THE WALL STREET JOURNAL.

WSJ.com

OPINION | FEBRUARY 23, 2011

Watson Doesn't Know It Won on 'Jeopardy!'

IBM invented an ingenious program—not a computer that can think.

By **JOHN SEARLE**

The recent victory of an IBM computer named Watson over human contestants on the TV show "Jeopardy!" has produced a flood of commentaries to the effect that computer understanding now equals—or perhaps even exceeds—human understanding. Thinking computers, at last.

But this interpretation rests on a profound misunderstanding of what a computer is, how it works, and how it differs from a human brain.

A digital computer is a device that manipulates formal symbols. These are usually thought of as zeros and ones, but any symbols will do. An increase in computational power is simply a matter of increasing the speed of symbol manipulation. A computer's effectiveness is a function of the skill of the programmers designing the program.

Watson revealed a huge increase in computational power and an ingenious program. I congratulate IBM on both of these innovations, but they do not show that Watson has superior intelligence, or that it's thinking, or anything of the sort.

Computational operations, as standardly defined, could never constitute thinking or understanding for reasons that I showed over 30 years ago with a simple argument.

Imagine that a person—me, for example—knows no Chinese and is locked in a room with boxes full of Chinese symbols and an instruction book written in English for manipulating the symbols. Unknown to me, the boxes are called "the database" and the instruction book is called "the program." I am called "the computer."

People outside the room pass in bunches of Chinese symbols that, unknown to me, are questions. I look up in the instruction book what I am supposed to do and I give back answers in Chinese symbols.

Suppose I get so good at shuffling the symbols and passing out the answers that my answers are indistinguishable from a native Chinese speaker's. I give every indication of understanding the language despite the fact that I actually don't understand a word of Chinese.

And if I do not, neither does any digital computer, because no computer, qua computer, has anything I do not have. It has stocks of symbols, rules for manipulating symbols, a system that allows it to rapidly transition from zeros to ones, and the ability to process inputs and outputs. That is it. There is nothing else.

This thought experiment carries over exactly to Watson. But

IBM's 'Watson' computing system

instead of working in Chinese symbols, Watson has proven adept at responding to "Jeopardy!" questions phrased in English.

All the same, as in the original Chinese room, the symbols are meaningless to Watson, which understands nothing. The reason it lacks understanding is that, like me in the Chinese room, it has no way to get from symbols to meanings (or from syntax to semantics, in linguistic jargon). The bottom line can be put in the form of a four-word sentence: Symbols are not meanings.

Of course, Watson is much faster than me. But speed doesn't add understanding. This is a simple refutation of the idea that computer simulations of human cognition are the real thing.

If the computer cannot understand solely by manipulating symbols, then how does the brain do it? What is the difference between the brain and the digital computer? The answer is that the brain is a causal mechanism that causes consciousness, understanding and all the rest of it. It is an organ like any other, and like any other it operates on causal principles.

The problem with the digital computer is not that it is too much of a machine to have human understanding. On the contrary, it is not enough of a machine. Consciousness, the machine process that goes on in the brain, is fundamentally different from what a computer does, which is computation. Computation is an abstract formal process, like addition.

Unlike computation, actual human thinking is a concrete biological phenomenon existing in actual human brains. This is as opposed to Watson, which is merely following an algorithm that enables it to manipulate formal symbols.

Watson did not understand the questions, nor its answers, nor that some of its answers were right and some wrong, nor that it was playing a game, nor that it won—because it doesn't understand anything.

IBM's computer was not and could not have been designed to understand. Rather, it was designed to simulate understanding, to act as if it understood. It is an evasion to say, as some commentators have put it, that computer understanding is different from human understanding. Literally speaking, there is no such thing as computer understanding. There is only simulation.

Mr. Searle is professor of philosophy at the University of California, Berkeley.