

Adimen artifizialaren nozioa, izan dituen lorpen berrietako batzuk, dauzkan mugak, eta humanitate digitalen nozio berriaren bidez humanitateekin batera nola ikus daitekeen eta nola ikusi behar den eztabaidatuko dugu lan honetan.

Giltza-Hitzak: Adimen Artifiziala. Sare Neuronalak. Ikaskuntza Sakona. Humanitate Digitalak.

En este trabajo discutimos la noción de inteligencia artificial, algunos de sus logros recientes, sus limitaciones y cómo puede y debe contemplarse junto a las humanidades, a través de la nueva noción de humanidades digitales.

Palabras clave: Inteligencia Artificial. Redes Neuronales. Aprendizaje Profundo. Humanidades Digitales.

Dans ce travail, nous discutons de la notion d'intelligence artificielle, de certaines de ses réalisations récentes, de ses limites et de la manière dont elle peut et doit être considérée conjointement avec les humanités, à travers la nouvelle notion d'humanités numériques.

Mots-Clés : Intelligence Artificielle. Réseaux De Neurones. Deep Learning. Digital Humanities.

## From Artificial Intelligence to Digital Humanities

**Bustince, Humberto**

Jakiunde, Zientzia, Arte eta Letren Akademia. Prim, 7. E-20006 – Donostia  
Universidad Pública de Navarra. Departamento de Estadística, Informática y Matemáticas. Campus Arrosadía, s/n  
E-31006 Pamplona  
bustince@unavarra.es

Recep.: 2022-06-20  
Acept.: 2022-11-21

## 1.-What is Artificial Intelligence

### 1.1 The idea of artificial intelligence

Nowadays, the notion of artificial intelligence pervades our common life. From last medical techniques to social life, from education to tourism, it seems that any new development should involve it. What is not so clear is whether our society actually has a clear idea of what artificial intelligence is and what artificial intelligence implies.

The idea of an intelligent machine originated probably at the same time as technology. However, it has always been an uncertain concept, which have been evolving along time. This is not surprising, since it is not easy to provide a definition of intelligence. Probably, for instance, a modern pocket calculator would have been considered intelligent, if not magic, by any person in the tenth century [Nilsson,09]. It is clear that the notion of artificial intelligence, as we consider it nowadays, goes further than those miraculous machines

Just omitting a very long history on the discussion about what intelligence is, and in order to fix a date, we may consider that the origin of the modern notion of artificial intelligence is at a conference which was held at Dartmouth, Hanover, USA, in 1956 [McCarthy et al, 55]. But we should not forget that modern artificial intelligence, as it is the case for any branch of knowledge, cannot be understood without previous developments, being specially relevant the works by Alan Turing in the thirties of the last century.

Let us go back to 1956. One of the researchers in Dartmouth was Marvin Minsky. Several years later, in 1968, he proposed which is, in our opinion, the most accurate definition of artificial intelligence [Minsky,68]. Paraphrasing his words:

Artificial intelligence is the science of creating machines to make things that, if they were made by human beings, they would be considered to involve intelligence.

Probably, this definition may be accepted by most of the people as comprising the modern notion of artificial intelligence. Note that it combines the notion of science and engineering, so it is assumed to live in the intersection of both fields. What does this mean? What is an intelligent machine? Or, going one step backwards, can a machine be intelligent? There is no simple answer to this question, not even a commonly accepted one, because we have a big original problem: we do not understand what intelligence is, and, although scientists in several different fields, including computer science, are devoting huge efforts in recent years to understand the human brain, we are very far from having an answer. Of course, we are not going even to try to provide an answer to this question on these pages. On the contrary, we would like just to make a review of some recent developments, in order to provide some insight on what we are speaking of when we speak of Artificial Intelligence.

At the same time, we would also like to focus on one aspect, which is implicit in the definition by Minsky. As we speak of science and engineering, does this mean that artificial intelligence is apart from humanities? This perception is implicitly shared by many people in our society. From our point of view, it is not correct. There are several reasons to support this opinion. Firstly, as we have already said, at the very basis of artificial intelligence lays a philosophical question, namely, what is intelligence? Note that an answer to this question would have very strong ethical implications that we cannot and should not ignore. Besides, the work on artificial intelligence involves a permanent question on which are the admissible limits for research in the field. In the case of Medicine, for instance, this ethical consideration is accepted and assumed as inherent.

Apart from this, the notion of artificial intelligence as an area separated from humanities leads also to a sort of mystification of the idea of artificial intelligence. For many people, artificial intelligence is understood as a technique and knowledge that is able to make everything in a way that is not possible to understand. In other words, as a sort of magic. This magic, as anything that we do not understand, is felt as a threat and received with fear. Even more, this feeling is reinforced by the vision in films of a world dominated by machines, by not very accurate headlines and, sometimes, even by the opinion of people who actually do understand artificial intelligence. Just recall that Stephen Hawking considered that artificial intelligence could be either the worst or the best that has happened to humanity and he signed with several hundreds of colleagues a manifesto about the risks posed by artificial intelligence [Haw15].

What can we find of truth in all these risks and threats? In the same moment these lines are being written, a discussion has originated in Google (<https://www.bbc.com/news/technology-61784011>) . One of its engineers claims that LaMDA, an artificial intelligence specifically design to create conversations as close to human ones as possible, has evolved and developed feelings. According to this engineer, LaMDA has recognized this fact talking to him. LaMda would have also expressed its fear to be switched off if this is known. Yes, it sounds really close to a film argument, but the engineer, Blake LeMoine, is a real and identifiable expert in Google, LaMDA do exist and the news is not a prank. So, have intelligence evolved in our machines? Well, for the moment, almost every expert agrees this is not the case. In fact, both for Google and for scientists outside the company, it is much more plausible that the engineer has fallen into humanization: a tendency common to human beings to identify patterns of humanity where they do not actually exist. Note that, in this case, LaMDA is an artificial intelligence devoted to develop conversational skills that are indistinguishable from human ones. Feelings and sentiments are expected to be present in human dialogues, either explicitly or implicitly. So it may be reasonable that LaMDA has been able to create the illusion of having sentiments by means of the words, but it is not clear at all that this means that such sentiments actually exist.

The case of LaMDA, however, provides an excellent example of the problems that may arise from artificial intelligence. Because, of course, artificial intelligence might involve very serious risks. But we must take into account that most of them, and the ones to which Stephen Hawking and many other scientists refer to, are not related to killer robots, self-evolved consciousness or a humanity enslaved by machines. In fact, the danger is much closer and much more real. It comes from the way artificial intelligence works, the use that we can make of artificial intelligence, and the interpretation we give to it. Because artificial intelligence can be devoted either to good or to evil, as any other science and knowledge. That does not depend on artificial intelligence itself, but on humans creating artificial intelligence. In this sense, it is crucial for our society that we destroy the veil of mystery around artificial intelligence, that we get a glimpse of how it actually works and which its limitations are, of which are the risks involved on it and on how to apply it].

