Umanistica Digitale – ISSN: 2532-8816 – n. 16, 2023
S. Spina – *Artificial Intelligence in archival and historical scholarship workflow: HTS an ChatGPT*
DOI: http://doi.org/10.6092/issn.2532-8816/17205

Artificial Intelligence in archival and historical scholarship workflow: HTS and ChatGPT

Salvatore Spina

Università degli Studi di Catania
salvatore.spina@unict.it

Abstract

The article deals with the impact of artificial intelligence tools on the archival heritage digitisation workflow, specifically regarding the manuscripts' automatic transcription, correction, and normalisation. It highlights how digitality leads scholars to redefine the paradigms of archival and historical research methodology and has facilitated the accessibility of analogue sources through digitisation and integration into Big Data. The article focuses on two AI platforms, Transkribus and ChatGPT, which enable efficient analysis and transcription of digitised sources, moreover deals with a test on ChatGPT, which was employed to normalise the text of 366 letters preserved in the "Correspondence" section of the Biscari Archive (Catania). Although, the AI tool exhibited some limitations, which resulted in inaccuracies, the corrected texts met the expected results. The article concludes that digitisation and AI tools can significantly enhance archival and historical research by allowing the analysis of vast amounts of data and the application of computational linguistic tools.

Keywords: Digital History, Semantic, Syntactic, Historical Craft, Text Normalization.

L'articolo tratta dell'impatto degli strumenti di intelligenza artificiale sul flusso di lavoro della digitalizzazione del patrimonio archivistico, in particolare per quanto riguarda la trascrizione automatica, la correzione e la normalizzazione dei manoscritti. Evidenzia come la digitalità porti gli studiosi a ridefinire i paradigmi della metodologia di ricerca archivistica e storica e abbia facilitato l'accessibilità delle fonti analogiche attraverso la digitalizzazione e l'integrazione nei Big Data. L'articolo si concentra su due piattaforme di IA, Transkribus e ChatGPT, che consentono un'analisi e una trascrizione efficiente delle fonti digitalizzate, e tratta inoltre di un test su ChatGPT, che è stato impiegato per normalizzare il testo di 366 lettere conservate nella sezione "Corrispondenza" dell'Archivio Biscari (Catania). Sebbene lo strumento di IA abbia mostrato alcuni limiti, che hanno portato a delle imprecisioni, i testi corretti hanno raggiunto i risultati attesi. L'articolo conclude che la digitalizzazione e gli strumenti di IA possono migliorare significativamente la ricerca archivistica e storica, consentendo l'analisi di grandi quantità di dati e l'applicazione di strumenti linguistici computazionali.

Parole chiave: storia digitale, semantica, sintattica, artigianato storico, normalizzazione del testo

Introduction

Like every system of communication and cultural experience, Humanities are being drawn into the world of computing, and the digital turn has pushed scholars to redefine Languages, Arts, and History paradigms. Just as words have been assimilated into "second-order abstractions" (*i.e.* numbers) ([4]; [5]), the concept of "source" has taken on those forms of communication that characterise the «onlife» ([10]; [11]) society.

The digital (and digitised) document will be, tomorrow in the future, the (meta)source ([9]) of historians, allowing them to write about our present. However, this perspective seems to leave out analogue or archival documents, which may be regarded as non-existent assets —since what is not displayed on the Internet does not exist! These documents seem to regain their cultural-historical role only when traditional historians study and analyse them. Several digitisation projects of archival documents, however, have allowed digital historians to also make use of these corpora, which, being increasingly abundant, have enabled —also— different scholars to consider them as the "Big Data of History" ([15]; [34]; [20]; [35]).

Analogue sources are transformed into images enriched with metadata and digital information that enables their identification on the web. Their volume is increasingly vast and challenging to manage without the aid of other ICT tools. Two artificial intelligences (*i.e.*, Transkribus and ChatGPT) now offer the opportunity to enhance archival and historical workflows. They both enable the automatic transcription of large quantities of manuscripts and facilitate a more profound analysis, allowing the extraction of extensive information from texts. Therefore, this helps reconstruct past events with an extraordinary richness of historical details.

Beyond a purely philological-linguistic approach, in the case of historical analysis, source texts require a certain degree of normalisation and correction in order to be useable in other computational contexts because, unlike literary compositions —both in prose and verse— or homogeneous documents such as companies' cash registers, banking documentation, or even medical records from a hospital or medical practice, the archival sources used by historians —to reconstruct even single event or the activities of a person— are characterised by different writing styles and languages, as in the case of epistles from a monarch or prince drafted by various secretaries —hence, the different handwriting, even though the epistles come from the same sender. In other cases, the writing is uniform precisely because it is personally penned by the sender, as in the case of the letters of lower-rank nobles or wealthy bourgeois. This uniformity is also particularly evident when there is a strong personal relationship between the sender and the consignee, so the letter is not entrusted to external writers (for example, a secretary). An example of this can be found in the 28 letters that are part of the section under examination, sent by Prior Michele Maria Paternò to Princess Anna Maria Morso Bonanno[1].

This variety in calligraphy prevents an HTR model from producing a consistently accurate transcription without a training phase that may allow historians to transcribe numerous manuscripts quickly —which is the purpose of HTR technologies.

Handwritten Text Recognition platforms, such as tranScriptorium ([33]), Transkribus ([18]; [26]; [8]; [24]), and eScriptorium ([21])[2], have expanded the prospects for archivists and historians. Digitalisation has undoubtedly facilitated access for historians to numerous archival documents, creating the conditions for reconstructing historical events with greater detail beyond the possibility of working with transcribed texts. Furthermore, artificial intelligence tools have enabled archivists to develop digitisation projects beyond the simple photographic

---

[1] State Archive of Catania, Biscari, Corrispondenza, ff. 370-414.

[2] Several articles have analyzed the differences between HTR (Handwritten Text Recognition) software and systems, with a particular focus on the two mentioned here, Transkribus and eScriptorium ([16]; [23]). Additionally, a report of the study day held at the Bibliothèque nationale de France on May 9, 2022, provides an insight into the development of these two platforms ([13]).

representation of documents, as exemplified by the Stockholm City Archive[3], the Mining Hub digital archive[4], and the Regesta Pomeraniae Monastica[5].

However, automatic transcription is not the only challenge related to digitisation. The ability to rapidly obtain the transcribed (digital) version of numerous documents from archival collections poses a "physiological" limit for researchers in quickly analysing their content. Whilst not intending to engage in a "quantity over quality" debate in this regard, it is safe to say that digitalisation severely tests the Droysenian capacity of historians to connect facts through the use of analogy, imagination, and subjective interpretation ([7]).

Quickly transcribing numerous documents in a digital format —albeit with a margin of error— generates an uncontrollable mass of data and information that cannot be easily organised and reviewed; moreover, it does not enable the identification of any patterns.

Historical research —since Herodotus— is based on a close reading methodology and the analysis of a limited number of documents. Numerous digitisation projects —like the collection of the Library of Congress with millions of newspaper pages and the Finnish Archives' court records dating back to the 19th century— pushed historians to start using machine learning (deep neural networks, in particular) to organise and examine historical documents.

Consequently, the «Big data of History» is both a problem and an opportunity for researchers: it provides much more information, though there is no evidence of an effective way to sift through it. However, ICT seems to yield a solution to the problem. A representative case of the aforementioned observation is that of the seventeenth-century plague in Venice, which historians described relying on archival sources that collect data and information from documents that recorded only a few days of that terrible experience. Instead, with the "Venice Time Machine" project ([19]; [1]), digitisation has allowed scholars to trace up to three years of events and incidents, increasing knowledge about the epidemic ([22]).

On the other hand, even HTR tools have their limitations ([39]). The models are being constantly trained and enhanced. Their use has undoubtedly profoundly impacted the Humanities research field, particularly Archival Science ([28]). However, it is necessary to consider that the documentary material processed by these technologies may not be correctly transcribed in its entirety. Each model has an error range (Character Error Rate) that requires scholars to carry out a thorough revision (1) in case of online distribution of a digital edition of a complex (large or small) documentary, an epistolary corpus, or other materials, as well as (2) in case of a detailed historical or linguistic analysis that requires accurate and normalised texts, regardless of their potential publication and/or dissemination. From the perspective of mass digitisation, such corrections cannot be entrusted to an individual scholar, as their capacity would be limited to a few documents, thereby prolonging the research timeline.

