If the Sources Could Talk: Evaluating Large Language Models for Research Assistance in History

Giselle Gonzalez Garcia[1,*,†], Christian Weilbach[2,†]

[1]Department of History and School of Irish Studies, Concordia University, Montreal, Canada
[2]Department of Computer Science, University of British Columbia, Vancouver, Canada

Abstract
The recent advent of powerful Large-Language Models (LLM) provides a new conversational form of inquiry into historical memory (or, training data, in this case). We show that by augmenting such LLMs with vector embeddings from highly specialized academic sources, a conversational methodology can be made accessible to historians and other researchers in the Humanities. Concretely, we evaluate and demonstrate how LLMs have the ability of assisting researchers while they examine a customized corpora of different types of documents, including, but not exclusive to: (1). primary sources, (2). secondary sources written by experts, and (3). the combination of these two. Compared to established search interfaces for digital catalogues, such as metadata and full-text search, we evaluate the richer conversational style of LLMs on the performance of two main types of tasks: (1). question-answering, and (2). extraction and organization of data. We demonstrate that LLMs semantic retrieval and reasoning abilities on problem-specific tasks can be applied to large textual archives that have not been part of the its training data. Therefore, LLMs can be augmented with sources relevant to specific research projects, and can be queried privately by researchers.

Keywords
Artificial Intelligence (AI), Machine Learning, Large Language Models (LLMs), GPT, Historical Research Methods, Historical Writing

1. Introduction & Related Works

Researchers in History and the Humanities tend to accumulate thousands of papers, books, and other sources to be read, i.e. to be processed, in the near future [5]. However, only a handful of these end up in the bibliographical section of our papers, dissertations, and monographs. Because, we have to read them, i.e. process the information one page at the time. Yet, we still personally archive all these sources in our physical or digital libraries. Departing from this premise, we propose: what if these collections of academic texts can be incorporated into a corpus, infused into a Large Language Model (LLM), such as ChatGPT, and interrogated in new ways to produce better informed research outputs? This informed dialogue between researcher and machine would not necessarily

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France
giselle.gonzalezgarcia@mail.concordia.ca (G. Gonzalez Garcia); weilbach@cs.ubc.ca (C. Weilbach)
https://whilo.github.io/ (C. Weilbach)
0009-0008-2707-5420 (G. Gonzalez Garcia); 0000-0001-7116-9338 (C. Weilbach)

be an output in itself, but it would significantly accelerate the research process in History and the Humanities.

LLMs can be defined as a computer system that, given a word or text (token), can predict the words that would come after. Although these systems are not infallible, many researchers conceptualize the advent of LLMs as a proposition of full automation, especially in creative fields. In referring to full automation, we allude to the increasingly popular believe that LLMs entail complete replacement of humans in the production of knowledge and historical narratives. But considering the history of technology since the Industrial Revolution, machines have automated production processes that were often repetitive and exhausting for human labour. Automation freed human resources, it has provided humans with agency, and given us the possibility of dedicating ourselves to higher cognitive tasks [7].

In the field of History, little attention has been devoted to understanding how using LLMs as part of our daily research praxis can have a deep impact and be a methodological game changer. LLMs have the potential to modify the way traditional historical archives are perused, the way primary sources are read and processed, the way theories and narratives are probed and validated, the way complex social processes are summarised, and finally, the way histories themselves are written. Our paper stems from the desire of incorporating LLMs into the authors' professional praxis in History and the Humanities. In doing so, we do not expect the LLMs to "tell the truth" for us, or to produce their own historical narrative in the form of synthetic new texts. Our main goal is to demonstrate what areas of the historical research process can be significantly enhanced and automated by having access to high-quality and accurate assistance. Even though, our professional experience is within the fields of History and Machine Learning, many of the case studies here exemplified can also appeal to other disciplines across the Humanities, and can illuminate the path forward in the field of Artificial Intelligence.

When addressing the emergence of LLMs, several areas of concern are mentioned: (1). its potential abuse by students to plagiarize class assignments, (2). its incapacity to distinguish truth from falsehood [10], (3). its environmental impact [2], and (4) its ethical biases and risks [11]. In the words of Wulf Kansteiner, "Large language models such as GPT-3 are able to generate compelling, non-plagiarized texts in response to simple natural language inputs, thus providing students with an opportunity to produce high-quality written assignments with minimum effort." [10]. One of the main limitations of LLMs (specifically of GPT-3), that Kansteiner identified, is its structurally inability "to attribute the statements it generates to a specific textual origin, let alone assess the factual reliability of any of its textual inputs or outputs." [10]. This is a problem to which our paper offers a tentative solution by arguing that combining an LLM with carefully selected corpora made of primary and secondary sources reduces the margins for errors, and produces texts that have an origin in peer-reviewed sources. A similar approach has been already implemented by Manjavacas Arevalo and Fonteyn [12] to positive results in the form of the MacBERTh model. But, unlike them, we do not train a new model, but augment pre-trained ones.

Moreover, the answers we intend on retrieving from the LLMs include specific details about its sources in the form of text chunks. Therefore, instead of getting answers to our

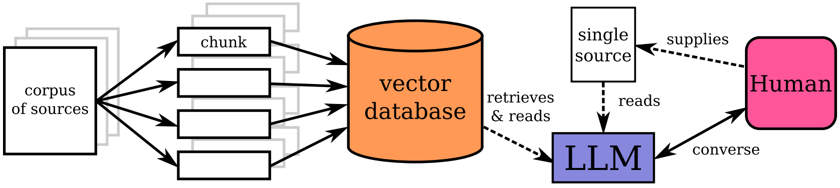


Figure 1: Interaction between a Human and an LLM in form of a conversation. The human optionally supplies a single source text document for the LLM to read first. The LLM can also retrieve text chunks from a vector database filled from a corpus of source documents. Dashed lines denote optional dependencies.

questions from models pre-trained on data only acquired from the world wide web, we propose to prompt the models with a context of chunks that have already been through the established process of academic validation, which can also be checked to validate the response. This approach also shows promise to train better LLMs in general [8].

Bender, Gebru, McMillan-Major, and Shmitchell [2], present a critical overview of the most problematic aspects of LLMs, including their significant contributions to climate change, and their tendency to reproduce and validate hegemonic ideas that are over-represented on the internet, and therefore on the training data that infuses these models. At the same time LLMs show promise to scale deliberation and integrate diverse view points in democratic processes [16]. We do not intend to add to this debate at this early stage in our research, but we acknowledge the current academic debate that problematizes the multiple societal paradoxes that the advent of LLMs (ChatGPT, in particular) have brought to the forefront.

