

Good morning to you all...

Thank You indeed!

I'm delighted to be here with you to hold this lesson. I would like to thank Enrica Salvatori and Vittore Casarosa for inviting me and for giving me the chance to explain, in detail, what Artificial Intelligence tools are, and how can we apply these technologies in Historical research.

Well, today I would like to introduce you to the field of Digital History.

The morning session will cover the definition of Digital History and **how digitality change our mind about historical sources**. Following that, I will demonstrate how to utilize the Artificial Intelligence platform known as Transkribus.

The afternoon session will focus on the use of Generative Artificial Intelligence in historical research. Following that, we will begin a practice session on Transkribus.

WELL...

Dealing with History is not easy. And it is not easy to be an historian in the digital age... probably, nowadays, it is not easy to be an historian!

During my academic and scientific experience, I had the opportunity to realize that the evenemential approach has taken over the reflection on History as the discipline which have to discover and describe the inner structure of society.

So, after a deep reasoning, accepting the fact that we are living in a new society founded on the concept of "digitality", I started a new path of my research. Someone define my scientific profile as "Digital humanist", others as "digital historian"... but, I like to define myself "Philosopher of History..." (I am unconcerned about whether scholars accept this definition).

As we know, there is much discussion about Digital Transformation. This has prompted historians to redefine the concept of History by incorporating the term "digital" into the field. But this new perspective has forced historians to consider various epistemological issues raised by Digital History. A key question is the reliability of knowledge generated through digital analysis. Interpreting data, particularly data obtained through computational methods, raises concerns about the objectivity of history and the unavoidable introduction of algorithmic biases.

Furthermore, the issue of the **representativeness** of sources becomes crucial. Not all historical information is available in digital form, and this raises doubts about the completeness and fairness of digital historical analysis. The Philosophy of History must therefore tackle the problem of how to interpret and integrate digital sources with traditional ones to achieve a comprehensive understanding of historical events.

The importance of how History is portrayed and understood becomes more significant in Digital History. Historians and philosophers need to be cautious about relying too heavily on digital tools, as this may distort the interpretation of historical events. It is crucial to critically assess digital methods, making sure they improve rather than replace current historical accounts.

This new perspective spread two viewpoints:

- 1) Digital History is the storytelling of the digital world.
- 2) Digital History involves using ICT in the research process.

In the first scenario, historians need to explain the transition of the world and society into the digital age, and how technological advancements have necessitated people to learn new skills to navigate in this new environment.

From the second scenario viewpoint: in the field of historical research, scholars continue to play a central role. With the help of Information Technologies, they are able to enhance their work. For instance, utilizing the internet and digital archives allows them to efficiently locate historical sources, reducing the time spent searching for and analyzing documents to address their historical inquiries.

Today, my speech will focus on this second scenario. However, to grasp this perspective, some clarifications are necessary!

Droysen proposed that sources can be classified into three main types: leftovers (Überreste), monuments (Denkmäler), and sources (Überlieferung). He believed that all documents reflect human reasoning and its connection to the world. These documents form the basis of concepts like "testimony" and the historian's profession.

The source material is crucial for historians and their research. Scholars must categorize all types of expression to accurately document historical events.

Today historians must deal with a new classification:

- 1) Analogue sources, enclosing Überreste (the leftovers), Denkmäler (the monuments), Überlieferung (the sources)
- 2) Digital sources
- 3) Digitised sources

If historians once had a precise definition of historical sources, the digital revolution has expanded the classification.

They now need to consider new types of documents like emails, social network posts, instant messaging, social media, and websites.

A day in the future, historians who try to describe our age... for example... imagine historians, in the future, wanting to reconstruct the COVID pandemic event that struck the entire world in 2019.

These historians will have access to a vast amount of information available on the internet.

Social media posts, YouTube videos, and messages sent through messaging services like WhatsApp will serve as new historical sources to document how governments and individuals responded to the pandemic.

Thanks to the video, we can describe political positions and the vaccine controversy. Additionally, we can use the video to depict the critical situation within hospitals.

Thanks to the government website, historians can accurately describe the nature of administrative decrees. And so on...

Well, according to Gustav Droysen, all documents represent the limitless value of humanity. As Chabod later explains, these documents (and, nowadays, we must consider also email, social network, websites) help in reconstructing and displaying civilization, indicating its presence or absence.

«Sono così fonti i resti di organismi umani (crani, ossa varie), le tracce di focolai che si rinvennero in caverne abitate nelle età preistoriche; gli avanzi di palafitte, il vasellame, le armi, i gioielli, le tombe, le monete ecc.: fonti che acquistano tanto maggior valore quanto più rare sono le fonti scritte, e che divengono l'unico nostro mezzo di conoscenza per le età per cui nulla di scritto ci è rimasto. Non solo quindi per le età preistoriche, ma anche per le età che già conoscono l'uso della scrittura, la importanza di questo genere di fonti può essere notevolissima e, in certi casi, decisiva — **Human remains such as skulls and bones, along with signs of ancient activities in caves inhabited during prehistoric times, are among the sources. They also encompass remnants such as stilts, pottery, weapons, jewellery, tombs, and coins.**

These sources become more valuable when written records are rare, serving as our primary means of understanding historical periods with no written documentation. Even in eras where writing was used, these sources remain crucial and can sometimes be pivotal in providing insights.

