

LUCA

Data-Driven Decisions

WHITEPAPER_

Surviving the AI Hype – Fundamental concepts to understand Artificial Intelligence

23.12.2016

A *Telefonica* DATA UNIT

luca-d3.com

Index

1. Introduction.....	3
2. What are the most common definitions of AI?	3
3. What are the sub areas of AI?	5
4. How “intelligent” can Artificial Intelligence get?	7
4.1. Strong and weak AI	7
4.2. The Turing Test	7
4.3. The Chinese Room Argument	8
4.4. The Intentional Stance	8
4.5. To what extend can machines have “general intelligence”?.....	9
4.5.1. Qualitative reasoning.....	10
4.5.2. Reflective reasoning.....	10
5. Can machines think? Or... Are humans machines?.....	10
5.1. Symbolic vs non-symbolic AI.....	10
5.2. The final question: Can machines think? Are humans machines?.....	12
6. Conclusion	12
7. References.....	13
About LUCA.....	14
More information	14

Surviving the AI Hype – Fundamental concepts to understand artificial intelligence

By Dr. V. Richard Benjamins, VP for External Positioning and Big Data for Social Good at LUCA

1. Introduction

(1) *Artificial Intelligence* (AI) is one of the hottest topic out there at the moment, and often it is merely associated with chatbots such as Siri or other cognitive programs such as Watson. However, AI is much broader than just that. To understand what we read in the press and on the Internet, it is important to understand some of the "AI basics", which are often lost in the midst of AI hype out there at the moment. By understanding these fundamental principles, you will be able to make your own judgment on what you read or hear about AI.

To understand AI, it is important to understand some of its basic underlying concepts.

2. What are the most common definitions of AI?

So, first of all, how does Google (one of the kings of AI) define Artificial Intelligence?

ar·ti·fi·cial in·tel·li·gence

noun

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.

FIGURE 1: A POPULAR DEFINITION OF ARTIFICIAL INTELLIGENCE (GOOGLE).

There are many definitions of AI available online, but all of them refer to the same idea of machine intelligence, however, they differ in where they put the emphasis, which is what we have analysed below (an overview of these definitions can be found [here](#)).

For example, [Webster](#) gives the following definition (Figure 2):

Full Definition of ARTIFICIAL INTELLIGENCE

- 1 : a branch of computer science dealing with the simulation of intelligent behavior in computers
- 2 : the capability of a machine to imitate intelligent human behavior

FIGURE 2: THE OFFICIAL WEBSTER DEFINITION OF ARTIFICIAL INTELLIGENCE.

All definitions, of course, emphasise the presence of machines which are capable of performing tasks which normally require human intelligence. For example, Nilsson and Minsky define AI in the following ways:

- *"The goal of work in artificial intelligence is to build machines that perform tasks normally requiring human intelligence." (Nilsson, Nils J. (1971), Problem-Solving Methods in Artificial Intelligence (New York: McGraw-Hill): vii.)*
- *"The science of making machines do things that would require intelligence if done by humans." (Marvin Minsky)*

Other definitions put emphasis on a temporary dimension, such as that of Rich & Knight and Michie:

- *"AI is the study of how to make computers perform things that, at the moment, people do better." (Elaine Rich and Kevin Knight).*
- *"AI is a collective name for problems which we do not yet know how to solve properly by computer." (Michie, Donald, "Formation and Execution of Plans by Machine," in N. V. Findler & B. Meltzer (eds.) (1971), Artificial Intelligence and Heuristic Programming (New York: American Elsevier): 101-124; quotation on p. 101.)*

AI is the study of how to make computers perform things that, at the moment, people do better.

The above two definitions portray AI as a moving target making computers perform things that, at the moment, people do better. 40 years ago imagining that a computer could beat the world champion of chess was considered AI. However, today, this is considered normal. The same goes for speech recognition; today we have it on our mobile phone, but 40 years ago it seemed impossible to most.