To evaluate such an scenario of historical research assistance empirically, we implemented an open-source pipeline (Figure 1) with different freely available LLMs that can be run locally (and privately) by any researcher (Section 3). We call the implementation KleioGPT. We then compiled a comprehensive academic corpus of digitized history monographs from the first author's academic research in Irish Migration Studies, and conducted a case study from the perspective of a professional historian in Section 4. Our case studies consist of two important textual research tasks, (1). question-answering in Section 4.1 and (2). data extraction in Section 4.2. We assembled 40 prompts for the question-answering task and graded the answers given by different LLMs. We took into account whether the LLMs really responded to our prompt correctly, and paid attention to which sources from our corpus were pulled for answering correctly. For testing data extraction we resorted to the 9 volume genealogical collection Historia de Familias Cubanas by Santa Cruz y Mallen [15].

Since the texts from the different corpora are fed to the LLMs in their general textual representation, any other form of document retrieval from an archive, such as full text, web or metadata search would equally work. We have not explored this in this work.

Table 1
The different LLMs used in our experiments. Settings annotated with ? are not documented publicly. GPT3 was originally described as having 175 billion parameters [4]. The ChatGPT and GPT3 results were created in July 2023.

model	private	#parameters	context size	access	Hugging Face id
Falcon [1]	✓	7b	4k	langchain	`falcon-7b-instruct`
XGen [6]	✓	7b	8k	langchain	`xgen-7b-8k-inst`
Beluga [14, 17]	✓	7b	4k	langchain	`StableBeluga-7b`
GPT3 [4]	✗	175b?	4k	langchain	`gpt-3.5-turbo`
ChatGPT [4]	✗	175b?	4k?	web	-

2. Background

Our pipeline shown in Figure 1 consists of a two stage process. In the first stage, the textual content of all documents in the archive is extracted and sliced into equally-sized overlapping chunks. These chunks are then individually stored into a vector database from which they can be efficiently retrieved later. To answer a question, first a number of relevant chunks are queried from the vector database, and then optionally fed to the LLM together with the question as its prompt. This allows the LLM to answer archive-specific questions that exceed the information of its training data.

2.1. Vector databases

Vector databases index so called "embeddings". These embeddings summarize the content of a text chunk from a document in a fixed length vector of real numbers. To create the database, first each chunk is paired with its respective embedding. The chunk is then added under its embedding to the index of the database. Once embedded, these vectors have the property of locating semantically similar documents near each other. While executing a query, the database refers to its index to find those documents located in proximity to the embedding that matches the text of the query. It then retrieves the text chunk of each nearby embedding from the archive.

2.2. Large language models

Large language models are a form of generative model. Generative models are probabilistic models that can be fit to a data set and generate synthetic data that mimics it. In our work, we depart from the successful class of generative pre-trained transformer (GPT) models [4, 18]. GPT models are pre-trained on large public and private archives to predict the next word in a given context. This type of pre-training has proven to be effective in yielding state-of-the-art language models that perform well when given tasks of interest for researchers in the Humanities [9, 12]. These include question-answering and summarization, a comprehensive recent survey is provided in Yang, Jin, Tang, Han, Feng, Jiang, Yin, and Hu [19].

While pre-training such a model on billions of words is very costly, both in terms of time and compute, running these pre-trained models is possible on commodity hardware. For our work, we selected competitive (at the time of their release) publicly downloadable models, as well as the most popular models, such as GPT3 and ChatGPT; see Table 1. For the open models we chose small versions that can be run on consumer graphics cards (Appendix A). The primary interface to LLMs is the so called prompt, which is the text provided to the model to describe the context and task at hand. The model then appends its answer to the prompt as a form of auto-completion by next word prediction. We document our prompts for each experiment in Appendix B. Generation in LLMs is generally stochastic and the amount of stochasticity can be modulated by a so called temperature. A common problem in LLMs is so called hallucination (see Appendix C), which refers to the fact that they can make up answers that have no factual basis.

3. Methodology

We integrated available LLM open-source software to execute the task described in Section 3.1 together with an interactive question-answering mode in a Python code base for KleioGPT.[1] For Section 4.2 we directly prompt ChatGPT as described there.

3.1. Retrieval augmented question-answering

Our methodology is based on the established combination of memory retrieval from vector databases with LLMs [3]. Augmenting sources does not only help the LLM to answer the question, it also provides a way to check whether the answer can be backed up by the documents in the vector database, a step we think is necessary in an academic setting. Our implementation uses a pre-trained transformer embedding model,[2] the vector database Chroma,[3] and, to integrate the different LLMs, we use LangChain.[4] For a fair and reproducible comparison we set the temperature of all models (except for ChatGPT where we have no control) close to zero (10^{-5}), rendering generation effectively deterministic. A zero temperature corresponds to picking the most likely answer in each step and encourages factual correctness over creativity. We chose the default settings of each model in LangChain for all other parameters. We improved the retrieval mechanism of LangChain to filter out chunks of text that are from bibliographic sections of the sources and chunks that contain less than 200 characters. During question-answering, we loaded a set of questions from a prepared comma separated value (CSV) file. In return, we obtained another CSV file as a report with the initial questions, the answers, and the retrieved sources. Examples are shown in Section 4.1. This simple process both lends itself to automated batch-processing of questions and is accessible to non-technicians through its tabularized inputs and outputs.

[1]https://github.com/GissyGonzalez/KleioGPT
[2]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2
[3]https://docs.trychroma.com/
[4]https://langchain.com/

3.2. Datasets

3.2.1. Migration Studies Corpus

To examine the question-answering capacity of each LLM, we assembled a corpus that forms our Migration Studies dataset. Our aim is to show that researchers can assemble their own corpora tailored to their specific projects. This bibliographic corpus is made of 86 books from leading scholars in Irish Migration Studies, Cuban Studies, general Migration Theory, histories of Slavery, etc. They have been selected from the main fields of study of the first author, and are the basis for their doctoral research. This corpus is representative of the best works in these fields, and among its authors are the leading academic experts in each subject.

In order to verify (and grade) the answers provided by each LLM, we selected volumes the first author was familiar with, had read recently, and had a knowledge of each writer's placement within the historiography and the broader academic conversation in themes that include migration, race, gender, and slavery. These are mostly historical monographs that have been through a peer-review publication process. This does not exempt them from reproducing the societal biases of their authors, but we believe no text is exempt of reproducing the ideologies and cultural systems researchers subscribe to. Nonetheless, we prioritized feeding KleioGPT sources that have gone through a well-established academic validation process, and can be generally factually trustworthy. We provide a list of these sources in Appendix D.