It was Chabod who introduced us to the digital age:

«La radio, già solo con i suoi programmi e le sue scelte; il cinematografo, con i suoi temi e il modo di presentarli, in continuo mutamento, costituiscono una “fonte” importantissima per comprendere la psicologia delle masse dei nostri tempi e il suo variare. E lo storico del secolo XX non potrà prescindere da questi “strumenti di ricerca”, che gli potranno consentire talora di capire un'età meglio di molti documenti — **The radio and cinema, with their programs and evolving styles, are vital resources for comprehending the psychology of the masses in our time. Historians studying the twentieth century must take into account these valuable research tools, as they frequently offer insights into an era that exceeds many documents**».

Chabod's insight, as well as that of numerous historians who witnessed the actual transition in computing, must serve as a crucial connection between the methods developed before the digital age, and the current era marked by technological advancements. This transition may help History regain the empirical aspect of observation and experimentation. This is a capability that Droysen did not have access to, and one that Giuseppe Galasso also emphasizes, highlighting the importance of memory adapting to the advancements in Science.

According to Chabod, every historical period uses the most appropriate resources that it has access to. The digital age is currently revealing new insights into our Knowledge, Sciences, and methods, prompting a major reassessment. The focus is on determining which technologies can propel different fields towards fresh perspectives.

This situation also compels humanists and historians to rethink their research approaches. They are not looking to discard the analogical tradition's foundations, but rather to embrace new tools that previous scholars did not have in their repertoire, which are now becoming integral to their academic pursuits.

Considerations of the technological level of historical research methodology are essential. Understanding the current state of research is vital, as is devising a way forward to combine various dichotomies like qualitative-quantitative and micro-macro perspectives.

In today's digital world, historians can achieve a holistic understanding similar to Topolski's notion of "nomothetic consciousness" in the study of History, which entails historians connecting diverse fields.

Indeed, Economics, Sociology, Mathematics, Biology, Archival Science, Paleography, and Numismatics support History.

Today, Computer Science plays a crucial role by providing a common language of algorithms and programming for interdisciplinary communication, enabling the establishment of universal laws.

Further narratives and debates are necessary before creating a comprehensive understanding, particularly as the concepts of computerization and digitalization require clear definitions. Historians encounter various challenges in defining the status of this field of Knowledge. The digital shift has not favored them, as Terry Kuny suggests that valuable information stored electronically may be lost, leading to a potential digital Dark Ages.

Hence, a new opportunity has arisen, pushing scholars to tackle a difficult task as pointed out by historians of the Annales school.

Until the early 1970s, there was a strong consensus among historians that the documentary source, whether typographic or not, was crucial to their work. This cornerstone had a long history but was abandoned with the rise of modernity. Today, historians are struggling to define terms like "document," "historical source," and "archive source."

The use of computers has significantly advanced historical research by expanding perspectives and enriching viewpoints. It has also prompted a reevaluation of the concept of "source" and the tools used for studying and analyzing it. However, the challenge arises from the coding tasks that scholars need to undertake to ensure accurate computations.

The historical method, which was developed during the nineteenth century, emphasized thorough analysis and the resolution of methodological controversies. This method allowed historians to make use of specific sources while maintaining their scientific autonomy, resulting in more refined classifications. This laid the groundwork for Gerard Labuda to propose a highly accurate definition of a source.

"We shall call «historical source» all psycho-physical and social remains that, being the product of human labor, and at the same time participating in the development of society's life, thus acquire the ability to reflect this development. As a result of these properties (*i.e.*, the product of labor and reflective capacity), the source is the cognitive means that make possible the scientific reconstruction of society's development, in all its aspects."

Braudel advocates for the significance of relics as representations of humanity's attempt to understand the vastness of the world. He stresses the importance of scholars adopting a multidisciplinary approach to gain a comprehensive understanding of evolving structures and human activity throughout history, by engaging with a wide range of sources.

It is essential in this context to regard computer technology both as a type of proof and as the main instrument for understanding it. Computers, software, and calculators act as documented information in their own right, conveyed through diverse digital media as opposed to traditional pen and paper. They reflect the evolving means of human expression.

Skepticism from traditional scholars towards digital tools, to safeguard scientific disciplines, impedes research progress by discouraging the analysis of these tools as significant advancements.

On one hand, technology serves as a tool in a digital environment, enabling us to access the sources Labuda defines as "psychotechnical." On the other hand, it allows analogue documentation to be converted into digital format, integrating it into the network.

