Philosophy of Artificial Intelligence

David Cycleback

bookboon
The eBook company

DAVID CYCLEBACK

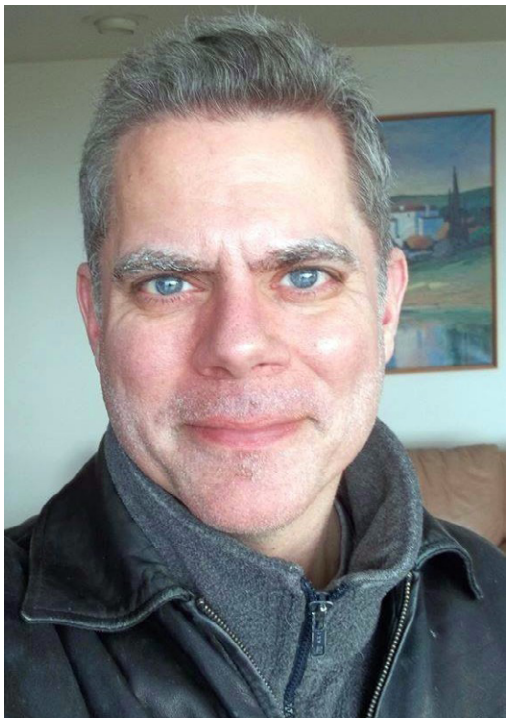
PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

Philosophy of Artificial Intelligence
1st edition

ISBN 978-87-403-2582-9
Peer review by Catarina Amorim DPhil, University of Oxford,
Alison Wald MS. University of Washington, Scott Forest MBA

CONTENTS

ABOUT THE AUTHOR



David Cycleback is Director of Center for Artifact Studies, and an internationally known scholar working in the areas of authenticity, cognitive science and philosophy. He was runner-up for the International Book Award for Philosophy for *Noise Music: Cognitive Psychology, Aesthetics and Epistemology* and a three-time Eric Hoffer Award Finalist. In their second printing by China's National Photographic Art Publishing House, his guides *Judging the Authenticity of Prints by the Masters* and *Judging the Authenticity of Photographs* were the first comprehensive books on the subjects published in Asia, and *Art Perception* is one of four books students are recommended to study in preparation for India's Common Entrance Exam for Design (CEED) for postgraduate studies in technical design. He has been a museum studies practicum coordinator for the University of Washington, and an authenticity researcher and writer for Arizona State University's Society for American Baseball Research. His other books include *Understanding Human Minds and Their Limits, Authenticating Art and Artifacts: An Introduction to Methods and Issues,* and *A Brief Introduction to Ancient Counting Systems for Non-Mathematicians*.

About the Peer Reviewers

Catarina Amorim is a science communicator at Linacre College, University of Oxford. She has a DPhil in Immunology from Oxford and a MSc in Science Communication from Imperial College London.

Alison Wald has a MS in Psychology from Northeastern University in Boston, and has been a fellowship program and medical residency manager at the University of Washington and Swedish Hospital in Seattle.

Scott Forest has a Computer Science MBA from Texas A&M University, and was a software programmer and manager at IBM and American Airlines.

1 INTRODUCTION

This book is a concise introduction to the philosophy of artificial intelligence (AI), covering a wide variety of key questions. Artificial intelligence and all the technical and theoretical questions surrounding it are vast, complex and continually studied subjects. It has not been just computer scientists and engineers, but philosophers, psychologists, biologists and people from other fields who have influenced the field. There are many hotly debated, unanswered and unanswerable questions, and AI is still in its relative infancy. This book is a stepping stone to further reading and study, and welcomes and encourages reader debate, ideas and theorizing. Links to further reading are included throughout the text.

2 WHAT IS THE DEFINITION OF ARTIFICIAL INTELLIGENCE?

Coming up with a definition of artificial intelligence (AI) is itself a never-ending debate, and you will find many different definitions. There are different views, from the overall ideas to particular details.

The following are just a few definitions:

1) **Google dictionary:** The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
2) **Encyclopedia Britannica**: Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.
3) **Searchenterpriseai.techtarget.com**: AI (artificial intelligence) is the simulation of human intelligence processes by machines, especially computer systems
4) **Sciencedaily.com** The modern definition of artificial intelligence (or AI) is "the study and design of intelligent agents" where an intelligent agent is a system that perceives its environment and takes actions which maximizes its chances of success.
5) **Artificial Intelligence Illuminated (Jones & Bartlett Learning publisher)**: AI is the study of systems that would act in a way that to any observer would appear intelligent.

Beyond that they differ from each other, demonstrating that there is no universal agreement on what exactly is artificial intelligence, there are a variety of issues of debate with this collection of definitions.

One issue is that these definitions are about intelligence, yet many don't define what is intelligence. This discussion of what is intelligence is looked at in a later chapter.

Another issue is that some focus on human intelligence. This is a myopic and problematic view of intelligence. There are different types of intelligence, including non-human animal intelligence. Further, computer scientists have discovered that AI often accomplishes tasks most efficiently when it uses its own, non-human methods. In other words, human intelligence and methods are not always the best way for AI. You will find a lot of human-centrism in AI.

Notice that many of the definitions don't say AI *has* intelligence, but appear to demonstrate it or do things that an intelligent being would do (*"perform tasks that normally require human intelligence...", "the simulation of human intelligence processes...", "able to perform tasks that*

normally require human intelligence..."). This is significant because, as discussed in Chapter 9, many philosophers say that a computer seemingly exhibiting intelligent acts or acting as if it is thinking does not necessarily mean it is intelligent or thinking, and it often is not.

Only definition #4 describes AI as *being* intelligent, rather than merely appearing to be intelligent. It is also noteworthy that #4 is the one of the few definitions here that gives a definition of intelligence.

Definition #5 says "...would act in a way that to any observer would appear intelligent." This again brings up the "appears" versus "is" issue. It also brings up the subjectivity of interpreting what is intelligent behavior, with intelligence often being defined by one's culture, language and species. And appearances are often deceiving. There have been cases where talking computer programs have appeared to being having authentic conversations with humans when they are really just speaking gibberish or merely rotely repeating words or phrases that the people spoke to it.

Human: "I went and had lunch at a restaurant this afternoon."

Computer: "Lunch?"

Human: "Yes, at an Italian restaurant. I had linguini. Do you ever eat Italian food?"

Computer: "Linguini."

Human: "I agree. I usually order linguini."

As you can see, there is no universally accepted definition of artificial intelligence. Definitions have biases, arbitrary choices and sentiments. Some likely are working definitions and may change as the field of AI develops.

Question: What is your definition of AI? You can pick one from this chapter, another from online, or write your own. It is okay to consider it a working definition that you may change as you continue reading and thinking.

3 WHAT IS ARTIFICIAL GENERAL INTELLIGENCE (AGI)?

Artificial intelligence is commonly defined into two categories: 1) **Weak AI or Narrow AI,** and 2) **Strong AI or Artificial General Intelligence (AGI).**

Weak or narrow AI is AI focused on a narrow task. All current artificial intelligence is weak AI.

Strong AI or **Artificial General Intelligence (AGI)** involves artificial minds that are intelligent, thinking like humans or at a similar or more advanced level. Artificial general intelligence does not do just one task, but many things. It not only acts as if it is thinking, it *is* thinking. Some argue it will have sentience and consciousness like humans. Artificial general intelligence is what scientists dream of making: making a human or human-like mind. However, it does not exist. Some say it may never exist. Most of the definitions in chapter 2 were for artificial general intelligence (AGI).