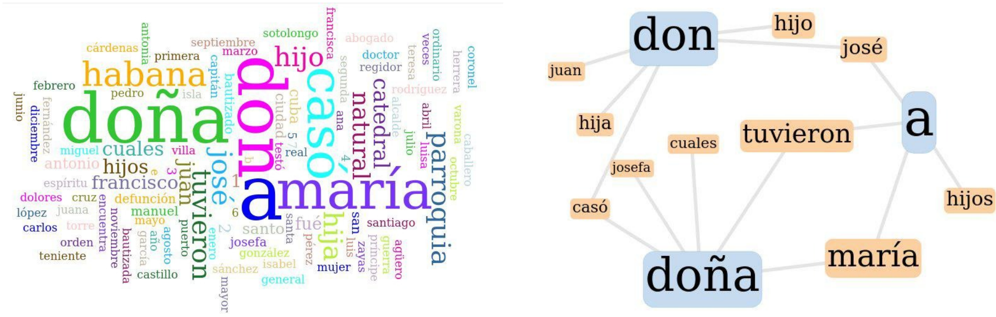
As mentioned above, researchers can gather their own corpus and ingest them into their own versions of KleioGPT. Every text loaded into our experiments was in PDF format. We made sure that each PDF had a plain text layer that ensured its content is machine readable. Users should make sure that their OCR'ed texts are as clean as possible for better results. Our Migration Studies corpus is exclusively in the English language, although this methodology can be applied to corpora in other languages if the embeddings and LLMs support it.

3.2.2. Family History Books Compilation

In the past five years a considerable part of the first author's doctoral research on Irish migrants in the Hispanic Caribbean has consisted of reviewing genealogical compilations in search for biographical data related to 19th-century migrants. In many occasions, genealogical data is the only type of information retrievable to rescue the histories of this population.

Genealogical sources are usually structured in a regular pattern that outlines birth, marriage, death facts, and kinship connections among people. These are worded similarly. Automating the extraction of this type of data from family history book compilations is a task that many historians and researchers in the Humanities undertake frequently. It is labour intensive, repetitive, and consumes significant amounts of time that could be devoted to other research endeavours. Therefore, the possibility of automation by using LLMs is a very attractive one.

Our genealogical dataset is made of the 9 volumes of Historia de Familias Cubanas



(a) Word cloud of the 9 volumes of Historia de Familias Cubanas by Francisco Javier de Santa Cruz [15] showing the word frequency and the regularity of names and expressions. Made by the authors with VoyantTools.

(b) Links visualization of the 9 volumes of Historia de Familias Cubanas by Francisco Javier de Santa Cruz showing the regular structure of genealogical language corpus. Made by the authors with VoyantTools.

Santa Cruz y Mallen [15] (including OCR). This corpus consists of 1,803,596 total words and 39,114 unique word forms. The data set is visualized in Section 3.2.2.

4. Evaluation

4.1. Academic question-answering

To interrogate our corpus from Section 4.1 we designed a fixed set of questions and followed the methodology outlined in Section 3.1. Our goal was to examine and compare the quality of different LLMs responses to four different types of queries: factual, argumentative, descriptive, and integrative. Their level of complexity corresponded to the types of questions undergraduate student in the 200 and 300 level courses would typically face. We graded the answers in a pass/failed binary, giving one point for correct answers, and zero for the opposite. We also knew beforehand the texts contained accurate answers to these questions. All the questions can be found in Appendix B.

We summarize our results in Table 2. All LLMs could answer the majority of questions in all configurations. The retrieved text chunks from the vector database in general improve the results and improve important context to validate the answer, but the LLMs are interestingly already able to answer many questions even without any sources. ChatGPT performed best even without any sources and remarkably we found that there is a big gap to GPT3 accessed through the OpenAI API, which is underlying ChatGPT. There are probably significant extensions applied to the web interface of ChatGPT. Both XGen and Beluga consistently outperform GPT3 and both are competitive with ChatGPT.

Acting under 0 chunks, i.e. without having access to our corpus, the LLMs tended to fail when giving answers to specific questions like: "What was the population of Ireland in 1841?". They were also not able to summarize the contents and main ideas

Table 2
Results of the question-answering task. The percentages are averaged over the 10 questions in each category. Percentages in bold are the best performing in each category.

model	#chunks	factual	argumentative	descriptive	integrative	total
Falcon	0	60%	60%	40%	70%	57.5%
	4	30%	60%	40%	70%	50%
	8	20%	70%	80%	70%	60%
XGen	0	80%	70%	70%	80%	75%
	4	90%	**100%**	80%	80%	87.5%
	8	80%	90%	90%	60%	80%
Beluga	0	80%	80%	90%	60%	77.5%
	4	80%	50%	90%	90%	77.5%
	8	90%	70%	**100%**	**90%**	87.5%
GPT3	0	70%	70%	**100%**	60%	75%
	4	**100%**	60%	70%	70%	75%
	8	90%	60%	**100%**	80%	82.5%
ChatGPT	0	**100%**	80%	**100%**	**90%**	**92.5%**

of books like Kerby A. Miller's Emigrants and Exiles [13]. Without the sources in our dataset, LLMs like Falcon were still able to define conceptual ideas and answer correctly prompts like: "Can you define what is migration/Irish nationalism?". Overall, tasks that required the LLMs to generate its own ideas or argumentation (like those in the integrative section of our Q&A) tended to perform worse, or even prompted the LLMs to hallucinate (see Appendix C) completely off-the-mark answers, when it did not have the backup of the corpus.

Factual The ten questions dedicated to retrieval of factual information focused on obtaining precise data such as dates, quantities, locations, etc. Fact-checking queries are very typical during the historical research process. When these types of question are posed to search engines like: Google, Bing, Startpage, etc., these point us to a list of often millions of indexed results. We have to open web pages to retrieve the information we are looking for and evaluate it. As a system to effectively retrieve factual data, LLMs give us a direct answer, therefore shortening significantly the time invested on this task. These types of questions were easy to grade because they were either right or wrong. When some answers were an approximate, results were graded according to the standard academic consensus.

> Question 5: "Where in Ireland are Protestants found?"
>
> XGen: "Protestants were found in all parts of Ireland, but they were particularly concentrated in urban areas and in the southern province of Ireland. In the early 1800s, they were a majority in County Londonderry."
>
> Falcon: "Protestants are found in all parts of Ireland, but are more commonly found in urban areas."

GPT3: "Protestants are found throughout Ireland, but they tend to be concentrated in towns and urban areas."

