FOCUS

ARTIFICIAL INTELLIGENCE IS LEARNING

24 | Not without a reason

32 | When machines get involved

38 | Software to scan the psyche

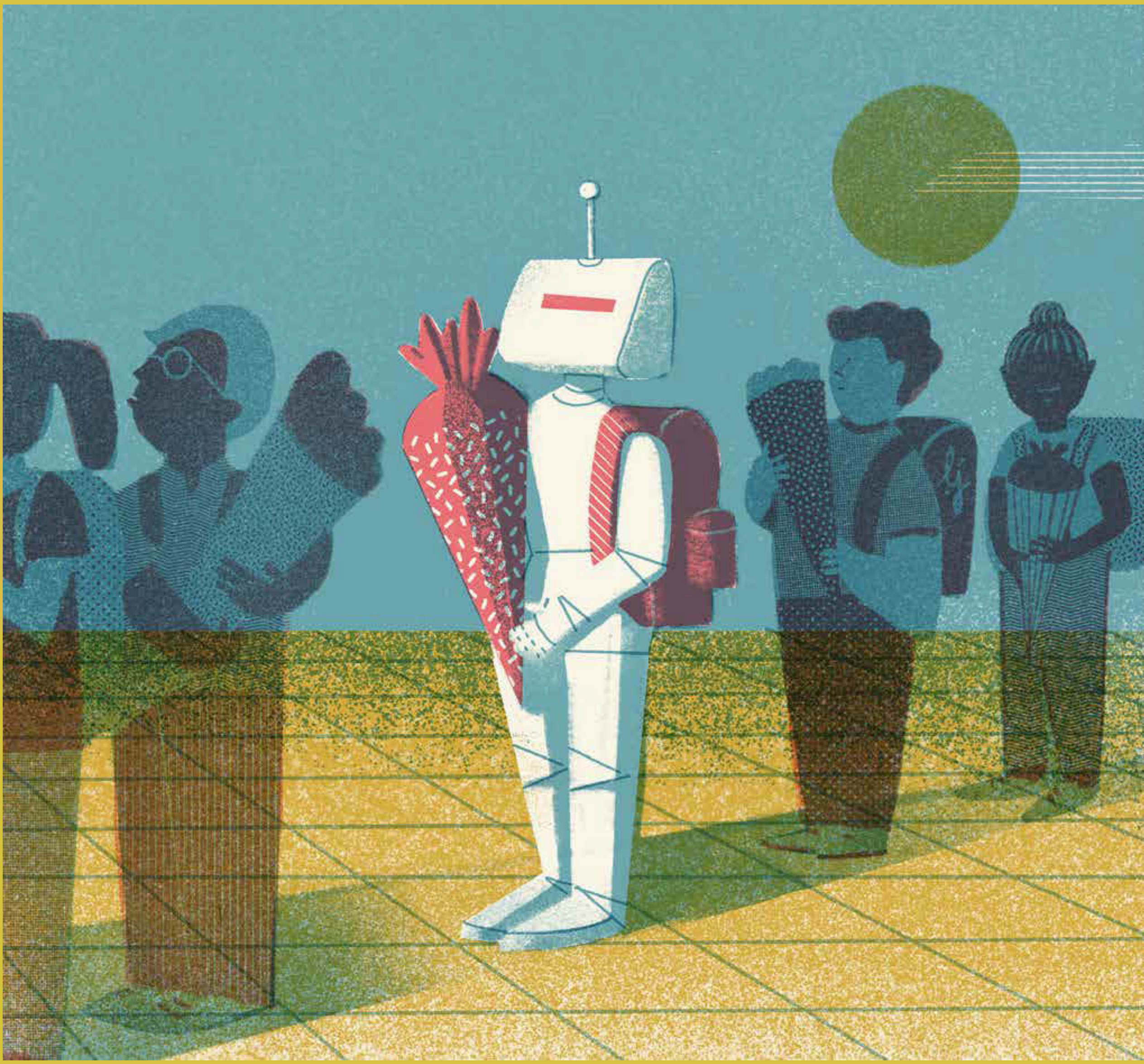


ILLUSTRATION: LUISA JUNG FOR MPG

Learning like a human: in the future, systems using AI will not only recognize patterns in large volumes of data but also make it possible to understand the criteria they use to make their decisions.

