Vision paper

# Towards Digital Life: A vision of AI in 2032

## The groundbreaking impact of secure data sharing and moral AI

**Peter Werkhoven**

**September 2022**

# World Economic Forum – the Fourth Economic Revolution

**TNO has lived up to its motto 'Innovation for life' by playing a significant part in this digital revolution, with major breakthroughs in computer simulation, network transactions, wireless communication, cyber security, optical satellite communication and quantum computing.**

The digital revolution (also known as the Third Industrial Revolution) has radically changed people's working and social lives – and it has taken place entirely within the 90 years since TNO was founded. With his information theory, American mathematician Claude Shannon (1916-2001) laid the foundations for digitisation, which involved recording electronic information in the form of discrete values, as early as 1948. Digitisation has revolutionised everything about information, as we can now capture it, process it, copy it and distribute it via networks without any loss of quality. The invention of the transistor also led to the development of the first digital computers and digital networks. This had an unprecedented impact on the economy and society.

In our anniversary year we are now in what the World Economic Forum has called the Fourth Industrial Revolution: The groundbreaking impact of secure data sharing and moral AI. The impact of this revolution will go beyond merely changing how we work and live. It will change who we are. Artificial intelligence (AI) is a key technology in this respect, which means we are talking about more than just an industrial revolution. It is actually a system transition in which computer intelligence will evolve from supporting humans to being their (autonomous) partner.

From an evolutionary perspective, human sensory and motor systems and social communication have developed to impressive levels. Cognition, however, is still in the 'embryonic' phase in evolutionary terms and is characterised by a relatively low speed of conscious information processing, by a great deal of 'forgetting' and by numerous biases. Compared to machines, we should be more proud of our ability to tie shoelaces than of our ability to think. Great thinkers such as Leibniz, Von Neumann, Turing and Minsky therefore concluded that additional artificial intelligence could be of help to mankind.

# Over the coming decade the next generation of AI will bring about a key breakthrough by integrating learning AI and reasoning AI in the form of 'hybrid AI'.

Today's AI technology is essentially a brain-derived 'neural' network that learns to recognise patterns (from faces to criminal behaviour) and does so in a 'self-learning' manner based on vast numbers of examples. AI is now often better than humans at spotting connections and patterns in the data we provide and linking the best action to them. For example, AI can successfully recognise cyber attacks and determine the best method of protection, diagnose illnesses and link them to the best treatment, analyse personal purchasing preferences and present the best product, and figure out the playing patterns of a champion of the board game Go and decide on the best countermove, but also discover new antibiotics. Although this form of artificial cognition has huge potential, AI is not (yet) able to reason. It is therefore not capable of explaining its choices, which means that AI cannot be verified or accounted for. AI learning also requires big data, which is often held in different databases that cannot yet be properly combined due to privacy laws – and sometimes due to the business models of data owners themselves.

Over the coming decade the next generation of AI will bring about a key breakthrough by integrating learning AI and reasoning AI in the form of 'hybrid AI': intelligent combinations of algorithms that learn patterns from big data (blind to the 'why') and algorithms that use symbols to reason about cause and effect. Hybrid AI is a prerequisite for socially responsible AI-driven innovation. The ability to reason is essential to be able to explain the choices made by AI and to allow ethical considerations to be taken into account, as well as for social communication and synergetic cooperation within combined AI/human teams. Recently developed secure data-sharing techniques will enable AI to learn from big data while maintaining privacy.

This socially skilled, ethical and verifiable form of hybrid artificial intelligence will allow responsible applications to be introduced in numerous sectors, particularly where artificial intelligence breaks free from a fixed 'workplace' and moves through our physical and digital worlds more autonomously in the form of robots and avatars. Robots that will transport us, take care of us, defend us, entertain us, or be companions. For the first time in its evolution, human society will then face the question of how to relate to these new forms of intelligence that it has created itself. This will involve striking a balance between fear and desire. On the one hand, the desire for what AI can do for us to improve traffic, transport, healthcare, safety and the sustainability of industry and the energy system. And, on the other, the fear that AI will disrupt our society and democracy if applied irresponsibly.

We are therefore also faced with major legal, ethical and social issues. How do we maintain 'meaningful control' over the goals of AI and the ethical achievement of these goals? How, as a society, can we develop a legitimate ethical framework for such intelligent systems? How do we make this comprehensible for machines? What role should the producer, user and government play here? Meaningful human control over AI will remain the starting point, with humans continuing to exercise judgement in determining legitimate uses of this technology. In many situations this can be achieved by keeping humans 'in the loop'. For essential, high-risk AI applications in which human cognition is inadequate due to time constraints or human/AI interactionde is limited, 'human before the loop' control is being developed. This 'human before the loop' control involves humans equipping machines with an adequate, legitimate moral code in advance.

TNO predicts that AI will become 'big' in the next ten years. Big in the sense of societal and economic impact, but also big in the sense of 'maturing' as far as responsible and moral behaviour is concerned. AI will thus develop into a 'trusted partner' of humans.

Moreover, AI will itself become an innovator in the field of research and innovation. For at least the next decade AI will 'obediently' follow the goals and working methods we impose on it. AI will therefore be able, in particular, to significantly speed up the analysis and validation phases of research and innovation and highlight new implications of existing knowledge. You could also refer to this as the 'perspiration phase' of the research and innovation process. The researcher's role will therefore shift to the 'inspiration phase', the creative phase focusing on asking the right questions, developing entirely new explanatory models and introducing completely different

## The researcher's role will therefore shift to the 'inspiration phase', the creative phase focusing on asking the right questions, developing entirely new explanatory models.

methods – in other words, deviating from the 'existing rules'. Having this freedom to be 'disobedient' is precisely what distinguishes humans fundamentally from AI.

At the same time, hostile AI will develop worldwide, in the same way that innovative nanotechnology, biotechnology and information technology intended to serve society can also be misused in a way we cannot control as a weapon against people. By analogy with the arms race, if you want to protect yourself against hostile AI, you need to have adequate AI of your own. The price of keeping the Netherlands secure is ongoing creativity and innovation in the field of AI.

Chapter 1

# Educating AI

## AI and human evolution

In humans, motor systems, perception and decision-making have evolved over time to reach impressive levels. We are supported in these processes by all kinds of tools: the axe and the wheel are an extension of our motor systems, the microscope and telescope build on our sensory perception, and the abacus, computer and AI improve our decision-making and actions in a world where information has become increasingly important.

Our brains evolved primarily to allow exceptionally fast processing of sensory information and to control our complex motor systems. From an evolutionary perspective, you could say that human cognition is still in its 'embryonic' phase. We have survived by thinking quickly and in the short term, rather than logically and with the long term in mind. The conscious processing of information by the human brain therefore does not exceed 60 bits per second (del Prado Martin, 2009). In addition, when processing information, we humans have only a limited attention span and our memory regularly lets us down.

Our thinking exhibits more than 200 experimentally proven biases (Korteling 2021). Some of these are inherent in our neural properties, such as believing in non-causal links, framing (not being able to forget) or judging based on what is perceivable at the time. Other biases have arisen as an evolutionary advantage from living (or surviving) in groups, such as favouring one's own group, parroting others and giving preference to action even though it may not be rationally necessary. All these biases have a negative impact on our cognitive abilities and cannot be unlearned, or only to a very limited extent.