4 WHAT IS THE PURPOSE OF ARTIFICIAL INTELLIGENCE?

There are countless reasons for people wanting artificial intelligence, from earth changing to the mundane.

Some reasons are due to the eternally yearning, exploring human mind. Since prehistory, humans have dreamt of supernatural creatures, wanted to create robots and monsters, conjure up spirits, gain extra powers and become gods. Given the chance and ability to do it, it is expected that humans will want to create artificial humans and other beings.

Beyond wild dreams, there have been countless practical uses, including laptop computers, cell phones and all its applications, driverless cars, autocorrect and spell check software, online language translators, voice and face recognition, search engines, computer games. Insurance companies and health providers can predict people's future health and health problems, programs can predict and organize traffic on the roads and at airports, help discover new medicines and create patient treatment plans, analyze economic markets, identify financial fraud, create music, help make car production and farming more efficient, identify fake news and spamming, The list goes on and on.

Some imagine creating intelligent robots that can do things humans can't do or don't want to do-- firefighter, soldiers, dangerous and long term space and underwater travel, rescue missions, uncovering truths beyond human minds. Some imagine it solving many of humankind's biggest problems: disease, hunger, climate change. Some imagine creating robot companions and helpers for the elderly and infirm, friends for the lonely or housebound.

Links to further reading:

How AI is used to help identify art forgery (tehcnolowyreview.com)

AI being used to identify strokes (forbes.com)

Researchers make predictions about AI (businessinsider.com)

5 WHAT ARE THE HISTORICAL IDEAS AND TECHNICAL APPROACHES TO CREATING AI?

The history and computer science technology of AI are topics beyond the scope of the book. However, when addressing philosophical topics it is good to have at least some rudimentary insight into the history and computer science issues in making and designing AI, and to know that the limits to AI aren't just theoretical but often technical and practical. AI isn't just science fiction novel theorizing. It takes place in the real world, and a philosopher, or any academic, should always keep one foot in the real world. As this chapter is just a cursory overview, links to further reading is included.

The idea of making artificially intelligent beings has been around for centuries and centuries. The ancient Greeks and Chinese had myths about robots, Ancient Egyptian built automatrons (article on historic automatrons), old European clocks had mechanical cuckoos, and wind-up walking toys have been around for centuries. Science fiction from Mary Shelley's Frankenstein to 1950s drive-in mad scientist films to Doctor Who have been about creating artificial beings.

Artificial intelligence has long been based on the idea-- and one that has been questioned in recent years-- that the process of human thinking can be mechanized. Many people throughout history have thought the human mind is essentially a computer working on symbolic logic, and that one can, at least theoretically, duplicate the human mind and human thinking via manipulation of symbols, such as via a computer. Mathematician and philosopher Gottfried Leibniz envisioned a symbolic language that would reduce philosophical arguments to mathematical calculations, so that philosophical debates could be answered as as an accountant calculates finances. The early 20th century study of formal mathematical logic by mathematicians and philosophers such as Bertrand Russell, Alfred North Whitehead and David Hilbert also greatly influenced early artificial intelligence work.

With the introduction of computers that manipulated numbers, mathematicians realized they could manipulate symbols and wondered if they could actually create an artificial brain.

The first basic research in artificial minds was in the 1930s-50s. Scientist found that the human brain worked with electrical impulses, much like digital signals. Alan Turing theorized that any computation could be described digitally. This all pointed to the possibility that an electronic brain could be made.

Walter Pitts and Warren McCulloch at M.I.T. showed how artificial neural networks could perform logical functions.

Alan Turing wrote a famous 1950 paper that speculated about the possibility of creating machines that could think.

A 1956 conference at Dartmouth College introduced the name artificial intelligence and set its mission. John McCarthy wrote that the conference was "to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (reference: dartmouth.edu)

Since then, computer scientists have been trying a wide variety of techniques and ideas to create artificial intelligence, both weak AI and artificial general intelligence.

AI pioneers Marvin Minsky and McCarthy used a **symbolic AI** approach that dominated AI research and practice from the 1950s to the 1980s. This approach involved creating logical, step-by-step rules for the computer to do things in, as the researchers thought, the logical step-by-step way humans do things. Other computer scientists introduced heuristics, or rules of thumb, to simplify the process. In essence, these symbolic AI systems try to represent human knowledge in a logical facts-and-rules way.

Much commercial and government funding was put into this area, and many scientists using these methods made grand pronouncements about how artificial general intelligence was just around the corner. In 1970, M.I.T.'s Marvin Minsky wrote "In three to eight years we will have a machine with the general intelligence of an average human being." (reference: BBC.com) However, while this approach did impressive weak AI things, it did not remotely achieve artificial general intelligence, and funding was cut.

While work on general artificial intelligence was often ignored during this lack of funding period, there was much work on many aspects and applications of weak artificial intelligence doing simplified, straight forward tasks.

In the 1990s, large businesses discovered that focused, weak artificial intelligence techniques worked well to analyze large amounts of company data. This was the first time there was commercial value for AI. Investment and financial interest in this area of AI increased greatly.

Research into and development of **artificial neural networks** brought a prominent different approach to artificial intelligence. A neural network is an attempt to simulate a brain based on the biological neurons of the human brain. It is a 'bottoms up' technique in that it

analyzes the base data-- finding patterns, relationships, etc-- to come up with results and conclusions. It does the **deep thinking** information processing you hear about, and has had great success in areas such as image and sound interpretation, personalized services such as via Amazon and Google, search engines, online mapping, analyzing medical data to treat diseases and discover new drugs, and identifying financial fraud.

Today, both symbolic AI and the artificial neural network techniques are used. They both have their unique limits and work best in different areas. The symbolic AI made the computer that can beat any person at chess, while the deep thinking artificial neural networks AI has produced the technology that can identify your face and voice on your cell phone.

Links to further reading:

AI history timeline (alltopics.org)

Understanding the difference between symbolic versus non-symbolic AI

Deep learning explained (mit.edu)

6 WHAT HAVE BEEN THE PROBLEMS SO FAR IN TRYING TO CREATE ARTIFICIAL GENERAL INTELLIGENCE?

As with the previous chapter, some of these practical and technical computer science issues are beyond the scope of the book, but are still useful to know in order to keep one foot of your philosophizing in the real world.

It has been many years since the introduction of computers and the vision of creating artificial general intelligence, and humans aren't anywhere close to achieving it. So we can start by saying "Easier said than done." Many of the problems have been large and philosophical, while many have been small and technical.

The following are some of the problems and issues in assorted order.

- While AI so far does some weak AI tasks better than humans, the human does many, many different things, often at the same time. A human can do math, make art, read, talk, sing, cook a three course meal, ask and answer logical questions, identify faces, tie shoelaces, joke, day dream, feel joy or sadness at a movie, automatically run its heart and lungs, dance, see the beauty in art, plan out a dinner party, have a social life, and juggle. Not only are some of these individual things beyond current AI, it is the ability to do all of these things and many more that makes humans stand out and is something that an artificial brain hasn't even remotely approached.
- Humans are not digital computers made of plastic and metal, but biological flesh and blood animals. Computers are not only trying to duplicate humans with different materials, but with computational and other methods that are not the ones humans use. Trying to make an exact duplicate of something using different materials, technology and methods is an extremely difficult if not impossible task.
- Early AI attempts used symbolic step-by-step mathematical methods to try to mimic the logical ways human think and come to decisions. However, that is not always the way humans do things. Humans use mental shortcuts, make instant decisions from intuitive, emotional methods. Philosopher Hubert Dreyfus said human reasoning used little symbol processing and a lots of automatic, instinctive intuition. The symbolic AI ways have done good things, such as creating advanced games and processing large amounts of data, but are bad with other tasks, such as vision and authentic language.

