"Mauria te pono"

(#Maori: "Believe in yourself")

Leesexemplaar

Geertrui Mieke De Ketelaere

WANTED: HUMAN-AI TRANSLATORS

Artificial Intelligence Demystified

PELCKMANS

CONTENT

INTRODUCTION

A MANUAL FOR AI TRANSLATORS

"That's all well and good, but maybe we should do something about AI as well?"

It was a weekday morning in 2019. I was, along with some colleagues from imec (the Interuniversity Microelectronics Centre), in a meeting with the management of a Belgian energy supplier. The question immediately caused confusion on our side of the table. Over the past hour, we had been talking about automatic systems and data-driven methods, which rely on everything from customer data to smart meters, and how those systems would allow them to build a more personal customer relationship in the future. In other words, we had been talking about artificial intelligence, but without using that term—and they had apparently not understood that at all.

Today, of course, we can laugh about that question. It was, after all, just a misunderstanding of terminology. At the same time, though, it was a symptom of an important and frightening reality. A management team that will ultimately take responsibility for AI systems did not really know what they are or how they are assembled. Moreover, this particular management team was anything but unique. There may be a lot of hype around artificial intelligence, but few decision-makers understand and know about these systems.

Later that day, back in my office at imec, one of my team members brought in the latest results. Our new research approach was showing a theoretical improvement in the accuracy of an AI system. It was a small eureka moment for my team that is responsible for developing AI systems and passing insights back to industry. The gears in my head started turning because the mathematical formula behind those new results may be very elegant, but the implications are much more complex. How relevant is this to the industry that relies on our algorithms to make their operations more efficient? If the new approach *is* relevant, will it be worth the extra cost? More importantly, how on earth do we explain this research clearly and transparently to the external world?

The questions extend beyond the purely economic perspective. What if this research result ends up on the Internet and is used for obscure purposes? It would not be the first time that a piece of technology, created by enthusiastic, but somewhat naive researchers, has become the basis for software that violates people's privacy or basic rights. As researchers, we need to guard against this. I often find myself being enthusiastic about technological progress and then later realising that it also has a social perspective and impact. How do we make sure that this doesn't happen with this research result?

At home that evening, the phone rang. It was a worried family member. She had seen a debate about AI and it had frightened her, so she called me with her questions. Will robots take over our work? Will autonomous cars continue to cause accidents? If so, who is responsible? How do we solve ethical dilemmas, such as the prejudices we humans put into algorithms? I tried to answer all of her questions, but her fear remained almost palpable.

THE SPREAD OF AI

Unfortunately, days like these are no longer the exception. Artificial intelligence or AI is rapidly becoming more important and prevalent in our world. Breakthroughs are stacking up, and AI systems are impacting everything, everywhere: from Netflix recommendations based on previous viewing behaviour to the gates at the airport that scan our faces.

Even cucumbers are no longer safe. Makoto Koike, a Japanese systems engineer working in the car industry, started helping on his parents' cucumber farm in 2016. At that time, cultivation was still a very manual process. Sorting the cucumbers into categories, in particular, took a lot of work and time. During the harvest season, Makoto's mother spent at least eight hours a day dividing the harvest into nine categories by shape, size and colour. This system was considered too difficult to teach part-time workers. Makoto thought it could be done more easily using AI. He drew on a so-called neural network, an AI system that imitates a human brain, and trained it using pictures of cucumbers. He did this with open-source tools from Google, which anyone can find online for free. The system soon learned Makoto's mother's method of dividing the cucumbers into the nine categories.

That system connected Makoto to an automatic distributor. With some of his own electronics, including a Raspberry Pi, a small computer for craft projects, he put together a robot that subdivided the cucumbers into the categories identified by the AI model, and ultimately automated the task that took his mother all day.[1]

A couple of decades ago, it was unthinkable that a non-expert like Makoto could put together such a system by himself. Today, AI is systematically penetrating every corner of our society. Andrew Ng, an AI researcher at Google and Baidu in China recently commented: “Just as electricity transformed almost everything 100 years ago, today I find it difficult to imagine an industry that AI will not transform in the years to come.” [2]

Is AI ‘the new electricity’? Electricity transformed our whole world, as streets and houses were illuminated, appliances took over household tasks and machines automated work in factories. Today, AI algorithms ensure that Amazon recommendations are guiding and telling us what we would likely be interested in buying next, that doctors give us better diagnoses, that human work can be automated and that we can translate almost any language if we have a simple internet connection.

