

# AI's Language Problem

By Will Knight.

---

Machines that truly understand language would be incredibly useful. But we don't know how to build them.

---

About halfway through a particularly tense game of Go held in Seoul, South Korea, between Lee Sedol, one of the best players of all time, and AlphaGo, an artificial intelligence created by Google, the AI program made a mysterious move that demonstrated an unnerving edge over its human opponent.

On move 37, AlphaGo chose to put a black stone in what seemed, at first, like a ridiculous position. It looked certain to give up substantial territory—a rookie mistake in a game that is all about controlling the space on the board. Two television commentators wondered if they had misread the move or if the machine had malfunctioned somehow. In fact, contrary to any conventional wisdom, move 37 would enable AlphaGo to build a formidable foundation in the center of the board. The Google program had effectively won the game using a move that no human would've come up with.

## 1 About the art

- One reason that understanding language is so difficult for computers and AI systems is that words often have meanings based on context and even the appearance of the letters and words. In the images that accompany this story, several artists demonstrate the use of a variety of visual clues to convey meanings far beyond the actual letters.

AlphaGo's victory is particularly impressive because the ancient game of Go is often looked at as a test of intuitive intelligence. The rules are quite simple. Two players take turns putting black or white stones at the intersection of horizontal and vertical lines on a board, trying to surround their opponent's pieces and remove them from play. Playing well, however, is incredibly hard.

Whereas chess players are able to look a few moves ahead, in Go this isn't possible without the game unfolding into intractable complexity, and there are no classic gambits. There is also no straightforward way to measure advantage, and it can be hard for even an expert player to explain precisely why he or she made a particular move. This makes it impossible to write a simple set of rules for an expert-level computer program to follow.

AlphaGo wasn't told how to play Go at all. Instead, the program analyzed hundreds of thousands of games

and played millions of matches against itself. Among several AI techniques, it used an increasingly popular method known as deep learning, which involves mathematical calculations inspired, very loosely, by the way interconnected layers of neurons fire in a brain as it learns to make sense of new information. The program taught itself through hours of practice, gradually honing an intuitive sense of strategy. That it was then able to beat one of the world's best Go players represents a true milestone in machine intelligence and AI.



Figure 1: Lawrence Weiner *A Rubber Ball Thrown on the Sea* 1970 / 2014 [\[Image Source\]](#)

A few hours after move 37, AlphaGo won the game to go up two games to nothing in the best-of-five match. Afterward Sedol stood before a crowd of journalists and photographers, politely apologizing for letting humankind down. "I am quite speechless," he said, blinking through a storm of flash photography.

AlphaGo's surprising success points to just how much progress has been made in artificial intelligence over the last few years, after decades of frustration and setbacks often described as an "AI winter." Deep learning means that machines can increasingly teach themselves how to perform complex tasks that only a couple of years ago were thought to require the unique intelligence of humans. Self-driving cars are already a foreseeable possibility. In the near future, systems based on deep learning will help diagnose diseases and recommend treatments.

Deep learning means that machines can increasingly teach themselves how to perform complex tasks that only a couple of years ago were thought to require the unique intelligence of humans.

Yet despite these impressive advances, one fundamental capability remains elusive: language. Systems like Siri and IBM's Watson can follow simple spoken or typed commands and answer basic questions, but they can't hold a conversation and have no real understanding of the words they use. If AI is to be truly transformative, this must change.

Even though AlphaGo cannot speak, it contains

technology that might lead to greater language understanding. At companies such as Google, Facebook, and Amazon, as well as at leading academic AI labs, researchers are attempting to finally solve that seemingly intractable problem, using some of the same AI tools—including deep learning—that are responsible for AlphaGo’s success and today’s AI revival. Whether they succeed will determine the scale and character of what is turning into an artificial–intelligence revolution. It will help determine whether we have machines we can easily communicate with—machines that become an intimate part of our everyday life—or whether AI systems remain mysterious black boxes, even as they become more autonomous. “There’s no way you can have an AI system that’s humanlike that doesn’t have language at the heart of it,” says Josh Tenenbaum, a professor of cognitive science and computation at MIT. “It’s one of the most obvious things that set human intelligence apart.”

