

Martin Ford

**Architects of Intelligence: The  
truth about AI from the people  
building it**

Made by Blinkist



These key insights in blinks were written by a team of experts at Blinkist. We screen the world of nonfiction to choose the very best books. Then, we read them deeply and transform them into this concise format that brings you the most inspiring ideas from the text.

Maybe these blinks will inspire you to dig deeper, or maybe they're enough to start you thinking and then on to something new. However you read blinks, we hope they help you become an even brighter you.

## What's in it for me? A big-picture look at the inner workings and implications of artificial intelligence.

Whether it's the latest sci-fi films or just everyday news, it seems that artificial intelligence can be found everywhere. But perhaps unsurprisingly, there's little agreement about the details surrounding AI: will automation lead to mass unemployment? Will traditional cars soon be replaced by self-driving vehicles? Are we destined to be subjugated by robotic overlords?

Despite the wide variance in opinions surrounding these issues, it's often the most pessimistic and melodramatic voices that take center stage when it comes to discussions about AI and its potential effects.

Of course, AI *does* pose some real threats to our economy. But by focusing only on the downsides, we ignore AI's

potential to reshape the world, redefine work, and provide major boons to healthcare, among many other benefits. These blinks provide a holistic view of AI, from the way it works to the ways it could help or harm us in the future. Along the way, you'll get a general sense of the opinions of the 23 AI experts the author interviewed in making his book.

In these blinks, you'll learn

- how AI learns what a cat is;
- why we should be afraid of killer robots; and
- why live concert tickets might become even more expensive.



## Different deep learning methods can train AI to complete tasks.

Think back to when you were a child. Do you remember the first time you saw a cat, or a picture of a cat? How many cats did you need to see before you fully understood exactly what *a cat* was?

You most likely only needed to see one or two before you could easily differentiate a cat from another animal. This sort of learning, which involves viewing a very small number of examples, comes easily and naturally to humans – but for an AI, it's very difficult. In order for an AI to understand what makes a cat a cat, it must be trained. Nowadays, that often happens through *deep learning*, a form of machine learning that has driven most of the major strides in AI over the last decade.

**The key message here is: Different deep learning methods can train AI to**

## complete tasks.

Whether an AI is being trained to recognize cats, dogs, or coffee cups, it all starts with a *neural network*. This is software with multiple layers of “neurons” that mimic the ones found in the human brain.

There are a few different methods scientists commonly use to train neural networks. One is *supervised learning*, a type of deep learning in which an AI is fed a set of training examples, each labeled with a description. After the AI has been trained, we could then show it a picture of a cat. Next, the collection of pixels in the picture flows through the neural network, after which the machine will confirm – we hope – that what it “sees” is, indeed, a cat.

Even if it guesses correctly, this AI still has absolutely no idea of what the word “cat” signifies – it doesn’t know what a cat does or whether it’s alive. For an AI

to develop this understanding, it needs to be taught via *grounded language learning*. This is a deep learning approach where sentences or words are associated with images, videos, or objects in the real world.

All of these techniques enable deep learning to have all sorts of potential applications. For one, grounded language learning could help develop AI's language skills, making it useful in personal assistants like Siri. And deep learning has already been used to train AI to play games. In one of the most famous instances, the AI AlphaGo was trained by observing many games of Go – and ultimately was able to beat the best human champion at his own game.



## Deep learning is limited.

It's hard not to be impressed when we see an AI beating the best human players at chess, Go, or shogi. But as monumental as that is, it doesn't mean we're any closer to achieving any kind of general intelligence. AI is still only good at completing specific, narrow tasks.

Take AlphaZero. It was trained through deep learning to play deterministic, two-player, and fully observable board games, like chess and Go. Despite AlphaZero's skill in these games, it would be utterly useless if asked to learn to play any *other* type of game – like, say, poker.

**The key message here is: Deep learning is limited.**

Unlike chess or Go, poker is a game of partial information. It's only *partially observable*, which means you can't see the full “board” while playing. So, a

poker-playing AI requires algorithms designed to estimate moves the machine can't actually see. AlphaZero can't do that. It's designed to assume that the game it's playing includes only the pieces on the board.

So, as of right now, AI is only able to complete one task that it's specifically trained for. But another major issue with deep learning and neural networks is in the data we use to teach them.

It's no secret that humans can be biased, even if we don't intend to be. For instance, when it comes to policing, statistics show that some neighborhoods are patrolled more than others, meaning we ultimately have more data about those neighborhoods than others. If an AI is trained using this skewed data, the system could end up making biased predictions about where crime is more likely to occur.

