

# Where will Artificial Intelligence come from?

Artificial Intelligence (AI) is making progress in great strides, or at least it appears so! Almost no week passes by without some major announcements of new challenges solved by AI technology or new products powered by AI.

Indeed many quantifiable factors attest an unprecedented level of activity: capital investments, number of academic papers, number of products involving AI technology, they all are on a steep rise in the past five years.

Computers are already very capable at some specialized tasks that require reasoning and other abilities that we typically associate with intelligence. For example, computers can play a decent game of chess or can help us order our holiday photos. Despite this genuine progress, we are still a long way from human level intelligence because our best artificial intelligence systems are not *general* purpose. They cannot quickly adapt to novel tasks the way most humans can do.

When talking about artificial intelligent systems there is a risk of emphasizing humans too much. Computers are *already* more capable than any human at many tasks, for example in numerical computation and search. Yet, in discussions about artificial intelligence we emphasize the shrinking set of abilities where humans still outperform machines. For a nice and more balanced recent discussion on issues surrounding artificial intelligence I recommend reading the [edge contributions towards the Edge 2015 question](#).

As artificial intelligence continues to make progress, I would like to ask the following question:

**Where will the next major advance towards general purpose artificial intelligence come from?**

Below I list seven possible areas which I believe could be the answer to this question; these answers are highly subjective and biased and they may be all wrong, but hopefully they do contain some interesting pointers for everyone.

The point of this exercise is to show that there are many strands of active research that could result in major AI advances. So here are they, the seven areas where a major general purpose AI breakthrough could come from.

## 1. Composable Differentiable Architectures (aka Deep Learning)

*Composable differentiable architectures* describes current state-of-the-art deep learning systems. Frameworks such as [Caffe](#), [Theano](#), [Torch](#), [Chainer](#), all allow the specification of function classes and to automatically compose and differentiate such functions. Because of this mix-and-match composability there is a frictionless and rapid diffusion of components and (sub-)models across application domains.

This *commoditizes* machine learning and allows customization to specific applications; it commoditizes machine learning because the level of knowledge required to leverage modern deep learning frameworks is low. These deep learning frameworks also allow for easy *customization* of the model to the application at hand. Years ago, this *was* the unattained dream for graphical models, but today it *is* achieved by deep learning frameworks where bespoke models are build for most applications.



But is it enough for general purpose AI? What is missing?

I believe there are two obstacles; *first*, almost all deep learning systems require large amounts of supervised data to work. For high-value industrial applications this may be okay because the required label data can be collected. However, there is a long tail of useful applications where label data is rare but unlabeled data is abundant. Future AI systems need to be able to leverage this abundant data source.

*Second*, what is missing are general architectures for reasoning, and an intense search for such building blocks is currently taking place. Maybe classic ideas from AI, such as *blackboard systems*, could be adapted and made differentiable to enable reasoning, or maybe some entirely unexpected new building block will appear.

Besides better models, the key novel technology to look out for in deep learning is custom hardware and novel engineering abstractions. Custom hardware could enable energy savings, or increased speed, or both. Current deep learning piggybacks on GPU development funded largely by the gaming

industry. This is a great thing because developing a new GPU generation such as Nvidia's new Pascal GPU requires very large research and development budgets. Novel engineering abstractions in the form of next generation deep learning frameworks could enable automatic scalability, distributed computation, or offer help in identifying the right architecture for the task.

*Scalability* is important beyond just training speed. For example, consider [basic estimates of the computing power of the human brain](#) or the following quote from a [recent interview with Geoff Hinton](#).

"So in the brain, you have connections between the neurons called synapses, and they can change. All your knowledge is stored in those synapses. You have about 1,000-trillion synapses-10 to the 15, it's a very big number. So that's quite unlike the neural networks we have right now. They're far, far smaller, the biggest ones we have right now have about a billion synapses. That's about a million times smaller than the brain."

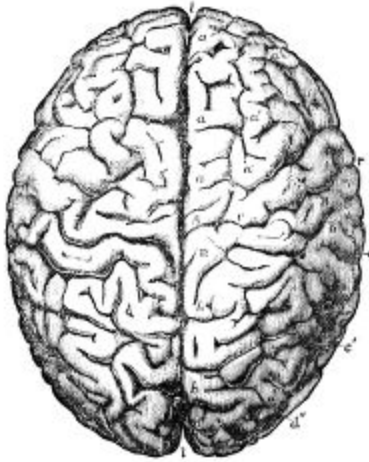
This puts up a ballpark estimate for the number of primitive computational units in the human brain, and it is quite reasonable to attempt to achieve this scale.