How can we lift this veil? As in so many other fields, the best tool to do it is knowledge. In these pages, we want to help in this task. For this reason, we will provide a short summary about how a machine can learn and what this means, in order to put in a real basis artificial intelligence capabilities. Then we are also going to consider some applications. Later, rather than focusing on examples which are repeated once and again, we will move into the intersection between artificial intelligence and humanities, in the recently appeared field of Digital Humanities [Berry, 2011]. We will see how artificial intelligence can help social sciences and arts, as well as how it can draw inspiration from it. We hope that, in this way, this work can help to break the, from our point of view, baseless dichotomy between Sciences and Humanities, and at the same time, open a discussion on the fuzzy boundaries of what machines can do.

In any case, and to finish this introduction, let us observe that the discussions to follow are only part of the solution to the problem of artificial intelligence. We need to understand the “machine side” of the

equation, to say so, but we also need to understand the other side, the human side of the same equation. If LaMDA has created the illusion of a feeling intelligence, it is because of not only an excellent programming and a very well-designed algorithm. It is also because human mind is biased to detect human patterns everywhere. And this is a good prove of the need of having a very deep knowledge of human beings, too, to avoid misunderstandings, or even worse, misuses of artificial intelligence.

## 1.2 How do machines learn? The case of neural networks

In this subsection, we do not intend to make an in-depth review of all the methods that have been designed for machines to learn. Firstly, because it would force us to devote much more space than the available one. But secondly, because in order to get the idea behind the learning methods in artificial intelligence, such a review is not necessary. In fact, most of them can be briefly summarized as follows: Machines learn in the same way as humans do: by means of examples (mathematical induction method) and a test and trial method. Let us explain this point with an example.

Imagine for a second that we are going to learn to cook. We have never tried it, but we have the opportunity of learning under the supervision of the best chef in the world. As a first step, the chef proposes you to start with something easy. You see that on the table in front of you, you can find all the ingredients you may need, even many you do not actually know what they are. And, of course, we also have pans, saucepans, ovens and any tool we may think of.

The chef tells us to prepare spaghetti bolognesa. We have never done it, and the chef does not tell any more. So we start trying to do our best: boiling water, frying some tomato, slicing meat... We struggle more or less for one hour and here it is our dish. It more or less resembles pasta bolognesa, although we already realize the smell is not the one we were expecting. In any case, the chef takes a mouthful from the dish and tastes it. After swallowing it, the chef looks at us smiling and starts saying: maybe you have put too much salt, it would have been better to fry the meat with garlic and tomato rather than just boiling it. Then, after a list of comments, he invited us to repeat the cooking.

Of course, in our next attempt the result will not improve so much. Neither in the following ten. But, after a large number of repetitions, we will be able to make a perfect dish of bolognesa.

Of course, the method we have explained is quite strange, and, in reality, no course on cooking develops in this way. However, most of the machines that we have around nowadays learn in a way very similar to this one. Because in fact, this example already considers the main elements for the learning process:

- The data (the ingredients) which should be available for the machine in all the necessary quantities and from which an answer (the result of the learning, the dish) must be obtained. Observe that, since we start our learning from scratch, the variety and quality ingredients at our disposal will determine what we finally do. The same happens in artificial intelligence: the variety and quality of data determines the final goodness of the learning.
- The supervisor (the chef). A tool that determines whether the answer that we have obtained is good enough or, on the contrary, if it must be improved. In the case of machines, this supervisor usually takes the form of a mathematical function that determines how different the obtained answer is from the one we were expecting to get, suggesting how the former should be modified in order to arrive at a better output.

- The test and trial method. The idea of repeating once and again the same process until an acceptable answer is attained. Note that this method implies the possibility of a huge number of repetitions, which probably will be necessary for having a good result.

Observe that going back to our example on LaMDA, this is precisely the way we would expect it to work. The machine will create some sentences, and an evaluator will decide whether these sentences are human enough or not. Then the machine will use this information to try to improve the next sentences it creates. But the key point is that the procedure we have described in such a rough way is so general that it can actually be applied to a huge variety of problems, from teaching a machine to speak to creating one for detecting cats in Facebook photographs. In fact, this is the procedure followed by what is nowadays the dominant paradigm of artificial intelligence: neural networks.

Neural networks are today the crucial tool in almost every artificial intelligence application, and since they also personalize some of those myths around artificial intelligence itself, we are going to focus on the, in the following lines.

From the beginning, the intention of neural networks was to imitate the behaviour of our brain. For this reason, they are based on the notion of artificial neuron, or perceptron [Rosenblatt, 58, Goodfellow et al, 16]. This notion first appeared on a journal of Psychology, and we show its basic schema in Figure 1. Note that the perceptron follows the model of neuron described by Ramón y Cajal [Ramon, 1892]

Fig. 1 Basic schema of a perceptron

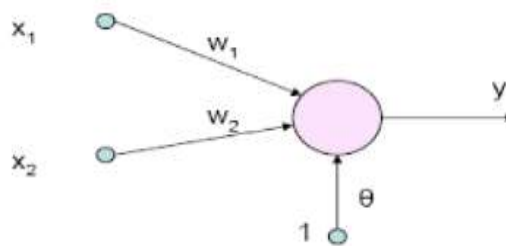
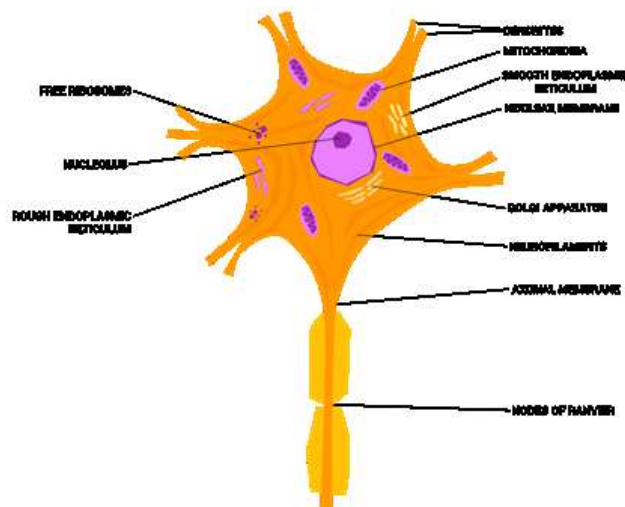


Fig. 2: Model of a neuron (Wikicommons). Note the similarity to the perceptron after a rotation of  $90^\circ$  (<https://commons.wikimedia.org/w/index.php?search=Model+of+a+neuron&title=Special:MediaSearch&go=Go&type=image>)

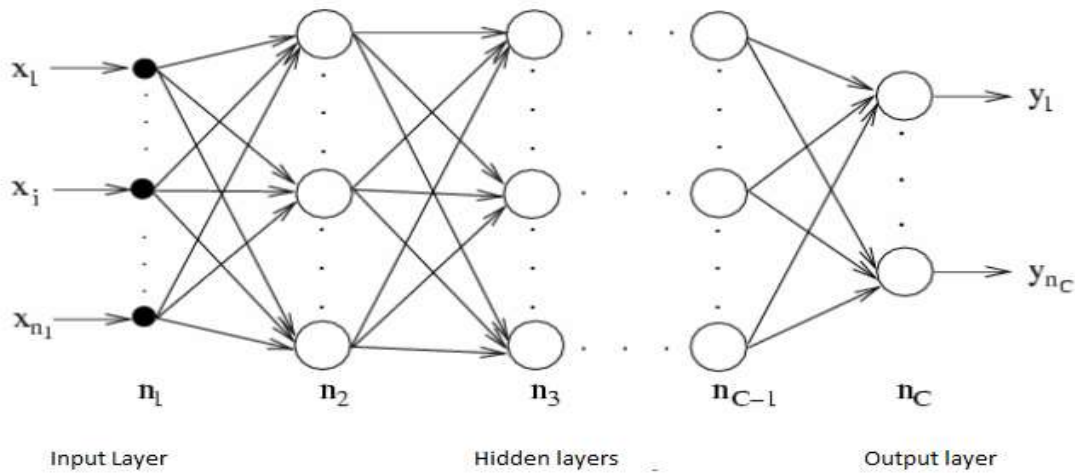


The perceptron takes some numerical inputs  $x_1, \dots, x_n$ , and modifies them according to some weights,  $w_1, \dots, w_n$ . Then, if the weighted mean of the inputs is above a threshold, it activates. If not, it does not

activate. The perceptron learns considering the difference between the obtained and expected activation (or not activation) for each specific example. Whenever the output is not the correct one, the weights are modified according to a well-known mathematical mechanism called the gradient rule.

