How to Make a Mind

By Ray Kurzweil

Can nonbiological brains have real minds of their own? In this article, drawn from his latest book, futurist/inventor Ray Kurzweil describes the future of intelligence—artificial and otherwise.

The mammalian brain has a distinct aptitude not found in any other class of animal. We are capable of *hierarchical* thinking, of understanding a structure composed of diverse elements arranged in a pattern, representing that arrangement with a symbol, and then using that symbol as an element in a yet more elaborate configuration.

This capability takes place in a brain structure called the *neocortex,* which in humans has achieved a threshold of sophistication and capacity such that we are able to call these patterns *ideas.* We are capable of building ideas that are ever more complex. We call this vast array of recursively linked ideas *knowledge.* Only *Homo sapiens* have a knowledge base that itself evolves, grows exponentially, and is passed down from one generation to another.

We are now in a position to speed up the learning process by a factor of thousands or millions once again by migrating from biological to nonbiological intelligence. Once a digital neocortex learns a skill, it can transfer that know-how in minutes or even seconds. Ultimately we will create an artificial neocortex that has the full range and flexibility of its human counterpart.

Consider the benefits. Electronic circuits are millions of times faster than our biological circuits. At first we will have to devote all of this speed increase to compensating for the relative lack of parallelism in our computers. Parallelism is what gives our brains the ability to do so many different types of operations—walking, talking, reasoning—all at once, and perform these tasks so seamlessly that we live our lives blissfully unaware that they are occurring at all. The digital neocortex will be much faster than the biological variety and will only continue to increase in speed.

When we augment our own neocortex with a synthetic version, we won’t have to worry about how much additional neocortex can physically fit into our bodies and brains, as most of it will be in the cloud, like most of the computing we use today. We have about 300 million pattern recognizers in our biological neocortex. That’s as much as could be squeezed into our skulls even with the evolutionary innovation of a large forehead and with the neocortex taking about 80% of the available space. As soon as we start thinking in the cloud, there will be no natural limits—we will be able to use billions or trillions of pattern recognizers, basically whatever we need, and whatever the law of accelerating returns can provide at each point in time.

In order for a digital neocortex to learn a new skill, it will still require many iterations of education, just as a biological neocortex does. Once a single digital neocortex somewhere and at some time learns something, however, it can share that knowledge with every other digital neocortex without delay. We can each have our own private neocortex extenders in the cloud, just as we have our own private stores of personal data today.

Last but not least, we will be able to back up the digital portion of our intelligence. It is frightening to contemplate that none of the information contained in our neocortex is backed up today. There is, of course, one way in which we do back up some of the information in our brains: by writing it down. The ability to transfer at least some of our thinking to a medium that can outlast our biological bodies was a huge step forward, but a great deal of data in our brains continues to remain vulnerable.

The Next Chapter in Artificial Intelligence

Artificial intelligence is all around us. The simple act of connecting with someone via a text message, e-mail, or cell-phone call uses intelligent algorithms to route the information. Almost every product we touch is originally designed in a collaboration between human and artificial intelligence and then built in automated factories. If all the AI systems decided to go on strike tomorrow, our civilization would be crippled: We couldn’t get money from our bank, and indeed, our money would disappear; communication, transportation, and manufacturing would all grind to a halt. Fortunately, our intelligent machines are not yet intelligent enough to organize such a conspiracy.

What is new in AI today is the viscerally impressive nature of publicly available examples. For example, consider Google’s self-driving cars, which as of this writing have gone over 200,000 miles in cities and towns. This technology will lead to significantly fewer crashes and increased capacity of roads, alleviate the requirement of humans to perform the chore of driving, and bring many other benefits.

Driverless cars are actually already legal to operate on public roads in Nevada with some restrictions, although widespread usage by the public throughout the world is not expected until late in this decade. Technology that intelligently watches the road and warns the driver of impending dangers is already being installed in cars. One such technology is based in part on the successful model of visual processing in the brain created by MIT's Tomaso Poggio. Called MobilEye, it was developed by Amnon Shashua, a former postdoctoral student of Poggio's. It is capable of alerting the driver to such dangers as an impending collision or a child running in front of the car and has recently been installed in cars by such manufacturers as Volvo and BMW.

I will focus now on language technologies for several reasons: Not surprisingly, the hierarchical nature of language closely mirrors the hierarchical nature of our thinking. Spoken language was our first technology, with written language as the second. My own work in artificial intelligence has been heavily focused on language. Finally, mastering language is a powerfully leveraged capability. Watson, the IBM computer that beat two former *Jeopardy!* champions in 2011, has already read hundreds of millions of pages on the Web and mastered the knowledge contained in these documents. Ultimately, machines will be able to master all of the knowledge on the Web—which is essentially all of the knowledge of our human–machine civilization.