It is therefore not surprising that additional artificial intelligence could be of help to mankind, provided that this intelligence is less fallible than our own. In 1951 Alan Turing said: 'It seems probable that once the machine thinking method had

## Our society therefore faces the unique challenge of making clear to AI systems the goals that we want to pursue as humans, and the ethical values that should underpin the choices made by AI.

started, it would not take long to outstrip our feeble powers', with 'not long' proving to be a rather elastic concept.

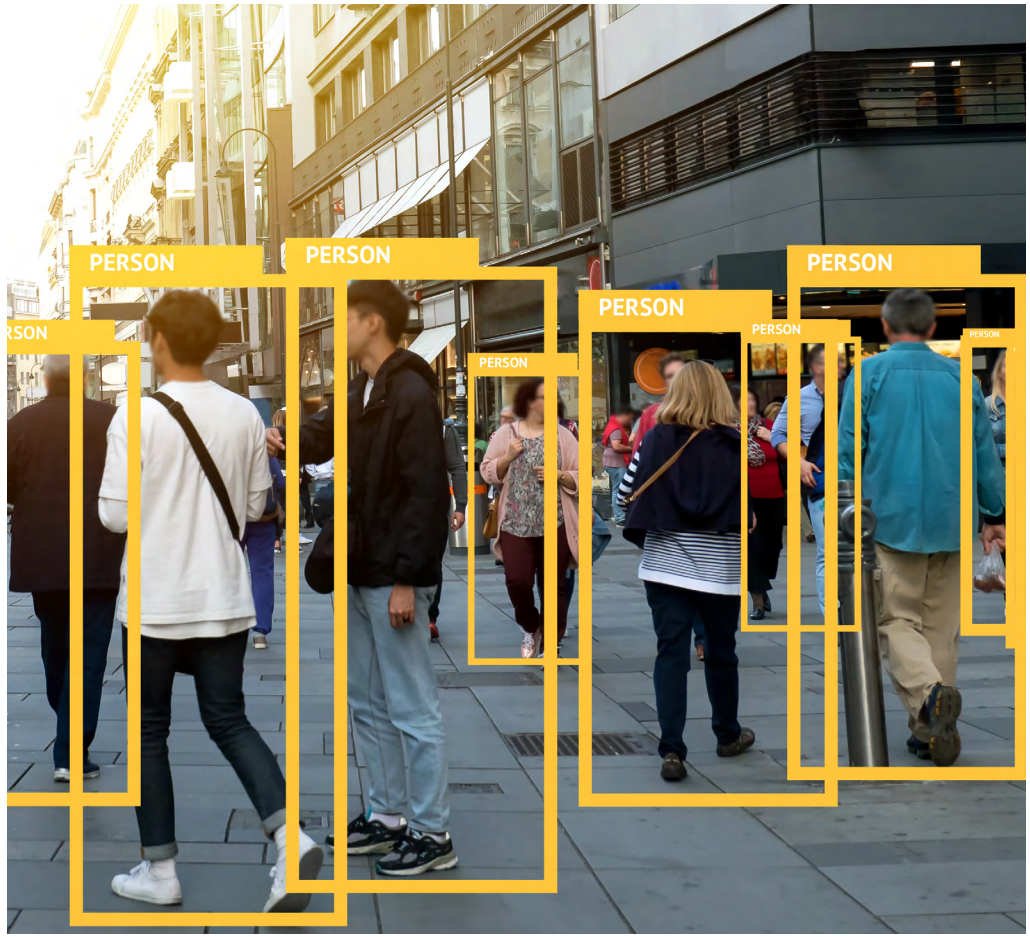
### AI in our society

Humans have managed to create artificially intelligent, data-driven machines that can make complex connections in large quantities of data and can be used in all kinds of decision-making processes. This presents both opportunities and threats for society. AI may be able to solve problems that are too complex for humans to handle. However, when AI pursues goals other than those society deems desirable or pursues those goals in an immoral way, a threat arises. Only with 'responsible' AI will important applications become socially acceptable and possible.

Our society therefore faces the unique challenge of making clear to AI systems the goals that we want to pursue as humans, the ethical values that should underpin the choices made by AI and the role that is reserved for AI systems. Artificially intelligent systems will always function in service of and in interaction with society. This means AI will have to be educated to understand human goals and values, to be socially competent and cooperate with humans, and to offer explanations and justifications in forms of communication that are natural to humans, such as language. Only then will AI and humans form a 'symbiotic' partnership in which they can learn from each other and develop further. A situation in which non-inclusive access to AI creates a divide within society must be avoided.

### Hybrid AI and learning from little data

Current AI technology (such as deep learning) is powerful when it comes to spotting connections based on very large quantities of data, ideally pre-sorted by humans (big data). However, AI is not yet capable of real understanding (such as distinguishing between cause and effect) and is a long way from being able to explain or justify its analyses in a clear and communicative way in all cases. Moreover, the availability of lots of sample data is not always guaranteed. To achieve greater maturity in this area, additional reasoning skills are needed (hybrid AI), but also the ability – just like humans – to learn from a limited amount of data (small data) with a deeper understanding of the semantic



# At present, AI lacks a memory function that could serve as a basis for human-AI team learning.

structure of that data. Only with hybrid AI can human goals and values ultimately be understood by AI and realised in collaboration with humans. Developing hybrid AI with a deeper understanding of data will also lead to safer forms of AI: artificial intelligence that is less susceptible to manipulation by malicious parties.

## Explanation and trust

On the one hand, humans want AI to be able to explain the knowledge it has acquired and the analyses it has made; on the other, we want to use AI to come up with solutions that humans may no longer be able to understand (or AI itself, for that matter).

AI models use the data we supply to build internal abstractions that are often incomprehensible to us. Just as we describe faces in terms of eyes, noses, mouths and ears, AI builds up abstract patterns of the face that allow better distinctions to be made, but have no meaning to a human. In the relationship between humans and AI, a balance will develop over the coming years, in each area of application, between the desired degree of explainability and the quality of the solutions produced by AI – just as a patient trusts a medical specialist when it comes to the effect of complex medical treatments that can only be explained to a limited extent. This also creates scope for new forms of explainability in the areas of cooperation and communication that are not necessarily focused on the operation of AI itself: forms of explanation that are context-dependent and range from conveying knowledge to people through to creating confidence in demonstrably good solutions; explanations that are limited to terms ('You are eligible for this drug because you meet the following medical profile') and contrasts ('If you earned more than €3,000, you would not be able to receive this benefit') that people can understand.

## Cooperation and communication

As mentioned above, living and working in groups is an evolutionarily important factor in human survival. Social skills are therefore an essential part of our maturing process. In fact, this also applies to cooperation between AI and humans in teams. Over the coming years, AI will 'mature' further in terms of social skills, such as sharing goals, giving feedback, explaining and communicating with humans in natural language.

At present, AI lacks a memory function that could serve as a basis for human-AI team learning. Inspired by our own cognitive functions, current research is focusing on the development of a semantic and episodic memory that can infer meaningful relationships between entities and identify corresponding meaningful patterns over time. This could serve as a basis to allow humans and AI to share experiences and provide mutual feedback for team learning. Implementing such a memory function will enable the development of applications in which humans and AI complement each other physically and intellectually, achieving a performance that is greater than the sum of their parts. It is possible to distinguish between the following levels of human-AI cooperation:

- Mutual interaction: Accurate analysis of human feedback during human-AI communication and other forms of cooperation
- Mutual explanation: Adaptive and inclusive communication and cooperation with humans taking human feedback into account and using the new dynamic forms of memory for learning in cooperation with human teams

# Making communicative AI systems sensitive to individual characteristics and context of humans is going to contribute substantially to inclusiveness.