Perhaps the same techniques that let AlphaGo conquer Go will finally enable computers to master language, or perhaps something else will also be required. But without language understanding, the impact of AI will be different. Of course, we can still have immensely powerful and intelligent software like AlphaGo. But our relationship with AI may be far less collaborative and perhaps far less friendly. “A nagging question since the beginning was ‘What if you had things that were intelligent in the sense of being effective, but not like us in the sense of not empathizing with what we are?’” says Terry Winograd, a professor emeritus at Stanford University. “You can imagine machines that are not based on human intelligence, which are based on this big-data stuff, and which run the world.”

### Machine whisperers

A couple of months after AlphaGo’s triumph, I traveled to Silicon Valley, the heart of the latest boom in artificial intelligence. I wanted to visit the researchers who are making remarkable progress on practical applications of AI and who are now trying to give machines greater understanding of language.

I started with Winograd, who lives in a suburb nestled into the southern edge of Stanford’s campus in Palo Alto, not far from the headquarters of Google, Facebook, and Apple. With curly white hair and a bushy mustache, he looks the part of a venerable academic, and he has an infectious enthusiasm.

Back in 1968, Winograd made one of the earliest efforts to teach a machine to talk intelligently. A math prodigy fascinated with language, he had come to MIT’s new AI lab to study for his PhD, and he decided to build a program that would converse with people, via a text prompt, using everyday language. It didn’t seem an outlandish ambition at the time. In-

credible strides were being made in AI, and others at MIT were building complex computer vision systems and futuristic robot arms. “There was a sense of unknown, unbounded possibilities,” he recalls.



Figure 2: Joseph Kosuth *Four Colors Four Words* 1966  
[Image Source]

Not everyone was convinced that language could be so easily mastered, though. Some critics, including the influential linguist and MIT professor Noam Chomsky, felt that the AI researchers would struggle to get machines to understand, given that the mechanics of language in humans were so poorly understood. Winograd remembers attending a party where a student of Chomsky’s walked away when he heard him say that he worked in the AI lab.

But there was reason to be optimistic, too. Joseph Weizenbaum, a German-born professor at MIT, had built the very first chatbot program a couple of years earlier. Called ELIZA, it was programmed to act like a cartoon psychotherapist, repeating key parts of a statement or asking questions to encourage further conversation. If you told the program you were angry at your mother, for instance, it would say, “What else comes to mind when you think about your mother?” A cheap trick, but it worked surprisingly well. Weizenbaum was shocked when some subjects began confessing their darkest secrets to his machine.

There’s an obvious problem with applying deep learning to language. It’s that words are arbitrary symbols, and as such they are fundamentally different from imagery.

Winograd wanted to create something that really seemed to understand language. He began by reducing the scope of the problem. He created a simple virtual environment, a “block world,” consisting of a handful of imaginary objects sitting on an imaginary table. Then he created a program, [which he named SHRDLU](#), that was capable of parsing all the nouns, verbs, and simple rules of grammar needed to refer to this stripped-down virtual world. SHRDLU (a nonsense word formed by the second column of keys on a

Linotype machine) could describe the objects, answer questions about their relationships, and make changes to the block world in response to typed commands. It even had a kind of memory, so that if you told it to move “the red cone” and then later referred to “the cone,” it would assume you meant the red one rather than one of another color.

SHRDLU was held up as a sign that the field of AI was making profound progress. But it was just an illusion. When Winograd tried to make the program’s block world larger, the rules required to account for the necessary words and grammatical complexity became unmanageable. Just a few years later, he had given up, and eventually he abandoned AI altogether to focus on other areas of research. “The limitations were a lot closer than it seemed at the time,” he says. Winograd concluded that it would be impossible to give machines true language understanding using the tools available then. The problem, as Hubert Dreyfus, a professor of philosophy at UC Berkeley, argued in a 1972 book called *What Computers Can’t Do*, is that many things humans do require a kind of instinctive intelligence that cannot be captured with hard-and-fast rules. This is precisely why, before the match between Sedol and AlphaGo, many experts were dubious that machines would master Go.