NOT WITHOUT A REASON

TEXT: THOMAS BRANDSTETTER

Artificial intelligence (AI) has long been able to recognize patterns much better than humans can. However, in order to truly be worthy of its name, it would also need to understand causal relationships. And that is precisely what researchers at the Max Planck Institute for Intelligent Systems in Tuebingen are working on.

Cause and effect are everywhere. But that asymmetrical pairing isn't always as easy to spot as in the domino effect in which one piece knocks over the next. Human decisions often have highly complex causes – and sometimes even more complex consequences. Sometimes things mutually influence each other. For example, our CO_2 emissions warm the Earth and the thawing permafrost releases even more greenhouse gas. And when you look at an idyllic landscape photo, you can usually see shadows caused by the light of the Sun.

While we humans usually recognize causal relationships intuitively thanks to our brains, they still present considerable difficulties for intelligent machines. Machines may outperform us in finding patterns and correlations. However, the concept of cause and effect generally still eludes them. Researchers at the Department of Empirical Inference at the Max Planck Institute for Intelligent Systems see this shortcoming as a challenge. Led by Bernhard Schölkopf, they are trying to impart a sense of causality to learning machines in areas as diverse as the search for exoplanets, climate change, and the granting of loans.

Experiments bring understanding

One starting point for their work is the quality of the data available. Most Big Data sets arise from purely passive observation and contain no additional information about how they were created. For example, account transactions that include only the monetary amounts and the associated times say nothing about why the payments were made. "If a system can only passively collect data, that's generally not enough to detect a causal relationship," explains Julius von Kügelgen, who is working on his dissertation in the Department of Empirical Inference. "Unfortunately, however, 99 percent of the data available today is only passively collected – or at least treated that way when it's analyzed." One example of this is statistical data from which interrelationships such as the correlation between chocolate consumption per capita and the number of Nobel Prizes won by a country can be derived. In this case, the common cause principle applies. This ultimately attributes many such patterns to a causal relationship. Either one variable influences the other, or there is a third variable that causally influences both.

It's easy for a machine to determine the correlation. To do this, the algorithm simply needs to sift through relevant statistics and identify any patterns. However, in order to be able to understand the connection and find a possible explanation, additional information is needed. We humans are usually helped by our general understanding of how the world works. We have eaten chocolate and know that it has no effect on intelligence. And why should winning a Nobel Prize boost chocolate consumption? Our thoughts are thus quickly steered toward a third variable. For example, a strong economic system that leads to prosperity and chocolate on one hand and a good education system on the other. But because a computer algorithm lacks this general understanding, it is inevitably groping in the dark given the sheer amount of data. "What is more interesting is data already collected in experiments in which someone actively intervenes in a system and changes something," says von Kügelgen. For example, close observation quickly reveals a correlation between wet soil and rain. However, an experiment in which the soil is made wet with a garden hose and not by rain debunks the supposed causal relationship.

In his current research, von Kügelgen is dealing with even more difficult cases. He is driven by "what if" questions, specifically: would I have been granted the loan if my income had been higher? "Artificial intelligence is increasingly making decisions for humans. For example, when assessing creditworthiness," says von Kügelgen. "Our system can help people answer the question of what they should do in order to achieve a positive outcome." To do this, he and his colleague Amir Hossein Karimi construct parallel worlds. While these generally bear a strong resemblance to the real world, there are also differences, such as a higher income for a particular person. But because this difference is not real, there is, of course, no experimental data to help answer the "what if" question. Thus, in order to estimate what the parallel world might look like, researchers make assumptions about how the relevant variables that led to the algorithm rejecting the loan relate to each other.