- Symbiosis: Evolution into a fully-fledged, adaptive, communicative and reliable cooperation partner for humans

Of these, interaction and, to some extent, explanation will be achieved by 2030 and will enable applications in which humans and robots work together, such as in healthcare and security operations.

To realise the communicative functions of AI, for some time we have had access to very powerful AI-based language models for generating human-like language and speech. However, these models have not yet been tailored to the characteristics of the human interlocutor. They do not have a clear picture of who this person is: what knowledge, values and goals does he or she have, what are his or her knowledge needs and is he or she satisfied with the information offered? Making communicative AI systems sensitive to individual characteristics and context of humans is going to contribute substantially to inclusiveness: making relevant information available to people with varying characteristics and backgrounds in a communicative and comprehensible way.

**Morality**

To a large extent, a functioning society is formed by the set of behaviours that are considered desirable and correct in a social context (morality). If we are to accept AI in our society, it is therefore very important to educate AI in moral behaviour. However, morality varies from culture to culture, from person to person and even through time. The challenge for our society, on the one hand, is thus to specify our ethical principles and values for AI in legitimate terms (as far as possible). On the other hand, AI must be able to (technically) make choices while respecting those ethical values and therefore be capable of ethical reasoning. This too will have to become part of the hybrid AI that will be developed over the coming years. These developments will bring applications within reach in which humans and AI work together in teams, such as in doctor-AI teams in the medical sector and human-robot teams in manufacturing or safety and security operations. The potential effectiveness of human-AI cooperation is

# Games of chess are won more often by human-computer combinations than by humans or computers on their own.

demonstrated by the fact that games of chess are won more often by human-computer combinations than by humans or computers on their own.

### Autonomous AI

Above we have described various forms of interaction between humans and AI in which humans and AI were in the same loop. This means that goals and situational information are shared at all times and each could take over the task from the other. Under certain conditions, however, interaction between AI and humans is impractical or even undesirable, for example if there is not enough time for a human to understand and control a situation (e.g. in the case of a conflict situation involving a self-driving vehicle or a rapid cyber attack) or if human-AI communication is limited or even impossible (e.g. during rescue operations by robots in buildings or underwater environments). Such tasks can be performed as responsibly as possible using AI if the role of the human shifts from 'in the loop' to 'before the loop' and 'after the loop'.

The idea here is that we educate AI beforehand and have it explain and account for itself afterwards. In such a case AI must be capable of achieving a goal on the basis of a legitimate moral model that allows choices to be optimised in accordance with moral values. To give a highly simplified example, a self-driving car would have a moral model in which not just 'travel time', but also 'harm to people and the environment' and 'emissions', for example, are taken into account to determine direction and speed. Aspects such as intention and vulnerability will also be considered alongside these factors. The legislature, the executive, AI producers and AI itself each play distinct roles in the implementation of a moral model.

This moral model (laws and ethical values) is determined by the legislature. The executive decides on the operational goals, while the autonomous AI system optimises the pursuit of the stated goals on the basis of the moral model and provides an explanation of this to the executive and judiciary. The role of the AI producer is limited to making it technically possible to work with moral models and therefore does not involve determining the ethical values. Such moral models also play an important role on

# The idea is that we educate AI beforehand and have it explain and account for itself afterwards.

another level: in interactive communication and cooperation AI will have to adhere to certain ethical and legal norms and values, for example concerning manners and the quality and completeness of information.

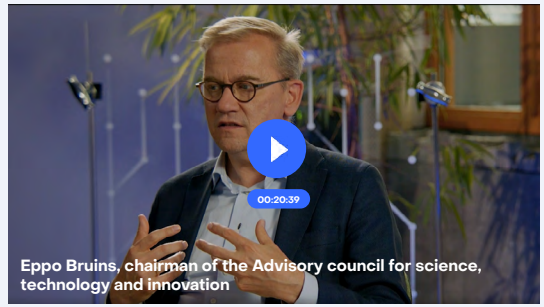
In addition to the interactive applications described above, this will also make more autonomous applications of AI possible in which humans are both before and after the loop, but not always in the loop. AI will thus eventually be able to drive us from A to B independently and responsibly or help us repel a cyber attack.

# “You create, what you shouldn’t want, second-class citizens.”

## Arnon Grunberg

Michiel van der Meulen, chief geologist for the Geological Survey of the Netherlands (GDN), speaks with **Eppo Bruins**. Bruins was educated as a nuclear physicist and has spent many years working in the world of science, innovation, and technology. Between 2015 and 2021, he was a Dutch member of parliament for the Christian Union. He was recently appointed chairman of the Advisory council for science, technology and innovation (AWTI). What will AI mean for the various government domains in the coming years?

**Watch full conversation**



Eppo Bruins, chairman of the Advisory council for science, technology and innovation

Peter Werkhoven, chief scientific officer at TNO, talks to **Arnon Grunberg** from his base in New York. Grunberg made his breakthrough in 1994 with his novel, Blue Mondays. He has since become one of the Netherlands’ best-known authors. The two talked about AI over dinner some years ago. Today, they finally get the chance to continue their conversation. What is Grunberg’s view on creativity? Can it be taught to machines? And how do humans morally relate to machines?

**Watch full conversation**

Arnon Grunberg, Dutch author

Chapter 2

# Innovation with AI: Developments over the next ten years


As AI matures and people are able (and willing) to share data securely, the impact on the economy and society will increase. AI will be a key driver of innovations within manufacturing, transport and logistics, healthcare, security and public services, for example. Across sectors, AI will accelerate the transition to circular materials and renewable energy. An analysis by PwC (Sizing the prize, 2017) predicts that AI adoption will increase global GDP by 14% by 2030 (10% for Europe, 15% for North America and 26% for China). The Netherlands AI Coalition anticipates potential growth of 1.2% per year for the Netherlands. According to the Economist Intelligence Unit (Intelligent Economies, 2018), the greatest transformation within the business world can be seen in the areas of manufacturing, retail, healthcare, financial services and the public sector.

The great strength of current AI systems is their ability to find patterns in large quantities of data and predict which actions will have the most effect. This leads to sector-wide applications, such as determining an action appropriate to a profile (e.g. security, insurance, retail), determining a treatment appropriate to a diagnosis (healthcare) or determining a route appropriate to a situation (logistics, gaming). This category of applications is in fact supportive decision-making, where the

There are also AI applications in which humans and AI cannot interact quickly enough due to time constraints (e.g. the need for rapid choices by autonomous vehicles in the event of traffic conflicts), due to cognitive limitations of humans (e.g. defence against large-scale cyber attacks) or because it is unsafe for humans to be present (e.g. during rescue operations or robot inspections in dangerous or unhealthy environments). Such tasks can be performed as responsibly as possible using AI if the role of the human shifts from in/on the loop to before and/or after the loop. We refer to this category as autonomous decision-making.

# AI has an essential role to play in the development of safe, low-personnel transport, in the realisation of 'mobility-as-a-service' and in policymaking.

Based on TNO's wide-ranging experience of AI applications in various sectors, over the coming decade we expect to see AI applications in the following sectors.

