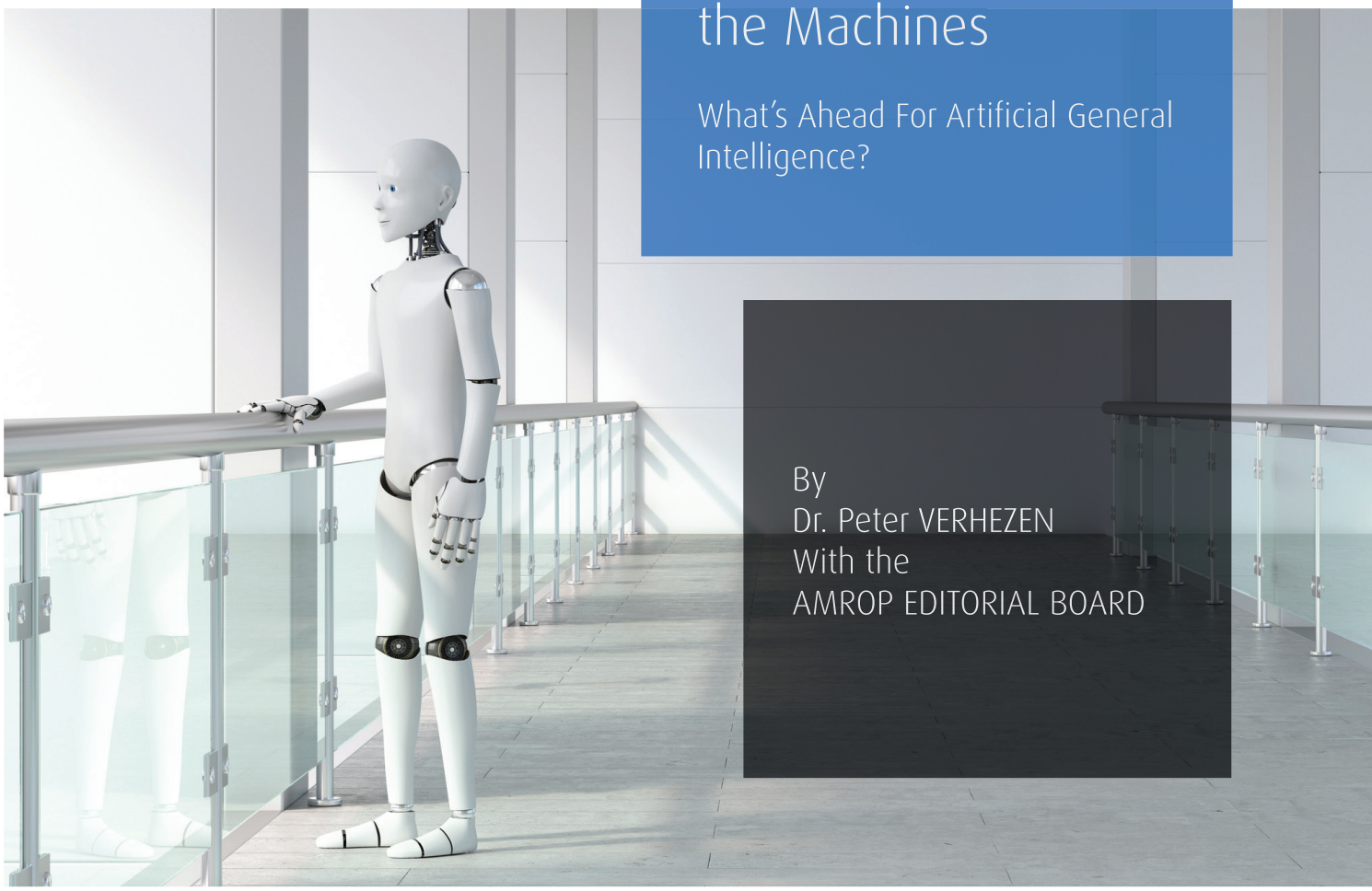
Leadership

Wise Leadership and AI

Chapter 3 | Behind the Scenes of the Machines

What's Ahead For Artificial General Intelligence?



By
Dr. Peter VERHEZEN
With the
AMROP EDITORIAL BOARD

Amrop
Leaders For What's Next

Putting the G in AI | 8 points

True, generalized intelligence will be achieved when computers can do or learn anything that a human can. At the highest level, this will mean that computers aren't just able to process the 'what,' but understand the 'why' behind data — context, and cause and effect relationships. Even someday chieving consciousness. All of this will demand ethical and emotional intelligence.

1 We underestimate ourselves

The human brain is amazingly general compared to any digital device yet developed. It processes *bottom-up* and *top-down* information, whereas AI (still) only works bottom-up, based on what it 'sees', working on specific, narrowly defined tasks. So, unlike humans, AI is not yet situationally aware, nuanced, or multi-dimensional.

2 When can we expect AGI? Great minds do not think alike

Some eminent thinkers (and tech entrepreneurs) see true AGI as only a decade or two away. Others see it as science fiction — AI will more likely serve to amplify human intelligence, just as mechanical machines have amplified physical strength.