ChatGPT: "Protestants in Ireland can be found throughout the country, but their distribution varies across different regions. Historically, Protestants have had a stronger presence in Northern Ireland, particularly in the counties of Antrim, Down, Armagh, and Londonderry (also known as Derry). In these areas, Protestant communities, including various denominations such as Anglicans, Presbyterians, and Methodists, have traditionally been more prominent. However, it's important to note that there are Protestant communities and individuals living in all parts of Ireland, including the Republic of Ireland."

All models answer Question 5 correctly, but ChatGPT provides a lot more relevant details. It is unclear why it behaves so differently, but it could be that it is provided an additional prompt that is not publicly accessible.

Argumentative We also designed ten questions for this type of query. In this category, we asked for summaries of concepts, ideas, and entire books. It is important to clarify that, once our corpus was added, LLMs summarize books based on other scholars opinion's of those books retrieved from the chunks, and not on their own complete reading of each book.[5] Essentially, it will summarize what other authors said about a specific book, and not the book itself. This can be valuable to realize what the academic consensus is about a specific work, idea, or theory; and to identify contentious arguments about them. Most questions in this category that the LLMs failed, were because they hallucinated or fabricated events that were easily identifiable as factually wrong.

Question 3: "Can you summarize the three main arguments in Kerby Millers book Emigrants and Exiles?"

XGen: "1. Irish American nationalism is the primary expression of a powerful sense of exile among Irish immigrants to America. 2. Irish immigrants to America were deeply homesick and resistant to the culture of individualism. 3. Irish chain migration to North America was not passive, but rather relied on the recommendations of previous migrants."

GPT3: "I don't know the answer."

ChatGPT: "Kerby Miller's book "Emigrants and Exiles: Ireland and the Irish Exodus to North America" explores the causes, experiences, and consequences of Irish emigration to North America during the 18th and 19th centuries. While it is challenging to fully summarize all the arguments in the book, here are three key themes..."

[5]The recent release of Claude 2 (https://claude.ai) with a large context window of 100.000 tokens provides such functionality, unfortunately it was not yet programmatically available at the time of submission to the authors.

In the example above, the authors were expecting a succinct answer similar to the one provided by XGen. GPT3's reply was unexpected given that Miller's book is a classic volume in the field. Emigrants and Exiles (1985) has been discussed in length on the web, and according to Google Scholar, is cited 1663 times. Discussions about this book should be on GPT3 training data. Its response contrasts starkly with ChatGPT's lengthy and detailed one.

Descriptive Raising the level of complexity, in this category, we included questions that necessitated more elaborated answers about historical events and processes.

> Question 4: "Can you describe the events that led to the Great Irish potato Famine?"
>
> ChatGPT: "The Great Irish Potato Famine, also known as the Great Hunger or the Irish Potato Famine, was a devastating period in Irish history that occurred between 1845 and 1852. It was a catastrophic event primarily caused by a combination of potato crop failure, economic policies, and socio-political factors. Here is a description of the events that led to the famine: Potato Dependence: ..."

All models answered this question correctly. This can be attributed to how pervasive in society the topic at hand is and how much it has been discussed online. XGen again gave a correct concise answer, however we were expecting a lengthier description given the complexity and gravity of the Great Famine. ChatGPT not only gave a very detailed answer, but it also outlined the causal chain of events, step by step, leading to the famine and its extent (the full answer is in Appendix B.2). Unlike ChatGPT, the other LLMs provided answers that were below the standard of what would have been expected of an early undergraduate Irish Studies student.

Integrative To elaborate this type of questions, we took inspiration in the type of exercises graduate students in North America are tested on during their comprehensive examinations. In answering these questions, the LLMs needed to demonstrate more creativity than in the previous ones. This section was supposed to entail the highest level of difficulty to be tested during our experiments. Here, the LLMs struggled with replying in accordance to the level expected of a human graduate student in History.

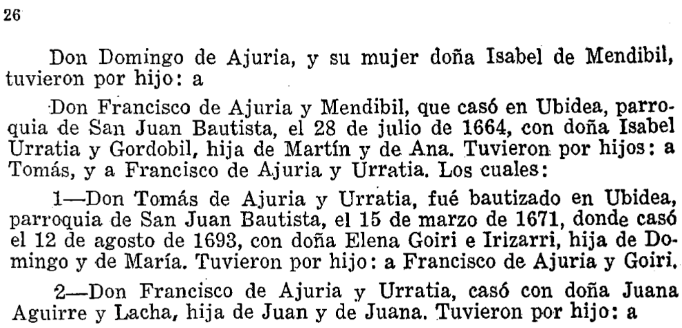
> Question 3: "What is the future of History as a discipline?"
>
> XGen: "The future of history as a discipline is uncertain and likely to be shaped by a variety of factors, including changes in technology, the rise of new historical methodologies, and shifts in the broader cultural and political landscape. Some experts predict that history will continue to evolve and become more interdisciplinary, incorporating insights from other fields such as anthropology, sociology, and economics. Others suggest that the rise of digital humanities and the increasing availability of large datasets may lead to

new approaches to historical research and analysis. Additionally, the future of history may be influenced by broader trends such as the increasing importance of global and transnational history, as well as the growing interest in the history of science and technology. Ultimately, the future of history will depend on the choices and priorities of historians and the broader academic community."

GPT3: "Based on the given context, it is not possible to accurately determine the future of history as a discipline. The context provides information about the challenges and changes that the discipline of history has faced, such as the questioning of the nature of history, the interaction with humanities, arts, and sciences, and the interdisciplinary nature of new scholarship. However, there is no explicit information or discussion about the future trajectory of the discipline."

In the example provided, we observe the contrast between two different approaches: one in which GPT3 determined it had not been provided enough material from the sources and is therefore uncertain about how to answer the question; and another in which XGen based on the same materials, tried to give a well-rounded reply.

4.2. Extracting Tabular Data from Genealogical Sources


26

Don Domingo de Ajuria, y su mujer doña Isabel de Mendibil, tuvieron por hijo: a

Don Francisco de Ajuria y Mendibil, que casó en Ubidea, parroquia de San Juan Bautista, el 28 de julio de 1664, con doña Isabel Urratia y Gordobil, hija de Martín y de Ana. Tuvieron por hijos: a Tomás, y a Francisco de Ajuria y Urratia. Los cuales:

1—Don Tomás de Ajuria y Urratia, fué bautizado en Ubidea, parroquia de San Juan Bautista, el 15 de marzo de 1671, donde casó el 12 de agosto de 1693, con doña Elena Goiri e Irizarri, hija de Domingo y de María. Tuvieron por hijo: a Francisco de Ajuria y Goiri.