### Digital society

Supportive and autonomous AI will play an important role in the further development of the internet into a network of 3D virtual worlds for our commercial, educational and social interactions (Social XR). The development of the metaverse is one example of this. Interactions in these 3D digital worlds will be influenced to a large extent by AI, as the appearance and language of people in these worlds can be adapted, as appropriate, by AI. Fully autonomous AI can also take the form of an avatar and interact with people in these virtual worlds. As a result, the number of journeys made to physically interact with people, and the associated environmental impact, will decrease further.

### Mobility and logistics

AI has an essential role to play in the development of safe, low-personnel transport (autonomous vehicles), in the realisation of 'mobility-as-a-service' (personalised matching of mobility supply and demand) and in policymaking (for example, predicting the effects of new forms of mobility).

Autonomous vehicles: Over the next decade great strides will be made in the use of autonomous systems in the mobility sector. Self-driving cars, in which people travel as passengers only (instead of also driving the vehicle), are currently undergoing extensive testing. However, this technology is not yet being used on a mass scale. As a result of various AI innovations, relating to the reliable assessment of the traffic situation, for example, this will change dramatically over the coming years. Initially, it will be introduced on routes that are reasonably predictable, such as motorways and provincial roads outside built-up areas. In addition to passenger cars, greater use of autonomous trucks is also anticipated, which will allow goods to be transported more efficiently and with fewer emissions. Only after the aforementioned development of hybrid AI and 'human before the loop' methods (machine-implementable moral models) will it also be possible to use self-driving cars in more unpredictable traffic situations.

Mobility-as-a-Service (MaaS): MaaS combines various transport options into a single service. Using one app, for example, travellers can get from A to B in accordance with their own preferences. They can take advantage of a wide range of options, such as public transport, bicycles, shared cars and new transport concepts, on a single multimodal platform that brings together all providers. This will lead to greater transport efficiency and reliability, greater convenience and ultimately – thanks to a decline in the use of personal transport (car ownership) – a reduction in emissions and in the number of parking spaces in the Netherlands. Data and AI play an important role here: for example, the various modes of transport are coordinated and user data (which modes of transport are chosen where and when?) leads to adjustments in urban planning and policy. However, there are also major hurdles to overcome in the course of the next decade, as who will 'own' this data and the MaaS platform itself? Without government intervention there is a significant risk that US or Chinese companies – with their own political and commercial interests – will take over the European market. Government intervention will also focus on formulating European values for AI applications, with which non-European AI will also have to comply.

Policymaking: AI can be used to predict the effects of new forms of mobility in different situations. Think of urban or rural travel, long or short distances, and travel with or without children, for instance. One example of a new form of mobility is the introduction of the shared scooter. This clearly offers people a new way of getting around, but it also creates inconvenience from scooters being left on the pavement. AI-driven simulations (including behavioural prediction) can be used to gain more insight into the positive and negative effects of various transport options. This allows the government to make better policies, for example by adapting the infrastructure, encouraging people to make more sustainable choices and allowing or banning certain transport options. AI-driven simulations thus lead to a better and more sustainable living environment.

# AI is being used to make the economy more circular, including for smart waste disposal and recycling of concrete and plastic.

**Construction and infrastructure**
AI will be increasingly employed in the design, construction, management and maintenance of buildings and civil infrastructure. During the lifecycle of an object 'digital twins' are used: virtual models incorporating up-to-date information about the properties, condition and performance of buildings or infrastructure. With the help of AI and digital twins, the quality, structural safety and energy efficiency of buildings can be tested even before they are actually built. The use of digital twins and AI can also optimise the management of buildings and energy demand and supply by adapting them to user needs and influencing user behaviour. When it comes to the maintenance and renovation of civil infrastructure, such as bridges, AI and digital twins will be used increasingly to make predictions about the scope and timing of the work, minimising the risks of maintenance and renovation being carried out too early or too late. For citizens this means fewer traffic jams and less disruption caused by road, bridge or tunnel works.

**Sustainable society**
Energy transition: AI contributes to current and future innovations in sustainability and the circular economy. The smart meters currently installed in households are being further developed to analyse the energy people are consuming in their homes and make suggestions for improvements. Over the next decade these energy management systems will be supplemented by smart batteries to optimise energy storage and relieve the burden on the energy grid. The introduction of a 'smart energy grid' will also allow supply, conversion and storage to be matched optimally to demand.

Circular materials: AI is being used to make the economy more circular, including for smart waste disposal and recycling of concrete and plastic. AI will also play an important role in the industrial transition towards circular and emission-free production by tracking scarce or polluting materials through the product lifecycle and across value chains.

**Health**
Vaccine and drug development: AI will be used for large, time-consuming tasks, such as performing smarter and faster searches in large datasets. It has already been used

# Thanks to secure data-sharing techniques, this can be done while respecting privacy and complying with the law.

in this way, for example, for the genome sequencing of the COVID-19 virus,[1] a technique that analyses the precise DNA structure of the virus to make an effective vaccine against it. A few years ago this process would have taken weeks or even months, but now, with AI, it takes no more than a few days. This has allowed the rapid development of vaccines that protect against variants of the virus. AI can also make drug development much faster and more specific. For example, an algorithm can help develop the molecule in such a way that the drug binds to specific locations in the body and therefore targets exactly the right spot. In addition, AI could help predict the effects of polypharmacy (taking several drugs at the same time), which are still difficult to measure in clinical trials.

Prevention: In the health sector, a shift from care to prevention is possible and necessary, especially when it comes to lifestyle-related diseases. Here it is essential to tailor behavioural interventions to individual lifestyles. AI is a powerful technology that allows this personalised assessment to be based on combinations of individual medical data (heredity, microbiome, immune system) and lifestyle-related data (ideally data on total exposures from diet, inhalation, and mental and physical stress, also known as the exposome). Thanks to secure data-sharing techniques, this can be done while respecting privacy and complying with the law. Early signs of unfavourable developments can therefore be used to support doctors and/or digital coaches in the guidance they provide with the aim of maintaining health.

Personalised medicine: AI will help bring together complex knowledge of interactions between symptoms and treatments to allow the most effective treatment to be identified. In ten years' time, the big dispenser containing various pills, where one drug is used to suppress the side effects of others, should be a thing of the past.

Care process: Social interaction is crucial for care. The Netherlands has a large ageing population, which means there are more patients and fewer healthcare workers. It has already been proven that labour-intensive tasks that are repetitive, precise or physically demanding could be taken over perfectly well by AI, such as in the area of surgery. This would allow healthcare workers to create more space for social interaction and improve care as a whole. AI could also support healthcare workers with the social side of their work, for example by allowing more frequent interaction through care robots or predicting which method of interacting will work best at a particular moment. Carebots will help humans report on the care provided, by automatically storing conversations between the healthcare worker and patient in the patient's medical record and making them instantly machine-readable.

Trust: Trust is incredibly important when it comes to the use and reuse of health data. By embracing decentralisation and using analysis methods that keep data at the source, it is possible to link data while maintaining autonomy over the individual data sources. This is made possible by secure data-sharing technology, standardisation, transparent management and supervision, and cooperation with citizens and patients.