This simple schema is able to classify data correctly in problems that are not very complex. But, as in the case of the human, the real power does not come from one single neuron, but from the combination of several neurons in a network in such a way that they are able to “collaborate” to solve complex tasks. This is the idea behind the notion of a neural network, as the one depicted below:

Fig. 3. Basic scheme of a multilayer neural network



In this case, the network learns in a way similar to that of the perceptron, calculating its output by means of a set of weights, one in each of the connections between different neurons. Then it compares the output to the one we were expecting to get, and changes the weights if the results is not good. However, the correction step is much more complicate in this case, and it was not until the eighties that an efficient method was developed to carry it on. This method, known as the backpropagation algorithm, is one of the basis of the actual developments in artificial intelligence [Rumelhart et al., 86]. The other main problem for the development of neural networks comes from the necessity of using a large amount of data and computational resources. Technical developments, however, led also to overcome this difficulty, open the way for the renowned deep neural networks.

Although they are technically difficult to implement, it is remarkable that deep neural networks basis are not complicate to understand from a theoretical point of view. In fact, they seem to be a rather weak imitation of human brain. It is worth to notice that the key element for a neural network to work is that of data. In fact, as we have already said, the quality of a network is given by the quality of the data at its disposal. In this sense, originally neural networks were able to work only with limited data of the type “temperature on a given region” or “amount of water in a dam”. Nowadays, however, deep neural networks are able to deal with a huge variety of data, ranging from photographs to words or sentences.

How good can be the results obtained by a neural network? In 2017, Brandt and Eagleman [Brandt et al, 2017] wrote in their book *The Runaway Species* that, while humans try to wonder or surprise the other, a computer is only able to give back exactly the data it has received. In fact, this seems to be the case for the neural networks that we have just explained. And, as Brandt and Eagleman suggested, apparently this fact excludes the possibility of sensibility, art, or the feelings we have commented before.

However, just five years after that book, we have artificial intelligence models that are able to create images, both hyperrealistic and abstract in many different styles and starting from a linguistic description. Are these creations art? Again, this is going beyond artificial intelligence itself, but at least the question is posed. So, are these networks creating something different from the data they use? Let us consider some examples.

### 1.3 GPT-3 and the revolution of language

In the previous subsection we speak about LaMDA as an artificial intelligence to create language. This kind of models have lived a huge expansion in recent years. In 2020, OpenAI presented its model GPT-3. This is a model that is able to “understand” sentences in English (really in many other languages) and provide an answer in the same language. It is an example of a so-called generative neural networks. That is, in this case, and in order to learn, the network does not only take into account external supervised data. It also competes with other neural networks to provide better results, in a similar way as we improve playing chess by playing against other people. In fact, this was the model behind the neural network that was able to win the Go game.

What is such artificial intelligence able to do? GPT-3 may keep a conversation with a surprising level of coherence and naturality. And it is also able to invent its own stories or even to provide explanations about how it works. In fact, here is an example of a poem writing by GPT-3 imitating Shakespeare and that can be found in <https://www.gwern.net/GPT-3#william-shakespeare>

Shall I compare thee to a summer's day?  
Thou art more lovely and more temperate:  
Rough winds do shake the darling buds of May,  
And summer's lease hath all too short a date;  
A winter's day, when beams of sun are few,  
And skies are grey with clouds intemperate,  
Is to thy beauty somehow sordid grown.  
Her lovely-lovely eye is quite shut up,  
Like windows pale before a sunny shower.  
The earth that's cold and moist, like to my mistress,  
When once that she unbrac'd, her lovely jewel,  
Grows as both ugly and uncourtly.  
Poor soul, the center of my motley host,  
See, now they request thy sovereign presence;  
The din of merry hearts hath brought me thus  
To greet thee, thou unsoil'd and saint-like beauty.

But how can GPT-3 do all this? In order to train GPT-3, millions of sentences from Internet were used. These included all types of sentences, from technical ones to familiar comments in social networks. The network was able to determine relations and links between this huge amount of data in such a way that it was even able to answer questions, something for which it had not been in principle designed. For instance, if it was asked, how can I cook a bolognesa, GPT-3?, the answer (and this is not invention, it is the real answer) is: “To cook bolognesa, first brown the meat in a large saucepan over medium-high heat. Add the garlic and cook for 1 minute. Add the tomato sauce and bring to a simmer. Cook for 10 minutes, or until the sauce is thickened. Serve over cooked pasta”. Of course, GPT-3 does not understand its answer. But it is able to relate concepts in a very surprising way, close to the way human do, or even better,



as it can handle much more data simultaneously. The large amount of data used allows the network to create these relations and provide its outputs.

#### 1.4 DALL-E and the generation of images

A year after the surprise of GPT-3, in January 2021, the same company presented a new revolution in the field of generative networks. They called their creation DALL-E, an homage both to Salvador Dalí and the robot WALL-E in the Pixar film. [Ramesh et al, 2021] And, at the beginning of this 2022, less than a year after, an even more surprising new version was introduced [Ramesh et al, 2022]

Both models are, in fact, similar to GPT-3, since they are able to understand a sentence in natural language in a very precise way. So, which is the novelty in this case and what has Dalí to do with this artificial intelligence? Well, DALL-E does not answer with a sentence or a paragraph, in the same way as GPT-3 does. What DALL-E makes is to generate an image in order to reflect the contents and the meaning of the provided sentence. In this case, so, the network has been trained not only with words or texts, but also with images of many different types. And due to the large number of images involved in its training, the network is able again to surprise us with some unexpected tricks. For instance, DALL-E has learnt to imitate the style of particular painters, as well as that of specific types of images.

Below we can find some examples of images created by DALL-E.



This first image correspond to the painting made by DALL-E-“ after providing the sentence: “a painting of a fox sit in a field in the sunrise in the style of Claude Monet”. Of course, this image corresponds to a very explicit and self-contained sentence, and the results fits the meaning of the proposal very accurately.



These are the results provided by DALL-E\_2 as output to the suggestions: “Portrait in a vibrant painting of Salvador Dalí with a robotic half-face” and “Publicity poster displays a cat dressed as the French emperor Napoleon and with a piece of cheese in its hand”. Note that the meaning of vibrant in the first image is not obvious, since it is not objective.