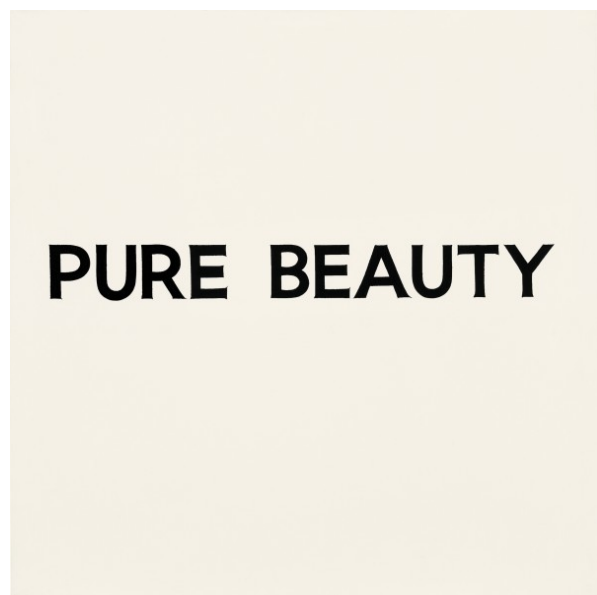


Figure 3: John Baldessari *Pure Beauty* 1966–68 [Image Source]

But even as Dreyfus was making that argument, a few researchers were, in fact, developing an approach that would eventually give machines this kind of intelligence. Taking loose inspiration from neuroscience, they were experimenting with artificial neural networks—layers of mathematically simulated neurons that could be trained to fire in response to certain inputs. To begin with, these systems were painfully

slow, and the approach was dismissed as impractical for logic and reasoning. Crucially, though, neural networks could learn to do things that couldn’t be hand-coded, and later this would prove useful for simple tasks such as recognizing handwritten characters, a skill that was commercialized in the 1990s for [reading the numbers on checks](#). Proponents maintained that neural networks would eventually let machines to do much, much more. One day, they claimed, the technology would even understand language.

Over the past few years, neural networks have become vastly more complex and powerful. The approach has benefited from key mathematical refinements and, more important, faster computer hardware and oodles of data. By 2009, researchers at the University of Toronto had shown that a many-layered deep-learning network could recognize speech with record accuracy. And then in 2012, the same group won a machine-vision contest using a deep-learning algorithm that was astonishingly accurate.

A deep-learning neural network recognizes objects in images using a simple trick. A layer of simulated neurons receives input in the form of an image, and some of those neurons will fire in response to the intensity of individual pixels. The resulting signal passes through many more layers of interconnected neurons before reaching an output layer, which signals that the object has been seen. A mathematical technique known as backpropagation is used to adjust the sensitivity of the network’s neurons to produce the correct response. It is this step that gives the system the ability to learn. Different layers inside the network will respond to features such as edges, colors, or texture. Such systems can now recognize objects, animals, or faces with an accuracy that rivals that of humans.

There’s an obvious problem with applying deep learning to language. It’s that words are arbitrary symbols, and as such they are fundamentally different from imagery. Two words can be similar in meaning while containing completely different letters, for instance; and the same word can mean various things in different contexts.

In the 1980s, researchers had come up with a clever idea about how to turn language into the type of problem a neural network can tackle. They showed that words can be represented as mathematical vectors, allowing similarities between related words to be calculated. For example, “boat” and “water” are close in vector space even though they look very different. Researchers at the University of Montreal, led by Yoshua Bengio, and another group at Google, have used this insight to build networks in which each word in a sentence can be used to construct a more complex representation—something that Geoffrey Hinton, a professor at the University of Toronto and a prominent deep-learning researcher who works part-time at Google,

calls a “thought vector.”

By using two such networks, it is possible to translate between two languages with excellent accuracy. And by combining this type of network with one designed to recognize objects in images, it is possible to conjure up surprisingly plausible captions.

### The purpose of life

Sitting in a conference room at the heart of Google’s bustling headquarters in Mountain View, California, one of the company’s researchers who helped develop this approach, Quoc Le, is contemplating the idea of a machine that could hold a proper conversation. Le’s ambitions cut right to the heart of why talking machines could be useful. “I want a way to simulate thoughts in a machine,” he says. “And if you want to simulate thoughts, then you should be able to ask a machine what it’s thinking about.”