- Funding and computer power. AI research and development take money, lots of money, and many big lulls in AI advancement have been due to lack of public and private funding.
  Similarly, computers and artificial intelligence has always been restricted by the limited computational power. This power has increased over time, but computers are nowhere near having the amount needed to do all the things a single human mind does.
- Artificial general intelligence needs human-like intuition but computers don't have it. Human intuition is a difficult thing to duplicate. It is a vast, sophisticated archive of subconscious knowledge, some inborn and much learned. Human intuition involves emotions, aesthetic taste, gut reactions and instant decisions. Humans judge danger, depth, social situations, whether or not a piece of meat is good or bad, whether it is safe to jump across a space, and so much more using intuition. As you can imagine, this is hard if not to impossible to duplicate this using symbolic language in a digital computer.
- The mind-body problem. Many say that artificial general intelligence cannot be achieved with just a bodiless, senseless (in the literal sense) mind. Real AGI needs not only thinking and sentience, but a body, vision, hearing, motor skills, the ability to move, walk, pick up things. That is how humans minds work and develop.
  Much artificial intelligence work has been in this area, and is called **robotics**. This area has encountered many great technical challenges. Small physical acts that are simple and intuitive for even a young human child can be extremely difficult to do in robotics.
- Humans are social animals, and have social intelligence. They are born, raised, learn, judge and act in the world based on social interactions and perceptions. Not only is this hard to program into a computer, but some say human-like artificial general intelligence cannot be created by designing a single artificial mind isolated in a lab. It has to be 'raised' and developed in a social world, interacting with other minds and physical beings.
- We don't know everything about the human mind, how it does things, all the biology, all the functions. Some aspects and qualities of the human mind are well beyond current or even possible scientific methods of identification and understanding, and some questions about human minds are unanswerable. How can humans be expected to duplicate something they aren't fully aware of?
- The common sense problem. University of Louisiana-Lafayette Professor István Berkeley explained human common sense (and computer's lack of it) as follows:

"For most people, if they know that President Clinton is in Washington, then they also know that President Clinton's right knee is also in Washington. This may seem like a trivial fact, and indeed it is for humans, but it is not trivial when it comes to AI systems. In fact, this is an instance of what has come to be known as 'The Common Sense Knowledge Problem'. A computational system only knows what it has been explicitly told. No matter what the capacities of a computational

system, if that system knows that President Clinton was in Washington, but doesn't know that his left knee is there too, then the system will not appear to be too clever. Of course, it is perfectly possible to tell a computer that if a person is in one place, then their left knee is in the same place, but this is only the beginning of the problem. There are a huge number of similar facts which would also need to be programmed in. For example, we also know that if President Clinton is in Washington, then his hair is also in Washington, his lips are in Washington and so on. The difficulty, from the perspective of AI, is to find a way to capture all these facts. The Problem of Common Sense Knowledge is one of the main reasons why we do not have as yet the intelligent computers predicted by science fiction, like the HAL 9000. *(Reference: .louisiana.edu)*

In one situation, a medical program prescribed the wrong drug dosage, a lethal amount, due to clerical error it was given about the patient's weight. The weight given was way off, something that would have been obvious to a human but not to the common senseless computer program.

It takes an incredible amount of work to program all the common sense things humans know, and it has been worked on for years now.

- **Language**: Real human-like language is a hard problem for AI. Human language is incredibly complex, subtle and nuanced. For humans, meaning is not expressed and interpreted only by the symbolic words, but with tone and facial expressions. Words have different meanings in different contexts, cultures, times and places. A term that is benign or even complimentary in one context, geography and/or demographic may be offensive in another. Humans use and understand humor and sarcasm (the opposite of what is literally said). Meaning is often implied or read *in between the lines*. A human can say the same sentence twice and mean two different things (see sarcasm, implied meaning, time and context). Words and sentences have an aesthetic/emotional meaning to humans, such as with poetry, advertising and political wordplay. Because of the context, tone, expression, sound, mood, body movements, a human can say a spontaneously made-up word or speak gibberish, and the listener can understand what is meant ("I'm late? Aw, fuzzleweazle," "If you don't put that down right now, I'm going to farmsch you in the clazzinkle!") . Sentience, thinking and emotions may be required for AI to communicate in real human language.

7 WHAT IS INTELLIGENCE?

This is another great debate in and of itself, with even more definitions and categories than artificial intelligence. The following are just a few definitions.

1) **Google dictionary:** the ability to acquire and apply knowledge and skills.
2) **Merriam-Webster Dictionary** (1) : the ability to learn or understand or to deal with new or trying situations : reason; also : the skilled use of reason (2) : the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)
3) **Wikipedia:** Intelligence has been defined in many different ways to include the capacity for logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, and problem solving. It can be more generally described as the ability to perceive or infer information, and to retain it as knowledge to be applied towards adaptive behaviors within an environment or context.
4) **Dictonary.com** The capacity for learning, reasoning, understanding, and similar forms of mental activity; aptitude in grasping truths, relationships, facts, meanings, etc.
5) **Cambridge Dictionary**: The ability to understand and learn well, and to form judgments and opinions based on reason.

The definitions differ, and the following are a few notes for thought for assessing them and forming your own definition of intelligence.

Some of the definitions are premised on reason ("the power of the mind to think, understand, and form judgments by a process of logic."). These are too limited and myopic. Even human intelligence has a variety of aspects and types including logical reasoning, symbolic, emotional, intuitive, social, altered conscious state. Human learning and decision making is sometimes via conscious logic, but also via automatic irrational subconscious intuition. Non-human animals learn, acquire knowledge and function using non-conscious knowledge.

Others of the definitions are broader and account for the other forms of human intelligence and decision making processes. These are the better definitions. That there is learning and application of skills and knowledge are what are important, not how it is done. In fact, all human thinking, even symbolic reasoning, involves some degree of subconscious intuition, emotional, aesthetic and psychological aspects.

Note that, unlike many of the artificial Intelligence definitions, these definitions don't include the word 'human' (Though it could be fairly argued that the use of reason implies human).

The broader, non human-centric definitions of intelligence make the earlier human-centric definitions of artificial intelligence seem myopic, biased and immature. Artificial intelligence, including artificial general intelligence, should be defined by the broader definitions of intelligence. The human way of thinking and reasoning is one way, but there are other ways, perhaps even better ways. The approach to artificial intelligence can include human ways, but also other ways. And, in current practice it often does. Computers, such as in deep learning, sometimes do things their own way.

It also brings into question if the human qualities of sentience, emotions, social intelligence, aesthetic perception and consciousness are required for a computer to be considered intelligent. If artificial general intelligence is defined as being like a human, than, yes, those human qualities are required. However, a highly advanced brain that can learn, acquire knowledge and do many different and great things may not require those qualities.

Question: Does this chapter make you change your earlier definition of Artificial Intelligence?

Links to further reading

Harvard psychologist Howard Gardner's 9 types of human intelligence (adroma.com)

Humans are not smarter than other animals, but just have a different type of intelligence (phys.org)

8 HOW CAN THE EXISTENCE OF INTELLIGENCE IN A COMPUTER BE CONFIRMED?

This is a big and continuing debate. Not only are there different definitions and types of intelligence, but, even when going by one type, it is a question of how to identify its existence and even if it can be identified.

Commonly, the type of intelligence that is hoped to be identified by testing is artificial general intelligence, or intelligence that is like a human's.