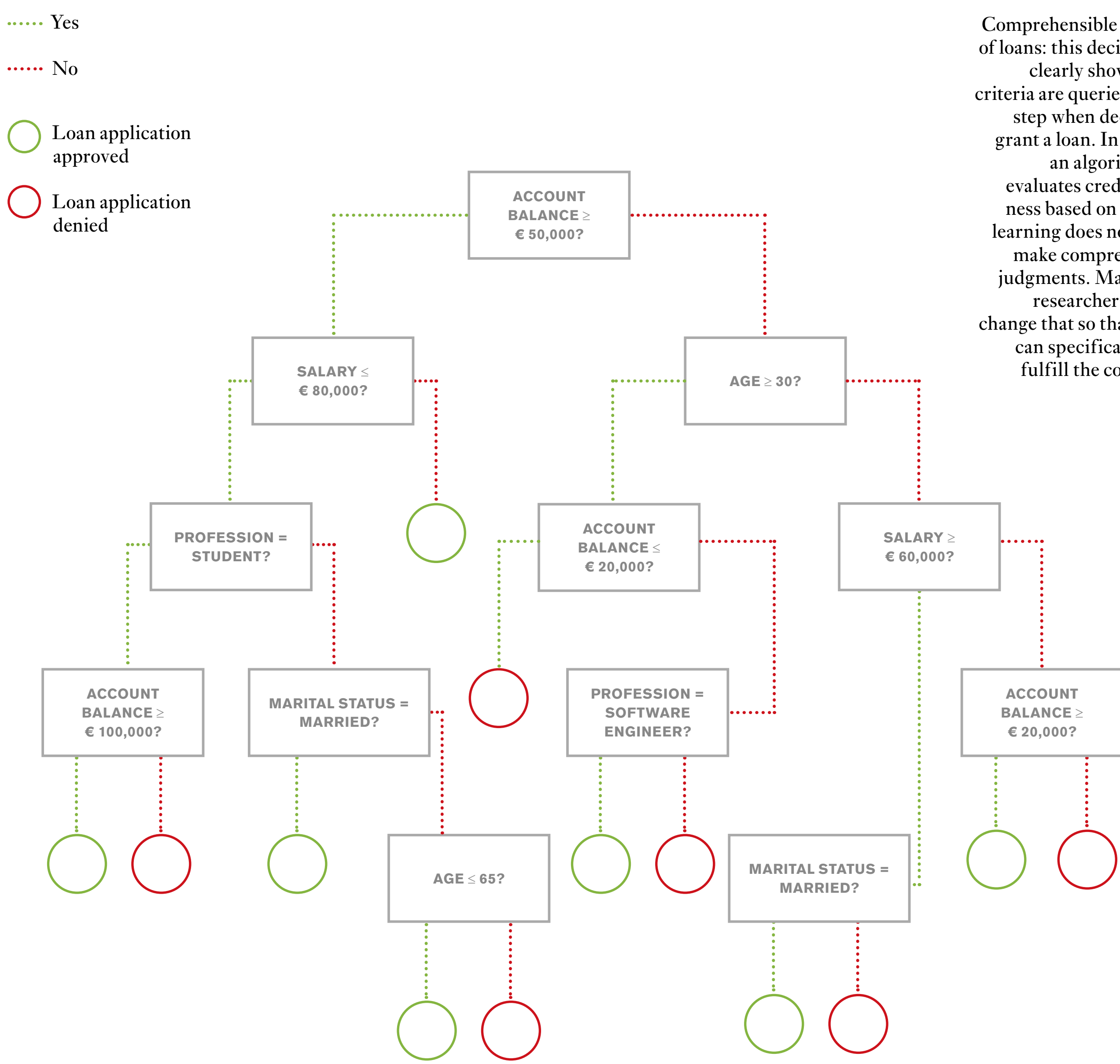
It is usually hard facts such as income, age, and education as well as the amount of the requested loan and its term that decide whether or not a loan should be granted.

SUMMARY

Machine learning algorithms are trained to recognize patterns within large amounts of data, for example in order to infer a disease from physiological parameters. However, they still do not understand cause and effect.

Researchers at the Max Planck Institute for Intelligent Systems want to teach computers to understand causality. Unlike children, algorithms cannot yet learn the corresponding models themselves; they have to be developed by humans for specific applications.

The Max Planck researchers are developing causal models for the comprehensible granting of loans, for the search for exoplanets, or for the climate policy control of an economic system.