The limits of deep learning techniques mean that we probably won't be able to use any of them to reach the next stage of AI development: Artificial General Intelligence, or AGI. Such a machine would need common sense, or the ability to make inferences about situations it's never encountered before.

But that challenge hasn't deterred researchers from exploring methods for doing this. There are a few ways of trying to give machines common sense. One way essentially involves sticking as many bits of knowledge into an AI's "brain" as possible in the form of logical rules. This technique, however, isn't very practical, since there is an infinite number of potential rules and situations!

Alternatively, some researchers hope that common sense will emerge through machines simply observing the world and learning how things work, in an unstructured way.

But there may be another option: scientists could create AGI by using a hybrid system that combines neural networks and traditional logical rules. We'll take a closer look at how in the next blink.



**Hybrid systems could be the key to further enhancements in AI.**

Over the years, different machine learning techniques have fallen in and out of favor. This phenomenon is exactly what happened with the concept of deep learning. The technique emerged as early as the 1950s but was dismissed as useless by the 1960s – yet today it's once again the dominant machine learning method.

Deep learning will undoubtedly play a role in the AI of the future, but its limitations mean that it can't be the only technique we use to create AGI. Rather, the development of AGI may require researchers to combine several different machine learning methods to create a hybrid system.

**The key message here is: Hybrid systems could be the key to further enhancements in AI.**

Human brains are built with the innate ability to learn – children are our only examples of intelligence reliably scaling up from something less than adult human intelligence. Because of this, many AI researchers are studying children in the hope of uncovering the underlying structures in the brain that allow us to learn.

One neuroscientist and researcher, Demis Hassabis, believes that combining *reinforcement learning* with other techniques offers the most viable path to AGI. In humans, this happens via our dopamine system, where the synapses in our brains are strengthened when they receive reward signals. We can mimic this in AI by asking a machine to repeatedly attempt a task and “reward it” every time it succeeds.

Of course, humans learn in lots of other ways, too. We do a lot of *unsupervised learning*, where we wander around and

gather knowledge through exploration. If scientists can figure out how to get AI to learn in this way, without us needing to provide it with tons of data, we'll almost definitely have a breakthrough in developing AGI.

Similar to human brains, AI could potentially be built with an underlying structure, and then trained with deep learning techniques on top of it. A hybrid system like this, though potentially complex in its application, isn't new. It's already being put to use in today's self-driving cars.

How so? Well, self-driving cars need to be able to understand what decisions to make on the road. Some of the cars' knowledge can come from data gained via deep learning, but not every situation is predictable. That means humans must build in rules that imagine situations the cars *could* potentially encounter, and what their responses should be.

Autonomous vehicles are indeed exciting, but there are plenty more applications for AI. Let's get into some of those next.



**Artificial intelligence has the potential to make life easier and better for everyone.**

In the recent past, there's been a lot of talk about AI's potential to reinforce the biases and stereotypes humans might unintentionally program into them. Lots of academic researchers understand the problem of bias when it comes to AI and are modifying algorithms to respond to it. But, this could be flipped and used to our advantage: we could use AI to eliminate biases.

It's much more difficult to zero in on and correct biases in ourselves. But, as computer scientist Fei-Fei Li points out, when we see our own biases reflected back at us through technology, we can find ways to correct them. And so, eliminating bias is just one example of AI's ability to improve people's lives.

**The key message here is: Artificial intelligence has the potential to make life**

**easier and better for everyone.**

Using AI to eradicate bias isn't unrealistic: in fact, it's already being put to use at Affectiva. Affectiva was created by computer scientist and entrepreneur Rana el Kaliouby, who believes that we need to start thinking about the *emotional* intelligence of machines instead of just their "smartness."

Affectiva has used algorithms and natural language processing techniques to create an anti-bias AI hiring tool. Instead of written resumes, candidates send in video interviews, which the AI ranks based on the candidates' non-verbal communication skills and question responses. The result? At Hirevue, a company that tested the system, hiring time was reduced by 90 percent, and the diversity of new hires increased by 16 percent.

Another of el Kaliouby's projects is designed to help children on the

spectrum of autism, who often have a difficult time interpreting other people's emotions. She and others in her lab created a special type of glasses that can "read" people's emotions, and give the wearer feedback on those emotions. Children who wore the glasses ended up making more eye contact and experienced greater understanding when looking at people's faces.

Aside from these specialized uses for AI, there are plenty of other ways for machine intelligence to help the average person. This might start with robots that can take over mundane, routine tasks like folding laundry, thereby freeing up our time.

But it doesn't stop there. According to Ray Kurzweil, the Director of Engineering at Google, we're likely to one day have nanorobots floating around in our bloodstreams. These microscopic bots could potentially aid our immune

systems, extend our lives, and even connect our brains to the internet.