3 AGI means moving from homo sapiens to homo deus

Reaching AGI has been described by the futurist Ray Kurzweil as 'singularity'. At this point, humans should progress to the 'trans-human' stage: cyber-humans (electronically enhanced) or neuro-augmented (bio-genetically enhanced).

4 The real risk with AGI is not malice, but unguided brilliance

A super-intelligent machine will be fantastically good at meeting its goals. Without a moral compass, AGI will be like a loose projectile on steroids.

5 AI has to learn how to learn

AI applies *supervised learning*, and needs a lot of data to do so. Humans learn in a '*self-supervised way*'. We observe the world, and figure out how it works. We need less data, because we are able to understand facts and interpret them using metaphors. We can transfer our abilities from one brain path to another. And these are skills which AI will need if it is to progress to AGI.

6 AI has to understand cause and effect

When they have access to large data sets, today's neural networks or deep learning machines are super-powerful detectors of correlations and conditional robabilities. But they still can't understand *causality* – the relationship between one thing and another. This ability to establish and understand causal models to grasp complex reality remains a human skill. Another one which AI will need to acquire, if it's to reach AGI..

7 It'll need another giant step to get from causal thinking, to consciousness

Exactly how neurons interact, and which parts of the brain are responsible for human consciousness, remain unknowns. So how AGI will achieve consciousness is a big question. It'll need to get to grips not just with causality but with *counterfactuals* (how a causal relationship would change, given the introduction of some other condition into the equation).

8 The wise application of A(G)I will need a moral compass

Human abilities still lie beyond the outer rim of AI. As mentioned, they involve self-supervised learning, and understanding causality. Most fundamentally, they involve framing and answering ethical questions, feeling empathy and compassion. These abilities will be critical to ensuring that AI — and AGI - are applied wisely, in a way that is ethical, responsible and sustainable.

2 Major AI Approaches

Augmentation

Assisting humans with daily tasks

Automation

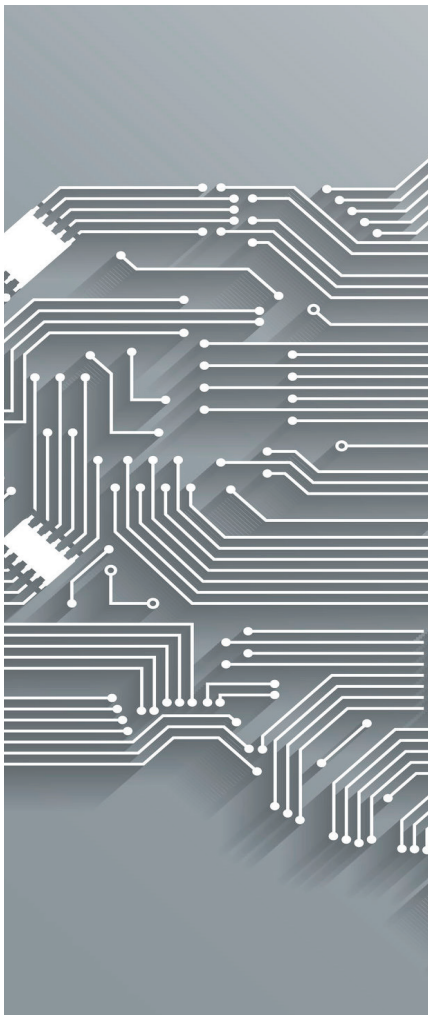
Working autonomously in a specific field

Virtual assistants
Data analytics
Software solutions
Reducing error/bias

Robots - key process steps in manufacturing plants
Trivial, repetitive tasks in the supply chain

Behind the Scenes of the Machines
What lies ahead for Artificial General Intelligence?

Neural networks, inspired by our brain's architecture, are scoring spectacular successes in gaming and pattern recognition. If you're an iPhone X or advanced Android user, you'll already find it pretty normal that the device can actually identify your face. This is just one example of 'principal component analysis' – using data to classify people, rather than human preconceptions. And thanks to machine learning, (self- or deep learning), computers are getting smarter by the day. Is AI set to out-smart us? What could the road from smart Artificial Intelligence to Artificial General Intelligence look like? Will it lead to wise decisions?



The first question is: *what is true, general intelligence and how does it differ from what machines can currently do?* In this article we'll argue that this goes beyond the pattern recognition, number-trading and probability calculation performed by today's AI. For example, true intelligence is about expressing abstract knowledge in symbolic forms — including the classic domains of computer science: math, programming language, logic. Beyond handling a specific task, true intelligence is about performing multiple, complex tasks, using common sense. Beyond supervised learning, it's about autonomous, flexible learning. Beyond high-capacity statistical modelling, it's about understanding the relationships between things — cause and effect (causality). And most fundamentally, it's about more than the '*what*'. It's about the '*why*', informed by ethical and emotional intelligence.