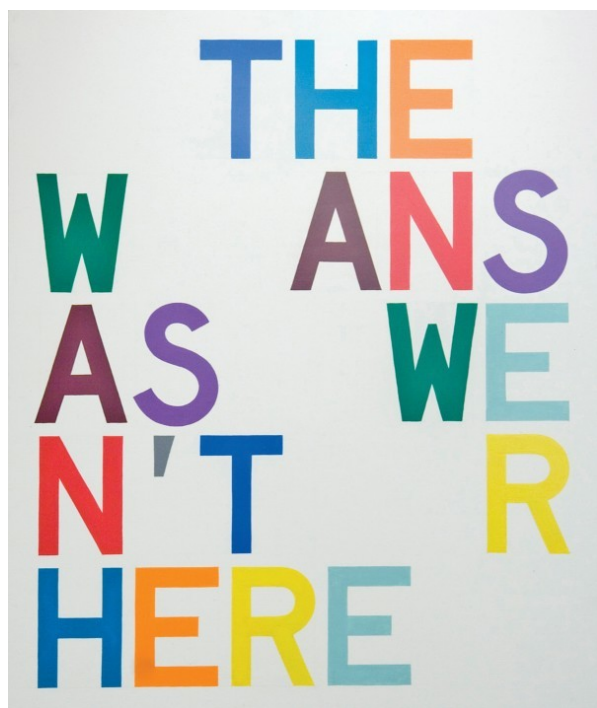


Figure 4: Tauba Auerbach *The Answer/Wasn’t Here II* 2008 [\[Image Source\]](#)

Google is already teaching its computers the basics of language. This May the company announced a system, dubbed Parsey McParseface, that can look at syntax, recognizing nouns, verbs, and other elements of text. It isn’t hard to see how valuable better language understanding could be to the company. Google’s search algorithm used to simply track keywords and links between Web pages. Now, using a system called RankBrain, it reads the text on pages in an effort to glean meaning and deliver better results. Le wants to take that much further. Adapting the system that’s proved useful in translation and image captioning, he and his colleagues built Smart Reply, which reads the contents of Gmail messages and sug-

gests a handful of possible replies. He also created a program that learned from Google’s IT support chat logs how to answer simple technical queries.

Most recently, Le built a program capable of producing passable responses to open-ended questions; it was trained by being fed dialogue from 18,900 movies. Some of its replies seem eerily spot-on. For example, Le asked, “What is the purpose of life?” and the program responded, “To serve the greater good.” “It was a pretty good answer,” he remembers with a big grin. “Probably better than mine would have been.”

There’s only one problem, as quickly becomes apparent when you look at more of the system’s answers. When Le asked, “How many legs does a cat have?” his system answered, “Four, I think.” Then he tried, “How many legs does a centipede have?” which produced a curious response: “Eight.” Basically, Le’s program has no idea what it’s talking about. It understands that certain combinations of symbols go together, but it has no appreciation of the real world. It doesn’t know what a centipede actually looks like, or how it moves. It is still just an illusion of intelligence, without the kind of common sense that humans take for granted. Deep-learning systems can often be wonky this way. The one Google created to generate captions for images would make bizarre errors, like describing a street sign as a refrigerator filled with food.

Le asked, “What is the purpose of life?” and the program responded, “To serve the greater good.”

By a curious coincidence, Terry Winograd’s next-door neighbor in Palo Alto is someone who might be able to help computers attain a deeper appreciation of what words actually mean. Fei-Fei Li, director of the Stanford Artificial Intelligence Lab, was on maternity leave when I visited, but she invited me to her home and proudly introduced me to her beautiful three-month-old baby, Phoenix. “See how she looks at you more than me,” Li said as Phoenix stared at me. “That’s because you are new; it’s early facial recognition.”

Li has spent much of her career researching machine learning and computer vision. Several years ago, she led an effort to build a database of millions of images of objects, each tagged with an appropriate keyword. But Li believes machines need an even more sophisticated understanding of what’s happening in the world, and this year her team released another database of images, annotated in much richer detail. Each image has been tagged by a human with dozens of descriptors: “A dog riding a skateboard,” “Dog has fluffy, wavy fur,” “Road is cracked,” and so on. The hope is that machine-learning systems will learn to understand more about the physical world. “The language part of the brain gets fed a lot of information, including from the visual system,” Li says. “An important part of AI will be integrating these systems.”



This is closer to the way children learn, by associating words with objects, relationships, and actions. But the analogy with human learning goes only so far. Young children do not need to see a skateboarding dog to be able to imagine or verbally describe one. Indeed, Li believes that today's machine-learning and AI tools won't be enough to bring about real AI. "It's not just going to be data-rich deep learning," she says. Li believes AI researchers will need to think about things like emotional and social intelligence. "We [humans] are terrible at computing with huge data," she says, "but we're great at abstraction and creativity."

No one knows how to give machines those human skills—if it is even possible. Is there something uniquely human about such qualities that puts them beyond the reach of AI?