However, the challenge presented is significant: the efforts we are making to conserve the traditional past in this digital age are focused on creating a documentary collection that can be easily understood by future generations who primarily rely on the Web and the Internet. Will our digital texts be as effective in conveying meaning as the ancient cave drawings were for researchers who did not have digital tools?

Computer Science language, in conjunction with Cybernetics, has enhanced the interconnection between diverse fields of study to guarantee comprehensive overarching theories. Nonetheless, this interconnection necessitates a formal arrangement that encourages historians to rethink their perception of sources and additional information.

According to Topolski, this supplementary information is vital for situating the examined records within their proper context. Historians must now rely not only on intuition but on a structured knowledge framework, which even those in the humanities cannot overlook.

In this context, the source is not only a solution to the historian's interpretive problem but also gains significance from other sources. The digital source is interconnected, carrying additional meanings beyond the scholar's initial reading. It becomes significant as a "node" within a network that historians cannot overlook.

Still, we are far from this concept. According to Benedetto Croce, every remnant of the past is a "document." This includes imperial decrees, historical accounts, terracotta artefacts, as well as personal stories, biographies, diaries, memoirs, and all other forms of narrative sources.

This fundamental idea prompts us to consider the modern definition of a source, especially in the context of the evolving communication landscape that has elevated the significance of personal storytelling in history. Websites and Social Network can be considered new medium for storytelling!

The Droysenian classification distinguishes between primary and secondary sources. Nowadays, this classification is further developed with a new distinction: analogue and digital sources. The key point for digital historians to focus on in their research methodology is that digital native documents are readily accessible for processing, while analogue historical sources need extensive processing and structuring to enable scholars to utilize computer technologies (such as Data Mining and Machine Learning) for extracting and analyzing the information within them.

In digital research, the way document analysis is initially approached is quite different from the traditional close reading often used in current studies, even if it's mistakenly done on a computer. The methodology aims to modernize traditional practices but still works within the analog domain.

Just using a computer instead of a typewriter or doing online research on digitized sources doesn't make a research project Digital History.

The term "Digital History" is only used when specific computer tools are used, fundamentally changing the epistemological and methodological framework into a truly digital realm.

This might be useful to consider the definition offered by William G. Thomas III, who sees Digital History as an approach «...to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems. On one level, digital history is an open arena of scholarly production and communication, encompassing the development of new course materials and scholarly data collection.

On another, it is a methodological approach framed by the hypertextual power of these technologies to make, define, query, and annotate associations in the human records of the past. To do digital history, then, is to create a framework, an ontology, through the technology for peoples to experience, read, and follow an argument about a historical problem.»

The proposal overlooks important aspects of the modern approach to history. Digital History involves utilizing computer technologies to collect, arrange, code, and analyze sources. The aim is to produce fresh perspectives on historical events, whether they are in analogue or digital form. This method is distinctive because of the abundance of documentary resources accessible, distinguishing it from Quantitative History, despite the methodological links between the two for goals and purposes.

So, let's go back to the classification, focusing our attention on the third point: 1) Analogue sources; 2) digital sources; 3) digitized sources.

To organize and analyze information, we must convert all sources into digital format. Computers and algorithms need "machine-readable" texts. Hence, the initial step in our digital historical research workflow is to digitize archival documents. To achieve this, we need to transcribe all documents and historical records kept in archives. To this day, historians have manually transcribed documents. As a result, they have reconstructed past events from a limited number of documents. However, digitization has transformed historical archives into "Big Data repositories", enhancing our understanding of history. Several significant digitization initiatives have enabled historians to uncover crucial information, leading to the reinterpretation of certain events. For instance, the digitization of the Venetian parish registers from the seventeenth century facilitated a detailed description of the Plague Pandemic of 1630, including counts, political interventions, and medical insights.

Can historians increase their transcription productivity? Can historians extract additional information from historical archives? Can historians narrate events by drawing on multiple sources for storytelling?

Is it possible to reach these goals?

Well, I'd like to ask you a question: How many documents can be manually transcribed, for instance, in a day, by scholars? How many documents in one hour?

Imagine You are working on an historical project dealing with the life of Queen Victoria...

Imagine finding 10 letters written by Queen Victoria...

Now, imagine finding one hundred letters...

Imagine finding one thousand letters...

Certainly, these quantities also challenge the most experienced **palaeographer**.

However, we have to point out that Humanists are "typographic scholars," meaning they work with texts and documents. At the conclusion of their research, they publish their findings in journals and books, thereby continuing their engagement with documents.

To write a history book, a deep analysis of each discovered source is required.

Is it possible to read 1000 letters? Certainly, yes. Historians may devote their entire life to research and complete it after many years.

To read, annotate, analyse, and extract information from 1.000 letters, historians typically require no less than 5-6 years.