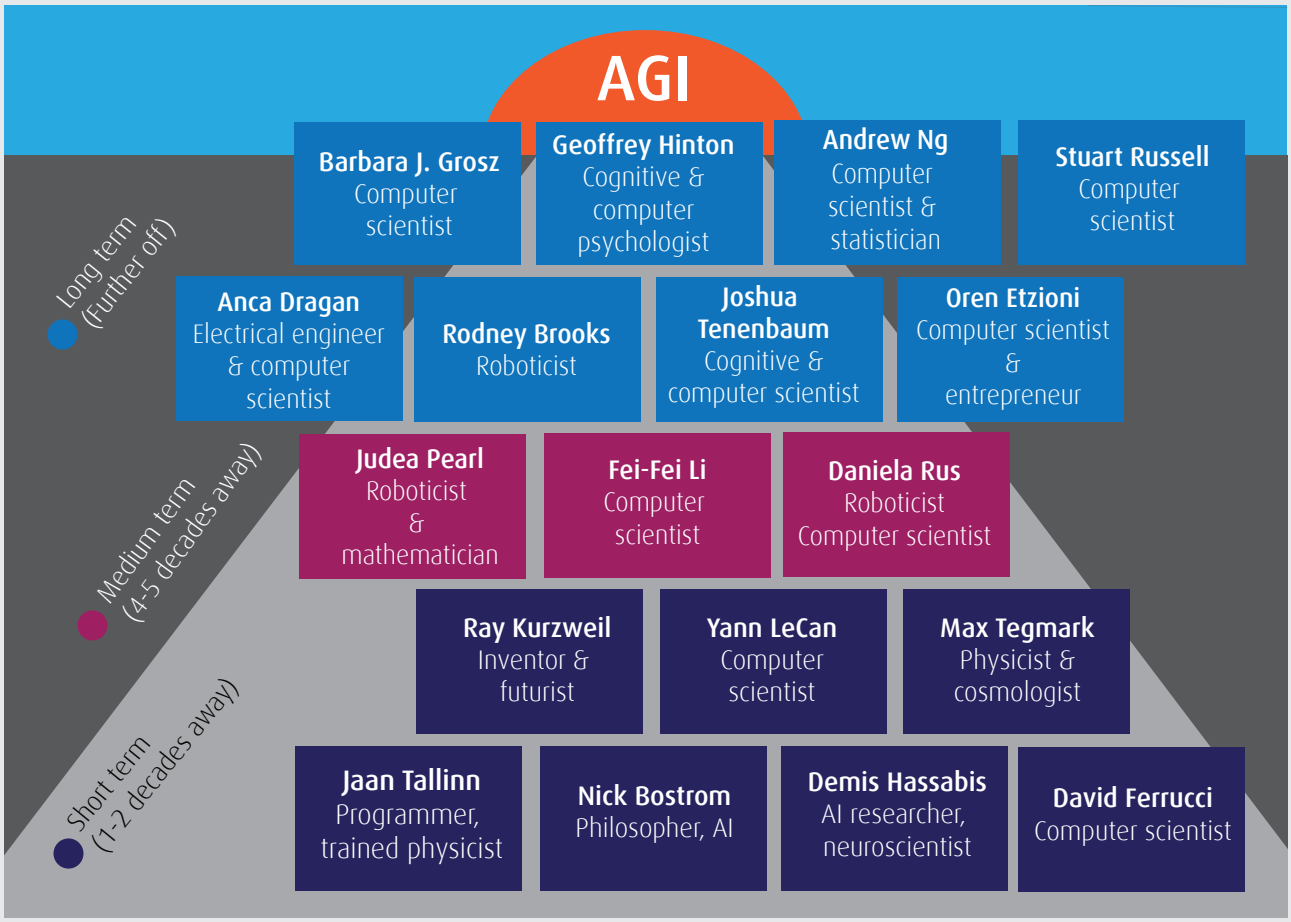
Putting the 'G' into 'AI'

True AGI would mean computers getting to grips with most of the above. Some thinkers are very optimistic about the timeframe: AGI is only about a decade or two away, they say. Others see it much further down the road *(see next page)*. Whatever the timing, it's hard to predict whether AGI, if and when it hits, will enrich our lives, make us subservient, or cancel us out. And progress is moving fast. Tesla uses neural networks in its car vision system to warn about potential collisions. Google Translate has made exponential advances. It took AI less than three years to find solutions to beat a human chess champion (our brains lack the processing power to think that many moves ahead). DeepMind's Alpha Zero (snapped up by Google) works by playing hundreds of millions of games, cutting out loss-making mistakes and elaborating on winning strategies. This, by the way, involves *'generative adversarial network techniques'* — in normal language, techniques that generate and observe data.

Neural networks like these are good at identifying objects in a picture, putting together sounds to make words, recognizing what to do in a game. But AI is not, yet, beyond those tasks. Even IBM's Watson, as we saw in our last article, remains highly specialized. It's still far off the power and versatility of the (100 billion) neurons and synapses of an average human brain. AI cannot yet make a plan, as a conscious being can.

When might we expect AGI?

The scientific community has different opinions...



Meanwhile, star entrepreneurs like Elon Musk (Tesla), Sergey Brin and Larry Page (Alphabet-Google), Jef Bezos (Amazon), or Mark Zuckerberg (Facebook), believe that we are only a decade or two away from some form of general artificial intelligence.

Time to wake up...

At their core, AI machine-learning algorithms are surprisingly simple. Some people even see them as 'dumb, fast machines on steroids', based on 'stupid little neurons'. They perform descriptive statistics using brute computing force: "you can get those little pieces to approximate whatever arbitrary function one wants", MIT social physicist Sandy Pentland, has said.