## Artificial intelligence aids scientific advances, particularly in healthcare.

Walk into any hospital or care facility in America today and you're likely to find a lot of stressed-out doctors and nurses. These days, healthcare professionals work long shifts, have tight schedules, and are highly susceptible to burnout and stress.

Not only that, rushed doctors and nurses mean patients are receiving lower-quality care than they otherwise would. In fact, the current third-leading cause of death in American hospitals is physician error.

Fortunately, healthcare is one area in which AI has the potential to be of major assistance. So, as entrepreneur Oren Etzioni argues – by choosing *not* to take active steps to integrate AI and healthcare, we are actually allowing more lives to be lost.

**The key message here is: Artificial intelligence aids scientific advances, particularly in healthcare.**

Healthcare is a massive industry, and there's a wide variety of ways AI can be used within it.

As of right now, neural networks can be trained to recognize when an image contains a picture of, say, a coffee cup. Those same neural networks could also be trained to recognize when a tumor is present in a radiology scan, for instance.

In a similar vein, diagnosing mental illnesses like depression is difficult right now because we rely mostly on self-reports of symptoms, like suicidal thoughts. But we also know that there are certain facial and vocal biomarkers of depression – they're just hard for humans to pick up on. For an AI with the ability to conduct facial and audio scans, though, it could be quite simple.

If robots could be made responsible for some portion of patient care, doctors' and nurses' time could be freed up and reallocated to where it's really needed. Algorithms could be used to interpret patient information and provide feedback to clinicians, patients, and family members, which would both save time, and improve communication.

But the benefits of AI when it comes to science don't stop with healthcare alone. It could also potentially help in the field of scientific research. Take the project Semantic Scholar, led by Oren Etzioni. In order to stay up-to-date on the latest research, scientists are required to wade through a huge number of publications. Semantic Scholar helps by showing them the papers they're likely to want to read, and locating important findings within those papers.

So far, we've discussed all the good AI can do for the world, but in the next

blink we'll look into some of the potential downsides.



## **Artificial intelligence could be weaponized.**

It's no secret that humanity's destructive capabilities have increased greatly over time, from the bows and arrows of the medieval age to the specialized bombs and drones most militaries use today.

These days, even average citizens have access to drones they could potentially weaponize with small bombs.

Fortunately, though, this sort of weapon isn't scalable. Each drone can only hold one bomb, and be piloted by a single person, meaning its destructive capabilities are limited.

Theoretically, of course, any weapon can be produced en masse and used to arm a country's military. But fortunately, we have international sanctions and military preparedness to prevent such occurrences. What we don't currently have is an international system of control

over autonomous weapons, which could give just one person remote control over an entire fleet of deadly drones.

**The key message here is: Artificial intelligence could be weaponized.**

One major reason autonomous weapons are so potentially dangerous is that they're highly scalable. A massive fleet of, say, 10 million autonomous drones could be overseen by just five people working in a control room, and launched at the mere touch of a button. Moreover, these autonomous weapons could be programmed to attack or kill specific people – say, every male between the ages of 12 and 60 in a given country.

And there are other major risks that come with the development of autonomous weapons. For one, there could soon be an arms race between countries all attempting to be the first one to create this technology.

Autonomous weapons would also be

vulnerable to enemy hacking, meaning a country could end up being attacked by its own weapons.

Given these conceivable issues, it's essential that as long as we continue to develop autonomous weapon technology, we understand the risks, and take the necessary preventative measures. Government regulations are one way of guarding against the misuse of the technology, and researchers should ensure that they design the safest possible systems from the get-go.

Of course, there are other ways AI could be weaponized aside from actual weapons. Advertising, for instance, could use machine learning techniques to influence the way people vote. This has, in fact, already happened, when Cambridge Analytica used Facebook user data to aid the 2016 Trump campaign.

But there may be something more worrying still than these kinds of

weaponization: the threat of job loss on a massive scale.



**Universal basic income or stipends for education could solve the problem of job automation.**

Imagine that, in ten years, you never have to work another day in your life. Does that sound like a utopia or a nightmare?

Whatever your answer, the growing presence and skill of AI – especially when it comes to efficiently completing repetitive tasks – may mean that many jobs will soon be automated. A few commonly-cited examples of the people who could soon be out of a job are cashiers, truck drivers, accountants, and factory workers.

So, how will people survive when they can no longer work to generate income? Well, there are a variety of solutions to this problem, though most of the people the author interviewed agreed that some form of a universal basic income, or UBI, will be necessary in the future.

