The Rise of Artificial Intelligence through Deep Learning | Yoshua Bengio |

manual labour	mental labour	to master something	tens of millions of moves	to cram into
knowledge enables intelligence	to contribute	to discover	key principles	laws of aerodynamics
neural networks	**speech recognition**	**computer vision**	**to approach human performance on some benchmarks**	**to understand and generate natural language**
intuitive physics	differential equations	**unsupervised learning**	**representations of data**	a bunch of pixels
high-level meaning	individual meaning of words	characters	to make sense of something	**neurons**
biological neural networks	deep neural networks	**crowdsourcing**	autonomous	to project themselves into the future
to generate plausible futures	to reason and plan ahead	**it's not going to fly**	to apply to do **graduate work**	**a research grant**
to push the boundaries of something	the launch of a factory	to be paid crazy salaries	to guide the next generation	**the social implications of AI**
social value added applications	legal services	**jargon**	to participate in	a lab
a profound impact	collective choices			

1.1.Questions

1. What two methods were used to "train" AlphaGo to master the game of Go?
2. What is the key characteristic of AI as a learning system?
3. What, according to the speaker, is the connection between knowledge and intelligence?
4. Explain the logic of the analogy the speaker draws between the principles od aerodynamics and those of intelligence. What does this analogy suggest about how the speaker perceives the goal of developing AI systems?
5. What are we to understand by the term "unsupervised learning"?
6. What does the speaker mean by "high level meaning" and "highest level of representation"?
7. What are the five lower levels of representation that the speaker identifies?
8. What is crowdsourcing used for in the development of AI?
9. Why are current systems inadequate to the task of training AIs to control self-driving cars?
10. In what ways can "creative" systems overcome the current inadequacies of AI training?
11. How fundamentally do these "creative" systems work?
12. Do you share the speaker's optimism about the benefits of AI?

1.2 Enter the highlighted terms and phrases into the online glossary and explain them in your own words.

Deep learning could reveal why the world works the way it does

At a major AI research conference, one researcher laid out how existing AI techniques might be used to analyze causal relationships in data.

by **Karen Hao**
May 8, 2019

This week, the AI research community has gathered in New Orleans for the International Conference on Learning Representations (ICLR, pronounced “eye-clear”), one of its major annual conferences. There are over 3,000 attendees and 1,500 paper submissions, making it one of the most important forums for exchanging new ideas within the field.

This year the talks and accepted papers are heavily focused on **tackling** four major problems in deep learning: fairness, security, generalizability, and causality. If you’ve been following along with MIT Technology Review’s **coverage**, you’ll recognize the first three. We’ve talked about how machine-learning algorithms in their current state are **biased**, **susceptible to attacks**, and incredibly limited in their ability to generalize the patterns they find in a training data set for multiple applications. Now the research community is busy trying to make the technology sophisticated enough to **mitigate** these weaknesses.

What we haven’t talked about much is the final challenge: causality. This is something researchers have puzzled over for some time. Machine learning is great at finding ***correlations*** in data, but can it ever figure out ***causation***? Such an achievement would be a huge milestone: if algorithms could help us **shed light** on the causes and effects of different phenomena in complex systems, they would deepen our understanding of the world and unlock more powerful tools to influence it.

On Monday, to a packed room, acclaimed researcher Léon Bottou, now at Facebook’s AI research unit and New York University, laid out a new framework for how we might get there. Here’s my summary of his talk.

Let’s begin with Bottou’s first big idea: a new way of thinking about causality. Say you want to build a computer vision system that recognizes handwritten numbers. (This is a classic introductory problem that uses the widely available “MNIST” data set) You’d train a neural network on tons of images of handwritten numbers, each labelled with the number they represent, and end up with a pretty decent system for recognizing new ones it had never seen before.

But let’s say your training data set is slightly modified and each of the handwritten numbers also has a colour—red or green—associated with it. Suspend your disbelief for a moment and imagine that you don't know whether the colour or the shape of the **markings** is a better **predictor** for the digit. The standard practice today is to simply label each piece of training data with both features and feed them into the neural network for it to decide.

