Deep learning is a class of machine learning algorithms that has been applied to different fields, like face recognition, autonomous vehicles, and drugs discovery. Since deep learning will be of growing relevance to our daily lives and increasingly steer us in our behavior in the future, it is problematic that there is little understanding of how deep learning actually works. We see its workings as a black box and this often leads to mistrust and alienation of the way machines reason.

**Our observations**

- The applications of deep learning are numerous, e.g. machine vision, speech recognition, natural language processing, social network filtering, and advertising. The deep neural networks that enable deep learning, are a puzzle. Last month, a new theory has been coined by Naftali Tishby that explains how deep learning works: in squeezing data through a bottleneck, the neural network gets rid of the irrelevant data and preserves the information relevant to general concepts. This theory could further help to construct new objectives and architectures of networks.

- The European Union’s new General Data Protection Regulation will be implemented in 2018 and impacts the use of machine learning algorithms. It will restrict automated decision-making which significantly affects users and create the right for users to ask for an explanation of an algorithmic decision that was made for them.

- The digital traces we leave behind, contextual content such as emotional or unconscious data, are feeding deep neural networks which leads to insights. They can be used for commercial or political ends, e.g. through microtargeting—by measuring people’s personality using their digital footprints. Cambridge Analytica, for example, helped to increase the power of the Trump campaign. Deep learning algorithms also enable China to develop its “citizen score”. This uses people’s online activities to determine how loyal and compliant they are.

- Google and Microsoft have shown a significant growth of their AI’s IQ. AI is of increasing importance for companies that use technology to power their offerings. For example, Google uses a machine learning program called RankBrain, interpreter algorithms that power most search results.
The scope of AI applications has been widening over the past few years. This is linked to the growing availability of computing power, storage, and data (big data). Deep learning is a small step closer to the ultimate goal of actually creating AI as a general purpose technology.

Although deep learning has already proved powerful in its applications, it is often looked at as a black box because it establishes a new way of reasoning. Deep learning does not search information for validation, but instead, learns from incomplete information. It focuses on extracting, combining, and discovering patterns in data and learns from other algorithms. It establishes a meta-learning from the (for us) hidden layers of the network.

As deep learning applications become more ubiquitous, humans are increasingly being steered by decisions based on reasoning that they cannot retrieve. This is what Luciana Parisi calls technological decisionism. By recommending us what movie to watch or what route to take to work, these decisions influence our behavior, even though the internal decision-making process is not open to scrutiny. In some cases, like AI doctors or automated legal services judging us, we will want to examine the decision-making that took place, to open the deep learning black box.

Parisi writes that technological decisionism is often received with mistrust and that it leads to a loss of confidence because of its force, while it can actually be a powerful instrument that supports our own decisions. We should not emphasize the disparity between human and machine reasoning, not focus on our inability to grasp automated reasoning, but reinvent how this new reasoning can be instrumental in reaching human goals. Similarly, with Open AI, Elon Musk hopes “to advance digital intelligence in the way that is most likely to benefit humanity as a whole”. Until AI reaches its potential, we have the opportunity to find our way to work with the black box and to program technological decisionism so it helps us with reaching human goals.

**Connecting the dots**

**Implications**

- Advancements in deep learning in everyday life applications, e.g. healthcare, education, and well-being
- Stimulus for research and businesses to “open the black box”