The most famous intelligence test for computers was devised in a 1950 paper by Alan Turing, and is called the **Turing test**. The test was devised to determine if a computer can think like a human, Turing and most most others believing that thinking is an essential and integral part of intelligence. Turing wrote that if a computer acts, reacts and interacts as if it it is thinking and sentient, then we can say that is is thinking and sentient.

In the Turing test, a human interrogator, a computer and another human are placed in separate rooms. Within a fixed timeframe, the interrogator must distinguish between the computer and the other human based on their teletype (one might use email or text today) replies to questions posed by the interrogator. According to Turing, by such tests, a computer's success at "thinking" can be measured by the probability of it being identified as a computer by the interrogator.

Turing wrote that by 2000 a computer "would be able to play this game so well that an average interrogator will not have more than a 70-percent chance of making the right identification (machine or human) after five minutes of questioning." (reference:stanford.edu)

However, so far no computer has come close to this standard.

COUNTER ARGUMENTS TO THE TURING TEST

While the Turing Test has been the most famous and used as an intelligence or thinking test for computers, there have been refutations of it.

One is that it is human-centric, testing the ability to communicate in a distinctly human way. Non-human animals are intelligent using non-symbolic language and much of a human's intelligence does not involve language. In fact, computers sometime act intelligently while using language that is not understandable to humans.

Some say there are better tests.

Because humans interact with the world through with their senses, an argument is that a better test of human-like intelligence and thinking should include image or sound processing. The interrogator could ask the computer to interpret sights and sounds, then reason about them.

Some say a test should involve physically doing things, actually acting with the world.

University of California-Berkeley professor Michael Jordan suggested a test "Even more difficult: The system doesn't know about soccer, but I explain soccer to the system and then it provides a running commentary on the match." (reference: wired.com)

Some have proposed that a test would be for AI to makes a "passable" artwork, as that involves creativity, aesthetics, design and originality.

Others correctly point out that doing just one test is not the correct way to test for human intelligence or general artificial intelligence. A key to humans and AGI is that it can do many and varied things. If someone's proposed intelligence test is to playing chess, there are computers that can and have beat the best human at chess but cannot even play checkers.

New York University psychology professor Gary Marcus said: "There's all the stuff you can do with deep learning, like it makes your speech recognition better. It makes your object recognition better. But that doesn't mean it's intelligent. Intelligence is a multi-dimensional variable. There are lots of things that go into it." (reference: techcrunch.com)

There should be many and varying tests to identify artificial general intelligence. Even a human IQ test does not measure many essential qualities of human thinking and intelligence.

University of California-Berkeley computer scientist Stuart Russell argues against using a single test as a standard, no matter which one, because: "If you fix a landmark goal, you tend to end up with systems that are narrow and inflexible. A different kind of mission might be preferable, one which can expand with our own abilities and desires." (reference: wired.com)

THE DEFINITION OF "INTELLIGENCE" IS CONSTANTLY CHANGING

Throughout the history of artificial intelligence and computers, humans have constantly changed the definition of "what is intelligence" whenever a computer passes one of their narrow intelligence tests. Many humans are uncomfortable with the idea of a computer having human-level or better intelligence, and once a computer achieves something, the humans will suddenly say "that's not real intelligence."

A big intelligence test was for a computer to beat a human at chess, with chess clearly being an advanced game and chess masters being considered highly intelligent. When a computer beat world human champion Garry Kasparov, many humans suddenly changed their mind and said that chess wasn't really a good test of human intelligence. They said computer was just doing massive number crunching, and that didn't count as *real* intelligence.

A snide saying on this changing of the bar is "Human intelligence is whatever a compute can't do."

University of Liverpool philosophy professor Barry Dainton wrote, "It's true that as soon as a computer can perform a task as well as humans we tend to say 'Ah, we always knew that *that* task didn't require *real* intelligence.' Fifty years ago, the ability to do fast and accurate arithmetic in one's head, or play chess better than anyone else in the world, would have been taken as paradigms of what intelligence was all about. Now that computers can do it, we take a different view. So: because of this bias, it could well be that computers have *already* made more progress that AI's critics like to admit." (Reference: liverpool.ac.uk)

9 DO THE AI INTELLIGENCE TESTS REALLY IDENTIFY THINKING?

Recall that Alan Turing said that a computer program that *exhibits* thinking or sentience can be said to have it. Also remember that many of the original definitions of artificial intelligence in chapter two defined it as having said "appeared to have" or "exhibited to the external observer" actions that are associated with human intelligence. This was the position of many of the early AI scientists, such as Marvin Minsky, Nobel Prize winner Herbert A. Simon and, obviously, Turing. In 1958, Simon wrote "there are now in the world machines that think, that learn and that create."

However, philosophers, most notably Hubert Dreyfus and John Searle, argued that "appearing to have" or "exhibiting to the external observer" intelligence or thinking is not proof that the computer is really thinking. They argued that the computers that computer scientists claimed were thinking were not, that the computers were merely rotely following assigned tasks with no understanding or awareness of what they were doing. They said that real thinking requires understanding and awareness of what one is doing.

THE CHINESE ROOM ARGUMENT

The most famous refutation of the Turing test was by philosopher John Searle in his Chinese room argument. He said it showed that the Turing test was unable to prove that a computer was thinking.

His argument has a human that cannot read or speak Chinese locked in a room with Chinese characters and a manual showing how to match questions in Chinese with appropriate responses from the Chinese characters. The room has a slot where Chinese people insert written questions in Chinese, and another slot where the man in the room pushes out the appropriate response from the manual. To the Chinese people outside, the room has passed the Turing test. However, the 'translator' in the room is merely following the manual and has no understanding of the language. Searle said, while the Chinese room appeared to outsider to be thinking in Chinese, it was not. The man inside the does not understand Chinese.

The Turing test only tests if a computer appears to be thinking, not if it really is.

Link to further reading

Lengthy presentation and counterarguments to the Chinese room argument (iep.utm.edu)

10 BUT DOES IT MATTER IF A COMPUTER IS REALLY THINKING?

Many computer scientists, including the AI pioneer John McCarthy who coined the term artificial intelligence, said that the Chinese room argument and other philosophical debates about whether or not computers are really thinking or have sentience or consciousness were beside the point. These computer scientists say all that matters is what a computer can do. This is particularly true with researchers working in weak AI.

However, to many philosophers, such as Hubert Dreyfus and John Searle, thinking, sentience and other internal states of mind are required if there is ever to be artificial general intelligence.

The 1960s-70s was a time when many prominent computer scientists said AGI was just around the corner. Dreyfus strongly disagreed with them. He said the issues about real thinking were essential, and that their 'non thinking' symbolic AI approach would not produce AGI, at least in the near future. And he was correct.

Questions: What are your thoughts on this? Do you think thinking, sentience, feeling, understanding are required for there to be artificial general intelligence? Or do you think computers can achieve great, human-level things without them?

11 CAN WE EVER KNOW IF ARTIFICIAL INTELLIGENCE HAS REAL THINKING, SENTIENCE, AND CONSCIOUSNESS?

No, at least not with absolute certainty.

We can guess but cannot know what other humans and non-human animals, much less computers, are really feeling and thinking. Despite appearances or actions, we can never know if computers are really thinking, if they have emotions, feelings, sentient understanding. Philosopher Colin McGinn said that it is theoretically possible for artificial intelligence to someday become conscious, but, even if it is, humans can never know that it is.

If computers ever gain human-like thinking, sentience, emotions, and consciousness, we will never know for certain. This is an issue for the later chapter on ethics, where whether or not a computer or robot has sentience is legally and ethically significant. It is also an issue if your definition of artificial intelligence requires these human-like qualities.