One does not need to be an AI expert to be moved by the performance of Watson on *Jeopardy!* Although I have a reasonable understanding of the methodology used in a number of its key subsystems, that does not diminish my emotional reaction to watching it—him?—perform. Even a perfect understanding of how all of its component systems work would not help you to predict how Watson would actually react to a given situation. It contains hundreds of interacting subsystems, and each of these is considering millions of competing hypotheses at the same time, so predicting the outcome is impossible. Doing a thorough analysis—after the fact—of Watson's deliberations for a single three-second query would take a human centuries.

One limitation of the *Jeopardy!* game is that the answers are generally brief: It does not, for example, pose questions of the sort that ask contestants to name the five primary themes of *A Tale of Two Cities.* To the extent that it can find documents that do discuss the themes of this novel, a suitably modified version of Watson should be able to respond to this. Coming up with such themes on its own from just reading the book, and not essentially copying the thoughts (even without the words) of other thinkers, is another matter. Doing so would constitute a higher-level task than Watson is capable of today.

It is noteworthy that, although Watson's language skills are actually somewhat below that of an educated human, it was able to defeat the best two *Jeopardy!* players in the world. It could accomplish this because it is able to combine its language ability and knowledge understanding with the perfect recall and highly accurate memories that machines possess. That is why we have already largely assigned our personal, social, and historical memories to them.

Wolfram|Alpha is one important system that demonstrates the strength of computing applied to organized knowledge. Wolfram|Alpha is an answer engine (as opposed to a search engine) developed by British mathematician and scientist Stephen Wolfram and his colleagues at Wolfram Research. For example, if you ask Wolfram|Alpha, “How many primes are there under a million?” it will respond with “78,498.” It did not look up the answer, it computed it, and following the answer it provides the equations it used. If you attempted to get that answer using a conventional search engine, it would direct you to links where you could find the algorithms required. You would then have to plug those formulas into a system such as Mathematica, also developed by Wolfram, but this would obviously require a lot more work (and understanding) than simply asking Alpha.

Indeed, Alpha consists of 15 million lines of Mathematica code. What Alpha is doing is literally computing the answer from approximately 10 trillion bytes of data that has been carefully curated by the Wolfram Research staff. You can ask a wide range of factual questions, such as, “What country has the highest GDP per person?” (Answer: Monaco, with $212,000 per person in U.S. dollars), or “How old is Stephen Wolfram?” (he was born in 1959; the answer is 52 years, 9 months, 2 days on the day I am writing this). Alpha is used as part of Apple’s Siri; if you ask Siri a factual question, it is handed off to Alpha to handle. Alpha also handles some of the searches posed to Microsoft’s Bing search engine.

Wolfram reported in a recent blog post that Alpha is now providing successful responses 90% of the time. He also reports an exponential decrease in the failure rate, with a half-life of around 18 months. It is an impressive system, and uses handcrafted methods and hand-checked data. It is a testament to why we created computers in the first place. As we discover and compile scientific and mathematical methods, computers are far better than unaided human intelligence in implementing them. Most of the known scientific methods have been encoded in Alpha, along with continually updated data on topics ranging from economics to physics.

In a private conversation I had with him, Wolfram estimated that self-organizing methods such as those used in Watson typically achieve about an 80% accuracy when they are working well. Alpha, he pointed out, is achieving about a 90% accuracy. Of course, there is self-selection in both of these accuracy numbers, in that users (such as myself) have learned what kinds of questions Alpha is good at, and a similar factor applies to the self-organizing methods. Some 80% appears to be a reasonable estimate of how accurate Watson is on*Jeopardy!* queries, but this was sufficient to defeat the best humans.

It is my view that self-organizing methods such as I articulate as the pattern-recognition theory of mind, or PRTM, are needed to understand the elaborate and often ambiguous hierarchies we encounter in real-world phenomena, including human language. Ideally, a robustly intelligent system would combine hierarchical intelligence based on the PRTM (which I contend is how the human brain works) with precise codification of scientific knowledge and data. That essentially describes a human with a computer.

We will enhance both poles of intelligence in the years ahead. With regard to our biological intelligence, although our neocortex has significant plasticity, its basic architecture is limited by its physical constraints. Putting additional neocortex into our foreheads was an important evolutionary innovation, but we cannot now easily expand the size of our frontal lobes by a factor of a thousand, or even by 10%. That is, we cannot do so biologically, but that is exactly what we will do technologically.

Our digital brain will also accommodate substantial redundancy of each pattern, especially ones that occur frequently. This allows for robust recognition of common patterns and is also one of the key methods to achieving invariant recognition of different forms of a pattern. We will, however, need rules for how much redundancy to permit, as we don’t want to use up excessive amounts of memory on very common low-level patterns.