However, nowadays, specific tools facilitate a quicker analysis process. Generative AI (GenAI) can extract data and patterns from texts and, with the aid of specific prompts, help scholars enhance their understanding of a particular subject or event, as in the study of History.

The problem is that Large Language Models (LLMs, *i.e.* ChatGPT) can be used to analyse "machine-readable" texts, *i.e.*, those that are digitized. This means that all material stored in historical archives must be transcribed before it can be analysed.

It is clear that digitizing an entire archival heritage is not a simple task, both due to the enormous costs involved and the lack of personnel capable of performing this work.

Transkribus comes to our aid! Well..... **What is Transkribus?**

Well, I would like to start the second part of my speech by proposing a question: What is Artificial Intelligence?

Or rather, I would better rephrase the question: What is Artificial Intelligence, if We compare-it to Human Intelligence? What is the difference between Artificial Intelligence and Human Intelligence?

Probably, it is easy to answer these questions, but if We "really" want to answer, we have to reason on a lot of concepts, paradigms and issues which show Us how it is difficult to find the right answer.

Nevertheless, during this session of my speech, we'll give a try to find an answer and redefine what we know about AI.

The first step, to start our reasoning, is to find the right answer to the question What is Human Intelligence?

As You can realize, to answer that question, we can refer to many philosophers who tried to describe and define this capability (Aristoteles, Kant, and others).

Intelligence is the **mental process** which allow living being with evolved brain (but even to those who do not have an evolved brain, such as bugs) to find an **acceptable** solution to any kind of problem.

"**Acceptable**" because we don't have the certainty that "that" given solution is the correct one, and "correct forever." We can only state that "that" solution represents the best option at that specific time, in that specific place, and with those specific conditions and information at our disposal.

However, this does not mean that we cannot think of finding a solution that could be valid in the long term. Then, perhaps, at time passes, we may realize that "that" solution is not the best or that we need to modify-it in part.

This process characterizes all living beings, including insects and animals. If we look at the question from the "Set Theory" viewpoint, within the set "Intelligence" we would find the subsets "Human," "Animal," and "Bug." And within these subsets, we would find the characteristics for each of them ... **Every living being with their own intelligence.**

We are therefore faced with the concept of a "Set," or, even better —borrowing the term that is most in vogue nowadays when we are talking about "intelligence"— **of "Box."**

Intelligence is configured almost like a "**container**," where living beings place and keep all the information, experiences, and knowledge they acquire during their life experiences, and from which they draw when they need to perform a task, activity, or solve a problem.

In recent years, technological development has taken us in a direction that, until recently, was only anticipated by science fiction literature. We are witnessing what represents the true digital revolution: **the engineering and creation of Artificial Intelligences.**

This is the cornerstone brought about by the digital revolution. When we talk about digitization and wonder what it is, we are always inclined to say that it is the process of transforming an analogue signal into a digital signal. Actually, the digital revolution is a much deeper process, where the "creation" of a digital version of a physical object represents one aspect.

Let's start with some clarifications.

I want to begin with the concept of "**digitalization**" in the Italian language. This term does not have a clear denotation, so much so that it is used similarly in different contexts. We speak of "digitalization," for example, when we scan a document, and we use the same word to denote the process of creating a digital edition of a text.

The English language comes to our aid...

In this language we find 3 terms:

- (1) **digitization**, which is the conversion of analogue information into digital form;
- (2) **digitalization**, which is the native creation of information in digital format;
- (3) **digital transformation**, which is the process aimed at the exclusive adoption of technology in production processes.

In all cases, however, **digitalization** does not take on the denotation that describes a process with distant origins: the creation of the **Digital Ecological Niche**, or the enhancement of the communication system. The **DEN** (which stands for **Digital Ecological Niche**) is an interplay dimension where Humans interact with other Humans and Machines.

This is the relation which the (DEN) is based on: person2persons2machines

To be part of the DEN, humans must make an anthropological leap: they must be **HOMO-LOGGATUS** (logged-in humans). With this term we denote an individual who adopts specific identifiers to belong to the digital ecological niche.

What commonly is referred to "**digitalization of our life**", is actually the process of acquiring an identity that digitally represents analogue personal data.

Only by acquiring this anthropological digital status can one participate in the Digital Ecological Niche, deeply interconnected with information and communication technologies (ICT).

Well... pay attention on this, now: I want to ask you a question:

What was the first example of digitalization in history?

The first act of digitization was the process that let humans create the alphabet, which is a code that let them create a readable version of their thought.

In our mind, we organize our thought in visual form.

Our inner life—regardless of the level of introspection, self-control, or self-monitoring— **is a mental script composed of images.** This inner flow, in order to be communicated or even simply archived more conveniently for our own use, requires being formatted into a more "discrete" and "easily manageable" system.

To achieve all goals, the first step is the **creation of the spoken language.**

We don't know how and when, but it is sure that, in a particular moment in ancient time, humans were forced to codify guttural expressions into words. However, to achieve this, humans need to create consonant and vowels to combine every word to create phrases and sentences.

