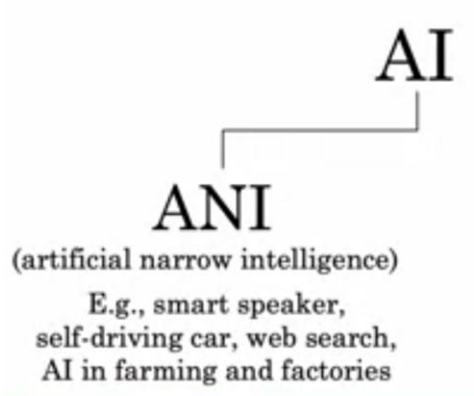
Is there a path to AGI?

Ever since I was a teenager starting to play around with neural networks, I just felt that the dream of maybe someday building an AI system that's as intelligent as myself or as intelligent as a typical human, that that was one of the most inspiring dreams of AI. I still hold that dream alive today. But I think that the path to get there is not clear and could be very difficult.

I don't know whether it would take us mere decades and whether we'll see breakthroughs within our lifetimes, or if it may take centuries or even longer to get there. Let's take a look at what this AGI, artificial general intelligence dream is like and speculate a bit on what might be possible paths, unclear paths, difficult paths to get there someday.

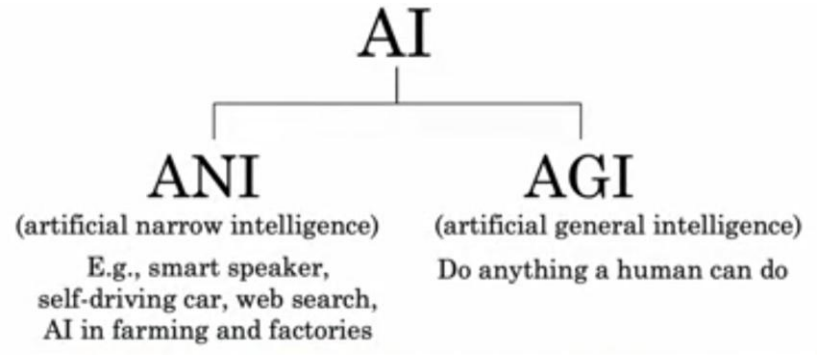
I think there's been a lot of unnecessary hype about AGI or artificial general intelligence. Maybe one reason for that is AI actually includes two very different things.



One is ANI which stands for artificial narrow intelligence. This is an AI system that does one thing, a narrow task, sometimes really well and can be incredibly valuable, such as the smart speaker or self-driving car or web search, or AI applied to specific applications such as farming or factories.

Over the last several years, ANI has made tremendous progress and it's creating, as you know, tremendous value in the world today. Because ANI is a subset of AI, the rapid progress in ANI makes it logically true that AI has also made tremendous progress in the last decade.

There's a different idea in AI, which is AGI, artificial general intelligence. There's hope of building AI systems that could do anything a typical human can do. Despite all the progress in ANI and therefore tremendous progress in AI, I'm not sure how much progress, if any, we're really making toward AGI.

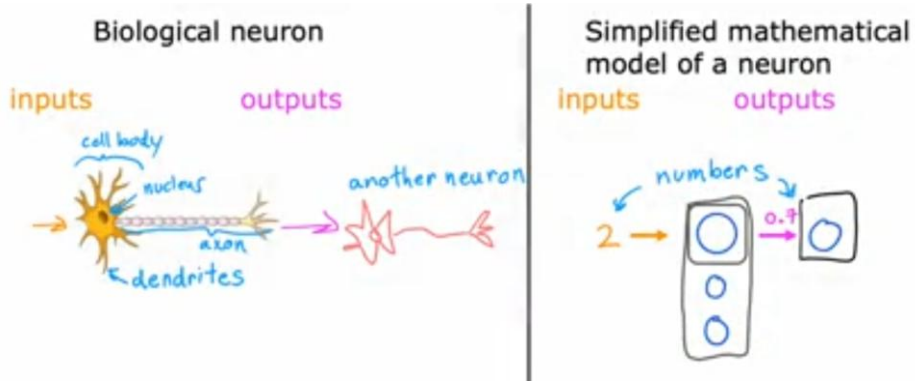


I think all the progress in ANI has made people conclude correctly that there's tremendous progress in AI. But that has caused some people to conclude, I think incorrectly that a lot of progress in AI necessarily means that there's a lot of progress towards AGI.