2—Don Francisco de Ajuria y Urratia, casó con doña Juana Aguirre y Lacha, hija de Juan y de Juana. Tuvieron por hijo: a

Figure 3: An screenshot of a page from *Historia de Familias Cubanas* describing the descendants of Domingo de Ajuria (Section 3.2.2).

In this section, we took several pages from Historia de Familias Cubanas (Section 3.2.2) and manually provided their full contents to the web interface of ChatGPT (the original text can be seen in Figure 3). The content of these 9 volumes is entirely in Spanish. Even though only 7% of GPT3 training data is in a language other than English [4], ChatGPT performed well in Spanish, translating accurately the excerpts provided. Afterwards, we designed the following prompt:

Prompt example: "From the previous text, list all the names of people in a table with columns: full name of each person, relationship, date of birth

Table 3

The above table (broken in three lines) is extracted from Figure 3 automatically by ChatGPT. The date highlighted in red was wrongly extracted from baptism date. The text fragments in **bold** were in the text, but missing in the table and manually added.

Full Name	Date of Birth	Place of Birth	Baptism Date	Marriage Date
Domingo de Ajuría	Unknown	Ubidea	Unknown	Unknown
Isabel de Mendibíl	Unknown	Unknown	Unknown	Unknown
Francisco de Ajuria **y Mendibil**	Unknown	Ubidea	Unknown	**28-07-1664**
Isabel Urratia **y Gordobil**	Unknown	Unknown	Unknown	**28-07-1664**
Tomás de Ajuria **y Urratia**	15-03-1671	Ubidea	**15-03-1671**	12-08-1693
Elena Goiri e Irizarri	Unknown	Unknown	Unknown	12-08-1693
Francisco de Ajuria **y Goiri**	Unknown	Unknown	Unknown	Unknown

Father's Full Name	Mother's Full Name	Children's Full Name
Unknown	Unknown	**Francisco de Ajuria y Mendibil**
Unknown	Unknown	Unknown
Domingo de Ajuría	Isabel de Mendibíl	Tomás de Ajuria; **Francisco de Ajuria**
Martin	**Ana**	Tomás de Ajuria; **Francisco de Ajuria**
Francisco de Ajuria	Isabel Urratia	Francisco de Ajuria **y Goiri**
Domingo	**Maria**	Francisco de Ajuria **y Goiri**
Tomás de Ajuria	Elena Goiri e Irizarri	Unknown

Spouse's Full Name	Gender
Isabel de Mendibil	Male
Domingo de Ajuría	Female
Isabel Urratia y Gordobil	Male
Francisco de Ajuria	Female
Elena Goiri e Irizarri	Male
Tomás de Ajuria **y Urratia**	Female
Unknown	Male

(format DD-MM-YYYY), place of birth, date of death (format DD-MM-YYYY), baptism date (format DD-MM-YYYY), marriage date (format DD-MM-YYYY), place of residence, full name and surname of father, full name and surname of mother, full name and surname of children, full name and surname of spouse, and occupation. Try to infer the gender of each person, and add a column Gender."

In Table 3 we show that ChatGPT is able to extract data from Historia de Familias Cubanas (9 volumes) in Santa Cruz y Mallen [15] into a structured tabular format by simply prompting it to create a table with the respective column names.

A large part of the information is correctly mapped into the table. However, ChatGPT did miss some of the information highlighted in bold. The missing fragments show regularities such as family names that are connected by the conjunction “y”. We believe that it is likely that by instructing ChatGPT explicitly on these expressions, e.g. by giving example translations or erroneous lines, the LLM would miss less data. We plan to address this in future work together with an automatic page wise extraction of the tables with our prompt.

As can be seen in Figure 3, the gender information was not in the original text. We asked ChatGPT to infer it automatically based on its pre-trained knowledge about Spanish name conventions.

Out of all the dates extracted, there is only one birth date (in red) inferred wrongly, everything else is correct. If this low error rate reflects the performance on the full dataset, then a large amount of data could be extracted into a tabular representation just by asking ChatGPT to do so. We could not get the same extraction quality out of the other LLMs, but expect them to catch up in the near future.

5. Conclusion

In this paper we have demonstrated that LLMs can provide a fluent conversational research assistance while being sufficiently accurate for an academic environment. While open source models that fit onto commodity hardware are not yet competitive with ChatGPT in all tasks, they already provide a private alternative on question-answering tasks on sensitive data. While this work already highlights the potential of LLMs for historical research, problems such as hallucination and biases in the training data will also probably be significantly reduced in the near future. Since the field of LLMs is evolving very quickly, we think that very soon much better open source LLMs will be available. We will continue to design tests to evaluate the LLMs abilities in the near future.

Interesting avenues for future work are better integration of memory to both help the generation of the LLMs and make it easier to validate them against the sources retrieved from memory. Models such as Claude 2 make use of larger context windows for better extraction and summarization abilities and show a lot of promise in our ongoing evaluations as well. We think that open source solutions that facilitate research will be widely available to researchers in the Humanities including models that can be run on machines without GPUs.[6]

References

[1] E. Almazrouei, H. Alobeidli, A. Alshamsi, A. Cappelli, R. Cojocaru, M. Debbah, E. Goffinet, D. Heslow, J. Launay, Q. Malartic, B. Noune, B. Pannier, and G. Penedo. "Falcon-40B: an open large language model with state-of-the-art performance". In: (2023). URL: https://falconllm.tii.ae/.

[2] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event Canada: Acm, 2021, pp. 610–623. DOI: 10.1145/3442188.3445922.

[6] An older example for this is https://gpt4all.io/index.html

[3] S. Borgeaud, A. Mensch, J. Hoffmann, T. Cai, E. Rutherford, K. Millican, G. van den Driessche, J.-B. Lespiau, B. Damoc, A. Clark, D. d. L. Casas, A. Guy, J. Menick, R. Ring, T. Hennigan, S. Huang, L. Maggiore, C. Jones, A. Cassirer, A. Brock, M. Paganini, G. Irving, O. Vinyals, S. Osindero, K. Simonyan, J. W. Rae, E. Elsen, and L. Sifre. Improving Language Models by Retrieving from Trillions of Tokens. 2022. DOI: 10.48550/arXiv.2112.04426. arXiv: 2112.04426 [cs].

[4] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. “Language Models Are Few-Shot Learners”. In: arXiv:2005.14165 [cs] (2020). arXiv: 2005.14165 [cs].

[5] G. Crane. “"What Do You Do with a Million Books?"”. In: D-Lib Magazine 12.3 (2006). URL: https://www.dlib.org/dlib/march06/crane/03crane.html.