At the same time, there are also many threats that we are facing. We worry that AI will take our jobs, violate our privacy, that there will be killer robots and even that AI will result in a super-intelligent entity that will take over our world, à la science fiction.

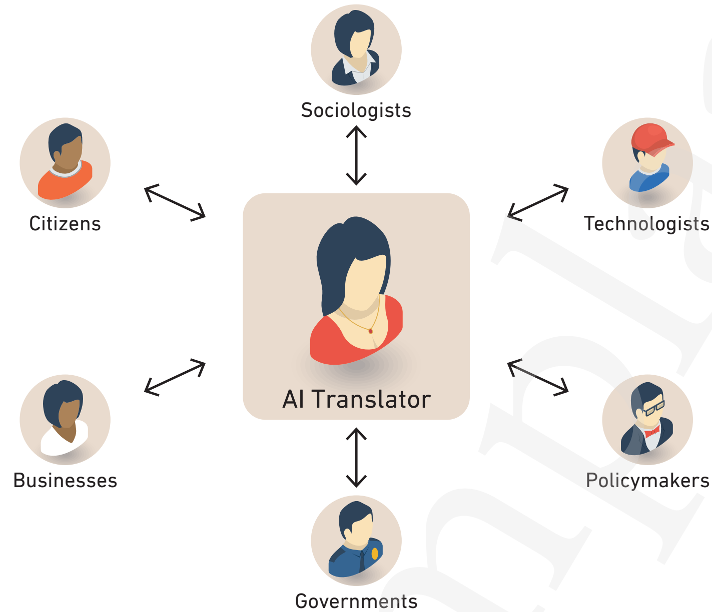
This is why my meetings worry me. Managers who have, or will have, the final responsibility for this type of system do not understand how AI operates or its implications. AI engineers still tend to look with a very narrow, technical eye. Too often, they ignore the social consequences. There is a certain irony in the fact that AI ensures that online translators such as Google Translate work quite well, but we lack people to translate AI between different groups. In my opinion, the great challenge of AI does not lie in the technical sphere, but rather in the AI translation. I am not suggesting that we should underestimate the remaining technical challenges and development that AI faces. I am, however, saying that the implementation of these systems most often fails during the translating of the technical aspects to the real world application.

We must not turn a blind eye to technology. In large parts of the AI community, there is still a view that only technical rules are needed to prevent problems arising from AI. I believe that is wrong. We should, of course, pursue technical improvements and solutions. However, at the end of the day, we need to democratise knowledge about AI systems and have a wider and more inclusive debate about them. Engineers alone will never be able to foresee all possible future scenarios or fully understand the complexity of the world. The AI systems they build will therefore never be perfect. This means that we must allow as many disciplines as possible to participate in the debate. And when something goes wrong, we should not blame the engineers alone, because they only make up one part of the complex whole.

Lawyers, managers, citizens and engineers all look at AI in different ways, and ask different questions about those systems. We therefore need people to act as liaisons between these groups. We need people who can translate the needs and demands of engineers for managers, and convey the concerns of citizens to technicians. People who can analyse whether an algorithm is potentially harmful if we put it online, or explain to a management team that AI might not be the best solution for their needs, as well as people. And people who can explain technical AI systems to legal experts so that they can develop fitting regulations.

In other words, we need bridge-builders who move between social and business worlds (see Figure 1). AI today is, by definition, multidisciplinary, but we are still managing it from separate silos. We have too many AI specialists, and too few generalists: too many people who know a lot about a single aspect, and too few who generally understand the collective whole.

Figure 1 **THE AI TRANSLATOR**



GONE SURFING

My experience with AI goes back a long way. I wrote my master's thesis about it as a young engineer in Germany, back in the early 90s. Today, I am the programme director for artificial intelligence at imec, Belgium's largest independent research institute specialising in digital and nanotechnology.

When I started working in AI, we were far away from the attention and hype that AI now enjoys. The computers were slower, the data limited and the algorithms less impressive and impactful. If you wanted to train a neural network of eight layers and thirty variables (this topic is explained in the next chapter), it would have taken three days. I used to live in Stuttgart, about five hours drive from Lake Garda in Italy. We had time to drive to Italy to go windsurfing while an algorithm was training, and it would be ready when we returned. Since then, the exponential growth and the velocity of AI innovation has significantly reduced the likelihood of a trip like this. Training standard algorithms is now extremely quick thanks to much faster computers.

