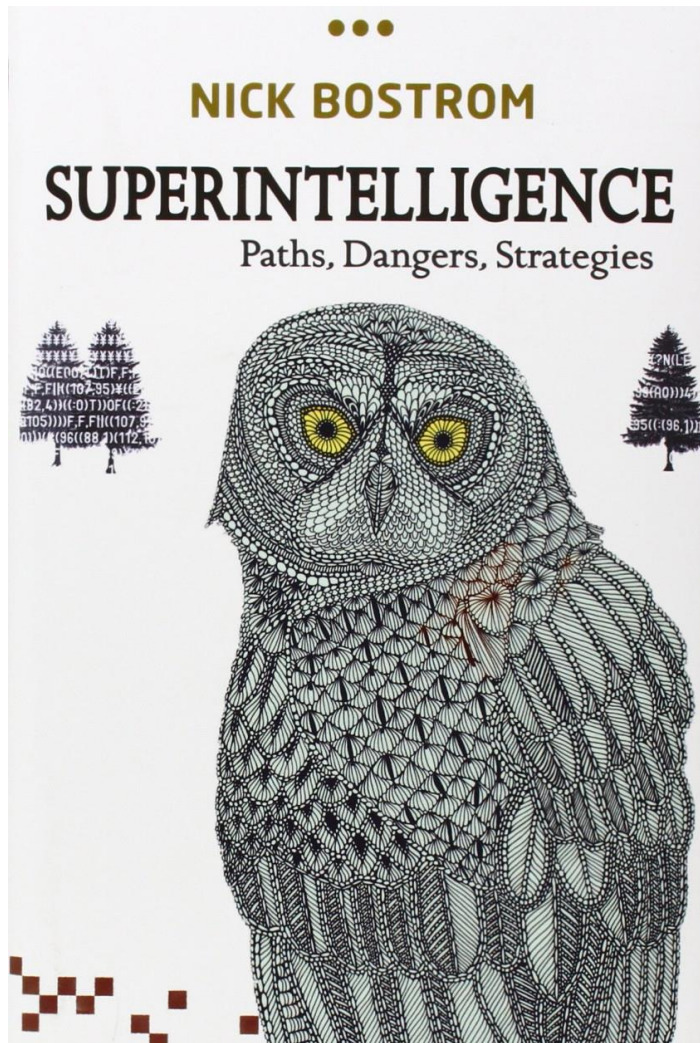# Superintelligence by Nick Bostrom

At Dartmouth College in the late spring of 1956, a gathering of researchers started charting another course for the world's future.

They started out with the idea that machines could recreate parts of human intelligence.

As you can already notice all around you, their effort evolved and "expert systems" thrived in the 80s, along with the promise of artificial intelligence.

However, at that point advances came to a plateau and hence, the funding subsided.

In the 90s, the "genetic algorithms" and "neural networks," pushed the idea to take off once again.

However, how do scientists measure the power of AI?

Well, for starters, by measuring how well specifically designed machines play games such as chess, poker Scrabble, Go and Jeopardy. For instance, a machine with successful algorithms and calculations will beat the best human Go player in about ten years.

But, that is not all – games are just the beginning.

AI's applications do not stop at games. They stretch out to listening devices, face and speech recognition, scheduling and planning, diagnostics, navigation, inventory management and a wide range of industrial robots.

It sounds nice, doesn't it?

In spite of AI's increasing fame and possibilities of utilization, indications of its confinements are rising.

For example, in the "Flash Crash" of 2010, algorithmic traders coincidentally made a descending spiral that cost the market a trillion dollars in moments.

However, we have to bear in mind that the innovation that caused the crisis in the first place was the one that ultimately helped to solve it.

In any case, the question remains: will AI's curve take after the transforming pattern of human intelligence?

As a matter of fact, AI's evolution may follow several paths.

Researchers believe that one day AI will evolve into "superintelligence," which would be a profoundly [different sort of intelligence](#).

This brings up another question:

Would such superintelligence be able to produce human feelings? And if that is the case, how?

A superintelligence could take up three forms.

First, "speed superintelligence" which could imitate human intelligence, but work more quickly.

Second, "collective superintelligence" which would be a network of subsystems that could autonomously take care of discrete issues that are a part of a larger undertaking.

The third is ambiguously defined as "quality superintelligence." It alludes to an AI of such high caliber that it is as superior to human's intellect as humans are too, say – dolphins.

With respect to how quick science could make another intelligence, the appropriate response relies upon "optimization power and system recalcitrance," or willingness to comply.

Key Lessons from "Superintelligence":

1. "Orthogonality"
2. AI Architecture and Scenarios
3. Moral Character

"Orthogonality"

Keep in mind that the character of superintelligence is not exactly human.

Do not get into fantasies about humanized AI. Although it may sound counterintuitive, the orthogonality thesis states that levels of intelligence do not correlate with final objectives.

In fact, more intelligence does not mean that the number of shared or collective objectives among different AIs will increase.

However, one thing is sure: an AI's motivation will inevitably consist of some "instrumental goals" such as achieving technological perfection.

AI Architecture and Scenarios

To study different possible scenarios in which the world will function after the widespread introduction of superintelligence, just think of how the new technologies influenced the horse.

Just some time ago, carriages increased the horse's capabilities, but when the cars were introduced, they almost completely replaced it. As a result, horse populations rapidly declined.

If that is the case, what will happen to people when superintelligence replaces many of their abilities? Humans have property, capital, and political power, but many of those advantages may become unimportant when superintelligent AIs enter the scene.

Moral Character

Scientists have practical strategies that could help them develop a moral character inside an AI.

When we say moral character, it does not necessarily mean that these values will match those of people. Instead, think of a moral which will be unique for superintelligence.