Comprehensible granting of loans: this decision tree clearly shows which criteria are queried step by step when deciding to grant a loan. In contrast, an algorithm that evaluates creditworthiness based on machine learning does not always make comprehensible judgments. Max Planck researchers want to change that so that people can specifically try to fulfill the conditions.

GRAPHIC: GCO BASED ON AMIR-HOSSEIN KARIMI/MPI FOR INTELLIGENT SYSTEMS, TUEBINGEN, GERMANY

There are also additional variables that are difficult to measure but which nevertheless influence the decision-making process. Examples of this would be factors such as the cultural background or charisma of the applicant that have crept into the algorithm. These are summarized as statistical noise and must be evaluated. In the causal model thus created, it is then possible to run through the conditions under which a person would be granted the loan. For this purpose, one variable at a time (e.g., income) is changed, while all others remain the same. "After all, everyone should be able to improve their situation by making some personal effort," says von Kügelgen. "But to do that, you first have to know how best to meet certain requirements. And the algorithms used so far can't tell you that."

However, the underlying causal model that describes how the individual factors are interrelated must come from an expert. It won't work without a person who has knowledge of the matter at hand. In the foreseeable future, there seems to be no way around using the expertise of humans to understand causal relationships – even when it comes to machine learning (ML). And our cognitive abilities can naturally serve as a model for the development of intelligent machines. "Our brains have evolved to make us highly cooperative and socially interactive animals," says Martin Butz, who leads the Cognitive Modeling group at the University of Tuebingen. "This requires extremely flexible behavior in different situations." That's why we also have an internal model structure that tells us how things interact

→

with each other in our environment, what the causal principle behind the interrelationships is, and what intentions drive our fellow humans. Humans would not be able to do this if we were purely reactive robots that merely recognize patterns and are driven by the prospect of rewards.

Artificial intelligence systems such as Alpha Zero, a program that taught itself to play chess in 2017, prove that with enough training time and computing power, even a reactive system can learn complex behavior. However, when compared with the real world with all its complexities, mastering the rules of a game with 32 figures on 64 fields is quite feasible. And although Alpha Zero may play better than a human, the program did not understand the game. It is therefore unable to explain to someone how to play chess. "In contrast, our brain is constantly trying to explain relationships," says Butz. Even young children intuitively understand the causal relationships of social situations. For example, developmental psychology experiments show them rushing to help an adult who is carrying a stack of books to open a door. They recognize both the intention of the other to open the door and the cause of the difficulty – the books – which is why the adult has no hands free. In order to react appropriately in such a situation, machines would first have to build their own kind of internal reality that is not simply overwritten whenever they are

For more transparency: Julius von Kügelgen (left) and Amir Hossein Karimi are developing models that make comprehensible lending decisions.

PHOTO: WOLFRAM SCHEIBLE FOR MPG

"Artificial intelligence is increasingly making decisions for humans."