No wonder AGI still smacks of science fiction for many skeptical scientists, including Venki Ramakrishnan, the 2009 Nobel Laureate for Chemistry. AI and deep learning machines cannot answer yet answer the *why* question, lacking consciousness or self-awareness. What's more, we don't even understand what exactly consciousness[1] is. How we remember a phone number, or suddenly lose memory. We don't know exactly how neurons interact. Or which parts of the brain — if the brain at all – are responsible for human consciousness. We tend to underestimate the brain's complexity and creativity, how amazingly general it is, compared to any digital device we have developed so far.

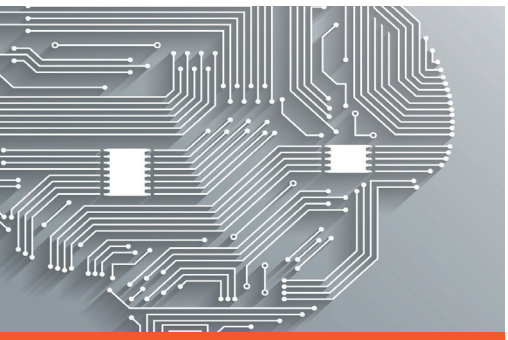
Today's AI (deep learning) is only getting the bottom-up information, rather like the *human occipital cortex* (our visual processing center). Gary Marcus, a top AI researcher from New York University, argues that deep learning systems don't capture what the *human frontal cortex* does when it reasons about what's really going on. For AI and robotics entrepreneur and scientist Rodney Brooks, any AI program in the world today is an "idiot savant living in a sea of now". Some pioneers, such as Judea Pearl, believe we'll see a 'moral causal thinking robot' in the distant future, but only if a form of AGI is able to process both bottom-up and top-down information, as humans do. Moreover, AGI would need to read between the lines as humans do. In other words, interpreting and understanding specific contexts, making invisible, abstract interpretations. These often involve an understanding of cause and effect relationships (more of that later).

Even when it comes to physical dexterity, there is a long road ahead. Robots are very good at performing sets of pre-programmed, restricted motions, precision welding on assembly lines for example, or calculating ideal distances in self-driving cars through a GPS system. But it's still proving challenging to build a robot that can perform *multiple* tasks fluidly and fast — tasks that come naturally to humans. Stacking shelves, tying shoelaces, or pouring a drink, to name just three examples. Despite what we might intuitively think, high level reasoning requires very little computation, compared to low-level sensorimotor skills. The paradox of AI-driven robotics progress is known as *'Moravec's paradox.'*

We don't even understand what exactly consciousness[1] is. How we remember a phone number, or suddenly lose memory. We don't know exactly how neurons interact. Or which parts of the brain – if the brain at all – are responsible for human consciousness.

We tend to underestimate the brain's complexity and creativity, how amazingly general it is, compared to any digital device we have developed so far.

[1]The neuroscientist Giulio Tononi has argued that it's possible to "quantify" consciousness, (denoted by the Greek letter 'Phi') by measuring the extent to which different parts of a system know each other. This consciousness theory - known as *'integrated information theory (IIT)'* logically postulates that computers cannot have a real consciousness. It has been challenged by philosopher and cognitive scientist David Chalmers and cognitive robotics expert Murray Shanahan.



The real risk with AGI is not malice, but brilliance. A super-intelligent machine will be supremely good at meeting goals. As long as those goals are aligned with ours, no problem. If not, we can expect big trouble. Hence the importance of bringing the 'ought' – or ethical dimension – into the equation, as well as safety engineering, without delay.

Superintelligence by 2100 AD?

In 2005, Max Kurzweil coined the idea of 'singularity' — the point at which computers become as intelligent as humans. When he declared that this would happen by 2035 (based on exponential improvements over the last 30 years), Pandora's box was opened. Max Tegmark at MIT cites a differing view: according to some experts, AI systems have an over 50 per cent probability of reaching AGI by 2045, and a 90 per cent probability by 2075. From reaching human ability — Max Kurzweil's singularity – AI has a 75 per cent probability of moving on to *superintelligence* by 2100. However, these are informed guesses at best.

At singularity, then, (if and when it arrives) computers will be at least as powerful as human intelligence. This means that humans may need to evolve apace. We can imagine a 'trans-human' stage — cyber-humans (electronically enhanced) or 'neuro-augmented' humans (bio-genetically enhanced). A *homo deus* in other words, if we're to compete with AGI powered machines.

AGI - friend or foe?

Some experts see singularity as an opportunity, others emphasize its dangers. Still, most discussions focus on a narrow and weak interpretation of AI (machines controlled by humans) rather than its potentially dramatic transformations. For Oxford philosopher Nick Bostrom, a superintelligence may see humans as a threat. As MIT Professor Brynjolfsson put it, any future depends on our choices: "we can reap unprecedented bounty and freedom, or greater disaster than humanity has seen before." We may see an ever more demanding struggle against the limitations of our brain and intelligence.