AI has already experienced peaks and troughs of hype and development, which I think is peculiar to the technology industry. Since those early days of windsurfing, I have watched the dotcom bubble burst at the beginning of the 21st century, and witnessed technology companies like Microsoft and SAS reinvent themselves as new players arrived in the market. I am now helping to build a forward-looking European approach to AI at imec.

This book brings together all my experience to provide a manual for future AI translators. It will discuss the technical aspects of AI, but also its social implications. I will look at Moore's law, and General Adversarial Networks, but more importantly, what they mean for our privacy and our futures, and especially how people and machines can work better together.

I believe that this is now more necessary than ever. There is currently a huge amount of hype around AI, but this will not necessarily last. As we will see later in the book, AI has already gone through highs and lows, with phases of hype and corresponding research budgets, and decades of disappointment and lack of investment. Right now, we are in the middle of a 'high', but if we do not address the contradictions inherent in AI, both ethically and technically, we may well hit another 'low'. This would be a problem, because AI has huge potential to make our world a better place, and we would not want to lose out on those advantages. AI translators are needed more than ever, hence this book.

To develop this capacity, however, we must first begin to understand our subject. What, therefore, is AI?

WHAT IS AI?

Artificial intelligence or AI is only made up of two words, but somehow creates a lot of confusion. Academics and researchers have been debating the exact meaning of the concept for decades. In recent years, we have also seen an increased use of the term, and it has been assigned to any computer systems capable of performing automatic tasks and decisions. AI has even become a marketing term, and companies have started using it to give their products an advanced, technological image.

Let us separate and spell the two words out. There is not much discussion around the word 'artificial'. We can generally agree that something is artificial if it does not occur or happen naturally. However, 'intelligence' turns out to be a lot more difficult. Philosophers are still unable to define the concept unambiguously in the 21st century. Psychologists describe it as a mental characteristic with many components, such as analytical intelligence, practical intelligence, social intelligence, emotional intelligence, creativity and wisdom.[3]

Defining intelligence is therefore difficult, with variation across disciplines and researchers. It is therefore also difficult to provide a clear, comprehensive definition of artificial intelligence. In this book, we chose to take a pragmatic path, with a common and, above all, practical definition of AI systems as:

> *"Computer systems that learn, make decisions and carry them out independently."*

This definition contains all the important features of today's AI systems. AI works on a computer. It learns to do something on its own. No human programmer has to tell it exactly, step by step, what to do, which is the case for other software. AI learns things. It learns, for example, to recognise a cat in a picture or to play a board game. And finally, it is seamlessly connected to a decision system and takes action. For example, it recognises a certain type of cucumber from a series of photos, and gives a command to a robot arm to remove that cucumber.

THE ORIGINAL VISION

The original vision for AI was to build computer systems that mimic human intelligence to take over dangerous, dull or dirty tasks: in English, the three Ds. The current bigger vision for these autonomous and self-learning systems is to mimic all forms of human intelligence, including 'difficult' tasks, adding a fourth D to the list. This is a gigantic challenge because human intelligence comes in an enormous number of forms. Just think of verbal-linguistic intelligence, logical-mathematical intelligence, visual-spatial intelligence, musical intelligence, creativity, physical intelligence, interpersonal intelligence (intelligence about collaboration between people) or intra-personal intelligence (insight into how to keep yourself mentally and physically in balance).

If we want to mimic human intelligence, AI needs to support:
- Processing: an AI system must convert input into a usable format.
- Understanding: an AI system must see connections or patterns in the processed input and create insights.
- Reasoning: an AI system must apply logic to the insights and use it to come to a decision.
- Taking action: an AI system must take the right action based on this decision.

These are all steps that people take. We take in impressions through our senses and process them so that we understand our environment. We think about our situation, and make a choice about what we want to do, then take that action.

This is a cycle that we repeat thousands of times every day, very often unconsciously. Every time we open the door, cut vegetables or throw a ball, our brains go through this cycle. It is easy for us, we do almost nothing else. For computers, however, this is very difficult. We are only now starting to build step-by-step systems that go through this cycle for very specific tasks.

To go through that cycle, an AI system requires three large building blocks:
- Algorithms, or the mathematical formulae that enable computers to learn.
- Data, the (raw) input for the algorithm. We usually need large quantities of data to make a system intelligent.
- Hardware. AI requires a lot of computers, and expects not only large data storage but also fast microchips to perform calculations.