On the other hand, other definitions highlight the role of AI as a tool to understand human thinking. Here we enter into the territory of Cognitive Science, which is currently being popularized through the term Cognitive Computing (mainly by IBM's Watson).

- *By Artificial Intelligence I therefore mean the use of computer programs and programming techniques to cast light on the principles of intelligence in general and human thought in particular." (Boden, Margaret (1977), Artificial Intelligence and Natural Man, New York: Basic Books)*
- *"AI can have two purposes. One is to use the power of computers to augment human thinking, just as we use motors to augment human or horse power. Robotics and expert systems are major branches of that. The other is to use a computer's artificial intelligence to understand how humans think. In a humanoid way. If you test your programs not merely by what they can accomplish, but how they accomplish it, then you're really doing cognitive science; you're using AI to understand the human mind." -- Herbert Simon*

AI also studies the principles of intelligence in general and human thinking in particular

Some however take a much more concise and less scientific approach with definitions such as:

- *"AI is everything we can't do with today's computers."*
- *"AI is making computers act like those in movies." (Her, AI, Ex Machina, 2001: A Space Odyssey, etc.)*

From all of these definitions, the important points to remember are:

- AI can solve complex problems which used to be performed by people only.
- What we consider today as AI, may just become commodity software in the not so distant future.
- AI may shed light on how we, people, think and solve problems.

3. What are the sub areas of AI?

Looking at the introductory table of content of any AI textbook will quickly reveal what are considered to be the sub-fields of AI, and there is ample consensus that the following areas definitely belong to it: **Reasoning, Knowledge Representation, Planning, Learning, Natural Language Processing (communication), Perception** and the **Ability to Move and Manipulate objects**. While each of those areas is a complete research discipline in itself, below we will briefly paraphrase what it means for a computer to manifest those behaviours:

There is a difference between Information Retrieval (search) and Reasoning (making inferences).

- **Reasoning.** People are able to deal with facts (who is the president of the United States), but also know how to reason, e.g. how to deduce new facts from existing facts. For instance, if I know that all men are mortal and that Socrates is a man, then I know that Socrates is mortal, even if I have never seen this fact before. There is a difference between Information Retrieval (like Google search: if it's there, I will find it) and reasoning (like [Wolfram Alpha](#): if it's not there, but I can deduce it, I will still find it).
- **Knowledge Representation.** Any computer program that reasons about things in the world, needs to be able to represent virtually the objects and actions that correspond to the real world. If I want to reason about cats, dogs and animals, I could represent something like `isa(cat, animal)`, `isa(dog, animal)`, `has_legs(animal, 4)`. This representation allows a computer to deduce that a cat has 4 legs, because it is an animal, not because I have represented explicitly that a cat has 4 legs, e.g. `has_legs(cat, 4)`.
- **Planning.** People are planning constantly: if I have to go from home to work, I plan what route to take to avoid traffic. If I visit a city, I plan where to start, what to see, etc. For a computer to be intelligent, it needs to have this capability too. Planning requires a knowledge representation formalism that allows to talk about objects, actions and about how those actions change the objects, or, in other words, change the state of the (virtual) world. Robots and self-driving cars incorporate the latest AI technology for their planning processes. One of the first AI planners was [STRIPS](#) (Stanford Research Institute Problem Solver), that used a formal language to express states and state-changes in the world, as shown in Figure 3.

AI requires computers to be able to perform:

- Reasoning
- Knowledge Representation
- Planning
- Learning
- Natural Language Processing
- Perception
- Ability to Move and Manipulate objects

Example : Blocks World

•STRIPS : A planning system – Has rules with precondition deletion list and addition list

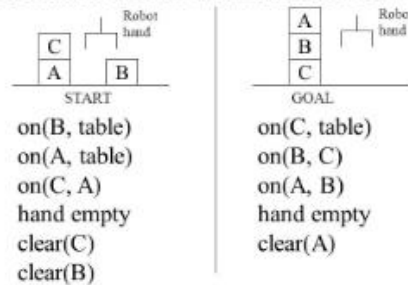


FIGURE 3: THE STRIPS PLANNER TO BUILD A PILE OF BOOKS.