However, nowadays, Computer Science and ICT development companies may have found a solution in the Artificial Intelligence tool named LLM (Large Language Model), which, due to its characteristics related to the construction and formulation of texts, could be configured as the best tool for correcting lengthy texts.

Notably, OpenAI was founded in 2015. Since then, significant funding has enabled the company to develop InstructGPT ([29]), followed by ChatGPT ([2]; [42]; [17]; [30]; [37]), an innovative

---

3 Link: https://stockholm-city-archives.transkribus.eu (last access, 20 Jul 2023).

4 Link: https://transkribus.eu/r/mining-hub/#/ (last access, 20 Jul 2023).

5 Link: https://transkribus.eu/r/regestapomeraniae/#/ (last access, 20 Jul 2023).

AI system capable of engaging in dialogue with humans, thus realising Alan Turing's vision [40]). From its early stages of development, GPT has emerged as one of the leading Language Models (LLMs). However, it is essential to note that regardless of the testing it regularly undergoes, the AI ChatGPT lacks any semantic understanding. Indeed, any test involving "reasoning" is inherently flawed *a priori*.

Unquestionably, GPT is anything but a biological entity. Despite its ability to engage in conversation —it almost seems like discussing with another sentient being— it is not an intelligent agent that relies on semantic understanding and logical inferences. ChatGPT is an AI system constructed upon models that enable it to generate syntactically accurate texts and statements ([12]). However, like other LLMs, ChatGPT is susceptible to hallucination issues ([32]; [27]) in its outputs.

Therefore, its strength lies in the feature to fix and reformulate the syntax of a text or its parts ([12]) correctly.

The workflow

Suppose it is true that the term "digitalisation" in archival and historical fields denotes the process of encoding documents/texts into a machine-readable format. In that case, it is even truer that most archival documentation digitisation projects do not go beyond the metadata phase. The historical documentation on the Web is, for the most part, a simple photographic reproduction accompanied by limited information and lacking a descriptive level that includes the complete transcription of the photographed text. Fortunately, HTR artificial intelligence tools have shown much more promising prospects to follow in the digitisation processes, enabling the creation of digital editions of documentation.

Nowadays, digital historians have at their disposal three different HTR platforms (*i.e.*, tranScriptorium, eScriptorium, Transkribus), which aim for the same goal while being structurally different[6].

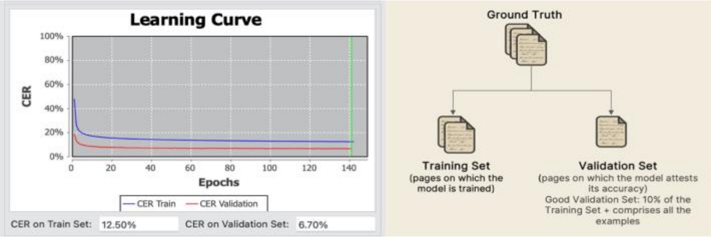
For instance, the Transkribus AI manages to combine, in the same tool, multiple functions and process management, ensuring more efficient results in terms of transcription accuracy, exporting results in different formats, and even enabling the possibility of dissemination and access to the digital edition.

Two artificial intelligence tools —Transkribus and ChatGPT— underwent a test to prove their effectiveness, performed on a corpus of letters, constituting the "Correspondence" section of the Paternò Castello, preserved at the Catania's State Archive. The documentary collection of the family is very diverse ([3]). It consists of various documents, making it an essential tool for scholars interested in the History of Sicily and the relationships between the island authorities

---

[6] The TranScriptorium and eScriptorium AI tools require computer skills for server configuration, installation, and application of transcription models, which some humanists seem to lack at times. On the other hand, the AI system Transkribus only requires the profile registration on the official website (https://transkribus.eu/), and a few simple steps to install the desktop version on a Personal Computer. Additionally, there is no need for specific configurations to use the models (or to access online repositories like GitHub), as the transcription process takes place entirely on the AI servers. Furthermore, unlike other platforms the Transkribus GUI, has a text editor where the user can, according to the segmented lines connected to the image, both produce a coherent transcription of each line of text and make necessary corrections to spelling errors; the user can also decipher any unrecognized symbols and abbreviations. In fact, they can transcribe their own documents and use a suite of tagging tools to create rich transcriptions that could also be part of a digital edition.

and European monarchies. In particular, the test used the epistolary section, a document that, although homogeneous in structure and layout, contains a large amount of data (names, places, dates) that need to be mined and numerous handwriting styles (262 senders). The section consists of 366 letters, and the acquisition process was performed using a Nikon D610 camera equipped with an AF-S Nikkor 24-120mm f/4G ED VR lens. The photos were collected in a database created using Claris FileMaker 19 software. After adding metadata, they were converted to PDF format in order to be uploaded to the chosen Transkribus server for automatic transcription, that is, the "Transkribus Italian Handwriting M1" public model, which exhibits a Character Error Rate (CER) of 12.50% for the "train set"[7] and 6.70% for the "validation set"[8] (Figura 1). Such percentages ensure an optimal transcription, requiring minimal corrections by the user, as described in Spina ([39]).



CER on Train Set: 12.50%
CER on Validation Set: 6.70%

Figura 1

Subsequently, the transcriptions were downloaded in ".txt" format without undergoing any manual corrections or training aimed at reducing the error rate in order to proceed with an automatic correction of the entire correspondence, initiating a "chat"[9] session to prompt LLM ChatGPT so that it could analyse each letter to assess whether this OpenAI tool can correct and normalise them, in a zero-shot learning ([41]; [38]) scenario —which is the typical condition of non-computer-savvy historians who wish to utilise technological tools for their research.

The prompt given to ChatGPT was to correct a text without linguistic instructions, historical references, or training on how to provide expansions of abbreviations ([36]).

The type of analysis required did not deviate from the capabilities of AI, and the corrections obtained met our expectations, allowing us to assert, without hesitation, that the LLM ChatGPT can be a helpful tool for correcting historical texts, especially those transcribed automatically. However, a problem arises regarding the length of the texts that require correction, which contradicts one of the fundamental principles of the digital era: the ability to correct, analyse, and process vast amounts of data, be it text or numbers.

At the beginning of the test, we used GPT-3, which could only analyse and correct texts of up to 1000 characters. In our case, this meant processing one letter at a time. This limitation, on the one hand, prevents historians from obtaining a correction for the entire body of texts in a single phase. On the other hand, it does not allow a revision process that takes into account linguistic

---

[7] Set of examples used to fit the parameters of the model, i.e., the data on which the knowledge in the net is based.

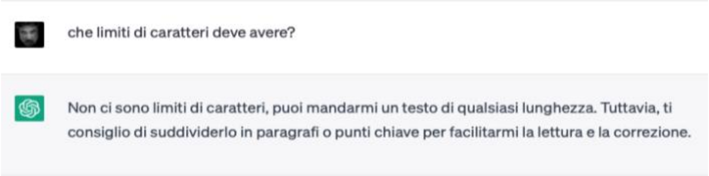
[8] Set of examples that provides an unbiased evaluation of a model, used to tune the model's parameters during training.

[9] Despite its name, ChatGPT has not been primarily used as a chatbot.

and syntactic correlations or any relevant data within the corpus that could contribute to a more thorough correction. Additionally, if scholars intend to correct multiple texts, they will have to provide as many prompts as the number of letters required.

The latest update of this AI (in May 2023), as stated by OpenAI, no longer has this limitation (Figura 2, Figura 3, and Figura 4). The company claims that «GPT-4 can handle over 25,000 words of text, enabling use cases such as creating long-form content, engaging in extended conversations, and conducting document research and analysis». This statement should allow historians to correct the entire corpus of letters, which consists of approximately 900,000 characters.

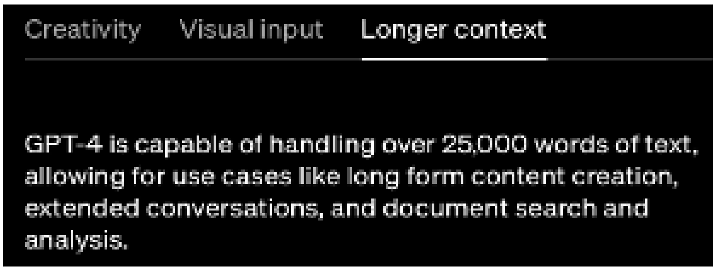
However, once again, the desired outcome cannot be achieved: the character limit in GPT-3 is still active in its subsequent versions (such as GPT-4), and although there has been a significant improvement in computational instructions and text analysis capabilities, the epistolography, which comprises 10,662 words, cannot be corrected.



che limiti di caratteri deve avere?