If you have else about AI and AGI, sometimes you might find drawing this picture useful for explaining some of the things going on in AI as well and some of the sources of unnecessary hype about AGI. With the rise of modern deep learning, we started to simulate neurons and with faster and faster computers and even GPUs we can simulate even more neurons.

I think there was this big hope many years ago that, boy, if only we could simulate a lot of neurons then we can simulate the human brain or something like a human brain and we've really intelligent systems. Sadly, it's turned out not to be quite as simple as that.

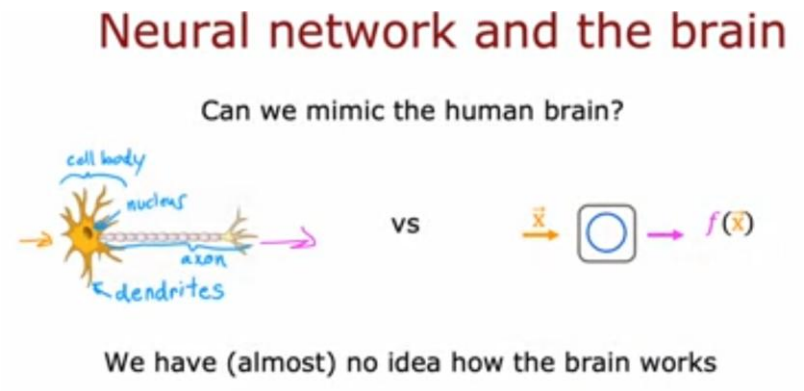
I think two reasons for this is first, if you look at the artificial neural networks we're building, they are so simple that a logistic regression unit is really nothing like what any biological neuron is doing is so much simpler than what any neuron in your brain or mine is doing.



Second, even to this day, I think we have almost no idea how the brain works. There are still fundamental questions about how exactly does a neuron map from inputs to outputs that we just don't know today. Trying to simulate that in a computer, much less a single logistic function is just so far from an accurate model of what the human brain actually does.

Given our very limited understanding both now and probably for the near future of how the human brain works, I think just trying to simulate the human brain as a path to AGI will be an incredibly difficult path. Having said that, is there any hope of within our lifetimes seeing breakthroughs in AGI?

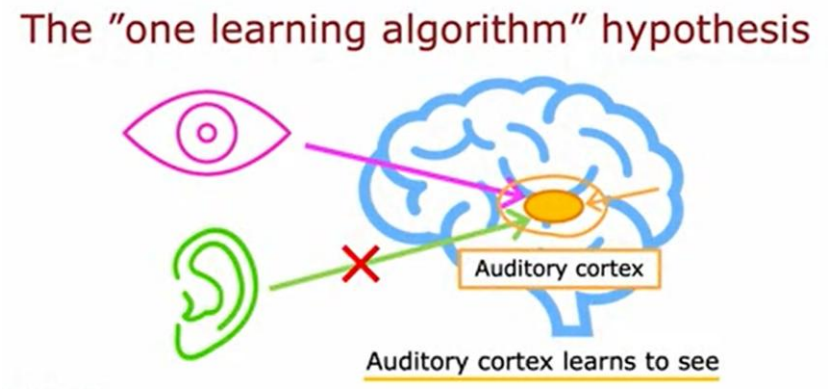
Let me share with you some evidence that helps me keep that hope alive, at least for myself. There have been some fascinating experiments done on animals that shows or strongly suggests that the same piece of biological brain tissue can do a surprisingly wide range of tasks.



This has led to the one learning algorithm hypothesis that maybe a lot of intelligence could be due to one or a small handful of learning algorithms. If only we could figure out what that one or small handful of algorithms are, we may be able to implement that in a computer someday.

Let me share with you some details of those experiments. This is a result due to Roe et al. from many decades ago. The part of your brain shown here is your auditory cortex, and your brain is wired to feed signals from your ears in the form of electrical impulses, depending on what sound your ear is detecting to that auditory cortex.

It turns out that if you were to rewire an animal brain, it's to cut the wire between the ear and the auditory cortex, and instead feed in images to the auditory cortex, then the auditory cortex learns to see. Auditory refers to sound, and so this piece of the brain that in most people learns to here, when it is fed different data, it instead learns to see.