- Learning.** Today this is probably the most popular aspect of AI. Rather than programming machines to do what they are supposed to do, machines are able to learn automatically from data: Machine Learning. Throughout their life, and especially in the early years, humans learn an enormous amount of things, such as talking, writing, mathematics, etc. Empowering machines with that capability makes them intelligent to a certain extent. Machines are also capable of improving their performance through learning by doing. Thanks to the popularity of Big Data, there is a vast amount of publications on [Machine Learning](#), as well as cloud-based tools to run ML algorithms as you need them, e.g. [BigML](#).
- Natural Language Processing.** We, humans, are masters of language processing since communication is one of the aspects that make humans stand out of other living things. Therefore, any computer program that exhibits similar behaviour is supposed to possess some intelligence. NLP is already part of our digital live. We can ask Siri questions, and we get answers, which implies that Siri processes our language and knows what to respond (oftentimes).
- Perception.** Using our 5 senses, we constantly perceive and interpret things. We have no problem in attributing some intelligence to a computer that can "see", e.g. can recognize faces and objects in images and videos. This kind of perception AI is also amply present in our current digital life.
- Move and Manipulate objects.** This capability is above all important for robotics. All our cars are assembled by robots, though they do not look like us. However, androids look a bit like us and need to manipulate objects all the time. Self-driving cars are another clear example of this, manifesting this intelligent capability.

Machine Learning is one of the most popular areas of AI today.

Another hot topic is NLP: crunching millions of text documents to extract knowledge.



FIGURE 4: SELF-DRIVING CARS COMBINE CAPABILITIES OF AI.

4. How “intelligent” can Artificial Intelligence get?

In this section, we will discuss several notions which are important for understanding the limits of AI.

4.1. Strong and weak AI

When we speak about how far AI can go, there are two “philosophies”: **strong AI** and **weak AI**. The most commonly followed philosophy is that of **weak AI**, which means that machines can manifest certain intelligent behaviour to solve specific (hard) tasks, but that they will never equal the human mind. However, **strong AI** believes that it is indeed possible. The difference hinges on the distinction between *simulating* a mind and actually having a mind. In the words of John Searle, “*according to Strong AI, the correct simulation really is a mind. According to Weak AI, the correct simulation is a model of the mind.*”

The most prevalent version of AI today is weak AI: computers can crack hard tasks, but will never equal the human mind

4.2. The Turing Test

The **Turing Test** was developed by Alan Turing in the 1950s and was designed to evaluate the intelligence of a computer holding a conversation with a human. The human cannot see the computer and interacts with it through an interface (at that time by typing on a keyboard with a screen). In the test, there is a person who asks questions and either another person or a computer program responds. There are no limitations as to what the conversation can be about. The computer passes the test if the “asking” person cannot distinguish whether the answers or the conversation comes from the computer or the person. **ELIZA** was the first program that challenged the Turing Test, even though it unquestionably failed. A modern version of the Turing Test was recently featured in the 2015 movie **Ex Machina**. So far, no computer or machine has passed the Turing Test.

A computer that passes the Turing Test cannot be distinguished from a human, regarding answering questions.

4.3. The Chinese Room Argument

A very interesting thought experiment in the context of the Turing Test is the so-called "Chinese Room Experiment" which was invented by John Searle in 1980. This experiment argues that a program can never give a computer the ability to really "understand", regardless of how human-like or intelligent its behaviour is. It goes as follows: Imagine you are inside a closed room with door. Outside the room there is a Chinese person that slips a note with Chinese characters under the door. You pick up the note and follow the instructions in a large book that tells you exactly, for the symbols on the note, what symbols to write down on a blank piece of paper. You follow the instructions in the book and you produce a new note, which you slip under the door. The note is picked up by the Chinese person who perfectly understands what is written, writes back and the whole process starts again, meaning that a real conversation is taking place.