The use of DALL-E and DALL-E 2 is not open to the public in general. Nevertheless, they are not the only existing generative models that are able to solve this kind of tasks. For instance, by means of a combination of two networks, VQGAN [Esser, 21] and CLIP [Radford, 21], we have generated the following image:



This is the image provided by VQGAN as an output for the sentence: “A delicious dish of pasta bolognesa”. It is relevant to note, again, that this sentence is rather subjective, but the network seems to be able to understand it. Actually, what the network does is to relate the sentence to similar ones in the web, as well as to images and pieces of images related to those similar images. Finally, the deep network puts everything together.

These examples may induce to think that artificial intelligence based on neural networks is not limited. Although it is always very risky to say it, these limitations do exist. So, after seeing some instances of what we can expect of deep learning, we are going to discuss what they are not able to do.

### 1.5 Limitations of artificial neural networks and deep learning

As we have already commented, some time ago applicability of neural networks was limited by computational considerations. The amount of calculations and resources necessary to solve some complex tasks made their solution impossible. In the same way as these problems have already been solved, the difficulties that we are going to comment below may also be solved in the future. But today they are undoubtedly the most relevant problems in the field.

The first main problem is that of interpretability or explainability. As we have seen in all the previous examples, a neural network takes as input a set of data and gives as its output another set of data. But once we have this output, can we justify or provide reasons for the answer which is given by the neural network? Can we say why one specific answer is obtained rather than another one? The answer, in general is no, since the network behaves as a black box: we put some data inside and we receive some result, but we can not see how his result has been obtained. For instance, if we consider the example of the painting of Dali, how was exactly the word vibrant reflected in the picture? We may discuss this point once the picture has been created, but the network itself provides no explanation or even hint, at all. From the point of view of applications such as DALL-E or even GPT-3, this may not be a serious problem: we do not get an explanation from the people we are talking to about why they have chosen some specific words. However, from the point of applications in fields such as Medicine or Economy, this is a very serious handicap. A doctor in order to make a decision needs to have reasons to do in some way and not in another one. Moreover, even in the case of DALL-E, the absence of any explanation about the results means that we do not know the relations the network has created in order to provide its output. Probably, many of these relations are not trivial or straightforward. So we are actually having a lack of knowledge which may help to boost or improve research in the field. And in fact, a whole area of research, called explainable Artificial Intelligence, is devoted to this problem. This research field is so relevant that one of every three euros the European Union is going to invest in Artificial Intelligence will be devoted to it.

One possibility to deal with the problem of interpretability is trying to understand the learning from outside in terms of rules, for instances. A rule is an expression of the type:

If the apple is red, THEN it is ripe

Note that apparently we are very close to classical syllogism. However, in order to understand learning, we should consider notions that are similar, but not identical, to the ones which appear in the rule. For instance:

The apple is very red

From a classical syllogism, we get no result, since the antecedents (the apple is red vs. The apple is very red) are not the same. However, it seem reasonable to expect an answer of the type The apple is very ripe. An even clearer example is provided by the rule:

If it is hot, THEN open the window.

No human being would have problems to understand and apply this rule. But for a machine, there is too much uncertainty involved: What does exactly mean hot? Neural networks solve this uncertainty by considering related or similar examples. This can be expressed in terms of what is known as approximate reasoning and natural language, which can provide a window to understand the internal processes in the network by providing a way to represent the uncertainty inherent to language [Zadeh, 65]

There is another problem that encompasses almost every existing artificial intelligence learning system: the lack of adaptability.

Recall our example at the beginning, when we were learning to cook. Imagine that we have learnt to make perfect spaghetti bolognesa. Now a friend asks us to prepare some macaroni bolognesa. Probably, we will not have problem to adapt our recipe to this case. But, if we were a machine, and if it has learnt in the same way, it will have no idea about how to change from spaghetti to macaroni. In fact, if we want the machine to learn how to prepare this dish, the learning would need to start again from scratch.

This limitation of adaptability seriously limits the possibilities of artificial intelligence, as it makes necessary to repeat once and again the learning after minor changes in the data. In some sense, it is linked to previous one. If we were able to understand how relations are established inside a neural network, we could learn an algorithm to modify them according to the changes in external data. But the problem of adaptability is also deeply related to the fact that learning must be done through examples. So the information that is not the examples can not be considered at least in principle. It is worth to mention that generative adversarial networks provide a possibility to go out from this backend, as they create their own examples. Basically, adversarial neural networks consist of two neural networks that compete to each other in order to provide a better solution for a given problem. In this way, they provide new sets of examples of solutions for the considered problem. But these new examples are still limited by the information available to the network for the original training. This adaptability problem was present in Deep Mind, the deep neural network developed by Hassabis in 2016 that was able to win over one of the best human players of go in the world [Koch, 2016]. The system was only able to play in a board of fixed dimensions, but it could not play at all for boards of different dimensions.

## **1.6 Some considerations**

Of course, there are many developments not considered in this brief discussion. However, we hope that the idea is rather clear: there is no magic in the way a machine learns. Even if the results can be absolutely surprising, the method behind is rather simple to understand and based on the use of examples. And the quality of the results will depend on the quality of the available data.

In any case, we have no idea on how the research is going to evolve. Ten years ago, GPT-3 was science fiction, for instance. What it is clear is that, as human beings, we must keep the control of this evolution by means of an ever deeper and better knowledge of the processes that are involved in the learning, as well as keeping an objective and realistic view both in possibilities and applications.



## 2. Digital Humanities

### 2.1. Why digital humanities?

In the previous section, we have seen that the learning process in artificial intelligence can many times be identified with creating relation between different elements. In order to get these relations, language plays a very significant role. Artificial intelligence cannot be understood without language, in the same way as our civilization makes no sense if language in every form is omitted.

However, as we commented at the beginning of the work, there is a tendency to consider artificial intelligence as a sort of scientific knowledge in the narrow sense, opposed to the idea of humanistic knowledge. From our point of view, this is a big mistake, and one that can have very serious implications: without ethics, the risks in artificial intelligence multiply. Even more, the further artificial intelligence reaches and the closer its results are from those we would expect from humans, the more necessary the understanding of human beings in all their facets is.

In this sense, artificial intelligence can be a very useful tool for the study of humanities. In the following, we intend to discuss how artificial intelligence can be used for helping to understand human beings, and how it can do so taking inspiration from humanities themselves. In this way, we arrive to the recent field of digital humanities, where both, artificial intelligence and humanities are combined to create new opportunities for understanding the world and ourselves.

The examples we are going to consider share one point. They are focused on the detection and use of relations between pieces of information. But, contrary to the case of DALL-E or GPT-3, we are now interested in artificial intelligence procedures which help us to determine which these relations are, rather than just giving us a final output. Note that in some sense, human beings can be explained by means of their relations, and the understanding of relations helps us to understand ourselves.

### 2.2. History as a basis for the analysis of social networks

Today it is difficult to find somebody who has never heard of the concept of social network. In fact, most of us have a life in them. Facebook, Twitter, Instagram, TikTok, LinkedIn,... are nowadays absolutely common for us, and for some people it is even difficult to imagine how the world was before all of them exist. All these networks have arrived at the same time as the explosion of artificial intelligence, because it is artificial intelligence which has made them possible.