## Defence

Defence is an excellent example of an area in which the application of AI will have a major

[1] https://www.sciencedirect.com/science/article/pii/S1876034122000144

## The application of AI by the Ministry of Defence will also help to make the work more 'labour-extensive', in anticipation of the impact of the tightening labour market.

impact. Soldiers must be able to build a detailed picture of their surroundings ('situation awareness') to allow them to take the right action, but they are facing a significant increase in the quantity and variety of data. They have to carry out risky and physically demanding tasks that would be better left to autonomous systems, in situations where this is possible. Autonomous systems can also be used to improve protection (against incoming missiles, for example) and to increase the effectiveness of weapons systems, within the applicable ethical and legal frameworks. Generally speaking, the application of AI by the Ministry of Defence will also help to make the work more 'labour-extensive', in anticipation of the impact of the tightening labour market.

Military decision-making. The decision-making process is often described by breaking it down into three phases: 'Observe', 'Orient' and 'Decide' (which, together with the 'Act' phase, make up the so-called OODA loop). Al can be applied in all these phases: data can be collected more intelligently and from more sources, and combined and processed in such a way that 'situation awareness' is increased. In addition, decision-making can be supported through the automatic generation of possible courses of action, including an assessment of the pros and cons of the various options. There is actually already a need for AI, as we are witnessing huge growth in the quantity and diversity of information, and because adversaries (who are also using AI themselves) are giving us less and less time to make decisions. Eventually, fully autonomous decision-making (in which the human is therefore before and/or after the loop) will be necessary in certain situations, although this will only be implemented under strict conditions.

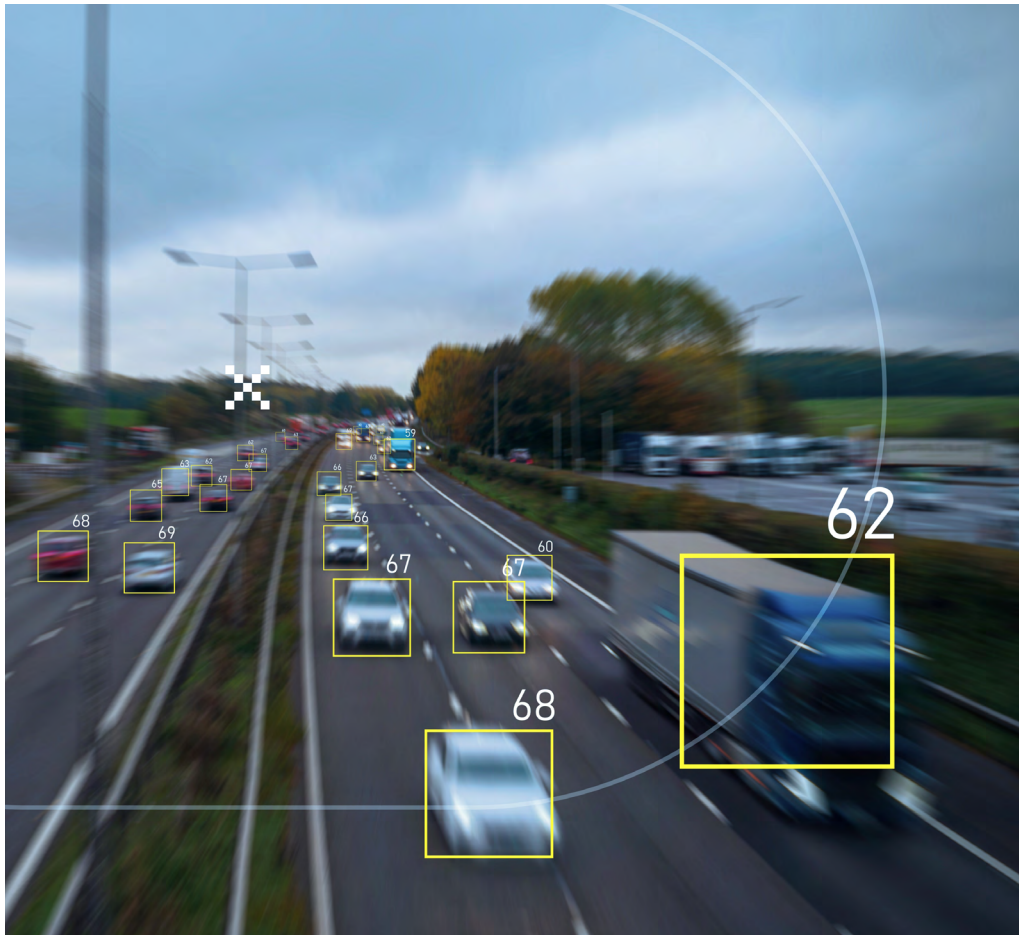
Robots for bomb disposal: Detecting and disarming explosives and mines on land and in water is an extremely dangerous task that should be taken over, wherever possible, by (semi-)autonomous systems. Robots are already being used for bomb disposal on land, but they are entirely remote-controlled and therefore highly labour-intensive. The next step in this development is 'supervisor' mode: the robot is given a command such as 'go there and use this procedure to dispose of the explosive'. Within a decade we expect to bring humans 'before the loop': the robot will be sent out with the task of 'disposing of the explosives' just like a human colleague, deciding autonomously how to execute this task and taking safety considerations into account. Humans are also increasingly being taken out of the 'loop' in the area of sea mine disposal. Autonomous surface and underwater vessels are being developed for sweeping, detecting and clearing tasks relating to various types of mine. Underwater systems in particular present technological challenges, as navigation is difficult and possibilities for communication are limited. Nevertheless, within a few years there are expected to be systems that will be capable of detecting and disarming mines largely autonomously.

Surveillance and battle drones: Today, we are seeing more and more drones being used on the battlefield. One important objective is reconnaissance: the collection of information. At present, this is very much a human-driven process, but over the coming years it will become increasingly autonomous. This will allow drones to be used more often and for longer, making military decision-making more effective. Armed drones are also being used more and more, not only those such as the remotely controlled 'Reaper' or the 'Bayraktar', which has been deployed in Ukraine, but also the so-called 'loitering munition drone', which can search for its target partly autonomously. It is crucial to ensure that these systems operate within the applicable ethical and legal frameworks. As stated in a recent report by the Advisory Council on International Affairs and the Advisory Committee on Public International Law[2], this is not possible if such systems are fully autonomous. That means it must always be possible to exercise meaningful human control over the system, with 'humans before the loop' being the most far-reaching option.

Transport and logistics: 'Amateurs talk tactics but professionals talk logistics' is a well-known statement that reflects the often underestimated importance of military logistics. As the war in Ukraine demonstrates, the ability to continue fighting (referred to in military circles as 'sustainability') largely relies on efficient logistics chains. Using autonomous systems for this task is a logical step, not only because considerable experience has already been

[2] https://www.adviesraadinternationalevraagstukken.nl/documenten/publicaties/2021/12/03/autonome-wapensystemen

# AI can be trained to detect anomalies that suggest potential criminal and/or terrorist activity.




acquired in using autonomous transport for civilian purposes, but also because military convoys are important targets and therefore extremely dangerous for the soldiers involved. Autonomous transport could also provide a solution for another high-risk transport activity: recovering wounded soldiers from the front line.

### Civil security

Security: Over the next decade AI will also play an important role in the area of civil security, for example for inspection and surveillance tasks, with autonomous systems increasingly being used to secure business parks, critical infrastructure or festivals.

Crime: AI is increasingly being used by law enforcement agencies to find anomalies in large quantities of data, the so-called 'needle in a haystack'. AI can be trained to detect anomalies that suggest potential criminal and/or terrorist activity without the need to examine citizens' privacy-sensitive information (secure data sharing). This way, crimes can be predicted without the government being able to turn into 'Big Brother'.