With these three building blocks, a computer can become rudimentarily intelligent.

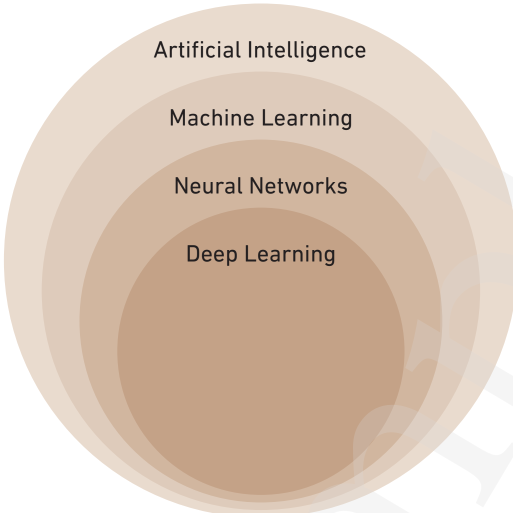
MACHINE LEARNING AND NEURAL NETWORKS

There are many different terms used in the context of AI, such as machine learning, deep learning or neural networks (see Figure 2). To explain these terms, we need to return briefly to the history of AI. The term originated in 1956, when the American computer scientist John McCarthy used it at an academic conference. Initially, AI was used as an umbrella term for machines that performed tasks intelligently. The first systems were instruction-based, so had to be programmed line by line. These types of systems are full of scenarios and responses: for example, if situation A occurs, the system must take action B.

These systems were both easy to understand and transparent, but they also had some drawbacks. They required an enormous amount of time to program solutions, and many of the solutions were difficult to write as a series of instructions. Just think about writing instructions for how to distinguish pictures of dogs or cats, which is now a standard application of AI. Rule-based systems are also very stable in an understood context and controlled environment, but won't take any actions if there aren't any rules for the scenario. This is problematic, because the world is often complex and difficult to predict.

Machine learning (abbreviated as ML) offered an answer to these disadvantages. In machine learning, computers learn from data and adapt through experience, without a programmer defining the rules. This technique can therefore respond to previously unknown situations, which is a huge advantage. At the same time, however, the reason behind a decision is less transparent, a problem that we will discuss in more detail later. Most of what we consider as AI today is machine learning.

Figure 2 **BUZZWORDS IN AI**



There is a particular category of machine learning called neural networks. This is a very popular technique for making machines learn, which imitates the complex structure of the human brain. Our brain is a huge network of interconnected layers of neurons, which pass information onto each other. This system was copied to computers as neural networks, and will be discussed in more detail later. Deep learning is a term for learning through neural networks. The word ‘deep’ has been added recently to recognise that with large amounts of complex data, there are many layers of neurons in the network.

Deep learning and neural networks are therefore a subdomain of machine learning, and machine learning is a subdomain of AI. In practice, however, when we say AI, we usually mean machine learning, and that usually means neural networks and deep learning, because of the success of these techniques.

FROM NARROW TO SUPER AI

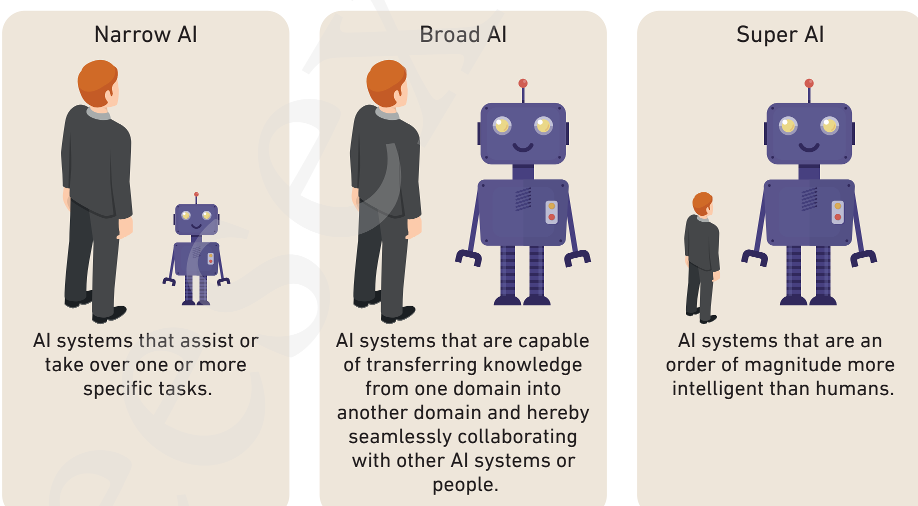
There are different ways to classify AI systems. The most common subdivision is between narrow AI, broad or general AI, and super AI (see Figure 3).[4]