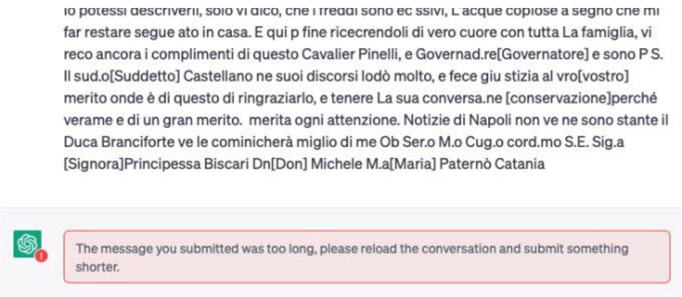
Non ci sono limiti di caratteri, puoi mandarmi un testo di qualsiasi lunghezza. Tuttavia, ti consiglio di suddividerlo in paragrafi o punti chiave per facilitarmi la lettura e la correzione.

Figura 2



GPT-4 is capable of handling over 25,000 words of text, allowing for use cases like long form content creation, extended conversations, and document search and analysis.

Figura 3



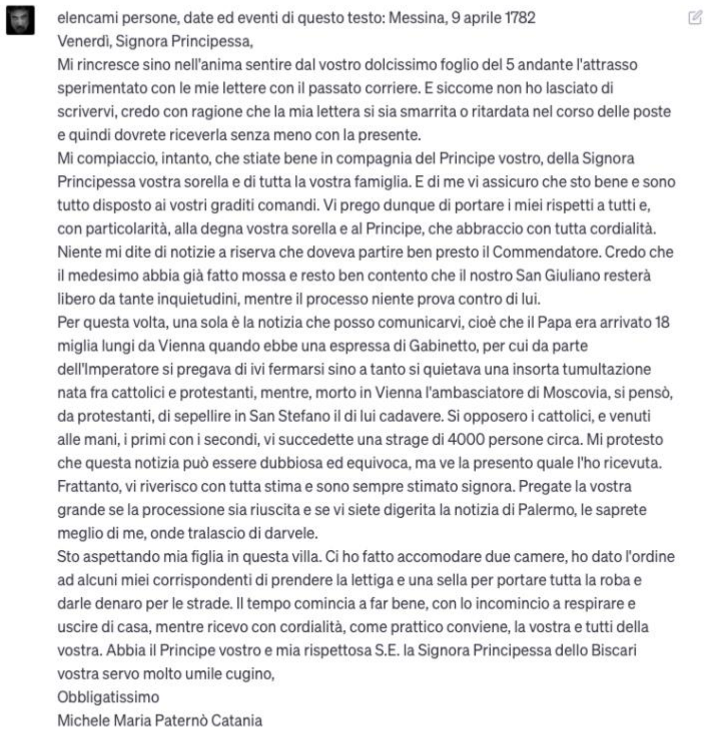
io potessi descriverli, solo vi dico, che i freddi sono ec ssivi, L'acque copiose a segno che mi far restare segue ato in casa. E qui p fine ricecrendoli di vero cuore con tutta La famiglia, vi reco ancora i complimenti di questo Cavalier Pinelli, e Governad.re[Governatore] e sono P S. Il sud.o[Suddetto] Castellano ne suoi discorsi lodò molto, e fece giu stizia al vro[vostro] merito onde è di questo di ringraziarlo, e tenere La sua conversa.ne [conservazione]perché verame e di un gran merito. merita ogni attenzione. Notizie di Napoli non ve ne sono stante il Duca Branciforte ve le cominicherà miglio di me Ob Ser.o M.o Cug.o cord.mo S.E. Sig.a [Signora]Principessa Biscari Dn[Don] Michele M.a[Maria] Paternò Catania

The message you submitted was too long, please reload the conversation and submit something shorter.

Figura 4

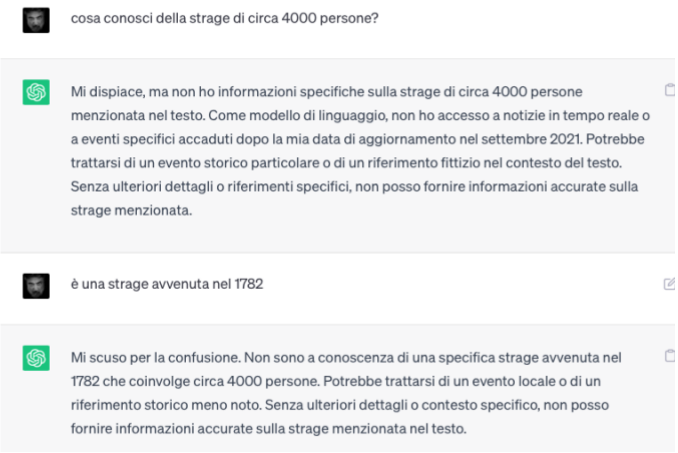
Although ChatGPT is a powerful language generation model, it has a limitation: it lacks access to external information and internet browsing. Consequently, it cannot provide accurate or updated information on a wide range of topics, and it may struggle to generate responses to complex or unconventional questions ([6]).

This constraint prevents scholars from analysing texts that contain information that could enrich the understanding of historical events, as in the case of the letters sent by Michele Maria Paternò to Anna Morso, wife of Ignazio, the fifth prince of Biscari. These letters cover a period (1781-1786) marked by various events related to the History of Sicily, Naples, and the Bourbon court (intrigues, attempts at reforms, management of seismic emergencies such as the reconstruction of the Catania Cathedral, and aspects of religious life). However, they cannot offer valuable insights into the History of Sicily in the modern era (Figura 5 and Figura 6).



elencami persone, date ed eventi di questo testo: Messina, 9 aprile 1782
Venerdì, Signora Principessa,
Mi rincresce sino nell'anima sentire dal vostro dolcissimo foglio del 5 andante l'attrasso sperimentato con le mie lettere con il passato corriere. E siccome non ho lasciato di scrivervi, credo con ragione che la mia lettera si sia smarrita o ritardata nel corso delle poste e quindi dovrete riceverla senza meno con la presente.
Mi compiaccio, intanto, che stiate bene in compagnia del Principe vostro, della Signora Principessa vostra sorella e di tutta la vostra famiglia. E di me vi assicuro che sto bene e sono tutto disposto ai vostri graditi comandi. Vi prego dunque di portare i miei rispetti a tutti e, con particolarità, alla degna vostra sorella e al Principe, che abbraccio con tutta cordialità. Niente mi dite di notizie a riserva che doveva partire ben presto il Commendatore. Credo che il medesimo abbia già fatto mossa e resto ben contento che il nostro San Giuliano resterà libero da tante inquietudini, mentre il processo niente prova contro di lui.
Per questa volta, una sola è la notizia che posso comunicarvi, cioè che il Papa era arrivato 18 miglia lungi da Vienna quando ebbe una espressa di Gabinetto, per cui da parte dell'Imperatore si pregava di ivi fermarsi sino a tanto si quietava una insorta tumultazione nata fra cattolici e protestanti, mentre, morto in Vienna l'ambasciatore di Moscovia, si pensò, da protestanti, di sepellire in San Stefano il di lui cadavere. Si opposero i cattolici, e venuti alle mani, i primi con i secondi, vi succedette una strage di 4000 persone circa. Mi protesto che questa notizia può essere dubbiosa ed equivoca, ma ve la presento quale l'ho ricevuta. Frattanto, vi riverisco con tutta stima e sono sempre stimato signora. Pregate la vostra grande se la processione sia riuscita e se vi siete digerita la notizia di Palermo, le saprete meglio di me, onde tralascio di darvele.
Sto aspettando mia figlia in questa villa. Ci ho fatto accomodare due camere, ho dato l'ordine ad alcuni miei corrispondenti di prendere la lettiga e una sella per portare tutta la roba e darle denaro per le strade. Il tempo comincia a far bene, con lo incomincio a respirare e uscire di casa, mentre ricevo con cordialità, come prattico conviene, la vostra e tutti della vostra. Abbia il Principe vostro e mia rispettosa S.E. la Signora Principessa dello Biscari vostra servo molto umile cugino,
Obbligatissimo
Michele Maria Paternò Catania

Figura 5



cosa conosci della strage di circa 4000 persone?