Investigation and collection of evidence: Another application is the AI-driven identification of causal relationships in textual material. This can be used for investigation and the collection of evidence, but also in the judicial system, for example to analyse case law. These innovations will make the work of the police and judiciary more efficient and effective.

### Industry

AI is widely used in tasks ranging from workforce planning through to product design, to optimise efficiency, product quality and worker safety. According to the Economist Intelligence Unit (2018), the main use cases of AI are being accelerated by AI-driven R&D (30%), predictive analytics (28%), real-time operations management (26%) and robotisation of manufacturing processes.

Maintenance and quality control: In factories AI is used to predict failures of critical industrial equipment, allowing preventive maintenance to be carried out and downtime to be avoided. In the area of quality control AI is used to warn of possible production errors (subtle deviations

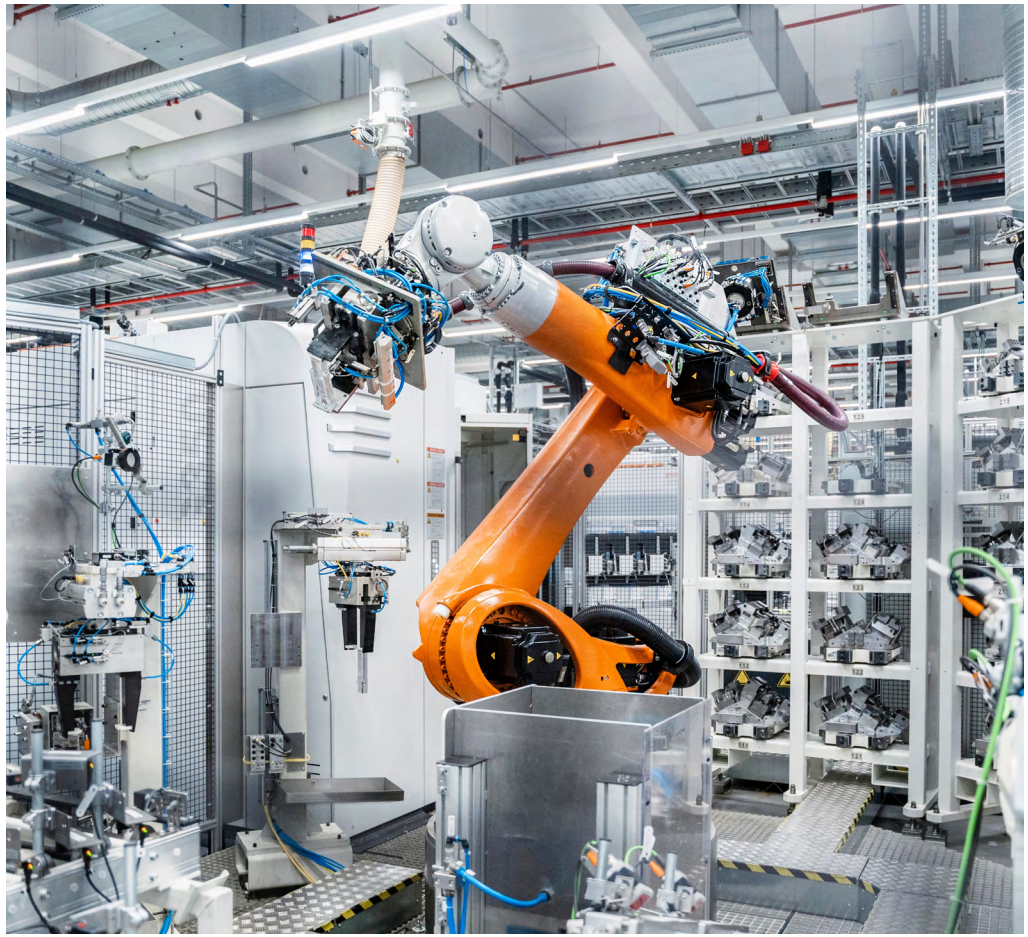
# Industrial robots are becoming an integral part of the production process and AI is ushering in the development of collaborative robots.

in processes, machine behaviour, raw material characteristics, etc.) in order to optimise product quality. By equipping production facilities with sensor systems (to detect deviations in materials, the production environment and machine behaviour) and applying AI to them, it is possible to make production processes adaptive and resilient.

Product design: Product design involves an increasing number of dimensions (in addition to performance, costs and safety, aspects such as sustainability, material availability, influence on supply chains, etc. are also being taken into account). AI can determine the effects of design choices along these dimensions and make rapid pre-selection possible. The designer's role therefore shifts from time-consuming 'trial and error' validation to the creative aspects only. AI is also used to allow the data-driven customisation of design based on specific customer characteristics.

Robotisation: Industrial robots are becoming an integral part of the production process and AI is ushering in the development of collaborative robots that can take instructions from humans and work safely and productively alongside and with them.

Circularity and emissions: Digitisation – and AI in particular – is a key factor in the transition to sustainable and circular industry. Thanks to digital tracking and tracing of materials, (worldwide) supply chains can be optimised with circular material use and emissions reduction in mind.



Hanneke Molema, senior consultant healthy living at TNO, interviews **Georgette Fijneman**, CEO of health insurer Zilveren Kruis since 2017. Both look at the same topic, health, from a completely different perspective. What is the promise of AI for one of the Netherlands' largest health insurers?

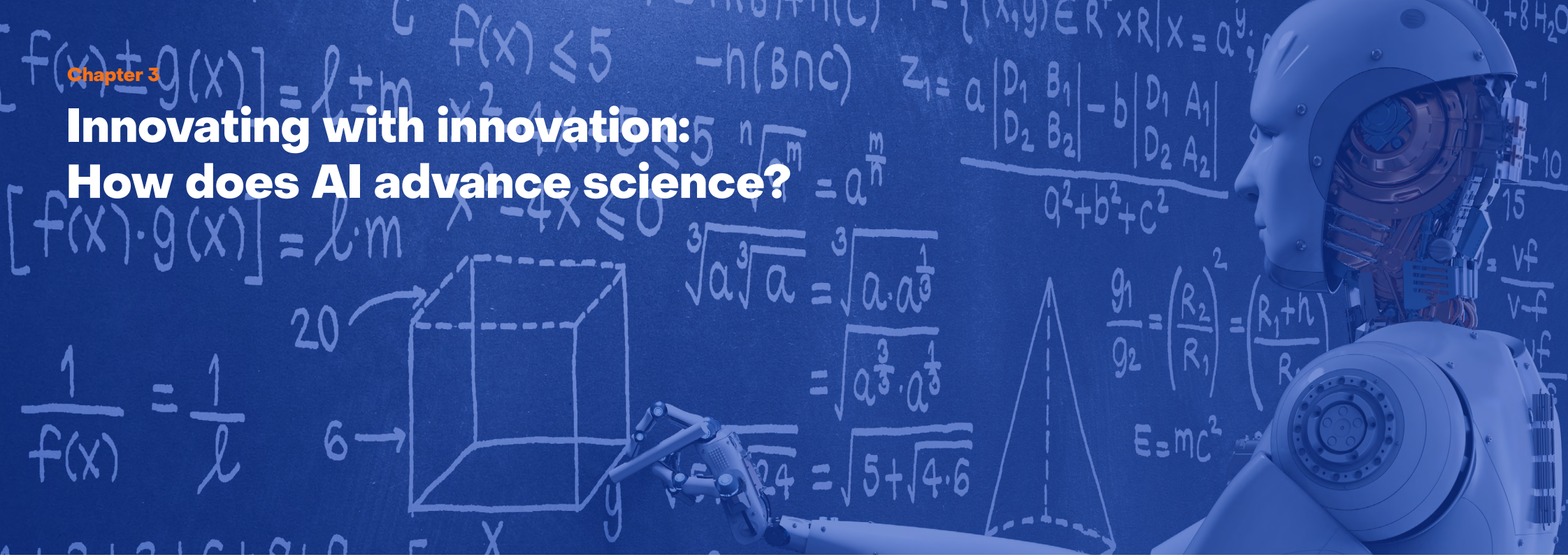
Watch full conversation


Anne Fleur van Veenstra, director of science at TNO's SA&P unit, interviews **Rob de Wijk**, emeritus professor of international relations in Leiden and founder of The Hague Centre for Strategic Studies. Rob is also a much sought-after expert who appears on radio and television programmes. What does the rise of AI mean geopolitically and in armed conflicts?

