I. About predicting the future

We will start by addressing what is known to be one of the hardest problems of all: predicting the future.

You may be disappointed to hear this, but we don't have a crystal ball that would show us what the world will be like in the future and how AI will transform our lives.

As scientists, we are often asked to provide predictions, and our refusal to provide any is faced with a roll of the eyes ("boring academics"). But in fact, we claim that anyone who claims to know the future of AI and the implications it will have on our society, should be treated with suspicion.

The reality distortion field

Not everyone is quite as conservative about their forecasts, however. In the modern world where big headlines sell, and where you have to dissect news into 280 characters, reserved (boring?) messages are lost, and simple and dramatic messages are magnified. In the public perception of AI, this is clearly true.

Note

From utopian visions to grim predictions

The media sphere is dominated by the extremes. We are beginning to see AI celebrities, standing for one big idea and making oracle-like forecasts about the future of AI. The media love their clear messages. Some promise us a utopian future with exponential growth and trillion-dollar industries emerging out of nowhere, true AI that will solve all problems we cannot solve by ourselves, and where humans don't need to work at all.

It has also been claimed that AI is a path to world domination. Others make even more extraordinary statements according to which AI marks the end of humanity (in about 20-30 years from now), life itself will be transformed in the "Age of AI", and that AI is a threat to our existence.

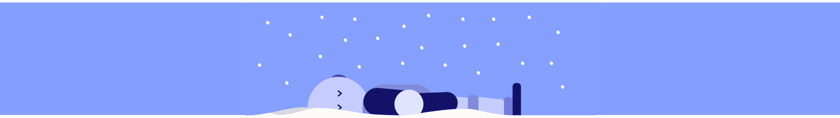
While some forecasts will probably get at least something right, others will likely be useful only as demonstrations of how hard it is to predict, and many don't make much sense. What we would like to achieve is for you to be able to look at these and other forecasts, and be able to critically evaluate them.

On hedgehogs and foxes

The political scientist Philip E. Tetlock, author of *Superforecasting: The Art and Science of Prediction,* classifies people into two categories: those who have one big idea ("hedgehogs"), and those who have many small ideas ("foxes"). Tetlock has carried out an experiment between 1984 and 2003 to study factors that could help us identify which predictions are likely to be accurate and which are not. One of the significant findings was that foxes tend to be clearly better at prediction than hedgehogs, especially when it comes to long-term forecasting.

Probably the messages that can be expressed in 280 characters are more often big and simple hedgehog ideas. Our advice is to pay attention to carefully justified and balanced information sources, and to be suspicious about people who keep explaining everything using a single argument.

Predicting the future is hard but at least we can consider the past and present AI, and by understanding them, hopefully be better prepared for the future, whatever it turns out to be like.



AI winters

The history of AI, just like many other fields of science, has witnessed the coming and going of various different trends. In philosophy of science, the term used for a trend is paradigm. Typically, a particular paradigm is adopted by most of the research community and optimistic predictions about progress in the near-future are provided. For example, in the 1960s neural networks were widely believed to solve all AI problems by imitating the learning mechanisms in the nature, the human brain in particular. The next big thing was expert systems based on logic and human-coded rules, which was the dominant paradigm in the 1980s.

The cycle of hype

In the beginning of each wave, a number of early success stories tend to make everyone happy and optimistic. The success stories, even if they may be in restricted domains and in some ways incomplete, become the focus on public attention. Many researchers rush into AI – or at least calling their research AI – in order to access the increased research funding. Companies also initiate and expand their efforts in AI in the fear of missing out (FOMO).

So far, each time an all-encompassing, general solution to AI has been said to be within reach, progress has ended up running into insurmountable problems, which at the time were thought to be minor hiccups. In the case of neural networks in the 1960s, the hiccups were related to handling nonlinearities and to solving the machine learning problems associated with the increasing number of parameters required by neural network architectures. In the case of expert systems in the 1980s, the hiccups were associated with handling uncertainty and common sense. As the true nature of the remaining problems dawned after years of struggling and unsatisfied promises, pessimism about the paradigm accumulated and an AI winter followed: interest in the field faltered and research efforts were directed elsewhere.

Modern AI

Currently, roughly since the turn of the millennium, AI has been on the rise again. Modern AI methods tend to focus on breaking a problem into a number of smaller, isolated and well-defined problems and solving them one at a time. Modern AI is bypassing grand questions about meaning of intelligence, the mind, and consciousness, and focusing on building practically useful solutions in real-world problems. Good news for us all who can benefit from such solutions!

Another characteristic of modern AI methods, closely related to working in the complex and “messy” real world, is the ability to handle uncertainty, which we demonstrated by studying the uses of probability in AI in Chapter 3. Finally, the current upwards trend of AI has been greatly boosted by the come-back of neural networks and deep learning techniques capable of processing images and other real-world data better than anything we have seen before.

Note

So are we in a hype cycle?

Whether the history will repeat itself, and the current boom will be once again followed by an AI winter, is a matter that only time can tell. Even if it does, and the progress towards better and better solutions slows down to a halt, the significance of AI in the society is going to stay. Thanks to the focus on useful solutions to real-world problems, modern AI research yields fruit already today, rather than trying to solve the big questions about general intelligence first – which was where the earlier attempts failed.

