

UC Irvine

UC Irvine Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/5ft7c1dh>

Journal

Leonardo, 27(1)

ISSN

0024-094X

Authors

Penny, Simon

Dreyfus, Herbert

Publication Date

1994

DOI

10.2307/1575958

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

BOOKS

WHAT COMPUTERS STILL CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON

by Herbert Dreyfus, The MIT Press, Cambridge, MA, U.S.A., 1992. 420 pp.

Reviewed by Simon Penny, Dept. of Art, College of Fine Arts, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, U.S.A.

E-mail: <penny@andrew.cmu.edu>.

When it was first published in 1972, *What Computers Can't Do* was "widely attacked and quietly studied," to quote the back-cover notes of this edition. Dreyfus's target was the then-nascent discipline of Artificial Intelligence (AI), or rather, the assumptions on which it was built. His voice was powerfully critical, working from a philosophical position informed by Heidegger and Merleau Ponty. His critique took issue with the notion that "common sense" would be acquired as systems with limited domains of expertise (i.e. "expert systems" like Newell and Simon's Logic Theorist) gathered more and more incremental "atoms" of knowledge about the world.

Central to this critique was the relationship between the body and intelligence, and the question of "understanding." As Dreyfus stated in 1979: "Intelligence requires understanding, and understanding requires giving the computer the background of common sense that adult human beings have by virtue of having bodies. . . ." In the

same essay Dreyfus quotes Marvin Minsky's deliberations on the subject: "We still know far too little about the contents and structure of common-sense knowledge. A 'minimal' common-sense system must 'know' something about cause-effect, time, purpose, locality, process, and types of knowledge. . . . We need serious epistemological research in this area."

Dreyfus responds: "Minsky's naivete and faith are astonishing. Philosophers from Plato to Husserl, who uncovered all these problems and more, have carried on 'serious epistemological research' in this area for two thousand years without notable success." He continues: "But Minsky seems oblivious to the hand-waving optimism of his proposal that programmers rush in where philosophers such as Heidegger fear to tread, and simply make explicit the totality of human practices which pervade our lives as water encompasses the life of a fish." (Quotes are from Herbert Dreyfus, *What Computers Can't Do*, Harper, 1979, Revised Edition, p. 3 and p. 36.)

The historic importance of Dreyfus's work should not be underestimated. His humanistic critique of the discipline and of the social function of computer technology preceded by more than a decade other works in a similar spirit, such as *Understanding Computers and Cognition* by Terry Winograd and Fernando Flores (Addison Wesley), Theodore Roszak's *The Cult of Information* (Pantheon), and J.D. Bolter's *Turings' Man* (Pelican)—all of which, curiously, were published in 1986.

What Computers Still Can't Do is not a new book, but rather it is the 1979 revised edition with a new essay at the front. This introduction has a curious tone that is simultaneously humble and crowing. In the first edition of *What Computers Can't Do*, Dreyfus rallied his forces; the second restated the argument with some updating. But the first lines of the new introduction establish that we are looking at AI as an event of the past. The demise of what he calls the "degenerating research program" of GOF AI (Good Old Fashioned Artificial Intelligence) are due (it would seem) precisely to the errors he originally noted.

In support of his argument that AI is all but dead, he describes the emergence of alternative approaches, alternatives that tacitly acknowledge the problems inherent to what used to be called Hard AI. These include the

connectionist model and neural network (and, one might add, Brooks's Subsumption Architecture), which sidestep the AI model of top-down, rule-based behavior. Dreyfus acknowledges and discusses perhaps the most ambitious AI project, the last remaining general expert system research project, that of Douglas Lenat. Dreyfus argues that Lenat's project will fall foul due to precisely the same philosophical dilemmas of previous attempts, though at a somewhat more subtle level.

What Computers Can't Do is an important book—no education, formal or informal, in the field of Artificial Intelligence is complete and well rounded without the inclusion of this work.

For those who have not read the 1979 version, *What Computers Still Can't Do* is a must. For those who have, the new essay alone is worth the time.