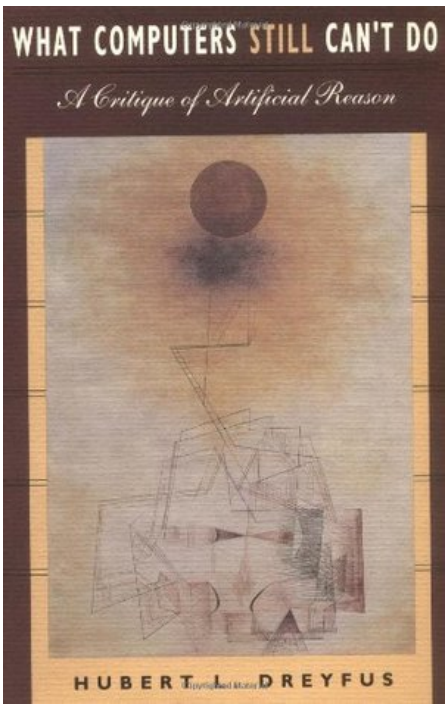

What Computers Still Can't Do: A Critique of Artificial Reason

Hubert L. Dreyfus

Download now

Read Online

What Computers Still Can't Do: A Critique of Artificial Reason

Hubert L. Dreyfus

What Computers Still Can't Do: A Critique of Artificial Reason Hubert L. Dreyfus
When it was first published in 1972, Hubert Dreyfus's manifesto on the inherent inability of disembodied machines to mimic higher mental functions caused an uproar in the artificial intelligence community. The world has changed since then. Today it is clear that "good old-fashioned AI," based on the idea of using symbolic representations to produce general intelligence, is in decline (although several believers still pursue its pot of gold), and the focus of the AI community has shifted to more complex models of the mind. It has also become more common for AI researchers to seek out and study philosophy. For this edition of his now classic book, Dreyfus has added a lengthy new introduction outlining these changes and assessing the paradigms of connectionism and neural networks that have transformed the field. At a time when researchers were proposing grand plans for general problem solvers and automatic translation machines, Dreyfus predicted that they would fail because their conception of mental functioning was naive, and he suggested that they would do well to acquaint themselves with modern philosophical approaches to human being. "What Computers Still Can't Do" was widely attacked but quietly studied. Dreyfus's arguments are still provocative and focus our attention once again on what it is that makes human beings unique.

What Computers Still Can't Do: A Critique of Artificial Reason Details

Date	:	Published October 30th 1992 by Mit Press (first published 1972)
ISBN	:	9780262540674
Author	:	Hubert L. Dreyfus
Format	:	Paperback 408 pages
Genre	:	Philosophy, Science, Nonfiction, Technology, Artificial Intelligence, Computer Science, Computers

Download What Computers Still Can't Do: A Critique of Artif ...pdf

Read Online What Computers Still Can't Do: A Critique of Art ...pdf

Download and Read Free Online What Computers Still Can't Do: A Critique of Artificial Reason Hubert L. Dreyfus

From Reader Review What Computers Still Can't Do: A Critique of Artificial Reason for online ebook

Manny says

Using philosophical arguments from Merleau-Ponty and Heidegger, Dreyfus convincingly demonstrates that there are things people can do, sometimes even without great effort, but which computers are simply incapable of ever being able to achieve. He ends with a list of 20 such items. Thirty-odd years after initial publication, computers still can't do 18 of them - it turns out that Dreyfus wasn't quite right about Grandmaster-level chess and large-vocabulary continuous speech recognition. Maybe there was a bug in Merleau-Ponty's conceptual analysis.

Oh well... if one out of two ain't bad, surely eighteen out of twenty is pretty darn good?

Carl says

Bought this a while back and keep meaning to get to it, but I a bit ignorant in Cognitive Science and computers, so it's been too intimidating so far. This is primarily a critique of AI research back 30-40 years ago, from what I hear, though it has been updated for this edition (though this edition is old by now as well, considering the speed with which research advances in the sciences compared to philosophy).

Danirainbow says

I'm probably biased towards Dreyfus' perspective in this book because I've grown fond of him from listening to his recorded Heidegger lectures at UC Berkeley. Despite his harsh and occasionally smug tone in this book, I've always found him a joy to listen to. He's clearly an expert on both AI technology and on continental European thought--not an easy mix to find!) What surprises me about this book is that it isn't more widely read, given that I believe it to be the most destructive critique on the possibility of human-level AI, hands-down. Maybe it's because the book is somewhat dated, but I find Dreyfus' critique far more convincing than Searle's famous Chinese Room, which is taught far more often.

Bookworm says

What Computers Still Can’t Do (1992) is an evolution of Hubert Dreyfus’s original work, *What Computers Can’t Do* (1972). Today, the ideas coming out of GOFAI research (*Good Old Fashion Artificial Intelligence*), which is based on the notion of using symbolic representations to replicate intelligence, are being replaced by more complex models of the brain/mind. In the revised edition, Dreyfus has added an introduction presenting an overview of the developments that have occurred in the field of Artificial Intelligence (AI) since the publication of his original manifesto. Dreyfus also assesses how the perspectives

of neural networks and connectionism have transformed the field. That said, *What Computers Still Can't Do* presents a similar philosophical analysis to the one contained in the original work, which triggered an avalanche of outrage in the AI community upon its release in the seventies.

