psychology writes programs as theories or models of *human cognitive behavior*. (3) AI as computational philosophy investigates *whether cognition in general* (and not restricted to *human* cognitive behavior) *is computable*.

Wang has two objections to defining AI as computational cognition. First, he suggests that some of the items included under cognition as characterized here are simply "other vague concepts" (p. 5), themselves in need of definition. But my proposal first refines 'I' to 'cognition', and then further refines 'cognition' to that (family resemblance) list above. Refining those further becomes one of the tasks of AI (along with the other cognitive sciences). To the extent that AI succeeds, each aspect of cognition will be made precise.

Second, Wang raises the specter of "fragmentation" (p. 12): separate solutions to each aspect of cognition, but no unified one such as we humans apparently have. This problem does need to be addressed: Various modes of cognition do have to interact somehow, but it doesn't follow that a single AI "master algorithm" is needed. Separate modules with a central coordinating system is also a possibility. Fragmentation in other sciences, such as math or physics, has not been a serious obstacle to progress.

References

Boden, M. A. (1977). *Artificial Intelligence and Natural Man*. New York: Basic Books.

Chow, S. J. (2015). Many meanings of 'heuristic'. *British J. Phil. Sci. 66*, 977–1016.

Fortin, J. (2019, 20 September). When dictionaries wade into the gender (non)binary. *New York Times*. https://www.nytimes.com/2019/09/20/style/they-nonbinary-dictionary-merriam-webster.html.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books, 2011.

McCarthy, J., M. Minsky, N. Rochester, and C. Shannon (31 August 1955). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. http://www-formal.stanford.edu/jmc/history/dartmouth.html.

Michie, D. (1971). Formation and execution of plans by machine. In N. Findler and B. Meltzer (Eds.), *Artificial Intelligence and Heuristic Programming*, pp. 101–124. New York: American Elsevier.

Minsky, M. (1968). Preface. In M. Minsky (Ed.), *Semantic Information Processing*, pp. v. Cambridge, MA: MIT Press.

Oommen, B. J. and L. G. Rueda (2005, May). A formal analysis of why heuristic functions work. *Artif. Intell. 164*(1-2), 1–22.

Rapaport, W. J. (1998). How minds can be computational systems. *J. Exp. Theor. Artif. Intell. 10*, 403–419.

Rapaport, W. J. (2000a). Cognitive science. In A. Ralston, E. D. Reilly, and D. Hemmendinger (Eds.), *Encyclopedia of Computer Science, 4th ed.*, pp. 227–233. New York: Grove's Dictionaries.

Rapaport, W. J. (2000b, October). How to pass a Turing test: Syntactic semantics, natural-language understanding, and first-person cognition. *J. Logic, Lang. & Info. 9*(4), 467–490.

Rapaport, W. J. (2003). What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. *Minds and Machines 13*(3), 397–427.

Rapaport, W. J. (2012, January-June). Semiotic systems, computers, and the mind: How cognition could be computing. *Int'l. J. Signs & Semiotic Systems 2*(1), 32–71.

Rapaport, W. J. (2017, Spring). What is computer science? *Amer. Phil. Assn. Newsletter on Phil. & Computers 16*(2), 2–22.

Rapaport, W. J. (2018). Syntactic semantics and the proper treatment of computationalism. In M. Danesi (Ed.), *Empirical Research on Semiotics and Visual Rhetoric*, pp. 128–176, 273–307. Hershey, PA: IGI Global. Revised version of Rapaport 2012.

Rapaport, W. J. (2019). Philosophy of Computer Science. http://www.cse.buffalo.edu/~rapaport/Papers/phics.pdf.

Romanycia, M. H. and F. J. Pelletier (1985). What is a heuristic? *Comp. Intel. 1*(2), 47–58.

[2]See also Rapaport 1998, 2000a, 2003.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences 3*, 417–457.

Shapiro, S. C. (1992). Artificial Intelligence. In S. C. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence, 2nd ed.*, pp. 54–57. New York: John Wiley & Sons.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge, UK: Cambridge University Press.

Wang, P. (2019). On defining artificial intelligence. *J. Artif. General Intel. 10*(2), 1–37.