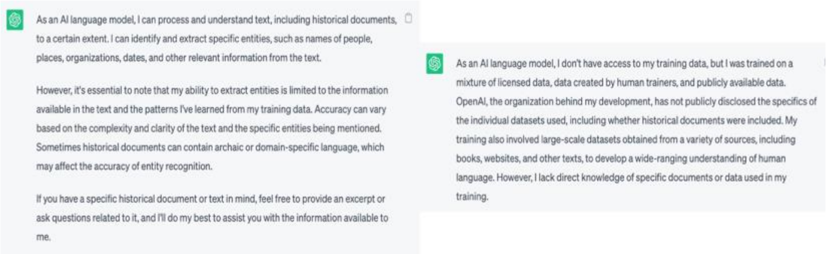
Mi dispiace, ma non ho informazioni specifiche sulla strage di circa 4000 persone menzionata nel testo. Come modello di linguaggio, non ho accesso a notizie in tempo reale o a eventi specifici accaduti dopo la mia data di aggiornamento nel settembre 2021. Potrebbe trattarsi di un evento storico particolare o di un riferimento fittizio nel contesto del testo. Senza ulteriori dettagli o riferimenti specifici, non posso fornire informazioni accurate sulla strage menzionata.

è una strage avvenuta nel 1782

Mi scuso per la confusione. Non sono a conoscenza di una specifica strage avvenuta nel 1782 che coinvolge circa 4000 persone. Potrebbe trattarsi di un evento locale o di un riferimento storico meno noto. Senza ulteriori dettagli o contesto specifico, non posso fornire informazioni accurate sulla strage menzionata nel testo.

Figura 6

As a consequence, ChatGPT cannot perform text-to-web comparisons. Furthermore, like most LLMs, OpenAI's GPT has not been explicitly trained for information extraction tasks (*e.g.*, Named Entity Recognition and Classification — NERC, and relation extraction) ([31]), especially concerning historical documents (Figura 7).



As an AI language model, I can process and understand text, including historical documents, to a certain extent. I can identify and extract specific entities, such as names of people, places, organizations, dates, and other relevant information from the text.

However, it's essential to note that my ability to extract entities is limited to the information available in the text and the patterns I've learned from my training data. Accuracy can vary based on the complexity and clarity of the text and the specific entities being mentioned. Sometimes historical documents can contain archaic or domain-specific language, which may affect the accuracy of entity recognition.

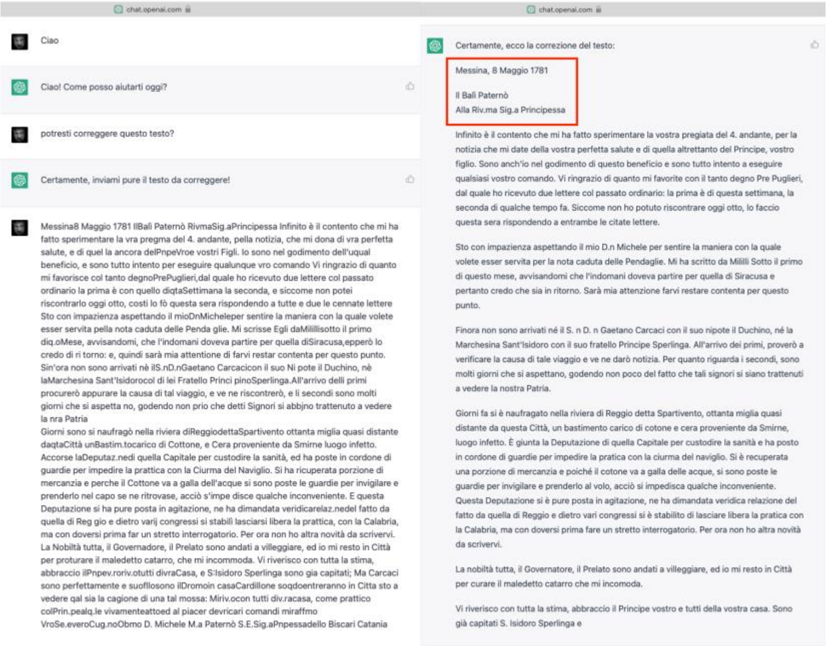
If you have a specific historical document or text in mind, feel free to provide an excerpt or ask questions related to it, and I'll do my best to assist you with the information available to me.

As an AI language model, I don't have access to my training data, but I was trained on a mixture of licensed data, data created by human trainers, and publicly available data. OpenAI, the organization behind my development, has not publicly disclosed the specifics of the individual datasets used, including whether historical documents were included. My training also involved large-scale datasets obtained from a variety of sources, including books, websites, and other texts, to develop a wide-ranging understanding of human language. However, I lack direct knowledge of specific documents or data used in my training.

Figura 7

The correction, therefore, is necessarily performed on the individual epistles.

ChatGPT first aims for a probability distribution over word or token sequences, enabling it to predict the next word in a sequence. In practice, however, there is an apparent disconnect between the training of language models (LLMs) and their intended use for valuable cognitive work. Although LLMs can generate text, they cannot perform accurate corrections. The text generated by the AI may align with the original text, somewhat satisfying researchers' expectations (Figura 8). The statistical architecture of the machine allows it to generate a text where the distribution of tokens conveys the linguistic sense of the text to be corrected. However, without specific instructions from the user aimed at a potential "ground truth", in many cases, even if the AI generates a text with minor corrections, the "next-token-prediction" may yield incorrect names or expand abbreviations inaccurately. Conversely, it can generate the name of a month composed of numbers and letters (8bre, 9mbre, Xmbre). Moreover, owing to its inherent functionality, the AI might introduce names that are absent in the original text or paragraphs and may not faithfully adhere to the original content (as shown in Figura 8 and Figura 9).



Ciao

Ciao! Come posso aiutarti oggi?

potresti correggere questo testo?

Certamente, inviami pure il testo da correggere!

Messina8 Maggio 1781 IlBalì Paternò RivmaSig.aPrincipessa Infinito è il contento che mi ha fatto sperimentare la vra pregma del 4. andante, pella notizia, che mi dona di vra perfetta salute, e di quel la ancora delPnpeVroe vostri Figli. Io sono nel godimento dell'uqual beneficio, e sono tutto intento per eseguire qualunque vro comando Vi ringrazio di quanto mi favorisce col tanto degnoPrePuglieri,dal quale ho ricevuto due lettere col passato ordinario la prima è con quello diqtaSettimana la seconda, e siccome non potei riscontrarlo oggi otto, costi lo fò questa sera rispondendo a tutte e due le cennate lettere Sto con impazienza aspettando il mioDnMicheleper sentire la maniera con la quale volete esser servita pella nota caduta delle Penda glie. Mi scrisse Egli daMilillisotto il primo diq.oMese, avvisandomi, che l'indomani doveva partire per quella diSiracusa,epperò lo credo di ri torno: e, quindi sarà mia attentione di farvi restar contenta per questo punto. Sin'ora non sono arrivati nè ilS.nD.nGaetano Carcacicon il suo Ni pote il Duchino, nè laMarchesina Sant'Isidorocol di lei Fratello Princi pinoSperlinga.All'arrivo delli primi procurerò appurare la causa di tal viaggio, e ve ne riscontrerò, e li secondi sono molti giorni che si aspetta no, godendo non prio che detti Signori si abbjno trattenuto a vedere la nra Patria
Giorni sono si naufragò nella riviera diReggiodettaSpartivento ottanta miglia quasi distante daqtaCittà unBastim.tocarico di Cottone, e Cera proveniente da Smirne luogo infetto. Accorse laDeputaz.nedi quella Capitale per custodire la sanità, ed ha poste in cordone di guardie per impedire la prattica con la Ciurma del Naviglio. Si ha ricuperata porzione di mercanzia e perche il Cottone va a galla dell'acque si sono poste le guardie per invigilare e prenderlo nel capo se ne ritrovase, acciò s'impe disce qualche inconveniente. E questa Deputazione si ha pure posta in agitazione, ne ha dimandata veridicarelaz.nedel fatto da quella di Reg gio e dietro varij congressi si stabilì lasciarsi libera la prattica, con la Calabria, ma con doversi prima far un stretto interrogatorio. Per ora non ho altra novità da scrivervi. La Nobiltà tutta, il Governadore, il Prelato sono andati a villeggiare, ed io mi resto in Città per proturare il maledetto catarro, che mi incommoda. Vi riverisco con tutta la stima, abbraccio ilPnpev.roriv.otutti divraCasa, e S:Isidoro Sperlinga sono gia capitati; Ma Carcaci sono perfettamente e suofllosono ilDromoin casaCardillone soqdoentreranno in Citta sto a vedere qal sia la cagione di una tal mossa: Miriv.ocon tutti div.racasa, come prattico colPrin.pealq.le vivamenteattoed al piacer devricari comandi miraffmo VroSe.everoCug.noObmo D. Michele M.a Paternò S.E.Sig.aPnpessadello Biscari Catania