Educating Our Nonbiological Brain

A very important consideration is the education of a brain, whether a biological or a software one. A hierarchical pattern-recognition system (digital or biological) will only learn about two—preferably one—hierarchical levels at a time. To bootstrap the system, I would start with previously trained hierarchical networks that have already learned their lessons in recognizing human speech, printed characters, and natural-language structures.

Such a system would be capable of reading natural-language documents but would only be able to master approximately one conceptual level at a time. Previously learned levels would provide a relatively stable basis to learn the next level. The system can read the same documents over and over, gaining new conceptual levels with each subsequent reading, similar to the way people reread and achieve a deeper understanding of texts. Billions of pages of material are available on the Web. Wikipedia itself has about 4 million articles in the English version.

I would also provide a critical-thinking module, which would perform a continual background scan of all of the existing patterns, reviewing their compatibility with the other patterns (ideas) in this software neocortex. We have no such facility in our biological brains, which is why people can hold completely inconsistent thoughts with equanimity. Upon identifying an inconsistent idea, the digital module would begin a search for a resolution, including its own cortical structures as well as all of the vast literature available to it. A resolution might mean determining that one of the inconsistent ideas is simply incorrect (if contraindicated by a preponderance of conflicting data). More constructively, it would find an idea at a higher conceptual level that resolves the apparent contradiction by providing a perspective that explains each idea. The system would add this resolution as a new pattern and link to the ideas that initially triggered the search for the resolution. This critical thinking module would run as a continual background task. It would be very beneficial if human brains did the same thing.

I would also provide a module that identifies open questions in every discipline. As another continual background task, it would search for solutions to them in other disparate areas of knowledge. The knowledge in the neocortex consists of deeply nested patterns of patterns and is therefore entirely metaphorical. We can use one pattern to provide a solution or insight in an apparently disconnected field.

As an example, molecules in a gas move randomly with no apparent sense of direction. Despite this, virtually every molecule in a gas in a beaker, given sufficient time, will leave the beaker. This provides a perspective on an important question concerning the evolution of intelligence. Like molecules in a gas, evolutionary changes also move every which way with no apparent direction. Yet, we nonetheless see a movement toward greater complexity and greater intelligence, indeed to evolution’s supreme achievement of evolving a neocortex capable of hierarchical thinking. So we are able to gain an insight into how an apparently purposeless and directionless process can achieve an apparently purposeful result in one field (biological evolution) by looking at another field (thermodynamics).

We should provide a means of stepping through multiple lists simultaneously to provide the equivalent of structured thought. A list might be the statement of the constraints that a solution to a problem must satisfy. Each step can generate a recursive search through the existing hierarchy of ideas or a search through available literature. The human brain appears to be only able to handle four simultaneous lists at a time (without the aid of tools such as computers), but there is no reason for an artificial neocortex to have such a limitation.

We will also want to enhance our artificial brains with the kind of intelligence that computers have always excelled in, which is the ability to master vast databases accurately and implement known algorithms quickly and efficiently Wolfram|Alpha uniquely combines a great many known scientific methods and applies them to carefully collected data. This type of system is also going to continue to improve, given Stephen Wolfram’s observation of an exponential decline in error rates.

Finally, our new brain needs a purpose. A purpose is expressed as a series of goals. In the case of our biological brains, our goals are established by the pleasure and fear centers that we have inherited from the old brain. These primitive drives were initially set by biological evolution to foster the survival of species, but the neocortex has enabled us to sublimate them. Watson's goal was to respond to *Jeopardy!* queries. Another simply stated goal could be to pass the Turing test. To do so, a digital brain would need a human narrative of its own fictional story so that it can pretend to be a biological human. It would also have to dumb itself down considerably, for any system that displayed the knowledge of Watson, for instance, would be quickly unmasked as nonbiological.

More interestingly, we could give our new brain a more ambitious goal, such as contributing to a better world. A goal along these lines, of course, raises a lot of questions: Better for whom? Better in what way? For biological humans? For all conscious beings? If that is the case, who or what is conscious?

As nonbiological brains become as capable as biological ones of effecting changes in the world—indeed, ultimately far more capable than unenhanced biological ones—we will need to consider their moral education. A good place to start would be with one old idea from our religious traditions: the golden rule.

About the Author

Ray Kurzweil is an inventor, writer, and futurist. Among his honors are the MIT-Lemelson Prize, the National Medal of Technology, and, in 2002, induction into the U.S Patent Office's National Inventor's Hall of Fame.

This article was excerpted from his most recent book, *How to Create a Mind* (Viking, 2012).

Source: The Futurist, Volume 49, No. 2, March – April 2013, pag 14 - 17