Stéphane Mallat – We talk a lot about the spectacular performance of artificial intelligence and its application but what interests me is rather the mysteries of the thematises that lies hidden behind. Artificial intelligence is based on statistical learning algorithms that learn to reply to questions relying on data. For example, in order to recognise if an image is a dog or a cat, the algorithm is trained on algorithms that know the answer – be it dogs or cats - and the game is to be able to generalise. In other words, the aim is to calculate a suitable answer no matter what the image. The main difficulty comes from the curse of extra dimensions that in some ways is the scoundrel of our tale. The reason is that a single datum includes a large number of variables. For example, an image is made of a million pixels that are all points on the image. This provokes an explosion of combinations of possible answers. And the mystery is that neural networks manage even so to recognise the image and therefore to step around this curse of dimensionality. Most astonishing is that very similar neural networks can learn to solve very different problems, like the recognition of an image, to play at go, synthesise speech, calculate the quantum energy of a molecule, make a medical diagnosis and so on. That means that these problems which seem so different have nevertheless similar properties that make it possible to evade the curse of dimensionality. And a question that fascinates me is to understand the mathematical nature of these properties.

The organisational hierarchy of the majority of these problems seems to play a fundamental role. It's a bit like in the *Discours de la méthode* by Descartes, we can first divide the global problem into little simpler problems then we build up bit by bit in order to attack the more complicated aspects of the initial problem. In physics, this hierarchy appears in interactions at different scales, going from little elementary particles that interact to form atoms which themselves interact to form molecules right up to macroscopic structures such as the Earth then, as you might imagine, the solar system and on up to the cosmos. And we notice that we find the same type of hierarchical aggregation in the information that is handled by neural networks, be they images, sounds, text or no matter what other type of data.

Now an understanding of how to separate information over different scales is at the heart of the mathematical theory of wavelets on which I have spent much effort. A wavelet is a little wave, a bit like a sinusoid but localised, tightly localised in time – think of a note in music. And if we take wavelets of different sizes and in different positions, we can reconstruct any kind of signal, for example a piece of music. We can write it on a score where we see the musical symbols, meaning little wavelets at different octaves, different instants and with different rhythms. And wavelet transform is very interesting from a mathematical point of view because one can indeed use it for music and sounds but also for images, for representing the energy of molecules and for any mathematical function whatsoever. It lets us separate data out over different scales and furthermore we have had the surprise to find these wavelets in the choclea of ears and in the visual cortex of both animals and humans. They appear also in the parameters learnt by neural networks. And it seems that by voyaging across hierarchies of scale, we may learn the parameters which best capture information.

For all this, we still lack a precise mathematical framework for understanding the nature of this voyage through many dimensions. Managing to explain the structures of information handled by neural networks or even by our brain is really an extraordinary mathematical problem but one that is very difficult. And I do not think that this research is a story that will finish quickly.

04 min 28 sec