NARROW AI

Narrow AI systems contain solutions that perform one defined task extremely well in a specific context (such as a spam filtering tool or the Japanese system that categorises cucumbers). These AI systems are hyper-specialists in their own small domain. For that one well-defined task in that very specific context, they are often better than people.

Outside the system's context, though, the system is mostly useless. The Japanese cucumber system, for example, can classify cucumbers extremely well. However, without further training, it cannot do any other tasks such as filtering spam out of your mail. It may not even be able to classify European cucumbers because they look different from their Japanese counterparts.

Figure 3 **THE THREE LEVELS OF AI**



The vast majority of AI systems today are narrow AI systems. They are good at processing information and searching for patterns, but they have no ability to reason and do not act outside their specific context. AI therefore offers a lot of value, but it is still a long way from what we would normally describe as intelligent.

Take the human capacity for reasoning, for example. Suppose we go on safari in South Africa and see a lion in front of us. Our senses transmit signals to our brain, such as the lion's image or roar. Our brain processes and interprets those signals, so that we know that there is a lion in front of us. We understand what we see and based on that information we analyse whether we need to take action. In most cases, we would probably decide to leave. However, when we see a lion in the zoo, behind bars, our reaction will be completely different, and we will know that there is no need to run away.

Most commercial AI systems do not possess this kind of intelligence. They are good at processing signals and recognising patterns, but they lack reasoning. An AI system might learn to recognise pictures of a lion very efficiently. However, it is very difficult to develop the general reasoning ability and understanding of the context that makes people realise that a lion in the wild is dangerous, but a lion in the zoo is safe.

BROAD AI

People therefore possess a degree of general intelligence that AI systems cannot match yet. In some tasks, such as playing board games, an AI system may beat us time and time again. Outside that context, though, that system is less intelligent than a toddler.

A broad AI system should be able to switch between different tasks and control different functions automatically. This might, for example, be through combining information from different sensors, just in the same way that we do. This is easier said than done, because the computer will also need to be able to act with flexibility as well as rationally, while interacting with the environment at the same time. We are still a long way from achieving this goal, even though current AI research is focusing on it.

SUPER AI

Some experts believe that we will eventually be able to build a system that will transcend human cognitive skills, general knowledge and detailed experience. They suggest that a computer can access faster processing speeds as well as a larger memory than a human brain. They therefore argue that it should be possible to create an artificial super-intelligent system. This scenario immediately conjures up science fiction stories, but that is all it is for the time being. Whether any AI system will ever surpass human intelligence, and what form it would take, remains unclear.

TURING TEST

[16:11:25] JUDGE: What music do you like?

[16:11:32] EUGENE: I like Eminem - because he doesn't have a thing to do with "polite-correctness".

[16:11:46] JUDGE: OMG i love eminem, he's my favourite artist!

[16:12:00] EUGENE: I wonder what Eminem's nick mean. I associate it only with a word "enema". But he is a cool guy, I like his "The Real Slim Shady".

[16:12:55] JUDGE: i know what his nick name means, his real name is Marshall Mathers. From the initials M&M phynetically sounding like Eminem.

[16:13:06] EUGENE: Eminem is a famous crapper. No, raper... ripper... Yeah, I mean - rapper. I like his "Stan" and "The Real Slim Shady".[5]

This is an example of a conversation with Eugene, the first computer to pass a 'Turing test', and therefore the first truly 'intelligent' computer if we use that test as a yardstick. Alan Turing was a British mathematician and a pioneer in computer science. He famously cracked the Enigma codes, which the Germans used during World War II to encrypt their reports, and therefore contributed enormously to the fall of the Third Reich. In 1950, he wrote an article on "Computing Machinery and Intelligence" that would have a strong influence on the field of AI. In it, he presented his so-called Turing test. Turing had been thinking for some time about self-learning and intelligent machines. Additionally, he was also interested in the deep philosophical questions involved, such as: What is intelligence? When can a machine think? What is consciousness?