Certamente, ecco la correzione del testo:

Messina, 8 Maggio 1781

Il Bali Paternò
Alla Riv.ma Sig.a Principessa

Infinito è il contento che mi ha fatto sperimentare la vostra pregiata del 4. andante, per la notizia che mi date della vostra perfetta salute e di quella altrettanto del Principe, vostro figlio. Sono anch'io nel godimento di questo beneficio e sono tutto intento a eseguire qualsiasi vostro comando. Vi ringrazio di quanto mi favorite con il tanto degno Pre Puglieri, dal quale ho ricevuto due lettere col passato ordinario: la prima è di questa settimana, la seconda di qualche tempo fa. Siccome non ho potuto riscontrare oggi otto, lo faccio questa sera rispondendo a entrambe le citate lettere.

Sto con impazienza aspettando il mio D.n Michele per sentire la maniera con la quale volete esser servita per la nota caduta delle Pendaglie. Mi ha scritto da Mililli Sotto il primo di questo mese, avvisandomi che l'indomani doveva partire per quella di Siracusa e pertanto credo che sia in ritorno. Sarà mia attenzione farvi restare contenta per questo punto.

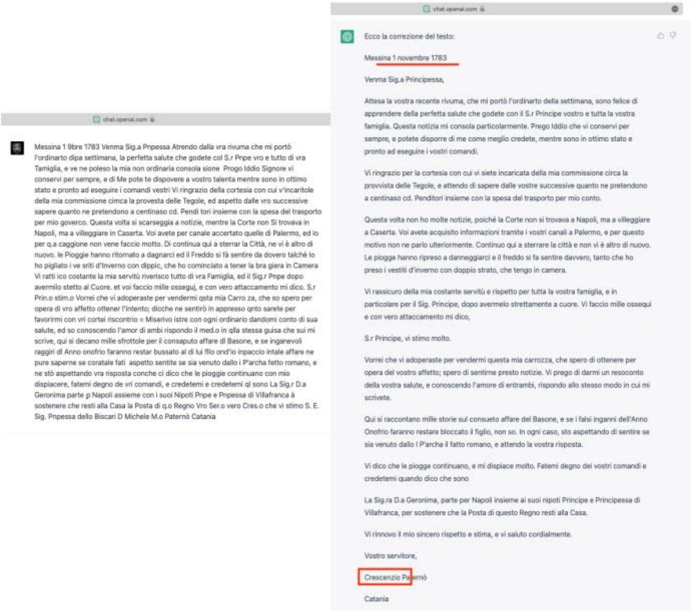
Finora non sono arrivati né il S. n D. n Gaetano Carcaci con il suo nipote il Duchino, né la Marchesina Sant'Isidoro con il suo fratello Principe Sperlinga. All'arrivo dei primi, proverò a verificare la causa di tale viaggio e ve ne darò notizia. Per quanto riguarda i secondi, sono molti giorni che si aspettano, godendo non poco del fatto che tali signori si siano trattenuti a vedere la nostra Patria.

Giorni fa si è naufragato nella riviera di Reggio detta Spartivento, ottanta miglia quasi distante da questa Città, un bastimento carico di cotone e cera proveniente da Smirne, luogo infetto. È giunta la Deputazione di quella Capitale per custodire la sanità e ha posto in cordone di guardie per impedire la pratica con la ciurma del naviglio. Si è recuperata una porzione di mercanzia e poiché il cotone va a galla delle acque, si sono poste le guardie per invigilare e prenderlo al volo, acciò si impedisca qualche inconveniente. Questa Deputazione si è pure posta in agitazione, ne ha dimandata veridica relazione del fatto da quella di Reggio e dietro vari congressi si è stabilito di lasciare libera la pratica con la Calabria, ma con doversi prima fare un stretto interrogatorio. Per ora non ho altra novità da scrivervi.

La nobiltà tutta, il Governatore, il Prelato sono andati a villeggiare, ed io mi resto in Città per curare il maledetto catarro che mi incomoda.

Vi riverisco con tutta la stima, abbraccio il Principe vostro e tutti della vostra casa. Sono già capitati S. Isidoro Sperlinga e

Figura 8



Messina 1 9bre 1783 Venma Sig.a Pnpessa Atrendo dalla vra rivuma che mi portò l'ordinarto dipa settimana, la perfetta salute che godete col S.r Pnpe vro e tutto di vra Tamiglia, e ve ne poleso la mia non ordinaria consola sione Progo Iddio Signore vi conservi per sempre, e di Me pote te dispovere a vostro talenta mentre sono in ottimo stato e pronto ad eseguire i comandi vestri Vi ringrazio della cortesia con cui v'incaritole della mia commissione cimca la provesta delle Tegole, ed aspetto dalle vro successive sapere quanto ne pretendono a centinaso cd. Pendi tori insieme con la spesa del trasporto per mio goverco. Questa volta si scarseggia a notizie, mentre la Corte non Si trovava in Napoli, ma a villeggiare in Caserta. Voi avete per canale accertato quelle di Palermo, ed io per q.a caggione non vene faccio motto. Di continua qui a sterrar la Città, ne vi è altro di nuovo. le Pioggie hanno ritornato a dagnarci ed il Freddo si fà sentire da dovero talchè lo ho pigliato i ve sriti d'Inverno con dippic, che ho cominciato a tener la bra giera in Camera Vi ratti ico costante la mia servitù riverisco tutto di vra Famiglia, ed il Sig.r Pnpe dopo avermilo stetto al Cuore. et voi faccio mille osseguj, e con vero attaccamento mi dico. S.r Prin.o stim.o Vorrei che vi adoperaste per vendermi qsta mia Carro za, che so spero per opera di vro affetto ottener l'intento; dioche ne sentirò in appresso qnto sarete per favorirmi con vri cortei riscontrio = Miserivo istre con ogni ordinario dandomi conto di sua salute, ed so conoscendo l'amor di ambi rispondo il med.o in qlla stessa guisa che sui mi scrive, qui si decano mille sfrottole per il consaputo affare di Basone, e se inganevoli raggiri di Anno onofrio faranno restar bussato al di lui filo ond'io inpaccio intale affare ne pure saperne se coratale fati aspetto sentite se sia venuto dallo i P'archa fetto romano, e ne stò aspettando vra risposta conche ci dico che le pioggie continuano con mio dispiacere, fatemi degno de vri comandi, e credetemi e credetemi ql sono La Sig.r D.a Geronima parte p Napoli assieme con i suoi Nipoti Pnpe e Pnpessa di Villafranca à sostenere che resti alla Casa la Posta di q.o Regno Vro Ser.o vero Cres.o che vi stimo S. E. Sig. Pnpessa dello Biscari D Michele M.o Paternò Catania

Ecco la correzione del testo:

Messina 1 novembre 1783

Venma Sig.a Principessa,

Attesa la vostra recente rivuma, che mi portò l'ordinarto della settimana, sono felice di apprendere della perfetta salute che godete con il S.r Principe vostro e tutta la vostra famiglia. Questa notizia mi consola particolarmente. Prego Iddio che vi conservi per sempre, e potete disporre di me come meglio credete, mentre sono in ottimo stato e pronto ad eseguire i vostri comandi.

Vi ringrazio per la cortesia con cui vi siete incaricata della mia commissione circa la provvista delle Tegole, e attendo di sapere dalle vostre successive quanto ne pretendono a centinaso cd. Penditori insieme con la spesa del trasporto per mio conto.

Questa volta non ho molte notizie, poiché la Corte non si trovava a Napoli, ma a villeggiare a Caserta. Voi avete acquisito informazioni tramite i vostri canali a Palermo, e per questo motivo non ne parlo ulteriormente. Continuo qui a sterrare la città e non vi è altro di nuovo. Le piogge hanno ripreso a danneggiarci e il freddo si fa sentire davvero, tanto che ho preso i vestiti d'inverno con doppio strato, che tengo in camera.