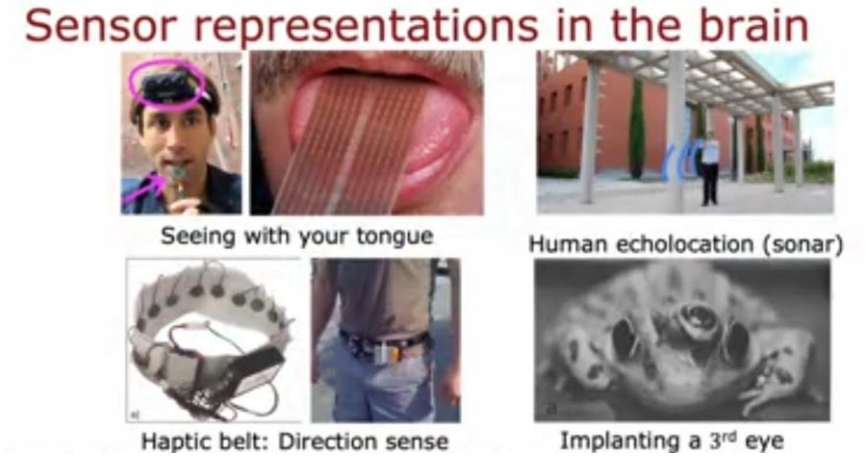


Here's another example. This part of your brain is your somatosensory cortex, somatosensory refers to touch processing. If you were to similarly rewire the brain to cut the connection from the touch sensors to that part of the brain and instead rewire the brain to feed in images, then the somatosensory cortex learns to see.

There's been a sequence of experiments like this, showing that many different parts of the brain, just depending on what data is given can learn to see, or learn to feel, or learn to hear as if there was maybe one algorithm that just depending on what data or this given, learns to process that inputs accordingly.

There happens systems built which take a camera can be mounted to someone's forehead and maps it to a pattern of voltages in a grid on someone's tongue. By mapping a grayscale image to a pattern of voltages on your tongue, this can help people that are not cited line individuals learn to see with your tongue, or they've been fascinating experiments with human echolocation or humans sonar, so animals like dolphins and bats use sonar to see, and researchers have found that if you train humans to make clicking sounds, and listen to how that bounces off surroundings, humans can sometimes learn some degree of human echolocation.

Or this is a haptic belt, and my research lab at Stanford once built something like this before as well, but if you mount a ring of buzzes around your waist and program it using a magnetic compass, so that say, the buzzers to the North most direction are always vibrating slowly, then you somehow gain a direction sense, which some animals have, but humans don't.



Then it just feels like you're walking around and you just know where North is, it doesn't feel like that part of my waist is buzzing, it feels like, oh, I know where that north is. Or surgeries implant a third eye onto frog and the brain just learns with you with this input.

There have been a variety of experiments like these just showing that the human brain is amazingly adaptable, neuroscientists say is amazingly plastic, they just mean adaptable to bewildering range of sensor inputs, and so the question is, if the same piece of brain tissue can learn to see, or touch, or feel, or even other things, what is the average of users, and can we replicate this algorithm and implemented in a computer?

I do feel bad for the frog and other animals, or which these experiments were done, although I think the conclusions are also quite fascinating. Even to this day, I think working on AGI is one of the most fascinating science and engineering problems of all time, and maybe you will choose someday to do research on it.

However, I think it's important to avoid over-hyping, I don't know if the brain is really one or a small handful of algorithms, and even if it were, I have no idea, and I don't think anyone knows what the algorithm is, but I still this hope alive, and maybe it is, and maybe we could, through a lot of hard work, someday discover an approximation to it. I still find this one of the most fascinating topics, I really think about it in my spare time and maybe someday, you are the one to make a contribution to this problem.

In the short term, I think even without pursuing AGI, machine learning and neural networks are very powerful tool, and even without trying to go all the way to build human-level intelligence, I think you find neural networks to be an incredibly powerful, and useful set of tools for applications that you might build. That's it for the required videos of this week, congratulations on getting to this point in the lessons.

After this, we'll also have a few optional videos to dive a little bit more deeply into efficient implementations of neural networks. In particular, in the optional videos to come, I'd like to share with you some details of how to vectorize implementations of neural networks. I hope you also take a look at those videos.