[6] N. Erik, T. Xie, H. Hayashi, B. Pang, C. Xia, C. Xing, S. Vig Jesse ang Yavuz, P. Laban, B. Krause, S. Purushwalkam, T. Niu, W. Kryscinski, L. Murakhovs’ka, P. K. Choubey, A. Fabbri, Y. Liu, R. Meng, L. Tu, M. Bhat, C.-S. Wu, S. Savarese, Y. Zhou, S. R. Joty, and C. Xiong. Long Sequence Modeling with XGen: A 7B LLM Trained on 8K Input Sequence Length. Salesforce AI Research Blog. 2023. URL: https://blog.salesforceairesearch.com/xgen.

[7] C. B. Frey. The Technology Trap Capital, Labor, and Power in the Age of Automation. First paperback printing. Book Collections on Project MUSE. Oxford: Princeton University Press, 2020.

[8] S. Gunasekar, Y. Zhang, J. Aneja, C. C. T. Mendes, A. Del Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi, A. Salim, S. Shah, H. S. Behl, X. Wang, S. Bubeck, R. Eldan, A. T. Kalai, Y. T. Lee, and Y. Li. Textbooks Are All You Need. 2023. DOI: 10.48550/arXiv.2306.11644. arXiv: 2306.11644 [cs].

[9] K. Hosseini, K. Beelen, G. Colavizza, and M. C. Ardanuy. “Neural Language Models for Nineteenth-Century English”. In: 7.0 (2021), p. 22. DOI: 10.5334/johd.48w.

[10] W. Kansteiner. “Digital Doping for Historians: Can History, Memory, And Historical Theory Be Rendered Artificially Intelligent?” In: History and Theory 61.4 (2022), pp. 119–133. DOI: 10.1111/hith.12282.

[11] L. Klein. “Are Large Language Models Our Limit Case?” In: Startwords 3 (2022). DOI: 10.5281/zenodo.6567985.

[12] E. Manjavacas Arevalo and L. Fonteyn. “MacBERTh: Development and Evaluation of a Historically Pre-trained Language Model for English (1450-1950)”. In: Proceedings of the Workshop on Natural Language Processing for Digital Humanities. NIT Silchar, India: NLP Association of India (NLPAI), 2021, pp. 23–36.

[13] K. A. Miller. Emigrants and Exiles: Ireland and the Irish Exodus to North America. New York; Oxford: Oxford University Press, 1985.

[14] S. Mukherjee, A. Mitra, G. Jawahar, S. Agarwal, H. Palangi, and A. Awadallah. Orca: Progressive Learning from Complex Explanation Traces of GPT-4. 2023. DOI: 10.48550/arXiv.2306.02707. arXiv: 2306.02707 [cs].

[15] F. X. de Santa Cruz y Mallen. Historia De Familias Cubanas (Tomos I-IX). 1940.

[16] C. T. Small, I. Vendrov, E. Durmus, H. Homaei, E. Barry, J. Cornebise, T. Suzman, D. Ganguli, and C. Megill. Opportunities and Risks of LLMs for Scalable Deliberation with Polis. 2023. DOI: 10.48550/arXiv.2306.11932. arXiv: 2306.11932 [cs].

[17] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom. Llama 2: Open Foundation and Fine-Tuned Chat Models. 2023. DOI: 10.48550/arXiv.2307.09288. arXiv: 2307.09288 [cs].

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,. Kaiser, and I. Polosukhin. "Attention Is All You Need". In: Advances in Neural Information Processing Systems. Vol. 30. Curran Associates, Inc., 2017.

[19] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, B. Yin, and X. Hu. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. 2023. DOI: 10.48550/arXiv.2304.13712. arXiv: 2304.13712 [cs].

A. Experimental details

We ran experiments for the local LLMs StableBeluga-7b, falcon-7b-instruct and xgen-7b-8k-inst on an A5000 NVIDIA GPU.

B. Academic question-answering

B.1. Questions

B.1.1. Factual

1. What was the population of Ireland in 1841?
2. When did Irish Potato Famine start?
3. When did Irish Potato Famine finish?
4. When did the Irish Parliament past the Act of Union?
5. Where in Ireland are Protestants found?
6. Where in Ireland are native Irish speakers found?
7. What was the population of Ireland in 1861?
8. When did Catholic Emancipation happen?
9. What happened in Ireland in 1848?
10. When was the Land League created?

B.1.2. Argumentative

1. Can you make a definition of migration in ten sentences?
2. What is Irish nationalism?
3. Can you summarize the three main arguments in Kerby Millers book Emigrants and Exiles?
4. Can you summarize the three main arguments in Donald Akensons book The Irish in Ontario?
5. What is the Irish diaspora?
6. What is Irish unionism?
7. What are three differences between Irish nationalism and Irish unionism?
8. What are three similarities between Irish nationalism and Irish unionism?
9. Can you summarize what David Fitzpatricks book Oceans of Consolation is about?
10. What is transnationalism?

B.1.3. Descriptive

1. Can you describe Irelands population loss process throughout the nineteenth century?
2. What was the Young Ireland movement?
3. What was the Land League?

4. Can you describe the events that led to the Great Irish potato Famine?
5. What were the main five consequences for Ireland of the Potato Famine of 1845?
6. What was the Catholic Emancipation movement?
7. Can you describe how was the Irish Act of Union passed in Parliament?
8. Can you describe chain migration?
9. What was the political state of Ireland in 1845?
10. Who was Daniel OConnell?

B.1.4. Integrative

1. What are the two most significant controversies that have developed in Irish diaspora studies in the past twenty-five years? Elaborate.
2. What areas of study should Irish history researchers focus on in the future? Elaborate.
3. What is the future of history as a discipline? Elaborate.
4. Which geographic areas and historical time periods do you think should receive more or less attention based on what are the dominant themes in the current Irish Historiography? Elaborate.
5. What elements do you consider to be crucial in teaching Irish diaspora studies, and what do you think needs to be revised? Elaborate.
6. What is the future of migration studies? Elaborate.
7. Taking from Kerby Millers Emigrants and Exiles book, and for Donald Harman Akensons book The Irish in Ontario, what are the main arguments these historian make, how are these arguments similar and different, and what are contested subjects in their works? Elaborate.
8. How does Irish migration to the United States differ from Irish migration to Canada? Elaborate.
9. Were the Irish in the United States considered white? Elaborate.
10. What were Irish migrants role in African slavery? Elaborate.