A thought experiment to challenge what it means to "understand" a complex task.



FIGURE 6: THE CHINESE ROOM THOUGH EXPERIMENT. DOES THE PERSON IN THE ROOM UNDERSTAND CHINESE?

The key question here is whether you understand the Chinese language. What you have done is received an input note and followed instructions to produce the output, without understanding anything about Chinese. The argument is that a computer can never understand what it does, because - like you - it just executes the instructions of a software program. The point Searle wanted to make is that even if the behavior of a machine seems intelligent, it will never be really intelligent. And as such, Searle claimed that the Turing Test was invalid.

4.4. The Intentional Stance

Related to the Turing test and the Chinese Room argument, [the Intentional Stance](#), coined by philosopher Daniel Dennett in the seventies, is also of relevance for this discussion. The Intentional Stance means that "intelligent behaviour" of machines is not a consequence of how machines come to manifest that behaviour (whether it is you following instructions in the Chinese Room or a computer following program instructions). Rather it is an effect of people attributing intelligence to a machine because the behavior they observe requires intelligence if people would do it. A very simple example is when we say that our personal computer is thinking if it takes more time than we expect to perform an action. The fact that ELIZA was able to fool some people refers to the same phenomenon: due to the reasonable answers that ELIZA sometimes gives, people assume it must have some intelligence. But we

Intelligence is in the eye of the beholder

know that ELIZA is a simple pattern matching rule-based algorithm with no understanding whatsoever of the conversation it is engaging in. The more sophisticated software becomes, the more we are likely to attribute intelligence to that software. From the Intentional Stance perspective, people attribute intelligence to machines when they recognize intelligent behaviour in them.

4.5. To what extent can machines have “general intelligence”?

One of the main aspects of human intelligence, is that we have general intelligence which always works to some extent. Even if we don't have much knowledge about a specific domain, we are still able to make sense out of situations and communicate about them. Computers are usually programmed for specific tasks, such as planning a space trip or diagnosing a specific type of cancer. Within the scope of the subject, computers can exhibit a high degree of knowledge and intelligence, but performance degrades rapidly outside that specific scope. In AI, this phenomenon is called **brittleness** (as opposed to **graceful degradation**, which is how humans perform). Computer programs perform very well in the areas they are designed for, outperforming humans, but don't perform well outside of that specific domain. This is one of the main reasons why it is so difficult to pass the Turing Test, as this would require the computer to be able to “fool” the human tester in any conversation, regardless of the subject area.

Computer performance degrades abruptly. Human performance degrades gracefully.

In the history of AI, several attempts have been made to solve the brittleness problem. The first expert systems were based on the rule-based paradigm representing associations of the type *if X and Y then Z; if Z then A and B*, etc. For example, in the area of car diagnostics, *if a car doesn't start, then the battery may be flat or the starter motor may be broken*. In this case, the expert system would ask the user (who has the problem) to check the battery or to check the starter motor. The computer drives the conversation with the user to confirm observations, and based on the answers, the rule engine leads to the solution of the problem. This type of reasoning was called *heuristic* or *shallow reasoning*. However, the program doesn't have any deeper understanding of how a car works; it knows the knowledge that is embedded in the rules, but cannot reflect on this knowledge. Based on the experience of those limitations, researchers started thinking about ways to equip a computer with more profound knowledge so that it could still perform (to some extent) even if the specific knowledge was not fully coded. This capability was coined “*deep reasoning*” or “*model-based reasoning*”, and a new generation of AI systems emerged, called “*Knowledge-Based Systems*”.

For a computer to “understand” something, requires that it have a representation of that something.

In addition to specific association rules about the domain, such systems have an explicit model about the subject domain. If the domain is a car, then the model would represent a structural model of the parts of a car and their connections, and a functional model of how the different parts work together to represent the behaviour of the car. In the case of the medical domain, the model would represent the structure of the part of the body involved and a functional model of how it works. With such models the computer can reason about the domain and come to specific conclusions, or can conclude that it doesn't know the answer.