Imagine this scene: two hunters who see a white rabbit running away.

Hunter1 says "UH" (guttural sound)

What should the Hunter2 do?

His action depends on the "meaning" of the guttural sound!

So, how many meanings can Hunter2 infer?

- (1) Hey, look in front of You, there is a running rabbit!
- (2) Take the arrow and try to hit-it!
- (3) Oh, look, what a beautiful white color!
- (4) I don't like rabbit! I would like to have a deer!

... and so on!

Four meanings, and only one guttural sound!

Thanks to **alphabet**, after a not determinable period, Hunter1 was able to

- (1) Describe: Hey, look in front of You, there is a running rabbit!
- (2) Give an order: Take the arrow and try to hit-it!
- (3) Create poems, songs, artworks: Oh, look, what a beautiful white colour!
- (4) Express views and opinions: I don't like rabbit! I would like to have a deer!

Obviously, the process was not as quick and simple. Peoples did not sit around the fire to create the alphabet. **The creation of this code took years, decades, centuries.** An adaptation was needed, resulting in the development of our physiological system. The vocal cords changed and adapted, among other things.

However, if We want to answer to the question: What was "digitized" for the first time?

The right answer is: THE FIRST THING THAT WE DIGITIZE WAS OUR THOUGHT!

The alphabet (the code) has thus allowed for the "formalization" of a concept. The archeologist Gardin emphasizes that the act of formalization is the first step towards obtaining an objectified knowledge of something.

Objectification means ensuring a proposition that can convey as much information as possible to the reader, who receives the message without any interaction with the writer. Therefore, the reader must grasp the message from the strictly "literal" level to the semantic level.

Throughout History, every Science has created a model of formalized "proposition," based on rules that aim for objectivity. Historical research, however, had to wait until the nineteenth century to construct its disciplinary Statute under the influence of the scientific method. At the centre of the methodology is the source, which serves as "the proof" that underpins the events reconstructed by the scholar.

This paradigm has led historians to the realization that source analysis must be conducted to uncover all the information that allows a source to be considered as such concerning the historical problem. We can refer to a document as a historical source **only** if that document can help us answer our historical question. Otherwise, it remains merely a simple document, without historical significance.

For an elucidation: If our research is dealing with the life of Queen Victoria of England, then all archival documents written by the Mayor of Pisa, in the same period, are meaningless papers ... unless there are written references to the queen.

Secondly, the information must be connected to the event and provide the "reasons" for it. "Why did the event occur?"

Well... beyond all the epistemological theories that emerged during the 19th century, we must consider that **the craft of historians is to dig into archives, extract data and information, gather them into a semantic dimension, analyze them, and then write a book or essay to "represent" an acceptable point of view to the scientific community about events that occurred in the past.**

Well, let's consider the Digital historians and the Digital Ecological Niche...

IF

→ the digital turn is an upgrade of the interplay dimension...

→ coding is the creation of a medium (graphemes) which let humans communicate with those who are part of their community (the biological niche)...

→ the digital turn is the creation of the Digital Niche, where Humans interact also with machines...

THEN

→ **the digital revolution is the creation even of a new language to let DEN's agents communicate among them (and so: machines communicate with us!)**

Unlike scientists, humanists consider computer and technology from an **untrustful** viewpoint.

This is because the object of History's research is the most elusive of all: humans.

Nevertheless, even for historians it is fundamental →

For this reason, historians are convinced to use computer systems to organize information and find common threads between them. This approach allows scholars to create order among sources and identify key aspects essential for understanding historical questions. Databases and digital archives assist historians in sifting through documents to determine which ones can be classified as "sources."

These "organizational" systems are now much more powerful and can provide more detailed results. This does not take away from the fact that their power is linked to the computer architecture they are based on, and the dataset. The final responsibility lies entirely with historians. It is their task to identify the data and provide specific instructions for their organization.

Over the past year, computer technologies have taken a different path, leading to the development of Artificial Intelligences. This has reshaped the D-I-EN, which now includes not only sentient beings but also platforms that gather, process, and provide detailed information across a wide range of knowledge.

The restructuring of the DEN leads scholars to focus on how they communicate with these machines.

How do we interact with algorithms and artificial intelligences? Is the "new knowledge" generated by these technologies truly "new"?

We are constantly bombarded by a communication that aims to portray AI as tools that will surpass humans in every aspect of life.

These technologies can kill us...

They can control our mind...

Bear in mind that this type of communication is not entirely new. Every invention has always been met with pessimism, reluctance and fear.

In 1863, the Italian poet Carducci wrote the poem **"Inno a Satana"**:

Humans likened the train to a demonic entity. Likewise, AI is envisioned as the technology poised to dominate the world, rendering humans subordinate and hindering personal worldviews free from external influences.

Is this true?

As We will realize, during this lesson, happily, the truth is very different!