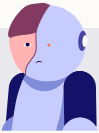
Prediction 1: AI will continue to be all around us

As you recall, we started by motivating the study of AI by discussing prominent AI applications that affect all our lives. We highlighted three examples: self-driving vehicles, recommendation systems, and image and video processing. During the course, we have also discussed a wide range of other applications that contribute to the ongoing technological transition.

Note

AI making a difference

As a consequence of focusing on practicality rather than the big problems, we live our life surrounded by AI (even if we may most of the time be happily unaware of it): the music we listen to, the products we buy online, the movies and series we watch, our routes of transportation, and even the news and information that we have available, are all influenced more and more by AI. What is more, basically any field of science, from medicine and astrophysics to medieval history, is also adopting AI methods in order to deepen our understanding of the universe and of ourselves.



Prediction 2: the Terminator isn't coming

One of the most pervasive and persistent ideas related to the future of AI is the Terminator. In case you should have somehow missed the image of a brutal humanoid robot with a metal skeleton and glaring eyes...well, that's what it is. *The Terminator* is a 1984 film by director James Cameron. In the movie, a global AI-powered defense system called Skynet becomes conscious of its existence and wipes most of the humankind out of existence with nukes and advanced killer robots.

Note

Two doomsday scenarios

There are two alternative scenarios that are suggested to lead to the coming of the Terminator or other similarly terrifying forms of robot uprising. In the first, which is the story from the 1984 film, a powerful AI system just becomes conscious and decides that it just really, really dislikes humanity in general.

In the second alternative scenario, the robot army is controlled by an intelligent but not conscious AI system that is in principle in human control. The system can be programmed, for example, to optimize the production of paper clips. Sounds innocent enough, doesn't it?

However, if the system possesses superior intelligence, it will soon reach the maximum level of paper clip production that the available resources, such as energy and raw materials, allow. After this, it may come to the conclusion that it needs to redirect more resources to paper clip production. In order to do so, it may need to prevent the use of the resources for other purposes even if they are essential for human civilization. The simplest way to achieve this is to kill all humans, after which a great deal more resources become available for the system's main task, paper clip production.

Why these scenarios are unrealistic

There are a number of reasons why both of the above scenarios are extremely unlikely and belong to science fiction rather than serious speculations of the future of AI.

Reason 1:

Firstly, the idea that a **superintelligent**, conscious AI that can outsmart humans emerges as an unintended result of developing AI methods is naive. As you have seen in the previous chapters, AI methods are nothing but automated reasoning, based on the combination of perfectly understandable principles and plenty of input data, both of which are provided by humans or systems deployed by humans. To imagine that the nearest neighbor classifier, linear regression, the AlphaGo game engine, or even a deep neural network could become conscious and start evolving into a superintelligent AI mind requires a (very) lively imagination.

Note that we are not claiming that building human-level intelligence would be categorically impossible. You only need to look as far as the mirror to see a proof of the possibility of a highly intelligent physical system. To repeat what we are saying: superintelligence will not emerge from developing narrow AI methods and applying them to solve real-world problems (recall the narrow vs general AI from the section on the philosophy of AI in Chapter 1).

Reason 2:

Secondly, one of the favorite ideas of those who believe in superintelligent AI is the so-called **singularity**: a system that optimizes and "rewires" itself so that it can improve its own intelligence at an ever accelerating, exponential rate. Such superintelligence would leave humankind so far behind that we become like ants that can be exterminated without hesitation. The idea of exponential intelligence increase is unrealistic for the simple reason that even if a system could optimize its own workings, it would keep facing more and more difficult problems that would slow down its progress, quite like the progress of human scientists requires ever greater efforts and resources by the whole research community and indeed the whole society, which the superintelligent entity wouldn't have access to. The human society still has the power to decide what we use technology, even AI technology, for. Much of this power is indeed given to us by technology, so that every time we make progress in AI technology, we become more powerful and better at controlling any potential risks due to it.

Note

The value alignment problem

The paper clip example is known as the value alignment problem: specifying the objectives of the system so that they are aligned with our values is very hard. However, suppose that we create a superintelligent system that could defeat humans who tried to interfere with its work. It's reasonable to assume that such a system would also be intelligent enough to realize that when we say “make me paper clips”, we don't really mean to turn the Earth into a paper clip factory of a planetary scale.

Separating stories from reality

All in all, the Terminator is a great story to make movies about but hardly a real problem worth panicking about. The Terminator is a gimmick, an easy way to get a lot of attention, a poster boy for journalists to increase click rates, a red herring to divert attention away from perhaps boring, but real, threats like nuclear weapons, lack of democracy, environmental catastrophes, and climate change. In fact, the real threat the Terminator poses is the diversion of attention from the actual problems, some of which involve AI, and many of which don't. We'll discuss the problems posed by AI in what follows, but the bottom line is: forget about the Terminator, there are much more important things to focus on.

Unanswered

Exercise 23: What is the perception of AI?

For this exercise, we want you to think about how AI is portrayed. Do an online **image search** for the term “AI” and see what kinds of pictures come up. If you are using Google search, you should choose "Images" in the top of the screen.

What's the general impression you get about AI from the image search results? Is this an accurate representation of AI? Why or why not?

Your answer

Words: 0

Submit

Next section

II. The societal implications of AI →

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Reaktor