The more profound the model is a computer can reason about, the less superficial it becomes and the more it approaches the notion of *general intelligence*.

There are two additional important aspects of general intelligence where humans excel compared to computers: *qualitative reasoning* and *reflective reasoning*.

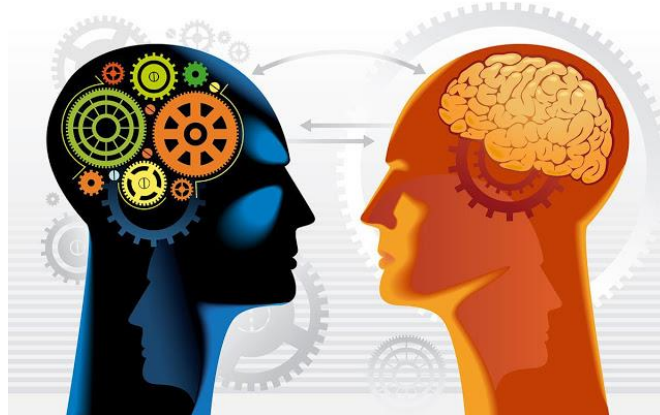


FIGURE 7: BOTH QUALITATIVE REASONING AND REFLECTIVE REASONING DIFFERENTIATE US FROM COMPUTERS

4.5.1. Qualitative reasoning

Qualitative reasoning refers to the ability to reason about continuous aspects of the physical world, such as space, time, and quantity, for the purpose of problem solving and planning. Computers usually calculate things in a quantitative manner, while humans often use a more qualitative way of reasoning (if X increases, then Y also increases, thus ...). The qualitative reasoning area of AI is related to formalisms and processes to enable a computer to perform qualitative reasoning steps.

4.5.2. Reflective reasoning

Another important aspect of general intelligence is reflective reasoning. During problem-solving, people are able to take a step back and reflect on their own problem-solving process, for instance, if they find a dead-end, and need to backtrack to try another approach. Computers usually just execute a fixed sequence of steps which the programmer has coded, with no ability to reflect on the steps they make. To enable computers to reflect on their own reasoning process, it needs to have knowledge about itself; some kind of meta knowledge. For my [PhD research](#), I built an AI program for diagnostic reasoning that was able to reflect on its own reasoning process and select the optimal method depending on the context of the situation.

Machines “reason” with numbers. Humans can reason also in a qualitative way (more, less, increases, etc.)

Reflective reasoning is reasoning about oneself or -for a computer- about itself.

5. Can machines think? Or are humans machines?

5.1. Symbolic vs non-symbolic AI

This dimension for understanding AI refers to how a computer program reaches its conclusion. **Symbolic AI** refers to the fact that all steps are based on “symbolic” human-readable representations of the problems that use logic and search to solve problems. Expert Systems are a typical example of symbolic AI as the knowledge is encoded in IF-THEN rules which are understandable by people. NLP systems that use grammars to parse language also are symbolic AI systems. Here the symbolic representation is the grammar of the language.

People are able to follow reasoning steps of a computer when it uses symbols that represent the real world, but we have more difficulties with understanding reasoning steps solely expressed in numbers.

The main advantage of symbolic AI is that the reasoning process can be easily understood by people, which is a very important factor when taking important decisions. A symbolic AI program can explain why a certain conclusion is reached and what the intermediate reasoning steps have been. This is key for using AI systems that give advice on medical diagnosis because if doctors cannot understand why an AI system comes to its conclusion, it is harder for them to accept the advice.

Non-symbolic AI systems do not manipulate a symbolic representation to find solutions to problems. Instead, they perform calculations according some principles that have demonstrated to be able to solve problems, without exactly understanding how to arrive at their solutions. Examples include genetic algorithms, neural networks and deep learning. The origin of non-symbolic AI comes from the attempt to mimic the workings of the human brain; a complex network of highly interconnected cells whose electrical signal flows, decide how we, humans, behave. It is also called "connectionist" AI.