Certainly, some considerations are mandatory.

As You can see in this image: there are two **set**: INTELLIGENCE and INVENTION. Within the second set, we can find the subset COMMUNICATION SYSTEM... within this latter there is the subset COMPUTER, and then LLM...

As You can realize, just because the LLMs are subset of the INVENTION set (*i.e.* they are created by humans), it is incorrect to use the term, which denote the set "Intelligence", to denote them!

However, during the last decade, and till to date, a huge part of the scientific community has been attempting to connect the subset "Analysis system" to the concept of intelligence. This is based on an assumption, that is not entirely supported: **the idea that computers could possess a form of "intelligence" which enables them to interact as agents within the Digital Ecological Niche.**

By the way, Artificial Intelligence (AI) platforms involve designing systems that can perform tasks at a level comparable to human abilities, such as problem-solving and understanding natural language.

AI is not a specific scientific field but rather a technological research area that combines disciplines like Psychology, Logic, and Linguistics, with a primary focus on Computer Science.

The objective of AI is not to study the human mind or replicate human cognitive processes but to develop systems capable of specific tasks.

While the goal is emulation, some AI researchers see their work as part of Cognitive Science, which does aim to study the mind.

Analyzing human cognitive processes can enhance the efficiency of AI systems. In the AI community, various projects coexist, blending simulative and emulative elements within the same system.

AI emerged in the '50s as a result of recognizing the processing capabilities of digital computers by Turing and Von Neumann. It became evident that computers were not limited to handling only numerical data but could process various types of data when appropriately encoded.

However, many years have passed since the famous question asked by Alan Turing in 1950:

I think that, to date, machines cannot think!

However, thanks to specific prompt given by human operators (scholars), the AI are able to provide outcomes which can let scholars overcome their "mechanical" mind limits!

Humans cannot analyze Big Data, rapidly! Our mind is not configured to process millions of data in few minutes. From a mechanical viewpoint, the AI overcomes our physiological structure. All invention were created to help ourselves to face all situations our mind and body are unable to face!

Why did we invent the wheel? To build the cart!

Why did we build the cart? To carry things, to carry the game, to carry ourselves.

As you know, it is tiring to walk, but with the cart we can rest... And we can store even more and make longer journeys if we put the animals to tow the cart.

But animals also get tired. What if we created an engine to make the wheels of the cart move?

This is not future... this is History!

Well, like the train and airplane, which can transport us from one part of the world to another, the AI technologies have the “mechanical” structure to enforce our reasoning about something, becoming our ally, to let us rest during mental process!

However, our reasoning starting point is that AI platforms are nothing more than a “calculator”.

When we deal with AI as ally, we have to **consider** that we’re not communicating whit someone... but we are handling an instrument to process data.

We do not speak with, BUT we give prompts to AI.

The prefix “CHAT” in the word “ChatGPT” deceive us! The term “Chat” denotes an intercommunication system where humans speak among them, through an informatic platform.

When we enter the ChatGPT portal, there are no humans behind the platform, **banging on a keyboard!**

So, it is meaningless to start a “conversation” saying “HELLO!”

AI is the most powerful **calculator** humans have ever engineered! It has the faculty to calculate and process numbers and linguistic “formulas”. This faculty lets AI to go beyond the simple mathematical problems, it can create knowledge extracting-it from all data and information it keeps within its “BOX”.

... as you can see... the term “box” is back!

However, when we refer to AI, we have to consider that also this “intelligence” has a box. But “that” box is closed to our curiosity (it is a black box!). We cannot look in it and we cannot know which structure underlies the algorithms that AI activates when we give it a prompt!

We have one certainty: AI always responds to our questions. We may not understand the process, but it can provide answers by following statistical guidelines and using a neural system.

However, we must point out that **this box is a STOCHASTIC environment.**

... a context where actions and events are influenced by random or probabilistic variables. In simpler terms, it's an environment where uncertainty and variability play a significant role in determining outcomes. This type of environment is common in various fields, including economics, finance, engineering, and natural sciences.

Uncertainty: The outcomes of actions or events are not entirely predictable. While it is possible to make predictions based on probabilities, determining the outcome of a single event with certainty is not possible.

Random Variables: Events or outcomes depend on variables that follow probability distributions. These distributions can be either known or estimated.

Probability: Outcomes are described in terms of probability. For example, it could be said that there is a 60% chance that a certain event will occur.

Examples of stochastic environments:

- **Financial markets:** Stock prices, bond prices, and other financial assets are influenced by a multitude of uncertain factors, making the market a highly stochastic environment.
- **Weather Forecast:** Weather conditions are influenced by various atmospheric variables and other factors that make weather predictions probabilistic.
- **Games of chance** like poker, roulette, and other gambling games are classic examples of stochastic environments, where the outcome is influenced by randomness.