It is clear that if we are interested in social relations, these are obvious examples. But we should not deceive ourselves. Even if these are the most refulgent examples, they are in no way the only ones. They share some features which were already presented in every society in the past, since basically they reduce themselves to a set of connections or links between subjects. But the same can be said of economic relations, of political relations or of the way a neighbourhood is arranged. And all of these are as old as human society.

What makes the present different is that we have a huge set of recent tools to carry out the analysis of these relations. This analysis has been considered in fields such as sociology or physics, and, as any relevant research field, it has its own name: social network analysis.

Note that for this social network analysis, many different techniques have been developed, from numerical analysis (that is, straight mathematical calculations) to artificial intelligence algorithms. As the capacity of the computers has increased, these techniques have been applied to more and more complex networks, ranging from text analysis or computer cybersecurity to marketing or counterterrorism [Newman 2018].

The field of social network analysis is extremely active, and extremely complex, too. However, and interestingly, it also a field where artificial intelligence and social sciences touch to each other. In order to show this fact, as well as to provide a view on how humanities may have an influence on artificial intelligence, we intend to devote the following pages to the discussion of two recent algorithms inspired by History and which can help to understand social networks.

The problem we are interested in is how to determine the different groups that exist in a given social network. Let us consider a school, for instance. Are we able to determine the different groups of friends just observing the way people relate to each other? It may seem not so complicate. What about a big company? Maybe there are hidden affinities, maybe there are people whose actions make other to act in one or another way, maybe there are people trying to hide their real relations And if we think of Internet or Twitter? Can we determine how participants relate to each other?

In order to get a possible solution, we look back into history. More specifically, let us go to Italy in 1497. At that time, Cesar Borgia was the commander in chief of the armies of the Pope in the Italy of the Renaissance. A Pope that was his father, Alexander VI. The papal armies were launching a campaign to conquer those territories which have usually been linked to Rome. As most of the campaigns of that time, in this fight there were many examples of gallantry and tactical genius. But, as every war, it was also an example of human interaction both between people and communities. This is the aspect we are interested from the point of view of artificial intelligence.

If we analyze the dynamics behind this campaign, there are some points that seem to be relevant and which, in simplified form, determine the division in groups of friends and foes and the final distribution of land and power. Specifically:

- The goal of every country is to grow. This is not a trivial fact, because in some situations this is not the case. On the contrary, some countries may prefer to create an equilibrium rather than pushing their boundaries. This is also the situation in a real social network: People tries to increase their influence (the people related to them).
- Some parts of Italy were more similar to one country, whereas some others were more affine to others. Naples, for instance, was very close to the Spanish culture, whereas the Spanish influence over Milano, let us say, was much smaller. These differences will have a great influence in future coalitions and dynastical unions. In a social network, some members are more similar between them than to other ones. This fact actually is very relevant in determining which different groups are going to exist. Nevertheless, it is not the only aspect to take into account, as we discuss now.
- The differences of size and power between the different countries make some alliances much more valuable than other ones. In fact, the Italian republics search of powerful friends out from Italy itself was finally leading the permanent war between France and Spain in those territories for two centuries. This behaviour originates the creation of opposed coalitions, which can even be against the cultural affinities of some of the communities. It is worth to mention that, contrary to what common sense may suggest, the smallest republics were not finally conquered by the

biggest countries involved in the fight (Spain and Italy), but by a rather small but able actor: the Papal states. If we think of the social network on a company, for instance, something similar may happen. Powerful individuals or groups will attract more people. Furthermore, the bigger a group is, the more probable it will attract even more members. Of course, this does not exclude the possibility of some smaller groups remaining.

The analogy between the social networks and the situation is clear. How can this analogy be translated into an algorithm, an artificial intelligence system to determine groups in a network? Basically, we can try to look for every possible distribution of the members in groups until we find which one reflects better the social network. In order to do so, and inspired by the historical analogy, we need to associate some “power” to each individual and group of individuals. Then, each of the individuals and of the groups will attract all the other ones, with a force proportional to their power and the similar they are. We will start with all the individuals separated and we will continue until all of them have joined in one single group.

The one we have just explained is not an original algorithm. It was proposed by Wright in 1977 [Wright, 77] to solve the problem of clustering. That is, the problem of determining automatically in which different groups a set of data is divided. His solution makes a clever use of a sort of gravitational force between the different data (just replace power by mass and to be similar by to be close). When we apply this algorithm with the extra information obtained from the historical analogy, we get the so-called Borgia clustering algorithm, which is the one we have briefly explained.

However, our explained is not complete yet. We have said that we continue until all the individuals are together in one single group. But of course, in a social network, as in the case of general clustering, we do not expect that only one single group exist. So, how can we decide which are the groups.

Following Wright’s original idea, we have a look into the stability of each possible distribution of individuals. We start, at time 0, with all the particles separated. So we have as many groups as particles, and this corresponds to a network where all the members are alone. The individuals start attracting and at some time, let us say,  $t_1$ , two of the particles will join in the same group. From instant  $t_1$  on, there are as many groups as individuals minus one. At some later time,  $t_2$ , two of these groups will join, so we will have as many groups as individuals minus two.

So, if we have  $N$  individuals in the network, we see that:

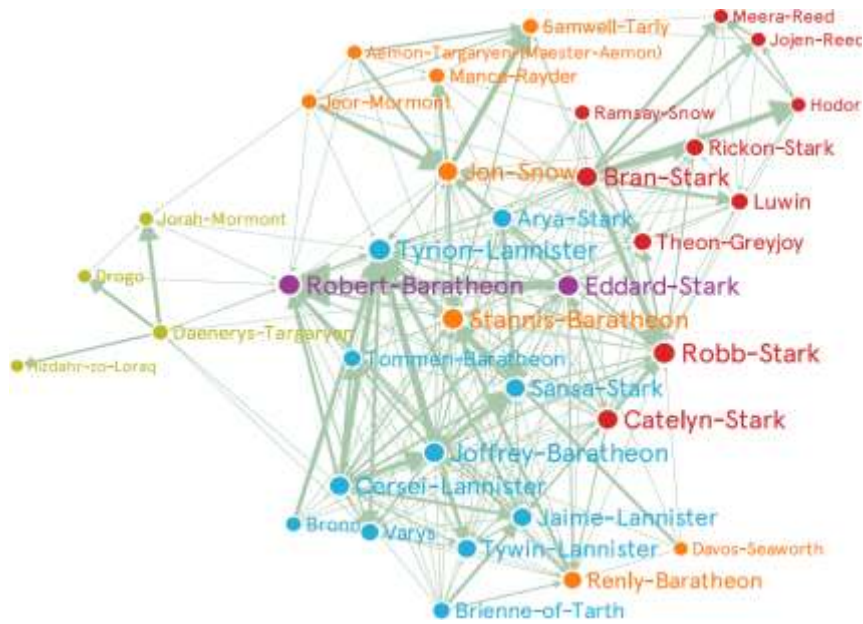
- At time  $t_0=0$  we have  $N$  groups.
- At time  $t_1$  we have  $N-1$  groups.
- At time  $t_2$  we have  $N-2$  groups
- ...
- At time  $t_{(N-1)}$  we have 1 group and we finish. We denote this time  $t_{(N-1)}$  as  $T$ .