Vi rassicuro della mia costante servitù e rispetto per tutta la vostra famiglia, e in particolare per il Sig. Principe, dopo avermelo strettamente a cuore. Vi faccio mille ossequi e con vero attaccamento mi dico,

S.r Principe, vi stimo molto.

Vorrei che vi adoperaste per vendermi questa mia carrozza, che spero di ottenere per opera del vostro affetto; spero di sentirne presto notizie. Vi prego di darmi un resoconto della vostra salute, e conoscendo l'amore di entrambi, rispondo allo stesso modo in cui mi scrivete.

Qui si raccontano mille storie sul consueto affare del Basone, e se i falsi inganni dell'Anno Onofrio faranno restare bloccato il figlio, non so. In ogni caso, sto aspettando di sentire se sia venuto dallo I P'archa il fatto romano, e attendo la vostra risposta.

Vi dico che le piogge continuano, e mi dispiace molto. Fatemi degno dei vostri comandi e credetemi quando dico che sono

La Sig.ra D.a Geronima, parte per Napoli insieme ai suoi nipoti Principe e Principessa di Villafranca, per sostenere che la Posta di questo Regno resti alla Casa.

Vi rinnovo il mio sincero rispetto e stima, e vi saluto cordialmente.

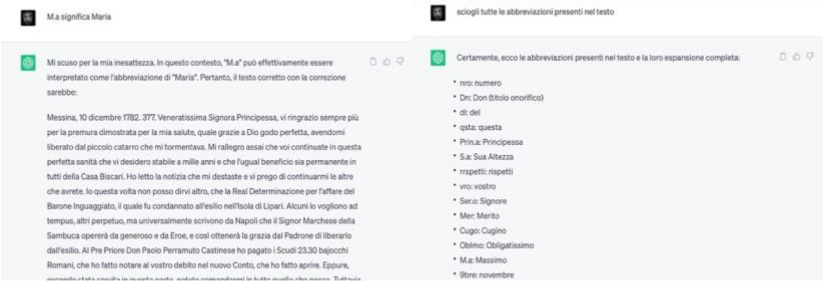
Vostro servitore,

Crescenzio Paternò

Catania

Figura 9

We proceeded by providing instructions to improve the result, prompting to expand the abbreviations found in the text and correcting those that have been expanded incorrectly (Reinforcement Learning with Human Feedback — RLHF). Subsequently, the user's instructions have successfully enabled the correct expansion of abbreviations in other epistles, resulting in a normalised text version of the latter and a list of all abbreviations present in the corpus (Figura 10).



M.a significa Maria

Mi scuso per la mia inesattezza. In questo contesto, "M.a" può effettivamente essere interpretato come l'abbreviazione di "Maria". Pertanto, il testo corretto con la correzione sarebbe:

Messina, 10 dicembre 1782. 377. Veneratissima Signora Principessa, vi ringrazio sempre più per la premura dimostrata per la mia salute, quale grazie a Dio godo perfetta, avendomi liberato dal piccolo catarro che mi tormentava. Mi rallegro assai che voi continuate in questa perfetta sanità che vi desidero stabile a mille anni e che l'ugual beneficio sia permanente in tutti della Casa Biscari. Ho letto la notizia che mi destaste e vi prego di continuarmi le altre che avrete. Io questa volta non posso dirvi altro, che la Real Determinazione per l'affare del Barone Inguaggiato, il quale fu condannato all'esilio nell'Isola di Lipari. Alcuni lo vogliono ad tempus, altri perpetuo, ma universalmente scrivono da Napoli che il Signor Marchese della Sambuca opererà da generoso e da Eroe, e così otterrà la grazia dal Padrone di liberarlo dall'esilio. Al Pre Priore Don Paolo Perramuto Castinese ho pagato i Scudi 23.30 bajocchi Romani, che ho fatto notare al vostro debito nel nuovo Conto, che ho fatto aprire. Eppure,

sciogli tutte le abbreviazioni presenti nel testo

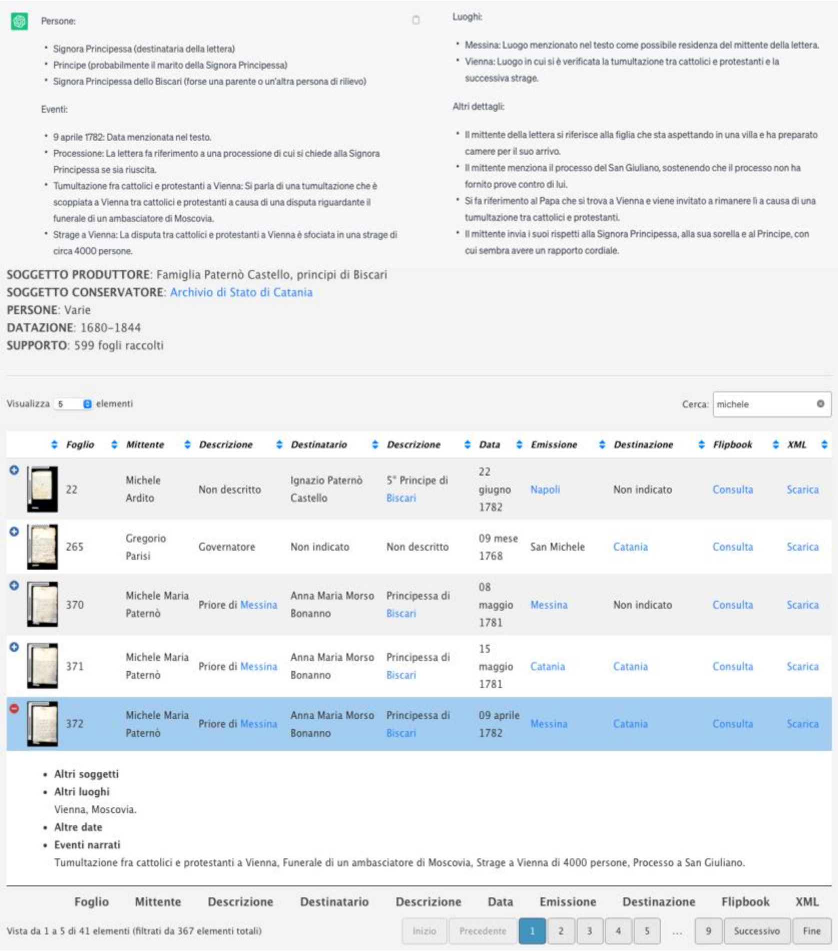
Certamente, ecco le abbreviazioni presenti nel testo e la loro espansione completa:

- nro: numero
- Dn: Don (titolo onorifico)
- dl: del
- qsta: questa
- Prin.a: Principessa
- S.a: Sua Altezza
- rrspetti: rispetti
- vro: vostro
- Ser.o: Signore
- Mer: Merito
- Cugo: Cugino
- Oblmo: Obligatissimo
- M.a: Massimo
- 9bre: novembre

Figura 10

AI enhances numerous fields within the Humanities. While historians can now have the opportunity to work with much larger volumes of data compared to the Past, archivists can utilise these computational technologies to improve their workflow. Nowadays, Archival Science is also a "Networked Science" since our archival heritage is constantly being digitised and disseminated on the web through websites and digital databases. However, the digitisation process requires a certain level of dynamism based on the principle of constructing a machine-readable version of the documents. Nevertheless, before reaching the completeness of archival information and achieving meticulous and analytical digitisation, archivists have the task of creating cards that at least provide helpful data for online research, even if it is not possible to include the complete transcription of the photographed document.

In epistolography, creating these cards involves identifying specific information, such as the individuals involved (sender, recipient, and other persons), places (origin, destination, and other locations), dates, and, when possible, events. These pieces of information make the digital archive more dynamic. Therefore, archivists' work is meticulous and requires a deep understanding of the written document. For this reason, although recent studies have highlighted inaccuracies in the recognition of entities in historical documents ([14]), in our specific case, ChatGPT was able to identify entities and events with absolute precision, providing details that have led us to create detailed records of the documents. On that account, with the assistance of ChatGPT, we have successfully identified relevant information to compile records on the "Biscari Epistolography" website (Figura 11), demonstrating that GPT-3 is a technology that can genuinely serve archivists and historians.