One important fact to consider: the driving force behind applications of deep learning is largely the industry, and this will remain the case as long as it pays dividends (it does so greatly at the moment).

## 2. Brain Simulations

Understanding the brain and simulating it is what I think of as the *safe route* to general AI. We do not know whether it will take 5, 50, or 500 years, but it is

likely that we eventually will get there and be able to accurately simulate an artificial brain which is functionally indistinguishable from a real human brain.



Novel technology and approaches to study neural systems, such as [optogenetics](#), [multi-electrode arrays](#) and [connectomics](#), eventually will enable us to obtain a high-fidelity understanding of the brain. Likewise, increase in computation and custom hardware will allow accelerated simulation of neural models.

Most of the investments in this area of research are government funds, for example through the large US [BRAIN initiative](#) and the [Human Brain Project](#), and more general neuroscience funding.

### **3. Algorithmic Information Theory and Universal Intelligence**

Whatever intelligence is, if we were to accept the possibility of a mathematical theory for it, the closest contenders for such theory are found in a field called algorithmic information theory. If you have not heard of algorithmic information theory before, [Gregory Chaitin](#) recently wrote [an excellent essay on the](#)

[conceptual roots of algorithmic information theory](#) and the general history of notions of complexity in science and mathematics.

One approach which leverages algorithmic information theory for general artificial intelligence is the [AIXI agent](#), a theory put forward by [Marcus Hutter](#) that attempts to be universal in the sense that it will successfully and optimally solve any solvable tasks. At its heart it is a [Bayesian reinforcement learning agent](#) where the hypothesis space are possible programs of a [Turing machine](#). It is an extension of an earlier idea to consider Turing machines for predicting future symbols in an observed sequence. This idea is [Solomonoff induction](#) proposed by [Ray Solomonoff](#). Because Turing machines are universal any computable hypothesis can be entertained. AIXI extends this idea from mere prediction of symbols to acting in an unknown environment, that is, to [reinforcement learning](#).



Grounding intelligence in Turing machines is very appealing: not only is it universal, but allows the [formal definition of universal intelligence](#) as well. In essence, reasoning and acting intelligently is reduced to formal manipulation of a notion of complexity defined by programs on a Turing machine. See also [Jürgen Schmidhuber's speed prior for Turing machines](#).

Despite this promise, so far we do not see impressive results achieved by AIXI agents. Why not? There are at least two obstacles:

1. Universal Turing machines are not practically implementable and [approximating AIXI is hard](#); there have been some approximation attempts, e.g. in the work of [\(Veness et al., JAIR 2011\)](#), but at best results have matched other reinforcement learning methods without enabling novel applications that were out of reach before. More recently [Jürgen Schmidhuber](#) proposed a more practical integration of recurrent neural network models of the world with algorithmic information theory in the form of [RNN-based AIs](#).
2. The choice of Turing machine is not clear. There is an infinite set of possible universal Turing machines and we could reasonably hope that the particular choice would not influence the agent efficiency except perhaps for some small overhead. (For a related example, the [Kolmogorov complexity](#) of a sequence is defined through a Turing machine, but whatever the choice of a Turing machine there is an invariance property in that only a constant overhead introduced compared to any other Turing machine.) [Unfortunately for AIXI this is not the case](#): the choice of Turing machine can determine the behaviour of the AIXI agent entirely. (This may also affects Bayesian reinforcement learning more generally: when using a non-parametric prior process the choice of prior may determine more than intended.)

This recent negative result leaves AIXI in an interesting state at the moment. It is clearly the most complete theory of universal agents we have at the moment, see e.g. [Hutter's own review from 2012](#), but it may turn out to be entirely subjective (if no "*natural*" Turing machine can be identified) or practically unworkable.

## 4. Artificial Life

In the above section on brain simulation I argued that by understanding the human brain and then simulating it we will eventually be able to attain human-level intelligence. However, we can start at a more basic level: by understanding and simulating a synthetic form of chemistry we may be able to simulate artificial life. Given a sufficiently rich environment such life may evolve to become intelligent.