**The key message here is: Universal basic income or stipends for education could solve the problem of job automation.**

In the long run, UBI might become necessary if a large enough sector of a country's economy is automated. AI could dramatically increase business productivity, generating revenue that could be circulated back to a country's citizens in the form of a monthly stipend.

The question is: Will UBI really be necessary? After all, whenever there has been a massive technological revolution throughout history, people have predicted job losses on a massive scale. Of course, certain jobs *have* indeed disappeared over time – but new ones have always appeared in their place. Just a few years ago, for instance, there were no jobs related to social media. Now, there are many!

For the people whose jobs *do* become automated, robust education programs

could be a solution. Countries could even adopt a policy of paying unemployed individuals to study and establish new careers – a proposal often referred to as *conditional basic income*.

Fortunately, it's likely that we simply aren't going to accept every aspect of our lives being run by robots. We'll still value experiences that involve human connection. Nowadays, for instance, it's possible to buy digital music for just a few dollars. But to attend a live concert, it might cost you a few *hundred* dollars. So, professions that focus on inspiring people and creating connections are likely to become even higher-paying as robots become more and more integrated into our lives.

Yes, job loss is worrying, but some researchers feel we should be focusing our attention on an even bigger threat posed by AGI. They believe this

technology has the potential to spark an  
apocalypse.



## **The potential downsides of Artificial General Intelligence are hotly contested.**

Without a doubt, the rise of AI has brought with it a whole host of concerns. But perhaps none has been as well-publicized as the threat posed by humanlike AGI. What happens if robots eclipse humanity, becoming faster, smarter, and stronger than all of us?

A world dominated by robot overlords is still probably a strictly science-fiction scenario. But according to philosopher Nick Bostrom and a few other researchers, it's worthwhile to consider how the mission to create AGI could end up going very wrong.

**The key message here is: The potential downsides of Artificial General Intelligence are hotly contested.**

If you've ever heard of the *paperclip problem*, first proposed by Nick

Bostrom, you've been introduced to one of the most famous thought experiments that illustrate the potential dangers of AGI.

Imagine that an AI is tasked with operating a paperclip factory. Over time, the AI progressively grows better and more efficient at operating the factory and creating paperclips. Eventually, the AI grows so intelligent that it realizes that the best way to make paperclips is to wrest control away from humanity and turn the entire world into paperclips.

This scenario is purposely cartoonish, but it's meant to show how AI could become *too* good at achieving the goals we program it with.

Most of the researchers the author interviewed believe Bostrom's example is unrealistic, and there are several ways for us to avoid the paperclip scenario. For one, you simply wouldn't give a paperclip-making AI the power to

control, for instance, an electrical grid. On top of that, you would be sure to design the AI with a value system and a conception of right and wrong, such that it isn't hardwired with the sole purpose of creating as many paperclips as possible.

There are yet more ways of safeguarding against an AGI takeover. One of the most radical has been proposed by Bryan Johnson, who believes that we need to upgrade humanity itself – not just machines. To do this, Johnson has created a company called Kernel, with the goal of using neuroscience to “hack” our brains and increase our cognitive abilities through a computer chip implant or another device.

The experts disagree about the exact timeline for when we'll achieve AGI. But what's important is that we take these issues seriously, and prepare for a world

we may soon be sharing with machines  
whose intelligence rivals our own.



## Final summary

The key message in these blinks:

Thanks to advances in deep learning and neural networks, artificial intelligence has become excellent at completing specific, narrow tasks. However, we're still a long way off from the age of Artificial General Intelligence, which will require advances in unsupervised learning, hybrid systems, and neuroscience. Even if we don't see AGI emerge in our lifetime, it's likely that AI's role in healthcare and the military will continue to grow, presenting new challenges *and* providing major benefits for humanity.

## Got feedback?

We'd love to hear what you think about our content! Just drop an email to [remember@blinkist.com](mailto:remember@blinkist.com) with *Architects of Intelligence* as the subject line and share your thoughts!

## What to read next: *The Future of Work*, by Darrell M. West

In these blinks, we've touched upon some of the effects AI will have on jobs and the workforce. Undoubtedly, the jobs we have – if we have them at all – and the work we do in the future will be different from what they are right now. But how exactly will things change? Whose jobs are likely to be taken over by robots? And what are some of the job-related ethical quandaries we'll soon be faced with?

To learn the answers to these questions and find out how the emerging supertechnology of artificial intelligence will influence our economy, check out our blinks to *The Future of Work* by Darrell M. West.



**Nice work! You're all done with this one.**

We publish new books every week at  
[blinkist.com](http://blinkist.com).

Come and see – there's so much more to learn.

Inspired to read the full book?

[Get it here.](#)

Copyright © 2014 by Blinks Labs GmbH.  
All rights reserved.