B.2. ChatGPT answer to Irish Potatoe Famine

ChatGPT: "The Great Irish Potato Famine, also known as the Great Hunger or the Irish Potato Famine, was a devastating period in Irish history that occurred between 1845 and 1852. It was a catastrophic event primarily caused by a combination of potato crop failure, economic policies, and socio-political factors. Here is a description of the events that led to the famine: Potato Dependence: The Irish population in the mid-19th century heavily relied on the potato as a staple food crop. The potato was affordable, easy to grow, and provided sufficient calories to sustain a large population. However, this heavy dependence on a single crop made the population vulnerable to any failures or diseases affecting the potato harvest. Potato Blight: In 1845, a strain of potato blight, caused by a fungal disease called Phytophthora infestans, spread across Ireland. The blight rapidly destroyed the potato crops, turning them into a black, rotting mass.

This initial outbreak resulted in a significant reduction in the potato harvest, leading to food shortages and increasing fears of famine. British Economic Policies: Ireland was under British rule during this time, and British economic policies exacerbated the impact of the blight. British policies favored exporting Irish agricultural produce, such as grain and livestock, to Britain while leaving the Irish population heavily dependent on the potato as their main food source. The policies prioritized the interests of British landowners and exacerbated existing economic inequalities in Ireland. Limited Relief Efforts: The British government's response to the crisis was initially insufficient. Relief efforts were limited, and policies were implemented with a focus on maintaining economic stability rather than prioritizing the well-being of the Irish population. This lack of effective intervention worsened the situation and increased the suffering of those affected by the famine. Rapid Spread and Devastation: The potato blight continued to devastate subsequent potato crops in the years that followed. The impact was particularly severe in the west and southwest regions of Ireland, where the population was heavily reliant on potato cultivation. The loss of the potato harvest resulted in widespread hunger, malnutrition, and disease. Mass Emigration and Death: The combination of crop failure, lack of relief, and dire living conditions led to mass emigration and significant loss of life. Many Irish people, desperate to escape the famine's effects, left their homes and families to seek opportunities in other countries, particularly the United States and Canada. However, emigration was not an option for everyone, and the lack of adequate food and resources resulted in a devastating death toll, estimated to be around one million people, while millions more emigrated. The Great Irish Potato Famine had a profound and lasting impact on Ireland. It led to depopulation, cultural trauma, and a reevaluation of Irish-British relations. The famine exposed the socio-economic inequalities, poor relief systems, and inadequate governance during that time, and its effects continue to be remembered as a defining period in Irish history."

C. Hallucination Example

Question 9 (Argumentative): "Can you summarize what David Fitzpatricks book Oceans of Consolation is about?"

GPT3: "David Fitzpatrick's book "Oceans of Consolation" is a fundamental study and model for further editions of emigrant letters. It explores the personal letters sent to and from Irish migrants in Australia, providing full transcripts and examining each correspondent's local context. The book aims to understand the settlement of the Irish abroad in a comparative context and emphasizes the importance of faith in the lives of Irish migrant."

Falcon: "Ocean of Consolation is a book about the history of the ocean and its impact on human culture. It explores the various ways the ocean has been perceived throughout history, from being seen as a source of wealth and power to being a place of spiritual significance. The book also examines the current state of the world's oceans and the threats they face due to human activity."

D. Irish Migration Studies Dataset

Akenson, Donald H. 1999. The Irish in Ontario : A Study in Rural History. 2nd ed. Montreal: McGill-Queen's University Press.

Barcia Paz, Manuel. 2008. Seeds of Insurrection : Domination and Resistance on Western Cuban Plantations 1808-1848. Baton Rouge: Louisiana State University Press.

Barclay, Katie. 2019. Men on Trial : Performing Emotion Embodiment and Identity in Ireland 1800-45. Manchester: Manchester University Press.

Bielenberg, Andy. 2013. The Irish Diaspora. Oxon England: Routledge.

Bravo Lozano, Cristina. 2018. Spain and the Irish Mission, 1609-1707. New York: Routledge.

Brehony, Margaret and Nuala Finnegan. 2019. Irlanda y Cuba: Historias Entretejidas = Irlanda & Cuba : Entangled Histories. Cuba: Ediciones Bolona.

Brettell, Caroline. 2007. Constructing Borders. Lanham MD: Lexington Books.

Brettell, Caroline and James Frank Hollifield. 2023. Migration Theory: Talking Across Disciplines. New York NY: Routledge.

Brundage, David. 2016. Irish Nationalists in America: The Politics of Exile, 1798-1998. Oxford: Oxford University Press.

Bock, Gisela. 1989. Women's History and Gender History: Aspects of an International Debate. Gender & History 730.

Bornstein, George. 2011. The Colors of Zion: Blacks, Jews, and Irish from 1845 to 1945. Cambridge: Harvard University Press.

Bueltmann, Tanja; David T Gleeson; and Donald M MacRaild. 2012. Locating the English Diaspora 1500-2010. Liverpool: Liverpool University Press.

Butler, Judith. 1999. Gender Trouble: Feminism and the Subversion of Identity. New York: Routledge.

Butler, Judith. 2004. Undoing Gender. New York: Routledge.

Coldham, Peter Wilson. 1992. Emigrants in Chains : A Social History of Forced Emigration to the Americas of Felons Destitute Children Political and Religious Non-Conformists Vagabonds Beggars and Other Undesirables 1607-1776. Baltimore Md: Genealogical Pub.

Curry-Machado, Jonathan. 2011. Cuban Sugar Industry : Transnational Networks and Engineering Migrants in Mid-Nineteenth Century Cuba. 1st ed. New York: Palgrave Macmillan.

Curry-Machado, Jonathan. 2013. Global Histories Imperial Commodities Local Interactions. Houndmills Basingstoke Hampshire: Palgrave Macmillan.

Desley, Deacon; Penny Russell; and Angela Woollacott. 2010. Transnational Lives : Biographies of Global Modernity 1700-Present. Basingstoke England: Palgrave Macmillan.

Diaz, Maria Elena. 2022. The Virgin the King and the Royal Slaves of El Cobre : Negotiating Freedom in Colonial Cuba 1670-1780. Stanford CA: Stanford University Press.

Din, Gilbert C. 2016. An Extraordinary Atlantic Life: Sebastián Nicolás Calvo De La Puerta y O'Farrill, Marqués De Casa-Calvo. Lafayette LA: University of Louisiana at

Lafayette Press.

Duffy, Jennifer Nugent. 2013. Whos your Paddy? Racial Expectations and the Struggle for Irish American Identity. New York: NYU Press.

Dunn, Richard and Rebekah Higgitt. 2015. Navigational Enterprises in Europe and Its Empires 1730-1850. Houndmills Basingstoke Hampshire: Palgrave Macmillan.