The field of *Artificial Life* (ALife) studies the formation and dynamics of life itself on top of artificial simulations of life. This life does not need to be intelligent, and in fact, so far no such simulation has produced life with the intelligence beyond that of a simple organism. But it is clear that [since the early 1990's](#), by any generally plausible definition of life (of which there are many and there is some controversy), artificial life does indeed spontaneously form in computer simulations and complex evolutionary dynamics such as symbiosis and parasites do occur in these simulations.

For a dated but inspiring introduction to the field of artificial life more generally, see [Christoph Adami's book on Artificial Life](#). Adami also wrote a recent [article on evolving artificial intelligence](#) that highlights current research issues for the goal to evolve artificial intelligence.

More fundamentally, in [another article](#) Adami argues that from theoretical results in a field called [integrated information theory](#) (and which I have not heard of before), one possible consequence may be that due to the complexity of general intelligence it is not possible to design it but instead an evolutionary approach is needed.

Given that our goal is to evolve intelligence artificially, fundamentally there are the following obstacles in producing useful general artificial intelligence through artificial life:





1. The *big intelligence filter* hypothesis. This hypothesis goes as follows: life may be abundant but intelligent life may be exceedingly rare. We currently do not know if intelligent life is rare or abundant in the universe, but if it is rare, it may also be exceedingly rare in any simulation of artificial life. A related point is what is known as the *Fermi paradox*, namely, that what science tells us about astrophysics implies we should have likely observed alien civilizations by now, but this has not happened yet. (See Tim Urban's wonderful article on the [Fermi's paradox](#).) Even for life on our own planet we are not sure what triggered intelligence to appear; one widely believed hypothesis is that it happened in a short time, akin to a phase transition, due to a change in ocean oxygen levels 540 million years ago, leading to the [Cambrian explosion](#).
2. Harnessing intelligence. One of our closest genetic relatives, the chimpanzee, are clearly intelligent, but harnessing this intelligence for something useful is difficult. Now imagine a giant squid, swimming a kilometer deep within the ocean. Likely they are also intelligent but we can hardly leverage this for anything useful. Who knows which form intelligent artificial life will take? Will we be able to recognize this life as intelligent? If we do, such life will likely be similar to encountering an alien species in our universe: unlike anything you can imagine or predict beforehand. That is to say, we may be able to achieve intelligent artificial life but may still struggle to make it useful. Even with full control over the simulation environment, a god-like state if you will, it seems necessary that to

make artificial intelligence life useful we will at least have to decode its representation or ``language" and understand the incentives sufficiently well in order to communicate with such intelligence and motivate it to work for us.



In summary, the evolutionary approach to constructing AIs is promising in the long run and there are now several labs working on it (the labs of [Adami](#), [Clune](#), and [Hintze](#)).

## 5. Robotics and Autonomous Systems

Autonomous robots are rapidly conquering novel applications in industry and consumer space, such as in self-driving cars, agricultural robotics, [industry 4.0](#), and drones.

The key enablers of this development are improved sensing technology (e.g. low-cost depth sensors), increased compute and memory capacities, and improved pattern recognition methods. As a result of the maturity of basic required technologies significant industry capital is invested in driving advanced autonomous robotics research.

Beyond the natural urge to feel scared by autonomous machines, how could this lead to a breakthrough in artificial intelligence that cannot be found in one of the constituent technologies?



One line of thought in the field of [embodied cognition](#) argues that an intelligent system is conditioned on its environment in a fundamental way, shaping the allocation of precious (evolutionary) resources in order to maximally exploit the types of sensors and actuators available to it. Therefore the specific nature of sensing and acting abilities is not ancillary to intelligence but the main driving force that enables intelligence in the first place.

If the above thesis were true, autonomous robots with modern sensors and actuators would provide a rich enough *embodiment* for artificial intelligence, and the lack of such an embodiment in other domains would likely impede the emergence of general intelligence.

