The Past and Future of Intelligence

Marvin Minsky

Workers in the field of Artificial Intelligence have several different goals. Some of them want to make machines do more of the things that people call intelligent. Others hope to understand what enables people to do such things. Yet other researchers want to eliminate the job of programming computers, by making machines that program themselves, and grow by learning from experience. Many computer programs already exist that do things most people would regard as requiring intelligence. Today, we already have machines that play chess, design electronic devices, prove theorems in geometry, and diagnose diseases, use geological surveys to find mineral deposits, and control robots working in factories. But none of those programs works at all well outside some limited specialty. They all lack the liveliness and versatility that people have. We now have computers that can "see" in factory environments — but none that do what children can — such as recognize the things in a typical home: furniture, kitchenware, clothing, people, decorations etc. This is why there are no robots that can clean a house, serve dinner, take care of an infant, or understand the simplest stories we tell our children. Why has it been harder to make machines do such childish things than to make them compete with trained experts? It is because we do not yet know how to provide our computers with common sense.

The field of Artificial Intelligence had roots in the Cybernetics movement of the 1940s, which was trying to combine "analog" ideas from the theory of servomechanisms with the new "digital" ideas that were emerging from the theories of brain scientists like Warren McCulloch and his associates Walter Pitts, Jerome Lettvin, and Oliver Selfridge. But it was only when computers first became available in the 1950s that those theories could actually be implemented. Soon some people started to try to make computers solve problems for themselves — and these first AI researchers set out in two directions. One approach was to try to develop "self-organizing systems" — programs that would improve themselves, through evolutionary processes. Most of these experiments failed because the theories were still too weak (and the computers were still too slow) — and that line of research went out of style for many years.

The other approach, called "heuristic search", was to write programs that try many different ways to solve a problem — and then simply see which one worked. Many problems were solved this way, but of course, this becomes impractical when the search-size grows exponentially. To get around this, the AI workers of the 1960s had to discover ways to predict which attempts might work best, instead of trying every possibility. One way to do this, developed by Allen Newell and Herbert A. Simon, used a "goal-directed" search technique that measured various "differences" between the present situation and the desired final state. Then, the program could often deal separately with each of those differences, instead of having to deal with the whole problem all at once. Another idea of that era was called "planning". Before going to work on an actual problem, a planning program first solves a simplified version of it — and then tries to adapt that solution to the original problem. These methods often worked so well that the programs were more proficient than their programmers. But to do all this required new ideas about programming itself, and the AI community contributed many advances in software technology, including basic ideas about multiprocessing and time-sharing, new debugging techniques, symbol-manipulation languages, graphical input and output devices, and word processing — as well as new fundamental theories about computation itself, and about natural language processing, visual pattern recognition, and robotic technology.

But the 1970s also revealed limits to what we could accomplish by using general principles. Even a very smart person cannot compete with experts in an unfamiliar specialty. So the next era in AI research was concerned with what kinds of knowledge a person — or a machine — must possess to be able to do solve various kinds of problems. Soon, AI researchers invented many new ways to "represent knowledge" inside computers. For example, libraries of "If-Then" rules were useful for simulating many kinds of expert skills. "Semantic Networks" were useful for reasoning about complicated networks of concepts and relationships. For mathematical types of problems, it was useful to represent knowledge using new "logic programming languages" that could automatically make logical deductions. Certain structures called "frames", which I invented in the early 1970s, represent knowledge in terms of typical situations so that we need only remember how each particular application differs from some more typical situation. In the 1980s, all this basic research culminated in the appearance of many commercially successful applications — usually called "Expert Systems" — each of which can solve certain types of problems. But none of these programs seemed to promise to lead to the kinds of versatility and resourcefulness that people have. Each program worked effectively only within some small domain or specialty.

One problem was that none of those systems were able to learn from experience — so the AI researchers of the 1980s now set out to look for ways to make machines learn. One approach was to develop knowledge representations of a type called "connectionist networks," which are intended to simulate how low-level concepts are encoded in brains. These have seen substantial success in learning to recognize many kinds of patterns, but seem unable to support the kinds of higher level reasoning and planning processes that must be involved in verbal reasoning. For such purposes, other researchers have been developing higher-level learning systems based on semantic networks, "case-based" reasoning, and on other techniques.

The other problem was that, while each of those schemes has its own advantages and limitations, none of them seemed to offer much promise of leading to machines that have "common sense". So this became the subject of my own research, as explained in my book, "The Society of Mind". My basic conclusion is that the human mind does not work on the basis of any single way to represent knowledge, or single way to do reasoning. Instead, the greatness of the human brain comes from having evolved hundreds of different kinds of machinery. Each of these mental "agencies" is specialized for solving certain kinds of problems — but more important, each agency is able to learn how to get help from other agencies! In this way, as each child develops from infant to adult, its mind can exploit the advantages, and escape the limitations, of many different types of low and high-level strategies and representations — and this I believe to be the secret of our versatility. There have not yet been enough experiments to find out how well such ideas will work — but if they work well, it seems possible that smart machines with common sense could arrive within the next century.

The effect of this could be so large that all of us should think about it. How will our lives become transformed, when we have artificial workers to do everything from mining, farming, house-cleaning, and factory production to basic scientific research? Will this lead to the decline and disintegration of what humankind has accomplished so far? Or will it lead, as I hope and expect, to a new kind of civilization, grander than anything we have ever imagined?