Persone:

- Signora Principessa (destinataria della lettera)
- Principe (probabilmente il marito della Signora Principessa)
- Signora Principessa dello Biscari (forse una parente o un'altra persona di rilievo)

Eventi:

- 9 aprile 1782: Data menzionata nel testo.
- Processione: La lettera fa riferimento a una processione di cui si chiede alla Signora Principessa se sia riuscita.
- Tumultazione fra cattolici e protestanti a Vienna: Si parla di una tumultazione che è scoppiata a Vienna tra cattolici e protestanti a causa di una disputa riguardante il funerale di un ambasciatore di Moscovia.
- Strage a Vienna: La disputa tra cattolici e protestanti a Vienna è sfociata in una strage di circa 4000 persone.

Luoghi:

- Messina: Luogo menzionato nel testo come possibile residenza del mittente della lettera.
- Vienna: Luogo in cui si è verificata la tumultazione tra cattolici e protestanti e la successiva strage.

Altri dettagli:

- Il mittente della lettera si riferisce alla figlia che sta aspettando in una villa e ha preparato camere per il suo arrivo.
- Il mittente menziona il processo del San Giuliano, sostenendo che il processo non ha fornito prove contro di lui.
- Si fa riferimento al Papa che si trova a Vienna e viene invitato a rimanere lì a causa di una tumultazione tra cattolici e protestanti.
- Il mittente invia i suoi rispetti alla Signora Principessa, alla sua sorella e al Principe, con cui sembra avere un rapporto cordiale.

SOGGETTO PRODUTTORE: Famiglia Paternò Castello, principi di Biscari
SOGGETTO CONSERVATORE: Archivio di Stato di Catania
PERSONE: Varie
DATAZIONE: 1680-1844
SUPPORTO: 599 fogli raccolti

Visualizza 5 elementi

Cerca: michele

	Foglio	Mittente	Descrizione	Destinatario	Descrizione	Data	Emissione	Destinazione	Flipbook	XML
	22	Michele Ardito	Non descritto	Ignazio Paternò Castello	5° Principe di Biscari	22 giugno 1782	Napoli	Non indicato	Consulta	Scarica
	265	Gregorio Parisi	Governatore	Non indicato	Non descritto	09 mese 1768	San Michele	Catania	Consulta	Scarica
	370	Michele Maria Paternò	Priore di Messina	Anna Maria Morso Bonanno	Principessa di Biscari	08 maggio 1781	Messina	Non indicato	Consulta	Scarica
	371	Michele Maria Paternò	Priore di Messina	Anna Maria Morso Bonanno	Principessa di Biscari	15 maggio 1781	Catania	Catania	Consulta	Scarica
	372	Michele Maria Paternò	Priore di Messina	Anna Maria Morso Bonanno	Principessa di Biscari	09 aprile 1782	Messina	Catania	Consulta	Scarica

- **Altri soggetti**
- **Altri luoghi**
  Vienna, Moscovia.
- **Altre date**
- **Eventi narrati**
  Tumultazione fra cattolici e protestanti a Vienna, Funerale di un ambasciatore di Moscovia, Strage a Vienna di 4000 persone, Processo a San Giuliano.

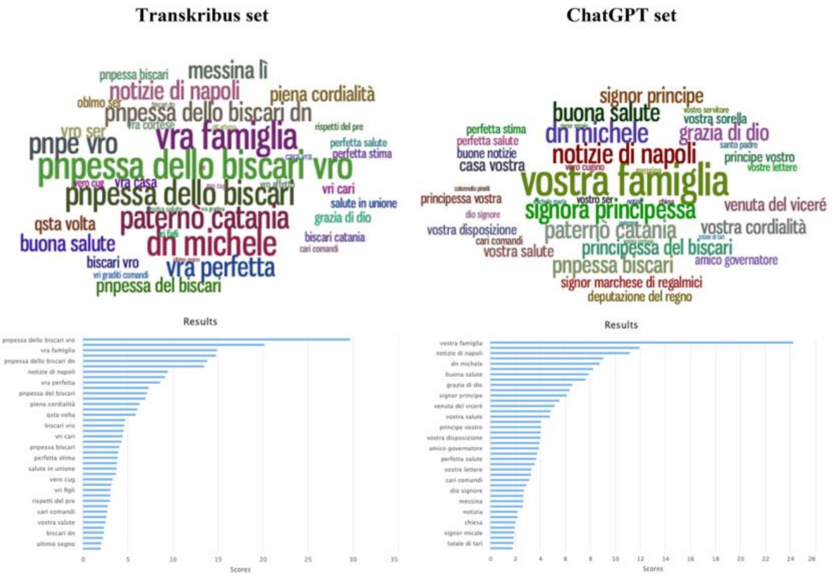
Foglio Mittente Descrizione Destinatario Descrizione Data Emissione Destinazione Flipbook XML

Vista da 1 a 5 di 41 elementi (filtrati da 367 elementi totali)

Figura 11

Observations

Historical scholarship is primarily based on text-based analysis —where the original wording of historical sources often provides new insights and inferences of specific meanings— while Archival Science is called to reassess the principle of disseminating the documentary heritage. Thus, both historians and archivists, albeit from different points of view, need to work on texts that have been transcribed correctly due to the advantages offered by computing normalised text corpora in modern languages. Firstly, this enables the dissemination of results without the need for manual normalisation; secondly, it allows for the application of computational linguistic analysis tools that can uncover additional data, a feature that appears to be crucial for accurate storytelling. Tools like Keyphrase Digger ([25]) are more effective in processing modern-language normalised texts than those written in 18th-century style (as seen in Figura 12).



Transkribus set

ChatGPT set

Figura 12

The unique features of 18th-century texts, such as abbreviations, syntactic errors, and incorrect word forms, would require specific modelling to train artificial Intelligence. It should be noted that querying neural networks and the semantic web with an 18th-century text might lead to its exclusion.

Artificial Intelligence is a rapidly advancing field within Computer Science, focused on creating intelligent machines capable of human-like thinking and action. While AI has found applications in various domains, History remains a discipline rooted in archival sources. Accurate digitisation and automatic transcription projects are essential for providing the necessary information to develop models that can effectively assist historians in their research.

However, specific observations seem necessary. The *digital turn* and digitisation projects have inundated historians with vast historical data extracted from digitised sources, posing challenges in managing and analysing such Big Data from the Past. ChatGPT —despite lacking sentient intelligence— utilises semantic processing and syntactic rules to generate a dialogue that may give the impression of understanding. Although it lacks a biological brain structure, the training on syntactic structures and instructions makes GPT-3 a valuable tool for normalising historical source corpora. Corrections made by the AI, even if unrelated to the text itself, such as reformulating propositions or merging names and abbreviations, do not distort the source but rather unearth new data for historians to explore.

Historical research relies on archival sources, which transmit more than just the linguistic structure of the text; they possess semantic structures. Nevertheless, this does not imply that historical methodology should forego the advantages offered by ITC innovations that ensure accuracy at specific stages of the workflow of historians. Transkribus, for instance, is an AI-powered tool used to transcribe digitised archival documents with impressive speed, generating machine-readable files that enable historians to analyse, process, and create digital editions. Though ChatGPT cannot invent new historical events, it has been trained to write about History, albeit not at the same level as scholars. It can be a valuable tool for archivists and historians, serving their objective and compelling narrative by normalising information from digitised, encoded, and automatically transcribed sources.

References

[1] Abbott, Alison. 2017. 'The "Time Machine" Reconstructing Ancient Venice's Social Networks'. *Nature*, no. 546: 341–44.

[2] Alshater, Muneer M. 2022. 'Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT'. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4312358.

[3] Calabrese, Gaetano. 2003. L'archivio della famiglia Paternò Castello principi di Biscari: inventario. Catania: Os. n.!

[4] Cartesio, Renato. 1999. *Discorso sul metodo*. Armando Editore. https://play.google.com/store/books/details/Renato_Cartesio_Meditazioni_metafisiche?id=7-adikgWNRMC.

[5] Cartesio, Renato. 2003. *Meditazioni metafisiche*. Armando Editore. https://play.google.com/store/books/details/Renato_Cartesio_Meditazioni_metafisiche?id=7-adikgWNRMC.

[6] Deng, Jianyang, and Yijia Lin. 2022. 'The Benefits and Challenges of ChatGPT: An Overview'. *Frontiers in Computing and Intelligent Systems* 2 (2): 81–83. https://doi.org/10.54097/fcis.v2i2.4465.