When we give a prompt to AI, for instance, **to describe the last will of King Luis the 16th (the last king of France)**, the platform creates an outcome which contains all information about the subject of the question, organized in a linguistic form that deceives us... We are forced to think that the outcome was written by historians. Actually, the outcome is based on a statistical system which guide the AI to create an “acceptable” answer.

However, look at these outcomes:

In the first case, the AI tells us that Luis met his family.

In the second case, the AI informs us that Luis did not meet his family.

This phenomenon is called “**ALLUCINATION OF THE AI**”!

Well, **did King Luis meet his family?**

The answers are not so clear! However, for the most curious, the answer is that King **Luis did not meet his family**. These examples invite us to reflect. AI is a powerful calculator, but we have to control-it and guide the algorithm to get a true and verifiable outcome.

In 1985, Ennals stated that →

We must control the machine and we have to create the best ontology to let AI became a powerful historical instrument in our hands.

Artificial intelligence is a game changer in many fields, including cultural heritage. It supports the planning and preservation of heritage sites and cities, enables the creation of virtual experiences to enrich cultural tourism and engagement, supports research, and increases access and understanding of heritage objects.

However: what is the goal of historians? They want to know WHY something has happened.

Imagine digging an archival repository. After some time, you found an Act of Sale.

This historical document can tell us **that a person sold something to another person**.

So, there are **2 persons**: the Seller and the Buyer.

The Act was signed in a **place**, in a specific day.

The seller received a sum of money

The buyer received **something** and, from that moment on, he/she is the legal owner.

1st W: Who... seller and buyer

2nd W: What... there was a buy and sell

3rd W: When... a day in the Past!

4th W: Where... a place in the World

5th W: Why...

Why??

Thanks to the four Ws (Who, What, Where, When), we can construct a concise narrative of a historical event. By establishing a network, we can determine the individuals involved and their actions.

In this research phase, the key focus is on identifying all entities mentioned in the documents, such as people, locations, dates, and events.

This is what scholars calls “ENTITY RECOGNITION”...

Thanks to this approach, historians can connect individuals and locations to establish a historical network...

This historical network illustrates connections among people, places, and events of the French Revolution.

Maximilien Robespierre is at the center, connected to the Reign of Terror, the guillotine, and Paris.

Surrounding him are other figures like Louis XVI, linked to the Palace of Versailles and his execution, and Marie Antoinette, associated with the Tuileries Palace and her trial. Georges Danton is connected to the Cordeliers Club and the Storming of the Bastille, while Jean-Paul Marat is linked to his assassination and his newspaper '*L'Ami du Peuple*.' Each person, location, and event are clearly labeled and connected with lines to show their relationships.

Software and tools like **Python** can accomplish this task, but nowadays, ChatGPT can perform the task even better. By providing ChatGPT with specific prompts, we can enhance its ability to identify all entities and extract a wider range of data due to its unique configuration.

This is the reason why, nowadays, ChatGPT can be the most powerful ally in historical research.

We can ask it to organize all data and create a dataset...

... and elaborate specific script (for example, in Python language) to analyze them.

In addition, we may ask you to explain an event based on historical information on the internet.

To clarify this assumption: in 1782, the prior of Messina, Michele Maria Paternò, wrote a letter to Anna Maria Morso Bonanno, the princess of Biscari. In the letter, the Prior mentioned that the Emperor of Austria had instructed the Pope to halt his journey through the Holy Roman Empire due to a revolution in Vienna.

This raises the question: why was the Pope traveling across the Empire? We have asked ChatGPT to gather all the necessary information to help us reconstruct the context.

This is the answer provided by the AI:

By leveraging its web exploration function, ChatGPT-4 redirected us to two websites housing a plethora of information about the event: the Pope's undertakings to obstruct the Emperor's religious reforms in his domain (this is the reason why!).

Then, we ask AI to identify publications dealing with this journey:

The first publication is a thesis submitted by Davide Franco Jabes for the Degree of Doctor of Philosophy at the University of York, Department of History, in January 2011. The title of the thesis is *A "Second Counter Reformation? Aspects of the pontificate of Pius VI reconsidered."*

So, let's go back to Collingwood statement: historians want to know why!

Consider the Act of Sale...

Why did agent A sell something to agent B?

Why did agent B decide to purchase agent A's property?

The only way to get answers is by asking questions. Unfortunately, we can't ask questions to someone who lived hundreds or thousands of years ago.

In addition, unlike the event involving Pio the Sixth, it is very challenging to determine the reasons behind these kind of historical documents.

The traditional historical methodology, as Droysen stated, led historians to solve the issue trying to explain the reason of an act, applying the **analogy**. Thanks to this method, historians filled the gap basing their reconstruction on similar events, where the reason why was clearly defined.

Starting from analyzing the context, to identify recurring similar aspects among events, historians aim to write a reasonable description of the new event.

What we know about the Past is a "reasonable description".