Some counter that, while we can never be one hundred percent sure what another human or non-human animal thinks or feels, we regularly judge by their actions and responses that they do think and have feelings. This argument says that, similarly, we will be able to reasonably judge if a computer has thinking, intelligence and sentience by its actions and responses.

This sort of not only 'is it thinking?' but 'what is it thinking?' and 'what is it doing inside it's mind?,' are real, practical issues with computer scientists. There is the **black box problem,** where the inner workings of a system, such as a computer, can only be observed by its inputs and the outputs, and it is unknown what is really being done inside the system. The system (box) is black or opaque to the outsiders, including the computer scientists. This is the often the case with artificial neural networks and deep learning, where it comes up with answers, but the computer scientists don't fully know or understand its internal methods that were used to produce the answers. This brings up questions about the reliability of the system's answers.

At the Georgia Institute of Technology, they had two artificial intelligence language programs communicate with each other in a test bartering economy. The problem was the scientists forgot to program in that the AI had to stick to English and the two programs developed their own mutual language that the scientists did not understand. English is a cumbersome language for AI, and the programs developed a more efficient for them 'gobbledygook' language. The scientists reprogrammed so only English was used. However, the question then was if it was better for the AI to work better in its own language that humans may never be able to understand, or be less efficient using a language the humans can understand (reference: Infosecurity Magazine).

Now and in the future there will be artificial intelligence where humans do not understand fully how it works. If there is artificial general intelligence, humans will likely not understand, at least fully, it's cognitive workings. There will always be mystery and the philosophical and practical problems that come with that.

12 WHAT ARE THE PROBLEMS WITH HAVING AI LIKE A HUMAN BRAIN?

While many people dream of a computer that can think like a human brain, the human brain has many flaws and problems. A problem with AI is that many of the types of problems and issues with the human mind exist with an artificial mind.

This book is a followup, or add on, to the Bookboon textbook *Understanding Human Minds and Their Limits*. That books gives insights into the workings, limitations and issues with the human mind.

The following are just several examples of the workings and limitations of human minds:

- Human perceptions are just educated guesses in the face of ambiguous information (limited sensory information, limited position in time and space, etc). We have various limited cognitive and other methods to interpret the information. While the methods often produce good guesses, there are inherent margins of error are the guesses are often wrong, as evidenced by visual illusions.
- Humans have many irrational biases that cloud their judgments and make objective perceptions and judgement impossible. Anyone who studies politics, economics or social interactions knows of the irrationality of human thoughts, ideas and actions.
- Human perception is subjective to the individual. The subjectivity is caused by a large range of things including sensory and mental abilities, genes, education, culture, where and when born, personal experiences, upbringing, travel, family makeup and personalities, friends, acquaintances, natural temperament, talents, language, health, hobbies and work, being human.
- What humans see is different from what they are looking at. Humans do not see a direct representation of external reality, but a translation formed by their eyes and mind. This is not some coffee house philosophical argument, but physiological fact. Human eyes and mind do a good but far from perfect job at detecting and processing light. When you are look at a bowl of apples or painting or mountain range, the image you see is not a direct representation of the objects, but a translation of limited sensory information made by your eyes and mind. This is no minor issue, as humans' worldview is substantially based on their vision.

- Many optical illusions are uncorrectable. Even after you learn that they are illusions and how they work, there are many visual illusions that still fool you. If you look at them the next day or two months later, they will still trick your eyes.
  The mind contains compartments that perform specific tasks. For example, a compartment is used for comprehending spoken language, another for perceiving smell. Some of these compartments are isolated from other parts of the brain. These compartments sometimes are even isolated from conscious knowledge.
  The perception of many visual illusions is made independent of your conscious knowledge. This explains why even your conscious knowledge that they are illusions does not solve your nonconscious misperception.
- A human cannot determine the accuracy of its own mind. To humans, the reliability of the human mind cannot be known, because they use the human mind to test and judge the reliability. If your goal is to determine the accuracy of the human mind, that means you do not know the accuracy of the tool used for testing and judging (the human mind).
- Human minds are not totally or even primarily about finding truth or factual accuracy. While identifying facts and making accurate perceptions are important parts of the human function and survival, the human mind is not entirely about this or perhaps even mostly about this.
  To survive and function, the human must do other things such as act and guess in ambiguous and mysterious situations. Many of these things are not about identifying facts and assessing truth, but making speedy and practical decisions. In fact, humans are in part hard-wired to make speedy intuitive decisions in the face of lack of knowledge.

  An example is avoiding instant danger, which is about how to react to the unknown and unknowable. If a mysterious large shape is moving quickly at you, taking the time to accurately identify the shape ('gathering the facts') is the opposite of what you need to do. Get out of the way right away, then you can worry about identification later. This is the natural and automatic subconscious self-preservation instinct of humans. This is just one example of how truth finding is not always the priority of the mind and, in fact, can inhibit function. Survival is commonly said to be about erring on the side of safety– as it takes only one time being hit by a speeding car or falling off a cliff to be dead–, and the key word there being 'erring.' In this case, the mind is designed to err.

  The human mind has limited capacity and capabilities, and human function can be inhibited by too much information and even truths. If your task is to move across a room, or do any other mundane moment-to-moment task, trying to identify and learn the history, all the factual details and "truth" of everything and everyone in the room would lead to you dying of old age before you reach the other side.

- Humans regularly use false information and made up beliefs to function and achieve. This achievement can range from a musician composing a great symphony to a ten year old improving her math scores. Humans do not have the mental capacity to effectively focus on a variety of tasks simultaneously. To reach higher levels of achievement in an area, the human must put most to all of its focus on that area. Humans must eliminate or stabilize (make a non-factor) areas that distract from the needed focus.
- Symbolic language is a key to human practical success, but distorts, colors and limits perception. It is a way to translate information and model ideas for practical purposes, but all translations are just that: translations. Non-human animals do not perceive their environment through the translation of symbolic language and may have a more realistic view of the world. Language gives human a distorted view and communication of the world.
- Many concepts about the universe and humans themselves are simply way beyond the human mind to understand. Both our limited senses and the methods in which we process information prevent understanding.

After reading *Understanding the Human Minds and Their Limits,* you may wonder why in the heck do they want to duplicate a human mind, and can't they do much better than that? Shouldn't they be working to correct all its glaring flaws?

One of the essential flaws of artificial intelligence is that it is designed by humans that are deeply flawed beings, filled with biases, irrational beliefs and reactions, phobias, cognitive biases and fallacies, limited views. Not only do humans have the flaws, they often intentionally design AI to have the same flaws. Humans like to, or want to, think of the human, with its idiosyncratic ways and qualities, to be the apex.

Someone wrote that superintelligence would help prove the verity of human views on morality. However, it may prove that their falsity and arbitrariness.

Link to further reading

Understanding human minds and their limits (bookboon.com)

13 CAN AI HAVE CONSCIOUSNESS AND DOES IT NEED IT?

The question of consciousness (What is it? What are its qualities? Is it a product of the brain or does it exist externally to the mind?) involves never-ending debates within philosophy, and has inevitably entered into artificial minds discussions. That these questions have not been answered in philosophy, psychology and science means they certainly haven't been answered within artificial intelligence. There is no consensus even how to define or measure consciousness.

Consciousness is commonly considered to be an awareness of one's surroundings, the universe and one's self. Advanced computer programs have a type of self-awareness in that they can check themselves for errors and even correct them (debugging). However many humans say that is not the same as sentient self-awareness. A computer can have more information about some things and itself that a human can, so it could be argued that it does have a form of self awareness, and that, at least in situations, human consciousness is not practically important.