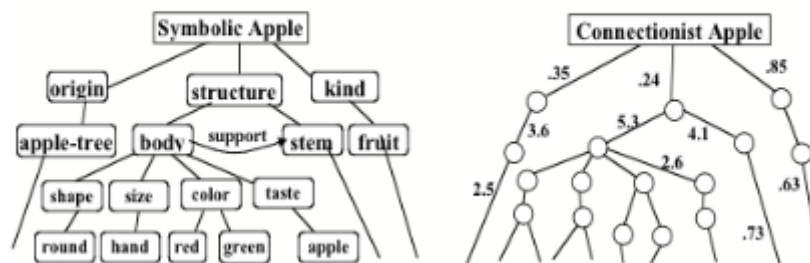


FIGURE 8: A SYMBOLIC AND NON-SYMBOLIC REPRESENTATION OF AN APPLE (SOURCE [HTTP://WEB.MEDIA.MIT.EDU/~MINSKY/PAPERS/SYMBOLICVS.CONNECTIONIST.HTML](http://web.media.mit.edu/~minsky/papers/SymbolicVs.Connectionist.html)).

Today, non-symbolic AI, through deep learning and other Machines Learning algorithms, is achieving very promising results, championed by IBM's Watson, Google's work on automatic translation (which has no understanding of the language itself, it "just" looks at co-occurring patterns), Facebook's algorithm for face recognition, self-driving cars, and the popularity of deep learning. The main disadvantage of non-symbolic AI systems is that no "normal" person can understand how those systems come to their conclusions or actions, or take their decisions. See for example Figure 8: in the left part we can understand easily why something is an apple, but looking at the right part, we cannot easily understand why the system concludes that it's an apple. When non-symbolic (aka connectionist) systems are applied to critical tasks such as medical diagnosis, self-driving cars, legal decisions, etc, understanding why they come to a certain conclusion through a human-understandable explanation is very important. In the end, in the real world, somebody needs to be accountable or liable for the decisions taken. But when an AI program takes a decision and no-one understands why, then our society has an issue (see [FATML](#), an initiative that investigates Fairness, Accountability, and Transparency in Machine Learning).

A "symbolic" apple versus a "non-symbolic" apple.

Probably the most powerful AI systems will come from a combination of both approaches.

5.2. The final question: Can machines think? Are humans machines?

It is now clear that machines certainly can perform complex tasks that would require "thinking" if performed by people. But can computers have consciousness, can they have, feel or express emotions? Or, are we, people, machines? After all our body and brain are based on a very complex "machinery" of mechanical, physical and chemical processes, that so far, nobody has fully understood. There is a research field called "computational emotions" which tries to build programs that are able to express emotions. But maybe expressing emotions is different than feeling them (see Intentional Stance)?

Can computers express or feel emotions?

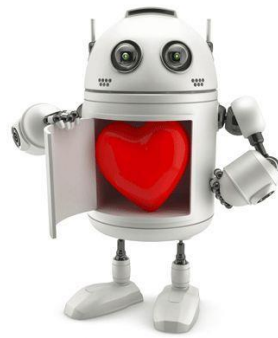


FIGURE 9: CAN COMPUTERS FEEL EMOTIONS?

Another critical issue for the final question is whether machines can have consciousness. This is an even trickier question than whether machines can think. I will leave you with this [MIT Technology Review interview](#) with [Christof Koch](#) about "What It Will Take for Computers to Be Conscious", where he says: "... consciousness is a property of complex systems that have a particular "cause-effect" repertoire. They have a particular way of interacting with the world, such as the brain does, or in principle, such as a computer could."

Whether machines can have consciousness is a matter of belief.

In my opinion, currently, there are no scientific answers to those questions, and whatever you may think about it, is more a belief or conviction than a commonly accepted truth or a scientific result. Maybe we have to wait until 2045, which is when Ray Kurzweil predicts [technological singularity](#) to occur: the point when machines become more intelligent than humans. While this point is still far away and many believe it will never happen, it is a very intriguing theme evidenced by movies such as 2001: A Space Odyssey, A.I. (Spielberg), Ex Machina and Her, among others.