Dreyfus's philosophical inquiry tracks the history of developments in AI and is concerned with exposing the incorrect assumptions (psychological, epistemological, and ontological) made by researchers working in the field. Throwing light on the kind of discursive moves used to understand the human brain/mind as reducible to the processes of a digital computer, Dreyfus concentrates on the two subfields of Artificial Intelligence: Cognitive Simulation and Artificial Intelligence. These two fields, he argues, have led to the examination of two distinct but interrelated questions: (1) Does a human being in "processing information" actually follow normal rules like a digital computer? (2) Can human behavior, no matter how generated, be described in a formalism which can be manipulated by a digital machine?

Dreyfus notes that given the difficulties AI experienced during Phase I and Phase II of its development (e.g. failure of GPS), cognitive simulation nevertheless assumed that the information processes of a computer revealed "the hidden information processes" of a human being. He also asks a pertinent question still relevant today: Why do those working on artificial intelligence assume that there must be a digital way of performing human tasks? He writes:

> Those who think that a formalization of intelligent behavior must be possible, seem to base their arguments on the ontological assumption that the world can be analyzed into independent logical elements and an epistemological assumption that our understanding of the world can then be reconstructed by combining these elements according to heuristic rules.

Dreyfus puts forward four models of human information processing to highlight the differences between human information processing and that of a computer: fringe consciousness, ambiguity tolerance, essential/inessential discrimination, and perspicuous grouping. In the context of GOFAI research, for example, each of these kinds of human processing includes a symbolic analogue that cannot be mapped directly onto the essential intelligence of human beings who are able to demonstrate the four models mentioned above. The symbolic analogues relevant to digital computers are as follows: heuristically guided search, context-free precision, trial and error search, and character lists. Dreyfus claims that descriptive or phenomenological evidence, when considered separately from traditional philosophical prejudices, suggests that non-programmable human capacities are involved in all forms of intelligent behavior.

The key point Dreyfus makes in this text is that the complex information processing humans are capable of doing, and human cognitive processes more generally, cannot be reduced to the systematic rule-driven workings of a digital computer. In this capacity, Dreyfus argues, AI is limited by its assumption that the world can be explained in terms of elementary atomistic concepts, a view that dates back to the Greeks (Plato). That said, today AI has evolved by leaps and bounds. Developing computer systems that are increasingly "intelligent" (e.g., Angelina), the field of AI has become one of the most significant components in technological research and development. Although Dreyfus was incorrect about the implementation of GPS, his arguments are still relevant for contemporary analyses of the tendency toward the mechanization of the human brain/mind in all things *neuro*.

What Computers Can't Do (1972) and *What Computers Still Can't Do* (1992) are necessary reads for those who are quick to jump on the "intelligence explosion" bandwagon (e.g., Ray Kurzweill) that predicts society is headed for a *technological singularity* (or the *singularity*), a hypothetical moment in history (2045) when AI has evolved to surpass human intelligence.

Jesse says

Best philosophy of cognitive science book I've read. Dreyfus is harsh, but his words proved prophetic.

Andrew says

If one earns one's bread in the world of Internet People too long, one will encounter a large number of people who seem inherently suspicious of the concept of humanity and go into long diatribes disparaging the weakness of the human mind without technological augmentation. Turns out that not only are they assholes who ruin your lunch break, they are also on very epistemologically shaky ground.

Dreyfus' argument, along with John Searle's critique, are both devastating attacks on the concept of artificial intelligence. Granted, Dreyfus has gotten some egg on his face as some of the things he considered impossible back in the '70s have since been proven attainable, but the majority of his argument remains sound: that a computer, while it can be trained to learn tasks heuristically, cannot conceive of meaning (among other failings) and is therefore not an intelligence.

It also casts serious doubt over the entire program of cognitive science as it is now practiced. Look out kids, a lot of those Ted talks aren't as accurate as they seem.

Ari says

Clever

Adriano Gaved says

Nobody should even talk about Artificial Intelligence without having read thi book!
Furthermore, I found very strong and original the way he uses both phylosophical arguments and historical facts to make his points across.

Seth Graham says

Dreyfus I think is correct, and he is the most endearing person in interviews.

Joshua Stein says

The book is a bit dated, and it really shows when Dreyfus talks about the conteporary limitations of

computers. He disparages chess playing computers in a time before Deep Blue, and so it is important to keep in mind that there are large portions of the critique that seem to have overstepped the appropriate boundaries, and that some of those criticisms have been scaled back in the wake of contemporary successes in certain forms of artificial intelligence.

For those interested in the take that many continental philosophers of mind have on the older-school understandings of A.I. and computational approaches to mind in general, Dreyfus is a good introduction. His colleague, John Searle, is arguably more widely read on the subject and more influential in the philosophical literature, but Dreyfus is much more clear in this particular area, expressing a level of comfort with and awareness of his philosophical history, as well as his current context in the continental tradition.

The context isn't totally articulated, though, in the sense that there has been some movement in what is meant by "A.I." since Dreyfus originally wrote the book. It used to mean something fairly overtly computational, centered on input and output; it used to mean, almost exclusively, software. IT no longer means that, and part of the reason is an understanding of the commentaries on context-free approaches. This is one such commentary, but Dreyfus often commits himself to the position that computers will, in principle, never be able to do certain things, which turns out to be predicated on an oversimplistic concept of "computer." This can hardly been seen as the fault of Dreyfus, though perhaps it is a failure of imagination.

On the other hand, Dreyfus manages to simultaniously provide an interesting view of continental philosophy of mind that is explicitly understood in terms contemporary technology. This makes for a deeply relevant account of an area of thought that is often disparaged as frustratingly limited in terms of its relevance; Dreyfus, and several of the thinkers who started pubishing around the same time or not long after this book, gave continental philosophy of mind a new sense of importance in the philosophical community, and in academia more broadly, and this is a nice way to become introduced to that in the context of some thoughts on technology.