Cognitive scientists like MIT's Tenenbaum theorize that important components of the mind are missing from today's neural networks, no matter how large those networks might be. Humans have the ability to learn very quickly from a relatively small amount of data and have a built-in ability to model the world in 3-D very efficiently. "Language builds on other abilities that are probably more basic, that are present in young infants before they have language: perceiving the world visually, acting on our motor systems, understanding the physics of the world or other agents' goals," Tenenbaum says.

If he is right, then it will be difficult to re-create language understanding in machines and AI systems without trying to mimic human learning, mental model building, and psychology.

### **Explain yourself**

Noah Goodman's office in Stanford's psychology department is practically bare except for a couple of abstract paintings propped against one wall and a few overgrown plants. When I arrived, Goodman was typing away on a laptop, his bare feet up on a table. We took a stroll across the sun-bleached campus for iced coffee. "Language is special in that it relies on a lot of knowledge about language but it also relies on a huge amount of common-sense knowledge about the world, and those two go together in very subtle ways," he explained.

Goodman and his students have developed a programming language, called Webppl, that can be used to give computers a kind of probabilistic common sense, which turns out to be pretty useful in a conversation. One experimental version can understand puns, and another can cope with hyperbole. If it is told that some people had to wait "forever" for a table in a restaurant, it will automatically decide that the literal meaning is improbable, and they most likely just hung around for a long time and were annoyed. The system is far from truly intelligent, but it shows how new approaches could help make AI programs that

talk in a more lifelike way.

At the same time, Goodman's example also suggests just how difficult it will be to teach language to machines. Understanding the contextual meaning of "forever" is the kind of thing that AI systems will need to learn, but it is a rather simple and rudimentary accomplishment.

"I want a way to simulate thoughts in a machine," he says. "And if you want to simulate thoughts, then you should be able to ask a machine what it's thinking about."

Still, despite the difficulty and complexity of the problem, the startling success that researchers have had using deep-learning techniques to recognize images and excel at games like Go does at least provide hope that we might be on the verge of breakthroughs in language, too. If so, those advances will come just in time. If AI is to serve as a ubiquitous tool that people use to augment their own intelligence and trust to take over tasks in a seamless collaboration, language will be key. That will be especially true as AI systems increasingly use deep learning and other techniques to essentially program themselves.

"In general, deep-learning systems are awe-inspiring," says John Leonard, a professor at MIT who researches automated driving. "But on the other hand, their performance is really hard to understand."

Toyota, which is studying a range of self-driving technologies, has initiated a research project at MIT led by Gerald Sussman, an expert on artificial intelligence and programming language, to develop automated driving systems capable of explaining why they took a particular action. And an obvious way for a self-driving car to do this would be by talking. "Building systems that know what they know is a really hard problem," says Leonard, who is leading a different Toyota-backed project at MIT. "But yeah, ideally they would give not just an answer but an explanation."

A few weeks after returning from California, I saw David Silver, the Google-DeepMind researcher who designed AlphaGo, give a talk about the match against Sedol at an academic conference in New York. Silver explained that when the program came up with its killer move during game two, his team was just as surprised as everyone else. All they could see was AlphaGo's predicted odds of winning, which changed little even after move 37. It was only several days later, after careful analysis, that the Google team made a discovery: by digesting previous games, the program had calculated the chances of a human player making the same move at one in 10,000. And its practice games had also shown that the play offered an unusually strong positional advantage.

So in a way, the machine knew that Sedol would be completely blindsided.

Silver said that Google is considering several options for commercializing the technology, including some sort of intelligent assistant and a tool for health care. Afterward, I asked him about the importance of being able to communicate with the AI behind such systems. “That’s an interesting question,” he said after a pause. “For some applications it may be important. Like in health care, it may be important to know why a decision is being made.” Indeed, as AI systems become increasingly sophisticated and complex, it is hard to envision how we will

collaborate with them without language—without being able to ask them, “Why?” More than this, the ability to communicate effortlessly with computers would make them infinitely more useful, and it would feel nothing short of magical. After all, language is our most powerful way of making sense of the world and interacting with it. It’s about time that our machines caught up.

*Will Knight is senior editor for AI and robotics at MIT –Technology Review. His feature “[The People’s Robots](#)” appeared in the [May/June issue](#).*

---

Published on 2016-08-10 11:08:23-04:00 in MIT Technology Review  
<https://www.technologyreview.com/s/602094/ais-language-problem/?set=602129>.