Here’s where things get interesting. The “coloured MNIST” data set is purposely misleading. Back in the real world we know that the colour of the markings is completely irrelevant, but in this particular data set, the colour is in fact a stronger predictor for the digit than its shape. So our neural network learns to use colour as the primary predictor. That’s fine when we then use the network to recognize other handwritten numbers that follow the same colouring patterns. But performance completely **tanks** when we reverse the colours of the numbers. (When Bottou played out this thought experiment with real training data and a real neural

network, he achieved 84.3% recognition accuracy in the **former** scenario and 10% accuracy in the **latter**.)

In other words, the neural network found what Bottou calls a "**spurious correlation**," which makes it completely useless outside of the narrow context within which it was trained. In theory, if you could get rid of all the spurious correlations in a machine-learning model, you would be left with only the "**invariant**" ones—those that hold true regardless of context.

Invariance would in turn allow you to understand causality, explains Bottou. If you know the invariant properties of a system and know the intervention performed on a system, you should be able **to infer** the consequence of that intervention. For example, if you know that the shape of a handwritten digit always dictates its meaning, then you can infer that changing its shape (cause) would change its meaning (effect). Another example: if you know that all objects are subject to the law of gravity, then you can infer that when you let go of a ball (cause), it will fall to the ground (effect).

Obviously, these are simple cause-and-effect examples based on invariant properties we already know, but think how we could apply this idea to much more complex systems that we don't yet understand. What if we could find the invariant properties of our economic systems, for example, so we could understand the effects of implementing **universal basic income**? Or the invariant properties of Earth's climate system, so we could evaluate the impact of various **geoengineering ploys**?

Idea #2

So how do we get rid of these spurious correlations? This is Bottou's second big idea. In current machine-learning practice, the **default intuition** is to amass as much diverse and representative data as possible into a single training set. But Bottou says this approach **does a disservice** to AI. Different data that comes from different contexts—whether collected at different times, in different locations, or under different experimental conditions—should be preserved as separate sets rather than mixed and combined. When they are consolidated, as they are now, important **contextual information** gets lost, leading to a much higher likelihood of spurious correlations.

With multiple context-specific data sets, training a neural network is very different. The network can no longer find the correlations that only hold true in one single diverse training data set; it must find the correlations that are invariant across all the diverse data sets. And if those sets are selected smartly from a full spectrum of contexts, the final correlations should also closely match the invariant properties of the ground truth.

So let's return to our simple coloured MNIST example one more time. Drawing on his theory for finding invariant properties, Bottou reran his original experiment. This time he used two coloured MNIST data sets, each with different colour patterns. He then trained his neural network to find the correlations that held true across both groups. When he tested this improved model on new numbers with the same and reversed colour patterns, it achieved 70% recognition accuracy for both. The results proved that the neural network had learned to disregard colour and focus on the markings' shapes alone.

Bottou says his work on these ideas is not done, and it will take the research community some time to test the techniques on problems more complicated than coloured numbers. But the framework **hints at** the potential of deep learning to help us understand why things happen, and thus give us more control over our **fates**.

2.1 Vocabulary Exercise

to tackle a problem	coverage	to be biased	to be susceptible to attacks
correlation	causation	to shed light on something	the workings of something
a predictor	to tank	the former… and the latter	a spurious correlation
invariant	to infer something from something	a default intuition	to do someone/something a disservice
contextual information	to hint at something	fate	

Complete the following sentences

1. AI developers are attempting to **tackle the problem** of causality by…
2. Algorithms designed to guide the behaviour and functioning of AI systems often display the unconscious **biases** of their developers, as demonstrated by the following example:…
3. Computer systems are often **susceptible to hacking attacks** because…
4. The difference between **correlation** and **causation** is…
5. The recent New York Times investigative report on Trump's tax returns has **shed light on**…
6. The shape rather than the colour of a digit is a better **predictor** of the correct choice because…
7. When data from different sets are combined into one general set the system's performance **tanks** because…
8. AlphaGo tends to perform better than Stockfish in contests between the two systems. This seems to be because **the former**… whereas **the latter**…
9. The stereotypical **correlation** between wealth and happiness could be a **spurious** one because…
10. Eratosthenes **inferred** the size of the Earth from…

2.2. Questions

1. What was Bottou's first big idea?
2. What is Bottou's second big idea?
3. Can you think of any as yet unexplained phenomena that Bottou's "invariant property" AI could find answers to?