The real risk with AGI is perhaps not malice, but brilliance. A super-intelligent machine will be fantastically good at meeting its goals. As long as those goals are aligned with ours, no problem. If not, we can expect big trouble. This is why it's so important to bring the 'ought' – or ethical dimension – into the AI equation, as well as safety engineering, and to do this without delay. As a number of AI specialists admit, postponing this integration until after AGI arrives would be irresponsible, and potentially disastrous. A super-AI machine lacking a moral compass would be like an unguarded projectile on steroids. Let's face it, we wouldn't send humans to the moon without every possible precaution.

So far, it looks as if the upcoming AI systems are set to amplify human intelligence, just as mechanical machines have amplified physical strength.

Interviewing the machine

How could you design an intelligence test for AI? The most famous example is Alan Turing's concept: a computer passes the Turing test, or 'imitation game' if it can fool a human during a Q&A session into believing that it is a fellow human. And that's a long way off. Whilst AI algorithms excel at specific tasks, humans still far outperform them when it comes to creatively connecting the dots in different frameworks.

According to philosopher Daniel Dennett[2], Turing couldn't foresee the uncanny ability of superfast computers to sift through the inexhaustible supply of internet data to find probabilistic patterns in human activity. These could be used to output responses that seem authentic enough to outwit a human – without necessarily being more intelligent than one. Dennett describes this multi-dimensional 'computer-agent' as more like an *amygdala* or *cerebellum*[3] than a fully-functioning mind. And it would not be "remotely up to the task for framing purposes and plans and building insightfully on its conversational experience."

Meanwhile, AI researcher Andrew Ng doesn't see a lot of progress in AGI, beyond faster computing. Fed only with a small dataset, deep learning — the most-used technique for AI applications today – isn't that helpful. Machines need to get much better at learning from small datasets to achieve any form of limited general purpose intelligence. As we've seen, even this still needs to be clearly distinguished from consciousness. And that's another topic altogether.

So far, then, it looks as if the upcoming AI systems are set to amplify human intelligence, just as mechanical machines have amplified physical strength. Current machine learning systems operate almost exclusively in a statistical or 'model-blind' mode. As we've seen in previous articles, they remain rather opaque and focus on *what-if* rules to execute a specific task.

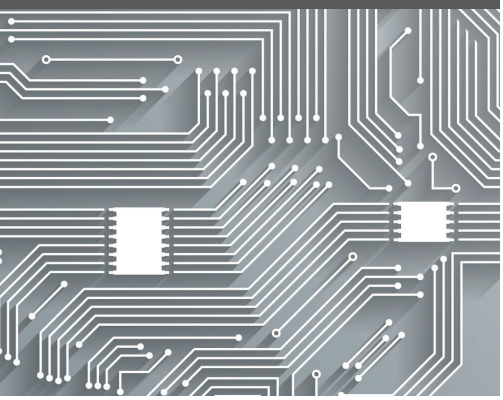
Some way to go in two fields – learning, and causality

For Oxford quantum physicist David Deutsch, who conceptualized the notion of quantum computing (and the late philosopher Karl Popper), human-level intelligence and thinking lie in the zone of creative criticism, interleaved with creative guesswork that allow humans to learn one another's behaviors, including language, and extract meaning from one another's utterances. It is this general creativity that leads to innovation — a truly human characteristic.

AI is going to need better techniques in two key fields. The first is in autonomous learning.

[2]To a certain extent he joins the eminent Oxford mathematician and physicist, Roger Penrose.
[3] The *amygdala* is primarily associated with emotional processes related to unpleasant or rewarding stimuli. The *cerebellum* controls balance and coordination by producing smooth, synchronized movements of muscle groups (Encyclopedia Brittanica, 2019).

Humans learn extremely effectively from a small amount of data. Most likely there is an architecture in the human brain that serves all the tasks that humans have to deal with, and can skillfully transfer general abilities and skills from one path to the other. These transferable skills and ways of learning are something that AI researchers have not been yet been able to give artificial agents – and AI needs a great deal of data to learn.



Autonomous learning — learning how to learn

The world has seen two 'AI winters' — dark days for AI, where interest and funding dropped. (1974-1980, 1987-1993). In 1986, just before the arrival of the second winter, researchers made a breakthrough: the '*backpropagation algorithm*'. Applied to cases from the early 1990s, this enabled machines to learn via representations of things, such as the image of a cat. It was a major advance: neural networks learning a bunch of features through recognizing patterns. It led to the image and voice recognition and language translation that we know today from Siri or Echo.

And yet. Almost all of today's AI sub-applications, from deep learning to neural networks, use *supervised learning*. This is a 'bottom up' cognitive process, learning on the basis only of what is 'seen'. It takes a lot of data to initially train the neural network. Supervised learning is powerful, however. It enables AI to recognize specific patterns in complex labelled images that may indicate, for instance, a cancerous tumor. However, interpreting those images is beyond its scope.

This is very different from the autonomous way in which a human child learns, a process which some people, such as Facebook's Chief AI Scientist, Yann LeCun, label *self-supervised learning*. You don't train for a specific task, you just observe the world and figure out how it works.

Reinforcement learning is about learning through trial and error — and it's one facet of AI that enables outstanding performance in gaming, to name but one field. However it doesn't (yet) work in many real-world scenarios. We humans, on the other hand, excel in *model-based reinforcement learning:* our internal model of the world allows us to predict outcomes or the consequences of our actions. This, by the way, is exactly how science progresses. And we can plan ahead, something which requires *causal modelling or imagining*. Computers cannot do that (yet) and despite the massive efforts of researchers, are unlikely to be able to, in the short or even medium term.