AMIR HOSSEIN KARIMI

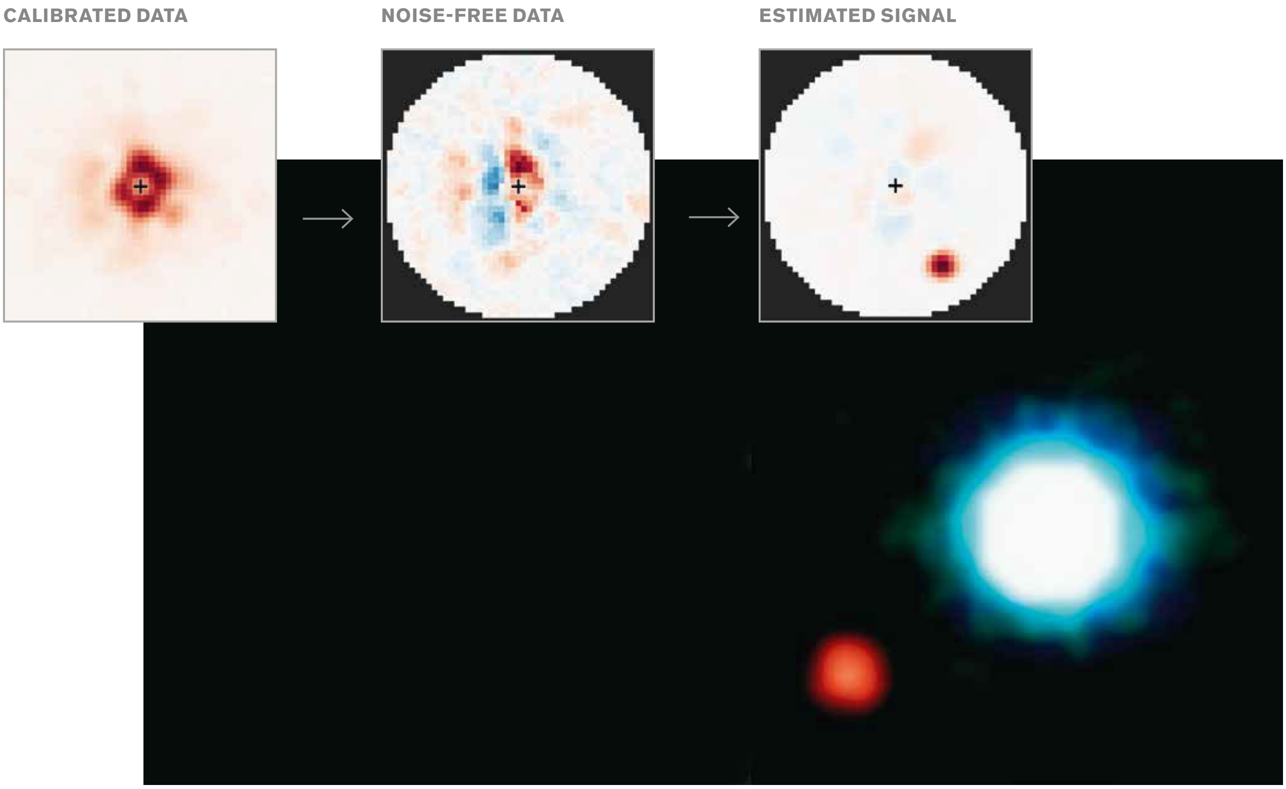
fed new data. Thus, in a way, the machines would have to form a consistent understanding of the world – in the same way that that understanding forms the basis of human perceptions. "We're still a long way from achieving that," says Butz. "On the other hand, there is also no evidence of any barrier that could prevent artificial systems from eventually reaching – or even surpassing – human cognition."

Not only social interaction and other temporal processes are characterized by cause and effect; even purely static representations – for example, photographs – have numerous causal relationships. They are in the mechanisms that make up the image, such as the perspective, the light used, and the distinction between the foreground and the background. Only someone who can recognize the connections and separate them from one another can achieve a robust understanding of the content of an image. For humans, it's easy. We can recognize a mug – even when we view it from an unusual perspective or in poor lighting conditions. And unlike machine-learning algorithms, no one had to show us 10 million pictures of mugs when we were children for us to learn what a mug is.

Similar, albeit much less intuitive, relationships also play a role in astronomy, such as in attempts to photograph exoplanets with large telescopes such as the Very Large Telescope (VLT) in Chile. As a rule, celestial bodies cannot be seen in the images that are created of them. Because the stars shine much brighter, the planets orbiting them are almost completely lost in the noise. "It's somewhat comparable to trying to photograph a firefly on a lighthouse several hundred kilometers away with its spotlight shining directly into the lens," explains Timothy Gebhard. To solve this problem, he is working in the Department of Empirical Inference to create an algorithm that takes advantage of causal relationships to elicit images of exoplanets from the images of the VLT.

In order to collect as much data as possible, astronomers focus their telescopes on a star which they suspect is in the vicinity of a planet. They then record a video over the course of several hours. At first, only a flickering around the star can be seen. This is caused mainly by turbulence in the Earth's atmosphere. Gebhard's task is to combine the thousands of images in the video in such a way that a single image in which the planet can be seen as clearly as possible is created. "Fortunately, in the field of physics, we have a good understanding of the causal relationships of the measurement process," says Gebhard. The telescope's sensor essentially counts individual particles of light. These can come either from the planet itself or from the star. Then comes the noise from the atmosphere and the measurement electronics. In addition, the planet travels in a circular orbit around its star as the video is recorded. Each pixel thus contains photons that can have different causes. "We try to use our causal knowledge of how the data is created to extract details from the recordings that are hidden within the noise," says Gebhard. Here too, the human brain ultimately helps the artificial system to recognize causalities.