Watch full conversation


Marieke Martens, science director at TNO and professor of automated vehicles at the Eindhoven University of Technology, talks to **Bram Schot**. Schot was the CEO of Audi until 2020, having previously held management positions at various car makers, including Mercedes and Volkswagen. Their conversation concerns the influence of AI on mobility. How will AI impact the production process? And what does a future with autonomous vehicles look like?

Watch full conversation



Chapter 3

# Innovating with innovation: How does AI advance science?

Aristotle (384-322 BC) once imagined that: '... if every instrument could accomplish its own work, obeying or anticipating the will of others, like the statues of Daedalus, or the tripods of Hephaestus [...], if, in like manner, the shuttle would weave and the plectrum touch the lyre without a hand to guide them, chief workmen would not want servants, nor masters slaves'.[3] In a world where most of the technologies we make use of today did not yet exist, Aristotle was already well aware of the potential of technology and automation. We have seen Aristotle's vision come true in many areas. What will this mean for artificial intelligence in relation to science? Will machines soon be able to learn faster and better than people, to the same extent that racing cars can outstrip us? Will they be able to produce knowledge? And, if so, is that a good or a bad thing? What impact will AI have on equality between people?

## Overcoming limitations

Our earliest and most basic tools made us 'stronger', enabling people to accomplish physical tasks beyond the limits of their physical strength. Think of tasks such as hunting big game, covering ever greater distances, constructing ever larger buildings, moving ever heavier loads, growing crops in ever greater quantities and extracting resources from ever deeper underground. Then there is a whole category of tools that have actually made us 'louder', that have allowed people to communicate with others far beyond the range of the human voice, such as the pen and paper, printing, telegraphy, telephony and, finally, the internet. Here we are not just talking about the ability to communicate as such, but about the opportunity to share valuable ideas, organise ourselves and our efforts and build modern societies. We have also developed tools that make us 'smarter'. The abacus, the slide rule, the calculator and, finally, the computer have enabled people to perform arithmetic and other analytical tasks beyond our cognitive capabilities. We now use this kind of tool for a vast range of applications, from the cash register through

[3] Politics, Boek 1, Sectie 1253b

# Where technology and creativity meet, controversy easily arises.

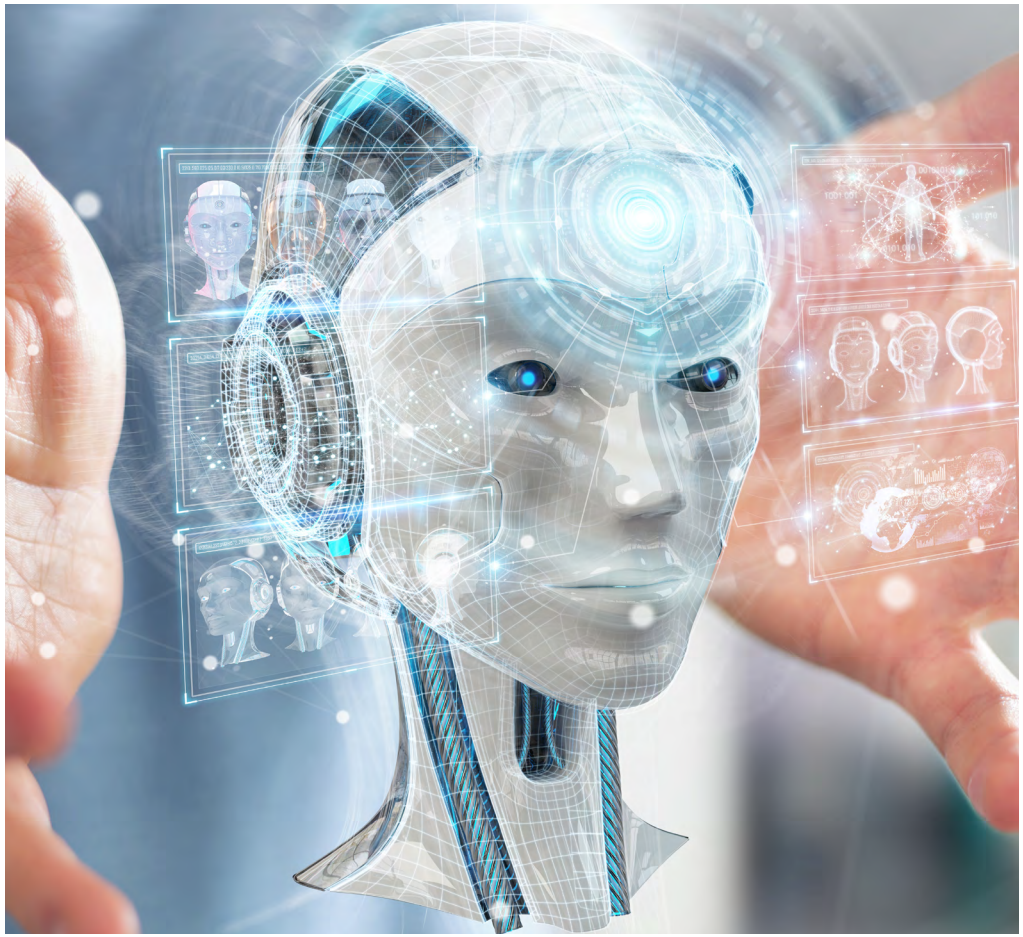
to advanced data science. Technology makes us more powerful, better connected and allows us to be the best we can be.

### An extension of human creativity?

Technology has so far enhanced all human attributes except our creativity. While computers, instruments and laboratories do, of course, support and unleash the creativity that drives science and art, the ability to acquire and develop knowledge is seen above all as an exclusively human trait. For many people the very idea of learning machines feels like being beaten at your own game. Consequently, where technology and creativity meet, controversy easily arises. For artistic reasons – but economic ones too, of course – painters protested against the new technology of photography, musicians went on strike when self-playing pianos and radios made their appearance in hotels and restaurants, and chefs and gourmands today look down on ready meals.

In such cases, technology was embraced by people whose autonomy and wealth it increased, but despised by those people on whom the former previously depended. Traditional craftsmen saw their work taken over by factories. Clerks, human computers and errand boys found themselves in the same situation due to office automation. With the advent of AI, which takes a considerable amount of brainwork off our hands, new occupations are now being affected – not just in the creative and science sectors, which are based around thinking and creativity, but also in the legal system and administration, where decisions are made. As with previous industrial and digital revolutions, we need to decide how to relate to the technology we develop and subsequently put into practice.