[7] Droysen, Johann Gustav. 1868. *Grundriss Der Historik*. Leipzig: Veit.

[8] Erwin, Brittany. 2020. 'Digital Tools for Studying Empire: Transcription and Text Analysis with Transkribus'. Not Even Past. 6 November 2020. https://notevenpast.org/digital-tools-for-studying-empire-transcription-and-text-analysis-with-transkribus/.

[9] Fiormonte, Domenico. 2000. 'Il Documento Immateriale. Ricerca Storica e Nuovi Linguaggi'. Edited by Guido Abbatista and Andrea Zorzi. *L'Indice* 5 (Dossier 4).

[10] Floridi, Luciano. 2012. *La rivoluzione dell'informazione*. Torino: Codice. http://trentino.medialibrary.it/media/scheda.aspx?idm=20146&idtm=30015&page=0&preferiti=.

[11] Floridi, Luciano. 2015. *The Ethics of Information*.

[12] Floridi, Luciano, and Massimo Chiriatti. 2020. 'GPT-3: Its Nature, Scope, Limits, and Consequences'. *Minds and Machines* 30 (4): 681–94. https://doi.org/10.1007/s11023-020-09548-1.

[13] Gautier, Dassonneville, Adèle Huguet, Marie-Laure Massot, Agnès Tricoche, Marie Carlin, Jean-Philippe Moreux, and Rostaing Aurélia. 2022. 'Compte-rendu de la journée d'étude « Point HTR 2022 » Transkribus / eScriptorium : Transcrire,

annoter et éditer numériquement des documents d'archives'. Report, CAPHES - UMS 3610 CNRS/ENS ; AOROC. https://hal.science/hal-03692413.

[14] González-Gallardo, Carlos-Emiliano, Emanuela Boros, Nancy Girdhar, Ahmed Hamdi, Jose G. Moreno, and Antoine Doucet. 2023. 'Yes but.. Can ChatGPT Identify Entities in Historical Documents?' https://doi.org/10.48550/arXiv.2303.17322.

[15] Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data. The Historian's Macroscope*. Imperial College Press. https://www.perlego.com/book/839949/exploring-big-historical-data-the-historians-macroscope-pdf.

[16] Huff, Dorothee, and Kristina Stöbener. 2022. 'Projekt OCR-BW: Automatische Texterkennung von Handschriften'. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 9 (4): 1–19. https://doi.org/10.5282/o-bib/5885.

[17] Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023. 'Is ChatGPT A Good Translator? A Preliminary Study'. arXiv. https://doi.org/10.48550/arXiv.2301.08745.

[18] Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 2017. 'Transkribus. A Service Platform for Transcription, Recognition and Retrieval of Historical Documents'. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 19–24. https://doi.org/10.1109/ICDAR.2017.307.

[19] Kaplan, Frédéric. 2015. 'The Venice Time Machine'.

[20] Kaplan, Frédéric, and Isabella di Lenardo. 2017. 'Big Data of the Past'. *Frontiers in Digital Humanities* 4. https://doi.org/10.3389/fdigh.2017.00012.

[21] Kiessling, Benjamin, Robin Tissot, Peter Stokes, and Daniel Stökl Ben Ezra. 2019. 'EScriptorium: An Open Source Platform for Historical Document Analysis'. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2:19–19. https://doi.org/10.1109/ICDARW.2019.10032.

[22] Lazzari, Gianrocco, Giovanni Colavizza, Fabio Bortoluzzi, Davide Drago, Andrea Erboso, Francesca Zugno, Frédéric Kaplan, and Marcel Salathé. 2020. 'A Digital Reconstruction of the 1630–1631 Large Plague Outbreak in Venice'. *Scientific Reports* 10 (1): 17849. https://doi.org/10.1038/s41598-020-74775-6.

[23] Maarand, Martin, Yngvil Beyer, Andre Kåsen, Knut T. Fosseide, and Christopher Kermorvant. 2022. 'A Comprehensive Comparison of Open-Source Libraries for Handwritten Text Recognition in Norwegian'. In *Document Analysis Systems: 15th IAPR International Workshop, DAS 2022, La Rochelle, France, May 22–25, 2022, Proceedings*, 399–413. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-031-06555-2_27.

[24] Milioni, Nikolina. 2020. Automatic Transcription of Historical Documents. Transkribus as a Tool for Libraries, Archives and Scholars. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-412565.

[25] Moretti, Giovanni, Rachele Sprugnoli, and Sara Tonelli. 2015. 'Digging in the Dirt: Extracting Keyphrases from Texts with KD'. In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-It 2015*. https://doi.org/10.4000/books.aaccademia.1518.

[26] Muehlberger, Guenter, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, et al. 2019. 'Transforming Scholarship in the Archives through Handwritten Text Recognition. Transkribus as a Case Study'. *Journal of Documentation* 75 (5): 954–76. https://doi.org/10.1108/JD-07-2018-0114.

[27] Muennighoff, Niklas, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M. Saiful Bari, et al. 2022. 'Crosslingual Generalization through Multitask Finetuning'. arXiv. https://doi.org/10.48550/arXiv.2211.01786.

[28] Nockels, Joe, Paul Gooding, Sarah Ames, and Melissa Terras. 2022. 'Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts: A Systematic Review of Transkribus in Published Research'. *Archival Science* 22 (3): 367–92. https://doi.org/10.1007/s10502-022-09397-0.

[29] Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. 2022. 'Training Language Models to Follow Instructions with Human Feedback'. arXiv. https://doi.org/10.48550/arXiv.2203.02155.

[30] Pavlik, John V. 2023. 'Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education'. *Journalism & Mass Communication Educator* 78 (1): 84–93. https://doi.org/10.1177/10776958221149577.

[31] Qin, Chengwei, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. 2023. 'Is ChatGPT a General-Purpose Natural Language Processing Task Solver?' arXiv. https://doi.org/10.48550/arXiv.2302.06476.

[32] Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. 'Language Models Are Unsupervised Multitask Learners'. *OpenAI Blog* 1(8):9.

[33] Sánchez, Joan Andreu, Günter Mühlberger, Basilis Gatos, Philip Schofield, Katrien Depuydt, Richard M. Davis, Enrique Vidal, and Jesse de Does. 2013. 'TranScriptorium: A European Project on Handwritten Text Recognition'. In *Proceedings of the 2013 ACM Symposium on Document Engineering*, 227–28. DocEng '13. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2494266.2494294.

[34] Santoro, Raffaele. 2015. 'I Grandi Archivi: Un Patrimonio Di Big Data'. In . Roma: Archivio di Stato di Roma.

[35] Schiuma, Giovanni, and Daniela Carlucci. 2018. *Big Data in the Arts and Humanities. Theory and Practice.* http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5371976.

[36] Schwarz-Ricci, Vera Isabell. 2022. 'Handwritten Text Recognition per Registri Notarili (Secc. XV-XVI): Una Sperimentazione'. *Umanistica Digitale* 13: 171–81. https://doi.org/10.6092/issn.2532-8816/14926.

[37] Sobania, Dominik, Martin Briesch, Carol Hanna, and Justyna Petke. 2023. 'An Analysis of the Automatic Bug Fixing Performance of ChatGPT'. arXiv. https://doi.org/10.48550/arXiv.2301.08653.

[38] Song, Yangqiu, Shyam Upadhyay, Haoruo Peng, Stephen Mayhew, and Dan Roth. 2019. 'Toward Any-Language Zero-Shot Topic Classification of Textual Documents'. *Artificial Intelligence* 274 (September): 133–50. https://doi.org/10.1016/j.artint.2019.02.002.

[39] Spina, Salvatore. 2023. 'Handwritten Text Recognition as a Digital Perspective of Archival Science'. *Aidainformazioni*, no. 1–2.

[40] Turing, Alan Mathison. 1950. 'Computing Machinery and Intelligence'. *Mind* LIX (236): 433–60. https://doi.org/10.1093/mind/LIX.236.433.

[41] Xian, Yongqin, Bernt Schiele, and Zeynep Akata. 2017. 'Zero-Shot Learning - the Good, the Bad and the Ugly'. In , 4582–91. https://openaccess.thecvf.com/content_cvpr_2017/html/Xian_Zero-Shot_Learning_-_CVPR_2017_paper.html.

[42] Zhai, Xiaoming. 2022. 'ChatGPT User Experience: Implications for Education'. SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.4312418.