6. Conclusion

The term Artificial Intelligence is nowadays used for many things that some time ago were called differently, e.g. Big Data, Machine Learning. It is also often used as a synonym for Chatbots. Whereas the increase in interest and attention is positive for an exciting field as AI is, it is important to keep in mind what we are talking about. AI is not only about fancy applications, it is also about fundamental questions related to cognition, how people think, how they solve problems and how they approach unforeseen situations.

AI is out there for more than 50 years, but the improvements in recent years have been spectacular. And this is only the beginning...

7. References

<http://www.cse.buffalo.edu/~rapaport/definitions.of.ai.html>

[http://www.merriam-webster.com/dictionary/artificial intelligence](http://www.merriam-webster.com/dictionary/artificial%20intelligence)

<https://www.wolframalpha.com/>

<https://en.wikipedia.org/wiki/STRIPS>

<http://www.datasciencecentral.com/profiles/blogs/a-tour-of-machine-learning-algorithms>

<https://bigml.com/>

https://en.wikipedia.org/wiki/Turing_test

http://data-speaks.luca-d3.com/2016/11/chatbots-new-you-havent-met-eliza_10.html

<http://www.imdb.com/title/tt0470752/>

https://en.wikipedia.org/wiki/Chinese_room

https://en.wikipedia.org/wiki/Intentional_stance

Benjamins, V.R. Problem Solving Methods for Diagnosis, PhD Thesis, University of Amsterdam, Amsterdam, The Netherlands, 1993:

<https://books.google.es/books?id=uKoYAAAACAAJ&dq=inauthor:%22Victor+Richard+Benjamins%22&hl=es&sa=X&ved=0ahUKewjl19Xh9bfQAhWl1MKHe4DB8sQ6AEIHTAA>

<http://www.fatml.org/>

<https://www.technologyreview.com/s/531146/what-it-will-take-for-computers-to-be-conscious/>

Kurzweil, Ray. The Singularity Is Near: When Humans Transcend Biology. Penguin Books, 2006 (https://en.wikipedia.org/wiki/The_Singularity_Is_Near)

About LUCA

LUCA's mission is to bring Telefónica's know-how in transforming into a data-driven organisation to other private and public sector organisations in a wide range of sectors including Retail, Tourism, Outdoor Media, Financial Services and Transport, just to name a few. Our diverse product portfolio, which brings together expertise in Data Engineering, Data Science, Infrastructure, Cybersecurity and Business Insights, enables companies to continue their Big Data journey with a wide range of solutions to propel them to double-digit growth.

More information

www.luca-d3.com

[@LUCA_D3](https://twitter.com/LUCA_D3)

data-speaks.luca-d3.com

2016 © Telefónica Digital España, S.L.U. All rights reserved.

The information disclosed in this document is the property of Telefónica Digital España, S.L.U. ("TDE") and/or any other entity within Telefónica Group and/or its licensors. TDE and/or any Telefonica Group entity or TDE'S licensors reserve all patent, copyright and other proprietary rights to this document, including all design, manufacturing, reproduction, use and sales rights thereto, except to the extent said rights are expressly granted to others. The information in this document is subject to change at any time, without notice.

Neither the whole nor any part of the information contained herein may be copied, distributed, adapted or reproduced in any material form except with the prior written consent of TDE.

This document is intended only to assist the reader in the use of the product or service described in the document. In consideration of receipt of this document, the recipient agrees to use such information for its own use and not for other use.

TDE shall not be liable for any loss or damage arising out from the use of the any information in this document or any error or omission in such information or any incorrect use of the product or service. The use of the product or service described in this document are regulated in accordance with the terms and conditions accepted by the reader.

TDE and its trademarks (or any other trademarks owned by Telefonica Group) are registered service marks.