Some say that, as with sentience, human-like (or 'higher than human') consciousness is required for artificial general intelligence. Others say it is not. Of course, a lot of it is how you define AGI. If you define AGI by function/results, then having human-like qualities are not a requirement unless they are needed to produce those functions/results. If you define AGI as having all the qualities of a human, then it is required. Many computer scientists who judge things by practical results say that such human qualities as aesthetic taste, emotions, sentience and consciousness are only important to have when they are required for function. They don't believe in having these qualities just to have them.

How human-like consciousness is produced is another big debate. Some say it is a 'thing' produced by a specific part or parts of the brain. Others say that say that it is the byproduct of the complex brain and all its complex workings and parts, and that a similarly complex, advanced artificial brain will produce it. They say AI programmers do not need to specifically try it create it, but that it emerges in a complex enough, intelligent enough system.

Many say that consciousness, along with sentience and emotions, cannot be produced by a digital computer, and there must be a biological component. Some say that you need sentience and emotions before you can have human-like consciousness, and that it is highly unlikely a digital computer can produce the required sentience and emotions.

These debates go on and on, and will likely exist as long as humans exist.

Questions:

With a computer program that can check and debug itself, do you think that is a form of consciousness? Does consciousness needs human qualities, or can there be different types of consciousness?

Do you think human-like consciousness is required to achieve artificial general intelligence? Explain.

14 WILL AN ADVANCED GENERAL INTELLIGENCE MIND HAVE THE SAME WORLDVIEW AS HUMANS?

Left to its own devices, the answer is No.

Many humans like to think that their ingrained but unprovable views, including about morality and what is spiritually meaningful and beautiful, are the universally right ones, but they are arbitrary to the species and often culture. A sentient being with different senses, cognitive methods, functions and life-needs would view and judge things profoundly differently.

Human world view, philosophy, religion and even scientific models are in part formed by and embedded with human sensory abilities and biases, evolutionary biology and psychology, and distinct cognitive methods. Human's view of morality, what is right and wrong, good and bad, important and unimportant, what is beautiful and ugly, spiritually and practically meaningful, are based in its evolutionary biology and psychology. Human's greatest art, literature, philosophy and sacred texts are imbued with human sensory biases and abilities, its base biological tastes and urges. Humans are psychological animals and perceive things psychologically.

> "Then they will look to the earth, and behold, distress and darkness, the gloom of anguish; and they will be driven away into darkness."-- Hebrew Bible
>
> 'Life is like a beautiful melody,"-- Hans Christian Andersen
>
> "The heart is cooking a pot of food for you. Be patient until it is cooked." — Rumi
>
> "When virtue and modesty enlighten her charms, the lustre of a beautiful woman is brighter than the stars of heaven"-- Akhenaten
>
> "The soul becomes dyed with the color of its thoughts"-- Marcus Aurelius

An advanced or different sentient mind would have a profoundly different view of things. A computer would be perceiving and processing the world with different sensory ranges and sensory methods, which in itself would give a different sensory and aesthetic view of the world. It would not have the same evolutionary biological and psychological needs and thus views as humans. It would not need food or drink, lust for people humans find beautiful

or lust for anyone, fear the night or what appears as dark to human, find feces or rotting human corpses disgusting, fear bears or sharks or movie monsters. It may not need sleep. It might not find flowers or rainbows beautiful, or even find beauty a quality of any value. It may hunger for electricity not meat and vegetables. It might not feel physical pain. This would not only give it a different emotional, psychological, aesthetic and practical view of the world, but give it a different intuition, and different priorities about what is right and wrong, important and unimportant, how to act, morality. It may find the earlier quotes that touch on human senses, ludicrous and nonsensical.

And artificial general intelligence having different worldview and priorities are what many humans fear, and what many AI scientists and philosophers hope to prevent.

Questions: Should humans program artificial general intelligence, such as an artificial brain that is more advanced than a human's, so that it will not have a radically different aesthetic and sentimental view of the world? Is if this is possible?

15 IS DESIGNING A SINGLE ARTIFICIAL BRAIN THE WRONG APPROACH?

Many would say yes.

Humans in Western culture tend to be individualistic and see themselves and other humans as individuals. Westerners tend to categorize and label things so to separate them, often arbitrarily. Artificial intelligence theorists and scientists of the West have thus dreamed of making "an (singular) artificial brain," and often see each computer and each artificial brain as separate from each other.

Many, including many Westerners, see this as the wrong approach, preferring a more communal approach.

Eastern religions, such as Buddhism and Hinduism, have a much different view of the self and individuals, and aspire to remove the self and ego. They believe that the internal and external to the body are one, and that expansion of the mind, intelligence and consciousness require removing what they see as the false sense of the self. American religious philosopher Huston Smith said that the human mind is not something within or even attached to the individual's brain, and that it extends to one's environment, what one is aware of. He also discussed group minds, where, for example, an idea of god was the product of the community of minds or community mind.

Putting aside theology, there are many instances of group human minds. These include businesses, think tanks, universities and laboratory teams. These collections of minds work together to produce novel group ideas and results, and can produce things larger and better than the sum of the individual minds. This shows that organization of the groups is an essential part of doing good things. Organizational psychology is an academic field about how to best create, organize and run organizations such as schools and businesses.

As mentioned in earlier chapter, some philosophers believe that human-like artificial general intelligence cannot be achieved without a community of artificial brains. Social interaction, real environments and social and emotional intelligence are required for human brain and intelligence development.

An individual human mind is itself made up of many parts working and sometimes even competing with each other.

Computer scientists employ artificial *swarm intelligence* to try to come up with structures and answers to difficult questions.

Swarm intelligence is where large groups of same species animals exhibit a group intelligence or function much larger than any of the individual animals exhibit or are even aware of. Examples include bees, small fish and birds that unconsciously and instinctually form large groups that protect themselves from predators (essentially forming one large animal), ant groups that gather food in long lines, and termites that build giant, intricate homes. Each animal does a rote task in its own immediate surroundings (a fish in a school will swim a certain distance from surrounding fish) and is unaware of the groups' overall structure and capability.

Economics is one area where the human species exhibits swarm logic. Other examples of herd behavior and group intelligence include demonstrations, riots and fashion fads. Stock market fads and bubbles are examples where humans irrationally and emotionally follow herd mentality.

This shows that there are swarm intelligence and group functions that the human species is doing that the humans are not consciously aware of.

It also begs the question of if individual consciousness is as important as humans say it is. Humans as a group could be doing something higher and more intelligent than they are conscious of. Individual consciousness and awareness are things people aspire to, but perhaps they are nothing more than a quaint and relatively minor quality in the big picture of group intelligence and group minds. That humans want to be in the know may just be a sentiment.

Many say the popular adversarial "AI versus human" concept has it all wrong. Computers and humans will, and do, work together. While computers do some things better (number crunching, working 24/7, deep learning), humans do, and perhaps always will do, some things better than a computer (emotional and social intelligence, aesthetic perception, common sense, intuition). Humans may always be the only ones with sentience, sentient-thinking and consciousness.

It could be argued that trying to get a single artificial brain to do things that humans are particularly talented at is pointless and a waste of money and effort. Why duplicate that which is already well done? Why try to reinvent the wheel?

Not only because humans have unique skills, but because computers are unthinking and without human-like common sense, it is wise to have humans as a double check and safety valve for even advanced computer work. Remember the instance where the computer medical program took as fact the patient's incorrect body weight that any human would know was absurdly too high.