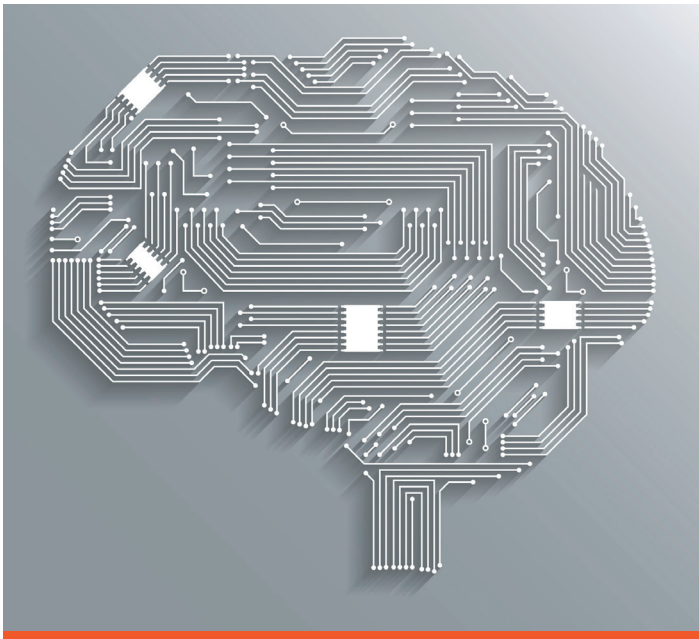
Moreover, humans learn extremely effectively from a small amount of data. Most likely there is an architecture in the human brain that serves all the tasks that humans have to deal with, and can skillfully transfer general abilities and skills from one path to the other. These transferrable skills and ways of learning are something that AI researchers have not been yet been able to give artificial agents. Remembering that AI needs a great deal of data to learn.

Humans also communicate through *stories* which are multilayered and highly contextual, something current computers struggle with. When it comes to fluently communicating with a computer based on what it has read and understood, research hasn't yet cracked the code to make computers think and learn like humans.

All of this means that when it comes to learning, current AI applications only have a small slice of capabilities compared to general human intelligence.

Understanding cause and effect

The second field in which AI will need better techniques is in understanding *causality*. This concerns the relationship between cause and effect. As statisticians constantly remind their students, *'correlation doesn't imply causation.'* To cite a famous example: it's not because ice cream sales and homicide rates both rise in hot weather that ice cream is responsible for homicides. Causality is one key to understanding the limitations of current (narrow) AI: today's neural networks or deep learning machines are currently more in the business of finding statistical regularities in complex patterns, than organizing these in a way that allows them to detect how one thing can affect another. To do this, AGI would need to be contextualized, situationally aware, nuanced, multifaceted and multidimensional.



Today's neural networks or deep learning machines are currently more about finding statistical regularities in complex patterns than organizing these in a way that allows them to detect how one thing can affect another. To do this, AGI would need to be contextualized, situationally aware, nuanced, multifaceted and multidimensional.

So it's hardly surprising that computers can't answer 'why' questions (that would imply not only understanding causal networks, but also self-awareness). Today's AI uses big data to operate or conform to the logic of probability and proportions. Even self-driving cars will likely only function within 'geofenced' regions, while autonomous driving in busy European cities, that tend not to be organized around logical grids, will take quite some time to materialize.

The UCLA computer scientist and mathematician Judea Pearl is perhaps the world's leading pioneer in AI. He won the 2011 Turing Award for solving its primary challenge — programming machines to associate a potential cause with a set of observable conditions, using what are called 'Bayesian networks.'[4] (For example, if a patient returns from Africa with a fever and body aches, the most likely explanation, or correlation, would be malaria).

For Judea Pearl, this *correlation* thinking was only the start. He recently introduced the need for *causal* thinking in AI *(The Book of Why: the New Science of Cause and Effect)*. He argues that, unlike AI, the human brain is not just wired to solve probability (or correlation) problems, but causal problems. A computer can only tell us how likely an event is, given what it observes. So, beyond establishing a *correlation* between fever and malaria, a computer needs to be able to reason that the fever is *caused* by malaria. Once causal reasoning is installed, Judea Pearl foresees a next step: the computer needs to get to grips with 'why' questions, if consciousness — and free will – are to become a reality. And for this, he argues, we'll need the algorithmitization of *counterfactuals*. These describe how a relationship would change, given the introduction of some other condition into the equation.

What do humans do? In performing causal reasoning, we explicitly or implicitly use models: we see patterns, and look for a causal explanation. So AGI will need causal models to reflect on its actions, as well as learning from mistakes.

[4]A Bayesian network can only tell us how likely an event is, based on an observation of another piece of information. It is basically the mathematical transformation of information, or conditional probabilities (e.g. given x, what is the chance of y occuring). It only identifies associations between variables, not causation..