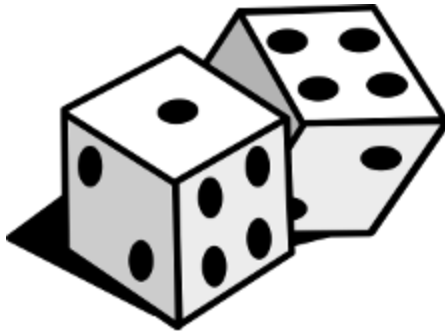
In the past decade, the European Union, through its robotics programme in the 7th Framework Programme (FP7, totalling more than 50 billion Euro for 2007-2013) has placed an emphasis on *combining* cognition with robotics at the exclusion of funding research on artificial intelligence not involving robotics. However, many of the resulting [large research projects of that time](#) are more reflecting the ample funding availability rather than representing progress on fundamental questions of cognition.

## 6. Game Playing

Games entertain humans; what could they do to enable artificial intelligence?

The answer: quite a lot! Games are designed to challenge our intellect, involve interactions between multiple agents, and are sufficiently abstract to be formalized. A computer-implemented game can be as simple and abstract

as tic-tac-toe, Chess, or Go, or as sophisticated and close to reality as the latest Grand Theft Auto game.



Therefore games are an almost ideal research vehicle to drive artificial intelligence research. [Julian Togelius](#) argues this point eloquently in [a recent article](#).

In fact, there are now popular game playing competitions and platforms which drive AI research: the Stanford [general game playing competition](#), the [Computer Poker Competition](#), the [StarCraft AI Competition](#), the [Atari 2600 Arcade Learning Environment](#), and, most recently [Microsoft's Minecraft AI environment \(Malmo\)](#).

It is likely that such platforms will provide diverse and challenging environments for testing the abilities of artificial general intelligence agents, thus accelerating research and enabling breakthroughs. Perhaps the next breakthrough will be in the form of mastering another game.

## 7. Knowledge Bases

A knowledge base is a discrete representation of basic facts and relations about entities. Large-scale knowledge bases constructed semi-automatically from the web are already incredibly useful commercially and they power search engine results and personal assistants.



In search results they provide highly accurate results for known entities in all major search engines (e.g. [Knowledge Graph](#) and [Knowledge Vault](#) in Google, [Satori](#) in Microsoft Bing). To see an example, search for a well-known person, e.g. "Stanislaw Ulam" ([results from Bing](#), [results from Google](#)) and observe that the details about the person displayed.

In personal intelligent assistants such as [Apple Siri](#), [Google Now](#), [Microsoft Cortana](#), or [Amazon Alexa](#) they are responsible for providing facts in basic reasoning abilities. For example, in order to answer queries such as "Who was the president following Thomas Jefferson?" a basic natural language understanding ability and a large knowledge base go a long way.

But can knowledge bases provide the substrate for artificial intelligence? The [Cyc project](#) started in 1984 and the [Open Mind Common Sense project](#) started in 1999 are both based on the belief that in order to enable artificial intelligence we need to encode common sense reasoning, particularly the entities and relationships of everyday life. The hope was that knowledge, encoded in this way, will make reasoning and discovery of novel knowledge simpler.

It is fair to say, that while the (commercial) usefulness of knowledge bases for intelligent applications is now well established, it is too early to say whether general artificial intelligence would require reasoning on top of an explicit

symbolic knowledge base. Perhaps a more continuous and non-symbolic representation of knowledge that supports reasoning is sufficient.

## Conclusion

The goal of artificial general intelligence (AGI) is challenging and exciting on many levels. In all likelihood artificial intelligence will make rapid progress in the next decade, perhaps along the directions we just discussed.

*Acknowledgements.* I thank [Jürgen Schmidhuber](#), [Neil Lawrence](#), [Pushmeet Kohli](#), [Chris Adami](#), and [Chris Watkins](#) for feedback, pointers to literature, and corrections on a draft version of the article.

*Image credits.* The tree image is licensed CC-BY-SA by [adoomer](#). The brain image is a drawing of the brain of Gauss and is [public domain](#). The Anomalocaris image is CC-BY-3.0 licensed art by [Nobu Tamura](#). The octopus image is CC-BY-2.0 licensed art by [Paul K](#). The robot image is licensed CC-BY-2.0 by [striatic](#). The dice image is public domain by Personeoneste. The couple image is [public domain](#).

- BLOG

- [Index](#)
- [Archiv](#)

- LINKS

- [About](#)
- [Microsoft Research Cambridge](#)
- [Edge.org](#)
- [Xians Og](#)

- FEED

- [Atom](#)

- [RSS](#)