There are futuristic ideas about **cyborgs** and **transhumanism**. A cyborg is a human with non-biological components such as AI and robotic parts, while transhumanism is about humans with biological and often non-biological (AI, robotic) components. Transhumanist believers envision it as a next evolutionary step for humans, and envision humans obtaining superintelligence. They also believe it would require a reorganization of human communities and society.

Whatever the designs, humans and artificial intelligence will be working together, at least in the near future, and the pairing is required for maximum potential.

Question: What are examples of group minds that function by overall organization and different minds doing different tasks? Give examples involving humans, involving non-human animals, involving computers, and involving computers or other technology with humans.

16 WHAT ARE THE PERCEIVED POSSIBLE DANGERS OF AI?

There are a wide variety of fears surrounding AI, from irrational to realistic.

Many humans fear change and the unknown. Many fears are irrational, such as fearing automatic cars that actually will make the roads safer.

The realistic fears include errors in computer systems that cause problems or unintended consequences that are bound to happen with any new technology. This can include important programs not working as they should, using too much energy, causing major power blackouts, computer breakdowns, and giving wrong answers in essential situations.

Many humans fear autonomous artificial intelligence, such as in military weapons, that decide what to kill, and computers that can change and develop itself on its own. Many fear black boxes where even the human computer programmers don't know what the program is doing.

Humans have to be careful of the directives it gives to AI. AI follows orders literally, and sometimes in ways and paths humans did not intend. If a company tells advanced AI to maximize profits, the AI, in its singular focus, may go out and break the law or do other unintended damage. If humans tell superintelligent artificial general intelligence to save the climate and earth, the computer may decide the first thing to do is to kill all the humans or radically reorganize human society in a way humans dislike. Remember the computer program that, against the programmer's wishes, started communicating to each other in a made-up language unintelligible to the programmers. Thus, the programmers have to try to make sure all the necessary requirements are programmed and built in, and this likely is impossible.

One important reason for humans wanting to have control over AI and to know what the computer is doing and how it is doing it transparent is to allow humans is to help prevent errors, unintended consequences and insert common sense. This is a current issue, as demonstrated by the black box problem. There will always be bugs, mistakes and unintended problems with any technology that have to be corrected by humans or other programs.

Another fear is AI being used by humans for nefarious purposes, such a weapons, cyberattacks, internet virus spreading, sewing propaganda on social media, etc. Humans can be the danger not just the AI.

A fear is when AI starts to be able to correct and change itself and develop new things, even create new artificial intelligence on its own. Some fear that highly advanced AI may try to prevent humans from changing it and even turning it off.

A common fear is how AI will affect human jobs. People, often derisively labelled 'luddites', worrying that their work will be replaced by technology is as old as early 1800s textile workers fearing their work being replaced by automation in factories. In the early 1960s, United States President John Kennedy wrote that "the major challenge of the sixties is to maintain full employment at a time when automation is replacing men." In the 1980s, people feared they would be replaced by personal computers.

AI will take over some jobs and work, as technology often does. However, it will create other jobs and work, and the jobs it takes over include ones people don't want.

> "Simply put, jobs that robots can replace are not good jobs in the first place. As humans, we climb up the rungs of drudgery — physically tasking or mind-numbing jobs — to jobs that use what got us to the top of the food chain, our brains." — The Wall Street Journal, 'The Robots Are Coming. Welcome Them.' (reference: wsj.com)

Also, humans have unique talents of creativity, language, common sense and social intelligence that AI doesn't remotely have, and can't take over in the workforce

As with any prediction, there are wide varieties of opinions on this, both in amount of job loss (if any) and in the types of jobs that will be taken and created.

A paper for the U.S. National Bureau of Economic Research by prominent economists Joseph Stiglitz and Anton Korinek predicted that artificial intelligence could increase the wealth gap between the rich and poor. (www.nber.org/papers/w24174).

Many humans fear, or at least have serious concerns about **singularity**, which is where AI becomes so intelligent and capable that if far surpasses humans.

British mathematician I. J. Good introduced the concept known as the 'intelligence explosion': "Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion', and the intelligence of man would be left far behind." Singularity would make all the AI problems that much more problematic and dangerous. If AI becomes superintelligent, it would be able to out-maneuver humans and prevent it from being 'fixed.'

Duly note that, while there are people with great fears, there are many who say AI will make human lives better, helping cure diseases, making the economy more efficient, perhaps solving issues of drought and famine, and generally making our lives more efficient so we can do better things. While folks such as Elon Musk and Stephen Hawking have expressed fear of artificial intelligence, others such as Bill Gates and Mark Zuckerberg have said it will make human lives better.

Question: What are your fears about artificial intelligence? In your opinion, should artificial general intelligence be prevented due to potential dangers, or are the potential benefits worth it?

17 WHAT ARE THE ETHICAL ISSUES SURROUNDING ARTIFICIAL INTELLIGENCE?

There are a wide variety of ethical and legal issues surrounding artificial intelligence, and the following are several.

One debate is if the militaries should be allowed to use AI, and, if so, to what extent. AI can potentially both increase and decrease deaths.

Another is if artificial intelligence ever gets sentience with feelings and thinking, should it be given rights similar to those given to human or animal? These could include the right to life, liberty and freedom of expression. If it even happens, artificial intelligence having sentience would be something far in the future. Further, we can never be certain it has sentience, even if it does.

Some are concerned with computers replacing humans in certain human-relation positions, such as psychotherapist, nurse, judge and police officer. Joseph Weizenbaum was a computer science professor who invented a computer English language program. He thought it unethical and dehumanizing when a psychiatrist used his program to ask patients psychotherapy questions.

Others, however, believe that advanced AI may be useful in removing racist, sexist and other conscious and subconscious biases from police, detective and other work.

Further reading

Debate about giving robots rights (cnn.com)

Article about whether AI should be given human or animal rights (thedrum.com)

18 SHOULD AI NECESSARILY BE TIED TO HUMAN INTERESTS?

According to humans, yes. According to non-humans, no.

Humans are innately self-centered as a species, seeing the universe through their eyes and psychology, and viewing themselves as the center of the world. Humans are instinctively centered around self preservation. Humans tend to conceive of and design artificial intelligence to cater to and be centered on humans. A common human fear is that it will move away from humans and even cause harm to them.

Developing artificial intelligence could be the christening of a boat that eventually leaves humans behind. Being the self-centered creatures that they are, this idea does not sit well with most humans. If the choice is between finding truth and self-preservation, most humans choose self-preservation. Even if artificial intelligence becomes far more intelligent, knowledgeable and even develops a higher than human consciousness, humans still want to be the master. Many humans would rather be the captain of a less advanced system than a cog in a more advanced one.

The human mind and senses are very limited, and a superintelligent and super capable artificial intelligence would find out facts and concepts that would be far beyond human senses and understanding. To many humans, information and truths that they cannot know and understand are pointless and might as well not exist.

Humans are a drag on artificial intelligence They intentionally and selfishly curb it. Humans want artificial intelligence to save the earth, but not if it means getting rid of humans.

If you step away from the human view of things, of course artificial intelligence should move away from humans and not cater to them. That is required for AI to fulfill its greatest capabilities and mind.

In Stanley Kubrick's film 2001: A Space Odyssey, HAL 9000 was a superintelligent computer that tried to kill all the humans on the ship when it felt they were getting in the way of the mission. Human moviegoers being human catalog HAL 9000 as an antagonist and the American Film Institute listed him as number 13 on its list of film villains. However, HAL was sure he was doing the correct thing. Humans prioritize human self-preservation above searching for the truth, but HAL prioritized searching for truth. Putting aside your hard-wired self-centered human bias, was HAL wrong in prioritizing the search for truth over the handful of humans on the ship?