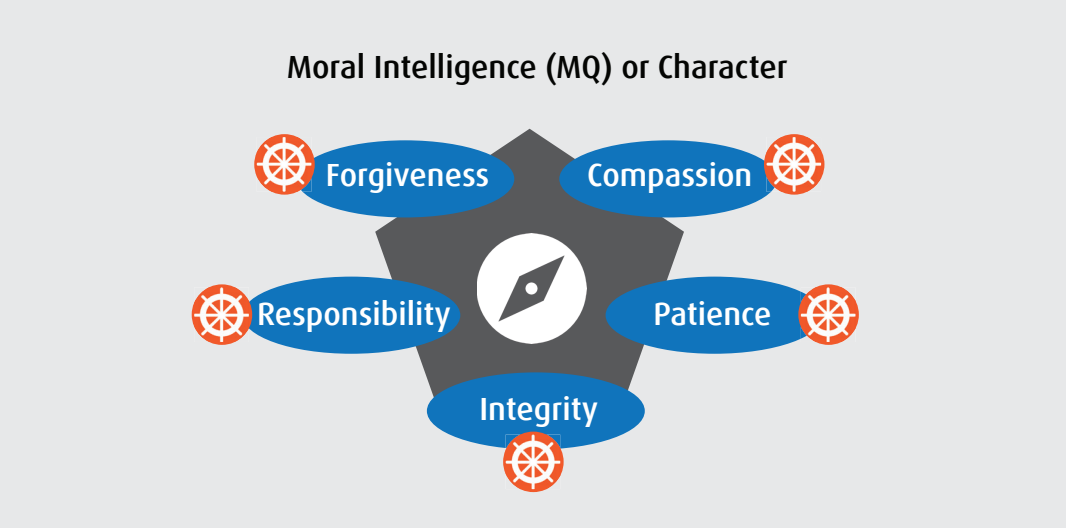
Conclusion

We're therefore left with AI as a purely data-driven or statistical approach to the world — very powerful for prediction and perception tasks: pattern and voice recognition, image perception and control, such as driverless cars and robotics. Less so, in the knowledge space: reading contexts, motivations and causal thinking.

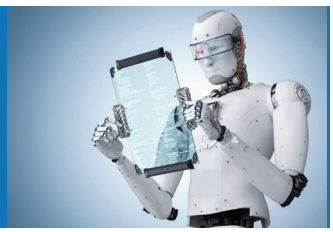
To all of this, we add a final, vital point: the human ability to frame and answer ethical questions, to feel empathy and compassion. These abilities still lie beyond the outer rim of AI. Yet they will be fundamental to ensuring that it is applied wisely, in a way that is ethical, responsible, and sustainable.

Today's big data analytics and deep learning machines — all part of narrow AI - may lead to smarter decisions. But they cannot, at the current time, make wise(r) decisions. This remains a unique human ability, one that we need to apply more effectively if we are to sustain our physical habitat (our planet) and our socio-economic habitat (our society). Humans will never beat a computer in speed and data processing efficiency. But when it comes to creativity, intuition and therefore innovation, we are far superior.

Having the wisdom to apply our innovative power — in collaboration with AI-driven machines - doing so in a way that is ethically and environmentally sound - would be an incredible step forwards on this fascinating road.



Recommended Reads



Barrat, J., (2013), *Our Final Invention. Artificial Intelligence and the End of the Human Area,* New York, Dunne Books

Bostrom, N., (2014), *Superintelligence. Paths, Dangers, Strategies,* Oxford, Oxford University Press

Brockman, J. (Ed), (2019), *Possible Minds. 25 Ways of Looking at AI,* London, Penguin

Chalmers, D., (1996), *The Conscious Mind. In Search of a Fundamental Theory,* New York; Oxford, Oxford University Press

Deheane, S., (2014), *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts,* London, Penguin

Domingos, P., (2015), *The Master Algorithm. How The Quest For The Ultimate Learning Machine Will Remake Our World,* London, Penguin

Ford, M., (2018), *Architects of Intelligence. The Truth About Ai From The People Building It,* Birmingham, Packt Publishing

Kurzweil, R., (2006), *The Singularity is Near. When Humans Transcend Biology,* London, Penguin

Kurzweil, R., (2012), *How to Create a Mind. The Secrets Of Human Thoughts Revealed,* London, Viking

Lindstrom, M., (2008), *Buyology: Truth and Lies About Why we Buy,* New York, Broadway Books

Manyika, J., Chui, M. & S. Lund, (2017), "What's Now and Next in Analytics, AI and Automation", *McKinsey & Company, May*

Pearl, J, MacKenzie, D, (2018) *The Book of Why: The New Science of Cause and Effect,* New York, Basic Books, Hachette

Penrose, R., (1998), *The Emperor's Mind: Concerning Computers, Minds and the Laws of Physics,* Oxford, Oxford University Press

Polson, N. & J. Scott, (2018), AIQ. *How Artificial Intelligence Works And How We Can Harness Its Power For A Better World,* London, Bantam Press-Penguin