Nowadays, historians can utilize computer technologies, like Artificial Intelligence, to **simulate** an event. This helps them analyze and comprehend the reasons behind the actions of the individuals involved. The simulation is based on hypothesizing behaviors that could make the choice to act in a certain way reasonable. By going beyond analogy and using simulation, we can identify the most likely reason behind an event.

When primary sources are unavailable to provide all the information about the reasons behind an action, we can describe an event from a logical perspective.

For example, why we built churches?

According to the "Analogical approach", we have to consider the devotional reason.

Instead, "Simulation" allows us to create alternative contexts, prompting historians to explore different narrative possibilities.

For example, no one thinks that it is possible to build a church for economic gain.

Nevertheless, this is the case of the foundation of the Church of the Sacred Letter of Riposto, built to allow the workers residing in that area to settle there permanently, creating a new town.

In Riposto, the church founder owned land where these workers were employed and had an interest in them staying there. And so, the solution was to provide them with the sacraments.

Well, in a simulation process, AI can connect various sources of data and information, guiding historians to explore new possible explanations...

One of the key features of ChatGPT is its ability to standardize and translate text. This feature is highly beneficial for historians who work with texts in various languages, especially Latin. Additionally, medieval and early modern texts often contain sentences that are not in perfect modern Italian. This aspect frequently requires historians to conduct a detailed linguistic analysis to extract the necessary information.

Obviously, this is the work that historians do when they have limited archival sources.

When we need to analyze a large number of documents, what are our options?

As we have observed, Transkribus enables us to transcribe numerous sources automatically in a brief period. However, it is also true that the transcriptions are frequently inaccurate and need to be corrected by the user. When compiling a digital edition of a historical document, it is crucial to ensure that corrections are made with great precision.

In historical research that does not focus on creating a digital edition, scholars can opt to use a normalized transcript. This type of transcript maintains the document's structure and key content.

ChatGPT can be helpful when we need to standardize multiple sources.

Here is an example. We imported the transcript from Transkribus into the prompt page and requested GenAI to standardize it.

The text now has a modern linguistic structure that enables us to extract the necessary data, such as names of individuals and locations.

We can also apply this process to a document in Latin. We can request ChatGPT to translate it into Italian and then extract the information.

And as we have observed, using normalized data enables us to utilize other features of generative AI, such as simulating historical contexts or generating datasets and networks.

Thanks to Transkribus and ChatGPT, I was able to quickly create the dataset and database for the website on the Prince of Biscari of Catania's Epistolography.

These features provide a chance for historians and archivists to **UNLOCK THE PAST** (this is the **motto** of Transkribus).

So, as you have seen, ChatGPT can simulate, analyze, describe, perform entity recognition, and provide accurate information to help historians contextualize historical issues. Also, it has the capability to generate non-questionable images and video, to let historians overcome their “typographical attitude” when they publish their findings. ChatGPT is a valuable tool for humanists to test their theories and offers historians the opportunity to experience an empirical approach.

Thanks to ChatGPT we can see past events as on a screen...

However, we have to point out that words, connected to Logic, can convey abstract concepts, complexities, and nuances. Images, however, often present a simplified and more immediate reality, which can distort or oversimplify complex truths.

Well..... **I would like to conclude my lesson by taking up a statement by Victoria Clark:** The future is here for historical research, but it is crucial to recognize that technological innovations often have unintended consequences. Understanding the limitations of generative AI is essential for historians. In many respects, our situation with generative AI is like that of early humans discovering fire — a powerful tool that requires careful and thoughtful handling. Therefore, a cautious approach is warranted.

We are at the start of our journey. **The future is uncertain to us.** As historians, we acknowledge our **inability** to predict what lies ahead. It is a misconception that historians can comprehend the present by studying the past, to manipulate the future. How many of you have ever heard that History is a «teacher of life»? **I want to tell you the truth:** every one of us fail the exams with this teacher!!

To understand the Present, **we need to embrace a new way of thinking**, and not everyone can grasp **the real meaning of the digital world**. While there is a lot of talk about digital technologies, how many traditional thinkers have the necessary mindset to truly understand its significance?

While it is true that History teaches us that revolutions are based on awakening, it is truer that **all revolutions need to overcome mores majorum and traditional mindsets.**

To comprehend Digital History, **you must start your reasoning from the Philosophy of History (a discipline often overlooked in today's Humanities Departments)**, then delve into the realm of Digital History to establish its foundation. **Traditional academics** (academic dinosaurs) **may not possess the necessary expertise to define this new Knowledge field**, but they can provide all paradigms, on which “traditional” History is based, to let you compile the ontologies' ontology... enabling you to take control of the Artificial Intelligence technology and teach it what ~~digital~~ historians do.

Remember: In his book "*The Humiliation of the Word*," Ellul explores the impact of technology on society. He stated that technology **tends to perpetuate itself** and impose its own logic and demands on human life.

Although, **hold firmly in your mind:** only humans can handle the power to switch off and stop all inventions they created.