"Thinking about cause and effect always ends up being the human's job," says Rüdiger Pryss, professor of medical informatics at the Institute for Clinical Epidemiology and Biometry at the University of Wuerzburg, Germany. In medicine, the goal is often to find patterns in patient data in order to divide patients into groups for which specific therapies can be found. However, if this is done using standard machine



GRAPHIC: ESO; GCO BASED ON TIMOTHY GEBHARD/ MPI FOR INTELLIGENT SYSTEMS, TUEBINGEN, GERMANY

Flashlight in front of a lighthouse: in order to identify an exoplanet next to a star, an algorithm learns to remove its bright light from the data. As a result, previously hidden structures become visible in the data with reduced noise. When the algorithm combines many such images – while also taking into account that the planet is moving in the data because of the Earth's rotation – the signal from the planet clearly emerges.

learning methods, the question of why a patient ended up in a particular group often remains unanswered. The machines cannot explain the reasons for their decisions. Also the medical professional gets too few points of reference to understand them. However, especially in medicine, it is essential to involve humans at the right point in the process. Only humans can make sense of the results of the algorithms and must ultimately be responsible for the therapeutic approaches derived from them. Medical informatics specialist Rüdiger Pryss thus warns against wanting to tackle every problem with ML. "There are extremely powerful statistical methods that have been tried and tested for a long time, and they are often better suited to a specific application system," he explains. Some of these are so clear that they make the cause-and-effect problem less likely to arise. However, because of the current artificial intelligence hype, many users are allowing themselves to be persuaded to put their trust in machine learning.

A healthy skepticism about machine learning as a supposed panacea is therefore certainly in order. Nevertheless, it would obviously be negligent to completely forgo its benefits, especially when it comes to perhaps the most pressing problem of our time: climate change. In order to be able to take causal relationships into account here as well, Michel Besserve of the Department of Empirical Inference is developing an algorithm that can automatically predict the effects of interventions in the global economic system. It should help politicians to find the optimal strategy that minimizes greenhouse gas emissions yet costs as few jobs as possible. "The big challenge here is that our economic system has complex interactions between a large number of actors with differing interests," says Besserve. In the process, each actor adapts their actions to those of the others. Therefore, if we want to describe the economy as a causal model, cycles occur in which individual factors influence each other in a reciprocal manner. "Calculating the resulting equilibrium is much more difficult than if causality always points in one particular direction," says Besserve. The underlying causal model comes from the field of economics and describes the dependencies and interactions of up to 50 different sectors – from power generation and metal processing to goods transport to the meat industry, and rice cultivation. This allows the algorithm to calculate the changes in equilibria that occur when the system is interfered with at a particular point. This would allow policymakers to consider more variables in their future deliberations while avoiding unforeseen and undesirable effects of their interventions on the economy.

"Time is running out, and important decisions must be made now," says Besserve. "That's why we aim to make the new tool available as quickly as possible and thus help develop a sustainable economy." As soon as machines are able to deal with the concept of cause and effect, they could also help solve the major problems facing humanity.

www.mpg.de/podcasts/kuenstliche-intelligenz (in German)

KEY CONCEPTS OF ARTIFICIAL INTELLIGENCE

ALGORITHM
Any kind of calculation rule with which computer programs solve a problem step by step.

ARTIFICIAL INTELLIGENCE (AI)
The name given to a non-natural system such as a computer program that mimics human cognitive abilities. Strong AI is defined as a system that, like humans, is capable of learning independently of a specific task. Conversely, weak AI is an algorithm developed specifically for one task.

MACHINE LEARNING (ML)
This is currently by far the most successful and widely followed AI approach in which an algorithm learns to recognize patterns in unknown data with the help of large volumes of training data. Examples include the identification of faces in images or computer-aided diagnoses based on physiological data.