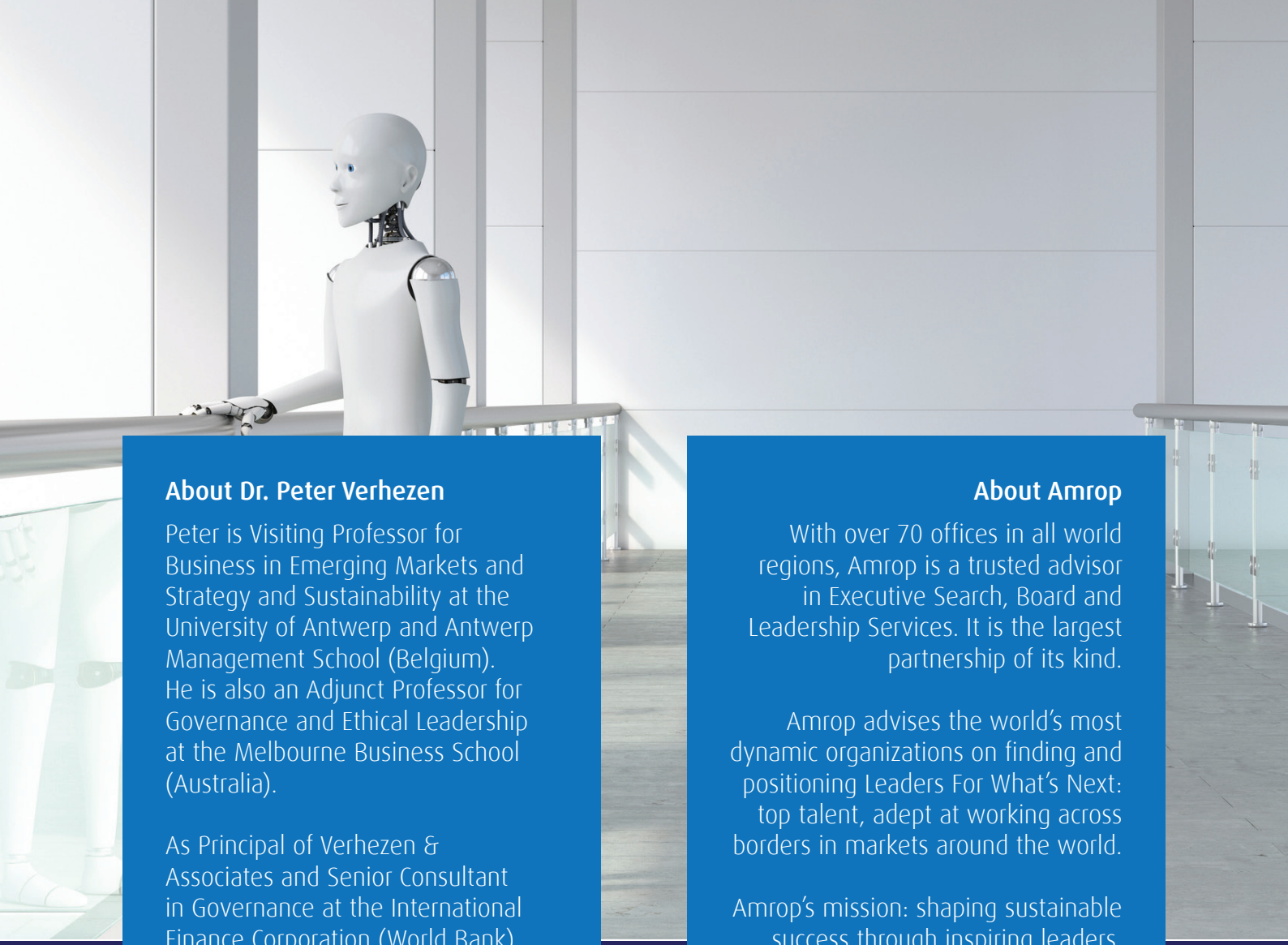
Satel, S. & S.O. Lilienfeld, (2013), *Brainwashed. The Seductive Appeal of Mindless Neuroscience,* New York, Perseus Books

Stevens-Davidowitz, Seth, (2019), *Everybody Lies. Big Data, New Data, and What The Internet Can Tell Us About Who We Really Are,* New York, Penguin

Tegmark, M., (2016), *Life 3.0. Being Human in the Age of Artificial Intelligence,* London, Penguin

Tononi, G., (2012), *Phi. A Voyage From The Brain To The Soul,* New York, Pantheon Books

Zarkadakis, G.,(2015), *In Our Image. Savior Or Destroyer? The History and Future of Artificial Intelligence,* New York; London, Pegasus Books



About Dr. Peter Verhezen

Peter is Visiting Professor for Business in Emerging Markets and Strategy and Sustainability at the University of Antwerp and Antwerp Management School (Belgium). He is also an Adjunct Professor for Governance and Ethical Leadership at the Melbourne Business School (Australia).

As Principal of Verhezen & Associates and Senior Consultant in Governance at the International Finance Corporation (World Bank) in Asia Pacific, Peter advises boards and top executives on governance, risk management and responsible leadership. He has authored a number of articles and books in the domain.

www.verhezen.net

About Amrop

With over 70 offices in all world regions, Amrop is a trusted advisor in Executive Search, Board and Leadership Services. It is the largest partnership of its kind.

Amrop advises the world's most dynamic organizations on finding and positioning Leaders For What's Next: top talent, adept at working across borders in markets around the world.

Amrop's mission: shaping sustainable success through inspiring leaders.

www.amrop.com/offices