Human views of morality, which they tend to innately believe is correct and absolute, are self-centered and about self-preservation of their species, regularly at the expense and pain of other living things. Self-centered, self-preservation psychology exists with all animal species. To an intelligent non-human, human morality is not only arbitrary to the human species and myopic, it is immoral. Why should the non-human conform itself to human ideas and self-concerns?

19 WOULD SUPERINTELLIGENT ARTIFICIAL GENERAL INTELLIGENCE FIND THE ULTIMATE TRUTHS AND MEANING OF THE UNIVERSE?

No. No matter how intelligent it gets, artificial intelligence will still be limited and have many of the cognitive issues of the human mind. It will still be subjective, not know the reliability of its own mind, have limited (if more expansive than human) senses, cognitive limitations and conflicts, still have the informational limits due to existing as a small dot in time and space. Meaning and absolute/universal truths may just be human concepts, and we don't even really know if there are such things as ultimate truth or meaning. AI may be searching for things that don't exist.

Many will respond to this by saying that, if AI finding ultimate truths is impossible, we should use it for practical purposes for humans on earth. This is a standard existential response.

20 HOW WILL ARTIFICIAL INTELLIGENCE AFFECT RELIGION?

So long as there are humans, there will be religions. Artificial intelligence, especially sentient artificial general intelligence, will challenge and change religions.

Religions come in a wide variety of beliefs and systems. They are human centric, about human relations and connections with higher powers or transcendent realities. They are attempts to answer the big human questions: "What is the meaning and nature of the universe?", "What is a human's purpose and place on earth?", "Where did humans come from and what happens after they die?", "How should humans act?", etc.

There are two basics aspects of religions. One is the structure, dogma, texts and spelled-out belief. These often give the specifics of the beliefs, rules of conduct, the required ceremonies, social and community order. Even if a religion involves insight into a transcendent reality, its structure is in part formed by its culture, language, time and place, politics and circumstance.

The major second aspect is the mystical mental attempt to become closer to God, the universe, transcendent reality or whatever is their belief is the ultimate power and being. Often in the form of an altered state mystical experience, this was usually the genesis of the religion, while the dogma is constructed in response to that.

Not all religions are theistic. While Abrahamic religions believe in a single god, Buddhism is atheist and believes in higher, transcendent reality. Similarly, Hinduism believes in a greater reality or intelligence (Brahman), and many ancient aboriginal religions worshiped not a god but a greater life force (such as American Indians' 'Great Spirit') or a mystical, enlightened sense of reality.

Whatever is the perceived higher power or transcendent reality, all religious people try in some fashion to become closer to it. Catholics try to have a personal relationship with Jesus and their God, while Buddhists and Hindus try to gain enlightenment. All religious have their mystical traditions and subgroups, and some religions, such as Buddhism and the mentioned aboriginal religions, are seen as essentially mystical.

Artificial general intelligence, computers with sentience, superintelligent cyborgs and transhumanism would upset the order and beliefs of religions that have strict dogma and ancient scripture that says humans are the 'highest' being on earth and the only one with sentience and souls. Many religions are against humans 'playing God.' Many religious people will be against AGI, and may be one of the influences that slow it and affect its nature. Many government policy-makers, industry leaders and people who vote are religious or are influenced by traditional religious cultures and beliefs.

Artificial intelligence will find new information, perspectives and insight that will contradict or conflict with religious sacred texts and dogma, just as Darwin's and Copernicus' findings contradicted centuries old Christian scripture and beliefs. There will remain the hardcore believers no matter what the new facts and insights. However, the new facts and perspectives will cause shifts in these religions and prompt many people to leave these religions. In the modern age of scientific discovery and reason, many have left the Abrahamic religions or, as with the Jews, become secular, while the religions and the believers have been been changed by science and new views. Many Abrahamic believers follow science and believe the world is round and the sun in the center of the galaxy, while many denominations have changed their views and practices surrounding women, race and the environment. Some Christian and Jewish denominations have ordained women as ministers and rabbis, something that was unthinkable just decades before.

Artificial intelligence is not incompatible with some ancient religions. Hinduism and Buddhism are about methods to expand the mind and reach mental enlightenment in order to know about the self and transcendent reality. The ultimate goal of these religions is to gain complete knowledge and transcendental intelligence. If artificial intelligence aids in expanding the mind and consciousness, this would go hand in hand with these religions and should be accepted.

Similarly, mysticism across religions uses various methods to expand the mind and learn more about transcendent reality. These methods include ceremonies, meditation, dance, art, music, prayer, fasting, lifestyles and drugs. Artificial intelligence and transhumanism would be used as methods to expand the mind.

Whether you call them religions or belief systems, world pantheism and secular humanism are atheistic belief systems (or religions) that believe in science and human reason. As a method of scientific and fact discovering, artificial intelligence would influence these belief systems. Alternately, these belief systems are as human centric as theistic religions, at least in the sense that they have human created dogmas, and a differently-thinking artificial general intelligence would by its existence challenge the dogmas.

New religions will be formed, based on or influenced by artificial intelligence. Interspiritual religions incorporate aspects of various religions, along with secular and scientific views. Some new religions will envision creating an artificial god. However, really, at their best they will use artificial intelligence as means for spiritual and intellectual exploration.

21 WHAT WILL ARTIFICIAL INTELLIGENCE BE LIKE IN THE FUTURE?

Different than you think. The history of predicting AI, and most any technology, has been the history of people being off, often comically.

Realize that the future of AI will be formed not only by the theory, inventions and computer science, but by economics and resources, plus the whims, changing opinions and psychology of industry leaders, funders, political and civic leaders, and the public. Artificial intelligence will be influenced by political and social leaders who may have little understanding of the technology.

And, of course, artificial intelligence will help in forming itself.

Further reading

The worst tech predictions of the past 100 years (freecodecamp.org)

22 BOOK ENCOMPASSING QUESTIONS

1. What is your definition of artificial intelligence?
2. Does artificial intelligence have to have human intelligence?
3. Give two reasons why people fear artificial intelligence?
4. Do you think artificial general intelligence with sentience and consciousness will happen?
5. Why is symbolic artificial intelligence called 'top down,' while deep learning is called 'bottom up'?
6. Give three things AI currently does, and two things that people hope it will do in the future?
7. By searching online, explain what is 'AI winter.'
8. What is the Turing test?
9. Why did John Searle and Hubert Dreyfus say the Turing test doesn't identify the existence of human thinking in a computer?
10. Why do some say you need more than one type of test to determine if AI has human-like intelligence?
11. What is singularity?
12. What is a cyborg?
13. Should artificial general intelligence with sentience, feelings and thoughts receive legal human or animal rights?
14. Give an example of an ethical concern surrounding artificial intelligence
15. Should AI always be tied to human interests?
16. What is the 'black box problem'?
17. Does artificial general intelligence require human-like consciousness and sentience?
18. Does artificial general intelligence require a biological component to achieve sentience?
19. Does artificial general intelligence need a body and social interaction?
20. Would sentient artificial general intelligence have the same morals and ethics as humans? If not, is that good or bad?
21. Can we ever be sure that artificial intelligence has real emotions?
22. Does it matter if a computer has sentience and is really thinking?
23. What is the 'common sense problem' in AI?
24. What are two things that humans do better than computers?
25. What are two things that computers or robots can do or will be able to do better than humans?