In this connection, what we think about the relationship between humans and machines becomes clearest when people fall victim to technology: then it is not the machine that is blamed, but the people who operated, maintained or designed it, or their managers. This says something fundamental about how we as humans view knowledge and the decisions we make based on it. Knowledge appears to be transactional. It is developed by people and shared between people. While machines may play a role, people remain responsible for results and applications. On the one hand, artificial knowledge is a triumph of computer science and human ingenuity in general. It is often



# From idealised models to a 'noisy' world.

presented as an almost magical solution to all sorts of problems, even by people who do not understand the technology itself. On the other hand, it makes us fearful, which prevents us from fully exploiting the potential of artificial intelligence.

### The potential of brute force

Machines can process and analyse more data in a more comprehensive and consistent way than humans, provided we teach them how, supervise them and do not take the results for granted. If we can manage to do this (and we can!), then the potential is enormous. Just like a hammer gives momentum to our arm, AI gives momentum to our mind. We can use it, for example, to carry out the laborious scientific work known as replication research. We can make machines crunch datasets that are too large for humans to handle. Machines could help us research how citizen science contributes to ornithology, ecology, palaeontology and astronomy. Another example is the re-analysis of old datasets in new contexts, such as oil and gas data being given a second life as part of the energy transition.

### From idealised models to a 'noisy' world

Until now, we have had to capture the world in the form of idealised models in order to express the fundamental laws of nature mathematically. Artificial intelligence is proving powerful enough to deal with the 'noise' of the real world. Japanese astronomers used deep learning to clean up data from the Subaru telescope's giant digital camera, making it possible to deduce the true shape of galaxies from images distorted by lensing.[4] Using data-driven techniques, Silva et al. not only succeeded in reproducing the law of gravitation, but also identified secondary mechanisms that come into play when you drop a real object rather than a perfect sphere in a vacuum.

The ability to handle the real world makes artificial intelligence a tool ideally suited to applied science, which is never about predicting the relative motion of the centres of gravity of two perfect spheres of different sizes, but about modelling the path of a tennis ball on planet Earth, for example. At TNO we are working, for instance, on models to make or keep people healthier. In healthcare in particular there is a great deal of mistrust towards AI, because people want to have control over why a certain treatment is proposed, but this limits the creativity of AI. What if we could really give AI the space to optimise our health? What function would we want to focus on? Often that would be a particular parameter or set of parameters that we know to be linked to a disease. This would require us, as human beings, to think differently about our bodies and our environment and ask different questions of the models. Would AI then be able to find where the cause of diseases lies or would the models be able to predict treatments that cure even chronic diseases?

### An attitude towards technology

We have had to learn to relate to all the technology that we now consider indispensable and take for granted. In the future we will need to learn about, mitigate and manage the negative effects of artificial intelligence. You could say that the technologies that have made us stronger collectively have had the opposite effect on individuals, or have simply had a negative impact. In many cases craftsmanship has been reduced to rather mind-numbing interactions with factory equipment. Communication technology is as successful at spreading disinformation and hate as it is essential to the functioning and organising of societies. Computers, which are ubiquitous in the modern world, have undoubtedly had a negative impact on individual numeracy and analytical skills. In the fields of politics and economics there are diametrically opposed views on the ownership of technology as a means of production, resulting in revolutions and wars. Whether intentionally or otherwise, technology can be harmful. Despite all this, we agree that technology makes us more powerful, for better or worse. On balance, technology has brought us prosperity.

Genius is 1% inspiration and 99% perspiration. What sets humans apart from AI in particular are intuition, unexpected flashes of inspiration and the 'disobedience' to think about and do things differently. We can create more scope for this typically human creativity in research and innovation by supporting the perspiration phase with AI, in particular by increasing the speed and scale of data analysis and helping to identify hidden patterns, which in turn will enable us to form new hypotheses. In this way artificial intelligence will change not only our view

[4] Shirasaki, M., Moriwaki, K., Oogi, T., Yoshida, N., Ikeda, S. & Nishimichi, T., 2021. Noise reduction for weak lensing mass mapping: an application of generative adversarial networks to Subaru Hyper Suprime-Cam first-year data. Monthly Notices of the Royal Astronomical Society 504(2): 1825–1839. doi:10.1093/mnras/stab982.

## What sets humans apart from AI in particular are intuition, unexpected flashes of inspiration and the 'disobedience' to think about and do things differently.

of the world around us, but also our view of ourselves. Through the way we are trying to get machines to learn we are gaining an increasing understanding of how human creativity works and how we do science, as well as of how we construct and get increasingly better at constructing moral and ethical frameworks.

Technology has made products and services that were once the preserve of the very wealthiest available for the masses. This process started with basic necessities and then moved on to things that today we can no longer imagine used to be luxuries, in the areas of communication and mobility for example. In a similar way, AI is transforming high-level thinking into a mass-market product. That may sound contradictory and at first may seem less desirable. However, more, better and faster thinking leads to better decision-making, about health, safety, living and working in our ever-changing environment. Seen in this light, AI is about improving everyone's quality of life. That's what we do it for!

# “AI has to be obedient, it has to do what it is programmed to do. Whereas a human is fundamentally disobedient.”

## David Deutsch

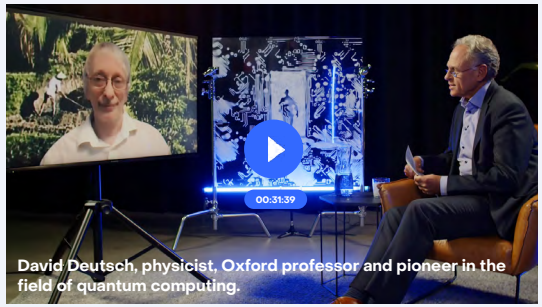
Michiel van der Meulen, chief geologist for the Geological Survey of the Netherlands (GDN), speaks with **Bas Haring**. Haring originally studied artificial intelligence, which at the time still fell under the umbrella of philosophy, which is why people started calling him a philosopher. He himself feels more like a ‘folk philosopher’: Haring tries to make science and philosophy accessible to a wider audience. In 2001, he published a children’s book about evolution, Cheese and the Theory of Evolution. What better springboard for a geologist and a philosopher to talk about AI?

**Watch full conversation**

Bas Haring, ‘Folk Philosopher’

Peter Werkhoven, chief scientific officer at TNO, joins physicist, Oxford professor, and pioneer in the field of quantum computing, **David Deutsch**, for a virtual discussion. Deutsch set out his vision in 1997 in the book, The Fabric of Reality. Together, they talk about the significance of quantum computing for the development and application of AI. Will AI ever be able to generate ‘explained knowledge’ or learn about ethics from humans?

**Watch full conversation**



David Deutsch, physicist, Oxford professor and pioneer in the field of quantum computing.

TNO innovation for life

**Courtesy of:**
Eppo Bruins
David Deutsch
Georgette Fijneman
Arnon Grunberg
Bas Haring
Bram Schot
Rob de Wijk

**Author:**
Peter Werkhoven

**Co-authors:**
Jildau Bouwman
Jurriaan van Diggelen
Judith Dijk
Michiel van der Meulen
Mark Neerincx
Stephan Raaijmakers
Anne Fleur van Veenstra

**www.tno.nl**