What does stability mean? For each possible distribution we define its relative life as the time it has existed divided by the total time until arriving at one single group. For instance:

- The distribution with  $N$  groups has a relative life given by  $t_1/T$ .
- The distribution with  $N-1$  groups has a relative life given by  $(t_2-t_1)/T$

And so on. Then we choose as the distribution in groups which best describes our network the one which has the largest relative life.

Does all this make sense. In the next figure, we see which is the result of applying this Borgia algorithm to detect communities in the series of books “A song of fire and ice”...



Communities detected using the Borgia Clustering algorithm for the main characters of “A song of Fire and Ice [Fumanal, 2020]

As we can see, the algorithm is able to correctly identify the main characters and the relations between them. Note that all the information provided to the algorithm is the text of the books. Here, in order to measure how similar two characters are, we just count how many times they appear close to each other in the text. In this sense, the algorithm does not have any real understanding of the text, but it is able to provide useful hints to understand it. More details on this algorithm can be found in [Fumanal, 2020]

### 2.3. Spartan training in artificial intelligence

We move now to Antiquity. Krypteia was an initiation rite in Sparta for young people in the upper levels of the state. It was rather brutal in its nature. According to Plutarco, every year young members of nobility were free to make whatever they want to ilotes, including stealing or even murdering them, without consequences. In order to act in this way, the Young people were given a knife and sent alone by night to ilota settlements, were they should survive by any mean they could.

The origin and the exact goal Krypteia are still the object of discussion. It is assumed that it fulfilled two objectives at the same time: On the one hand, it frightened and helped to keep under control the ilota population. On the other hand, it was a means to prepare the next generation of leaders. Actually, no Young man can expect to reach a power position without having gone through this test.

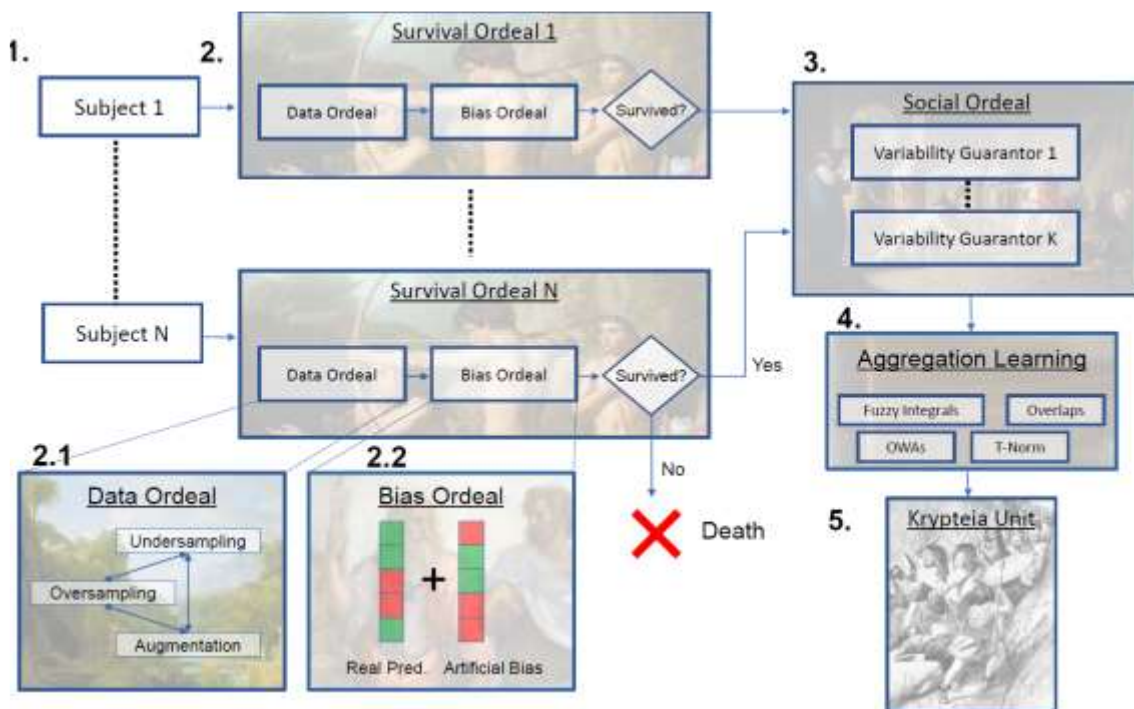
If we come back to the present, here is where the analogy comes. If we have several algorithms to solve a given problem, how can we determine which is the best one. One possibility is to make them go through their own Krypteia, forcing them to find the solution in the worst possible setting. Then we check the different solutions that we have obtained and we choose the strongest candidates for solving our problem.

This mechanism is not fully original. In fact, when we have several candidates for a position, for instance, and we submit them to some sort of test we are basically doing the same. The advantage is that, in the



case of artificial intelligence, there is no suffering. Weak solvers can be discarded without further consequences. This is the mechanism behind the so-called Krypteia algorithm.

In the following figure we depict a visual scheme of the full procedure.



Krypteia algorithm training scheme [Fumanal, 2022].

It is worth noting that an algorithm such as Krypteia can help to overcome the difficulties linked to the data dependence of the training in deep neural networks. By making different networks compete to each other, we can force them to increase their adaptation to the available data and to identify in a better way the existing relations. This is a work still in progress.

### 2.3. Using artificial intelligence to understand myths

The goal of this subsection is to show how the previously discussed Borgia and Krypteia algorithms can be used in a real problem. In particular, we are going to consider how they can be used in order to establish which relations exist between the main deities in different mythologies.

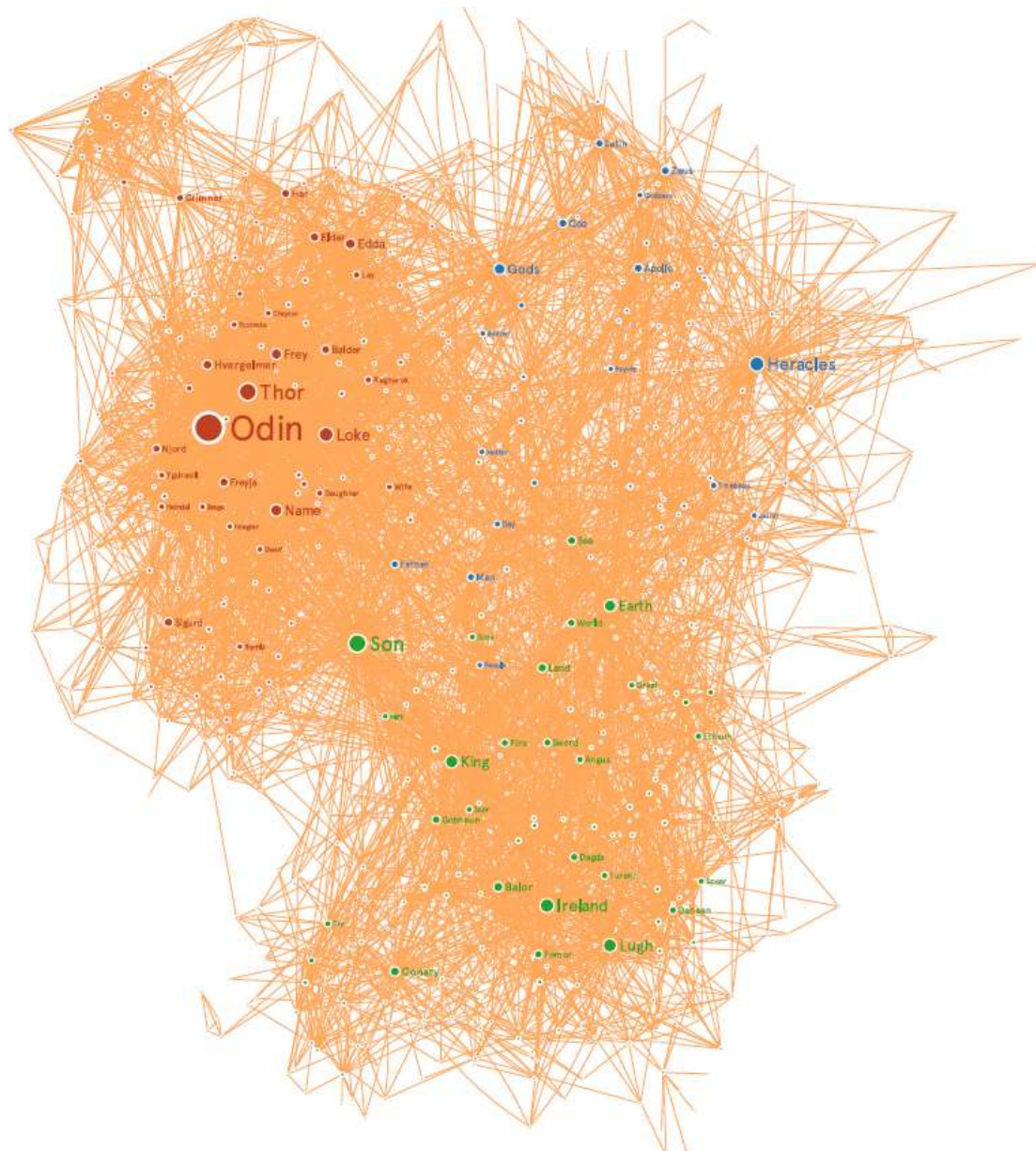
Myths are often created around characters (gods, devils...), and the way these characters relate to each others. In other words, myths can be seen as an instance of social network. The information we have about these networks is many times not complete, as we depend on the preserved texts, which by no way contain all the information about a given culture. But at least, even the smallest pieces of literature may provide a glimpse into the society that have created those myths.

Let us consider a very easy example to show the possibilities of artificial intelligence. Imagine that we want to extract a graph displaying the links between the main characters and concepts. This is not a complicate task to do. In fact, it can be done using the Borgia algorithm. We can also use other algorithms, and make use of Krypteia algorithm to get the best solution.

The idea is that if two concepts or characters are closely related, we will have a lot of connections between the words representing them in the graph. We are making a very simplistic approach, but even in this way, we get some surprising results.

For this example, we have considered the following three texts: Greek Tales, by Olivia Collidge, on Greek mythology, Celtic Wonder-tales, by Ella Young, on Celtic mythology and The Eddica Minora, on Nordic Mythology and considered to be a work by Snorri Sturluson.

We show in the next figure the result of our analysis.



Links between characters and concepts in Greek Tales (1949), Celtic Wonder-tales (1910) and The Eddica Minora (1220) ([Fumanal 2021])

The graph shows clearly that the culture and the context have a great influence on the significance of each symbol. For instance, the notion of “King” does not have the same value in the Nordic context as in the Celtic one. In the same way, the features associated to Greek heroes are very different to those

associated to heroes in other cultures. But the most relevant fact is that we do not get three isolated communities, one for each culture. On the contrary, a dense network of interrelations arises between them showing that each of the three cultures do actually share features. Of course, this is a well known fact, but note how the use of artificial intelligence techniques makes it explicit by means of a drawing without further analysis.

This is the beginning of a very recent work whose preprint can be found in [Fumanal, 2021].

#### **2.4. One final example: Artificial intelligence for the automatic analysis of art**

The process of interpreting and synthesizing the relations between an artistic work and its context has been usually done by experts in the topic along a process which may be years long in order to get a correct chronological dating and an author identification. However, this information is just given in terms of relation, so it is clear that artificial intelligence techniques can help in this sense.

Nowadays, artificial intelligence has been use, for instance, in the following tasks:

- Identifying the author.
- Identifying the style.
- Identifying the subject.
- Finding the correct data of the work.

In any case, it is important to remark that artificial intelligence techniques, even if they are becoming more and more accurate, have not overcome yet human experts. Note that there are some resources and methods that are used for these tasks that cannot be accomplished by a machine. For instance, determine the data of the materials used to create some work is commonly used to give a data for the work itself and check its possible authorship.

Taking this fact aside, the main difficulty in order to study an artistic image by means of artificial intelligence techniques is its difference from a real image. For instance, a given neural network which is able to recognize a face in a photograph, can accomplish this task regardless the background and the specific framework of each photo. But from an artistic point of view, there are very different movements and tendencies which represent the same face in very different ways. Just think on the relation between a photo of a sunrise and the sunrise in the painting *Impression, Sunrise*, by Monet. For human beings, this kind of distinction is not problematic, since we are able to adapt our understanding to each of them. However, as we have already discussed before, this is not the case for a neural network, for instance. So the problem of “explaining” to the computer that the *Menines* of Velazquez and the *Menines* of Picasso represent the same *Menines* is for the moment, beyond the capacity of the machines.

Note that in this case, we are in the opposite to DALL-E. Given the image, we need to extract the information on it, and not only the more or less obvious one, but specially the implicit one, ranging from interpretation to possible hidden meanings, etc.. For the moment, it does not exist an artificial intelligence which is able to receive a painting and to provide a full description of it, including those meanings which are not obvious or explicit in the image. We are not telling this is not possible. For instance, it is possible to analyse the text comments made by different experts on the painting, focusing on the most significant ones. Then we may use one algorithm to put together those comments closest to each other. In this way, we codify both the paintings and the most significant comments about those paintings. Finally, we can use this information to train a neural network. Observe that the comments on a given work by different authors can be seen as a social network, so we can apply social network analysis in this process in order



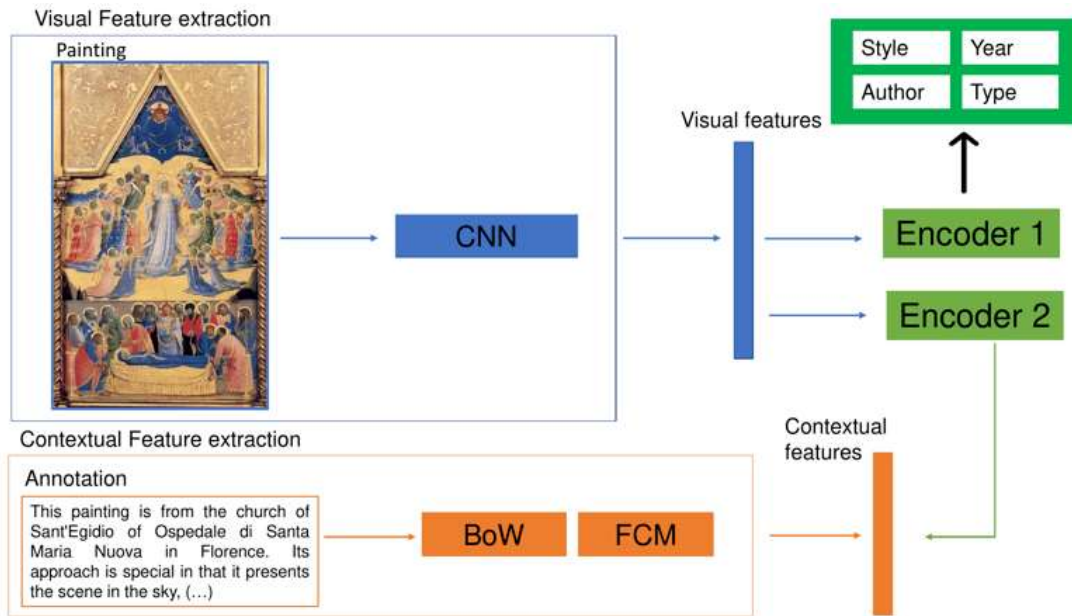
to determine which the most relevant ones are. But would this mean that this neural network understands art?

Let us consider a specific example, *Las Hilanderas*, by Velazquez. In this well known painting we can see a group of women spinning in a room. For the computer appropriately training it is not difficult to determine this fact. But if we go deeper into the painting, it comes out that some historians have realized that there exists a mythological content subjacent to this work. According to them, the work can be seen as a representation of the myth of Aracne. If we assume this interpretation, we see the painting under a new light: in the front we see a young girl, Aracne, spinning next to an old lady. This old lady is Athenea, since, regardless her face and hair, she is showing a leg without age signs. In the background three young women see how the goddess, with a helmet, is about to make the transformation into spider of Aracne. So the context, and the relations with different data, are providing a new view on the painting. But contrary to the case of building sentences, it is not clear where all this information can be extracted from in such a way that a computer is able to determine the appropriate relations!.



*Las Hilanderas*, by Velazquez (Wikicommons: [https://commons.wikimedia.org/wiki/File:Velazquez-las\\_hilanderas.jpg](https://commons.wikimedia.org/wiki/File:Velazquez-las_hilanderas.jpg))

How can this be done? One possible solution is discussed in the next figure [Fumana, 2022].



The explanation of this diagram is as follows: First we analyse the text comments made by different experts on the painting, focusing on those which are more relevant. Then we use one algorithm to put together those comments that are the closest to each other. In this way, we codify both the paintings and the most significant comments about those paintings. Finally, we can use this information to train a neural network. Observe that the comments on a given work by different authors can be seen as a social network, so we can apply social network analysis in this process. And also note that once the process advances, possibly the neural network will surprise us by detecting relations which are not obvious. So artificial intelligence and art experts can become a sort of colleagues, helping to each other to improve results, and without assuming any sort of sensibility by the part of the machine!

### 3. Conclusion

In these pages, we have made a review on how machines can learn, what they are able to do and which are their limitations. We have also insisted on the necessity of not divorcing artificial intelligence from Humanities. In fact, we have discussed some brief examples of the relevant new field of digital humanities.

Of course, much more can be said. For instance, we have not made a proper discussion on the problem of data. And we have not considered either some strong artificial intelligence developments, in the way to understand the brain. But we hope this work has at least fulfilled two goals:

- 1.- Demystify artificial intelligence by showing that its main guidelines are not so complex to understand.
- 2.- To show that artificial intelligence is not completely separated from humanistic disciplines, not is it an enemy to them. On the contrary, it can be a helpful tool, and humanism is absolutely necessary for a correct development of artificial intelligence.

We do not know which new developments artificial intelligence will bring in the nearby future. But we are sure that humanities are an essential part of it if we want to help to make a better world and society.

## References

- BERRY, David M. *The Computational Turn: Thinking about the Digital Humanities*. Culture Machine 12, 2011
- EAGLEMAN, David, BRANDT, Anthony, *The Runaway Species: How Human Creativity Remakes the World*, Catapult, 2017.
- ESSER, P., ROMBACH, R., OMMER, B., *Taming transformers for high-resolution image synthesis*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12873-12883), 2021.
- FUMANAL-IDOCIN, J., ALONSO-BETANZOS, A., CORDÓN, O., BUSTINCE, H., MINAROVA, M., *Community detection and social network analysis based on the Italian wars of the 15th century*, Future Generation Computer Systems 113, 25-40, 2020.
- FUMANAL-IDOCIN, J., CORDÓN, O., DIMURO, G., MINAROVA, M., BUSTINCE, H., *The Concept of Semantic Value in Social Network Analysis: an Application to Comparative Mythology*, arXiv preprint arXiv:2109.08023, 2021.
- FUMANAL-IDOCIN, J., CORDÓN, O., BUSTINCE, H., *The Krypteia ensemble: Designing classifier ensembles using an ancient Spartan military tradition*, Information Fusion, 90, 283-297, 2022.
- FUMANAL-IDOCIN, J., TAKÁČ, Z., BUSTINCE, H., CORDON, O. *Fuzzy Clustering to Encode Contextual Information in Artistic Image Classification*. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 355-366, 2022.
- GOODFELLOW, Ian, BENGIO, Yoshua, COURVILLE, Aaron, *Deep Learning*, MIT University Press, 2016.
- HAWKING, Stephen, MUSK, Elon, et al., *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter*, 2015.
- KOCH, Christof, *How the Computer Beat the Go Master*, Scientific American, 2016.
- MCCARTHY, John; MINSKY, Marvin; ROCHESTER Nathan; SHANNON, Claude, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, available at <https://web.archive.org/web/20080930164306/http://www.formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- MINSKY, Marvin, *Semantic Information Processing*, MIT Press, 1968.
- NEWMAN, Mark, *Networks*, Oxford University Press, second edition, 2018.
- NILSSON, Nils, *The Quest for Artificial Intelligence*, Cambridge University Press, 2009.
- RADFORD, A., KIM, J. W., HALLACY, C., RAMESH, A., GOH, G., AGARWAL, S., SUTSKEVER, I. *Learning transferable visual models from natural language supervision*. In International Conference on Machine Learning (pp. 8748-8763), 2021.
- RAMESH, A., PAVLOV, M., GOH, G., GRAY, S., VOSS, C., RADFORD, A, SUTSKEVER, I., *Zero-shot text-to-image generation*. In International Conference on Machine Learning (pp. 8821-8831), 2021.
- RAMESH, A., DHARIWAL, P., NICHOL, A., CHU, C., CHEN, M., *Hierarchical text-conditional image generation with clip latents*. arXiv preprint arXiv:2204.06125, 2022.
- RAMÓN Y CAJAL, Santiago, *Nuevo concepto de la histología de los centros nerviosos*, Revista de ciencias médicas de Barcelona, 18, 1-68, 1892.
- ROSENBLATT, Frank, *The perceptron: a probabilistic model for information storage and organization in the brain*, Psychological Review, 65, 386-408, 1955.
- RUMELHART, David E. HINTON, Geoffrey E., WILLIAMS, Ronald J., *Learning representations by back-propagating errors*, Nature, 323, 533-536, 1986.

**Bustince, Humberto:** From Artificial Intelligence to Digital Humanities

WRIGHT, W.E., *Gravitational clustering*, Pattern Recognition, 9, 151-166, 1977.

ZADEH, Lotfi A., Fuzzy sets, Information and Control, 8, 338-353, 1965.