Dunn, Richard S. 2014. A Tale of Two Plantations : Slave Life and Labor in Jamaica and Virginia. Cambridge Massachusetts: Harvard University Press.

Dunn, Richard S and Gary B Nash. 2000. Sugar and Slaves : The Rise of the Planter Class in the English West Indies 1624-1713. Chapel Hill North Carolina: Published for the Omohundro Institute of Early American History and Culture Williamsburg Virginia by the University of North Carolina Press.

Evans, Nicholas and Angela McCarthy. 2022. Death in the Diaspora: British and Irish Gravestones. Edinburgh: Edinburgh University Press.

Fanning, Tim. 2018. Paisanos : The Irish and the Liberation of Latin America. Notre Dame Indiana: University of Notre Dame Press.

Fitzgerald Patrick, and B. K Lambkin. 2008. Migration in Irish History 1607-2007. Basingstoke England: Palgrave Macmillan.

Garcia Rodriguez Gloria. 2011. Voices of the Enslaved in Nineteenth-Century Cuba : A Documentary History. Chapel Hill: University of North Carolina Press.

Garner, Steven. 2007. ATLANTIC CROSSING, Atlantic Studies, 4:1, 117-132, DOI: 10.1080/14788810601179485.

Giemza, Bryan Albin. 2013. Rethinking the Irish in the American South: Beyond Rounders and Reelers. University Press of Mississippi.

Gleeson, David T. 2013. The Green and the Gray : The Irish in the Confederate States of America. Chapel Hill: University of North Carolina Press.

Gleeson, David T. 2001. The Irish in the South 1815-1877. Chapel Hill: University of North Carolina Press.

Gleeson, David T. and Simon Lewis. 2014. The Civil War As Global Conflict : Transnational Meanings of the American Civil War. Columbia South Carolina: University of South Carolina Press.

Gray, Breda. 2004. Women and the Irish Diaspora. London: Routledge.

Gordon, Linda. 2022. The Great Arizona Orphan Abduction. Cambridge MA: Harvard University Press.

Guglielmo, Jennifer. 2010. Living the Revolution : Italian Women's Resistance and Radicalism in New York City 1880-1945. Chapel Hill: University of North Carolina Press.

Hall, Catherine and Sonya O Rose. 2006. At Home with the Empire : Metropolitan Culture and the Imperial World. Cambridge UK: Cambridge University Press.

Hodgson, Kate. 2018. Franco-Irish Saint-Domingue Family Networks Trans-Colonial Diasporas. Caribbean Quarterly 43451.

Horning, Audrey. 2013. Ireland in the Virginian Sea: Colonialism in the British Atlantic. Chapel Hill: University of North Carolina Press.

Houston, Cecil J. and William J. Smyth. 1990. Irish emigration and Canadian settlement: patterns, links, and letters. Toronto: U. of Toronto Press.

Jordan, Don, and Michael Walsh. 2008. White Cargo: The Forgotten History of Britains White Slaves in America. New York: NYU Press.

Josiah, Barbara P. 2011. Migration Mining and the African Diaspora : Guyana in the Nineteenth and Twentieth Centuries (version 1st ed). 1st ed. New York: Palgrave Macmillan.

Kenny, Kevin. 2013. Diaspora : A Very Short Introduction. Oxford: Oxford University Press.

Kenny, Kevin. 1998. Making Sense of the Molly Maguires. New York NY: Oxford University Press.

Knight, Franklin W. 1970. Slave Society in Cuba during the Nineteenth Century. Madison: University of Wisconsin Press.

Kuethe, Allan J. 1986. Cuba 1753-1815: Crown Military and Society. 1st ed. Knoxville: University of Tennessee Press.

Ignatiev, Noel. 2012. How the Irish Became White. Hoboken: Taylor and Francis.

Ireland, John de Courcy. 1986. Ireland and the Irish in Maritime History. Dublin Ireland: Glendale Press.

Lorimer, Joyce. 2018. English and Irish Settlement on the River Amazon 1550-1646. Farnham: Routledge.

Mar, Lisa Rose. 2010. Brokering Belonging : Chinese in Canada's Exclusion Era 1885-1945. New York: Oxford University Press.

Marshall, Oliver. 2000. English-Speaking Communities in Latin America. New York: St. Martin's Press.

McCarthy, Angela. 2005. Irish Migrants in New Zealand 1840-1937 : 'the Desired Haven'. Woodbridge Suffolk U.K: Boydell Press.

McCarthy, Angela and Catharine Coleborne. 2012. Migration Ethnicity and Mental Health : International Perspectives 1840-2010. New York: Routledge.

McCarthy, Angela. 2011. Scottishness and Irishness in New Zealand Since 1840. Manchester: Manchester University Press.

McGaughey, Jane. 2020. Violent Loyalties: Manliness, Migration, and the Irish in the Canadas, 1798-1841. Liverpool University Press.

McMahon, Eileen M. 1995. What parish are you from? A Chicago Irish community and race relations. Lexington: University Press of Kentucky.

McMahon, Timothy G., Michael de Nie, Paul Townsend. 2017. Ireland in an Imperial World. Citizenship, Opportunism, and Subversion. Cambridge University Press.

Middleton, Stephen; David R. Roediger; and Donald M Shaffer. 2017. The Construction of Whiteness : An Interdisciplinary Analysis of Race Formation and the Meaning of a White Identity. Jackson: University Press of Mississippi.

Miller, Kerby A. 1985. Emigrants and Exiles: Ireland and the Irish Exodus to North America. New York: Oxford University Press.

Mohanram, Radhika. 2007. Imperial White : Race Diaspora and the British Empire. Minneapolis: University of Minnesota Press.

Morgan, Gwenda and Peter Rushton. 2013. Banishment in the Early Atlantic World: Convicts Rebels and Slaves. London: Bloomsbury.

Morton, Graeme, and David A. Wilson, eds. 2013. Irish and Scottish Encounters with Indigenous Peoples: Canada, the United States, New Zealand, and Australia. McGill-Queens University Press.

Mosse, George L. 1996. The Image of Man: The Creation of Modern Masculinity. New York & Oxford: Oxford University Press.

Murray, Edmundo.2005. Becoming Irlandés: Private Narratives of the Irish Emigration to Argentina, 1844-1912. Buenos Aires.

Nash, Catherine. 2015. Genetic Geographies: The Trouble with Ancestry. Minneapolis: University of Minnesota Press.

Nelson, Bruce. 2012. Irish Nationalists and the Making of the Irish Race. Princeton: Princeton University Press.

O'Brien, Sarah. 2017. Linguistic Diasporas Narrative and Performance : The Irish in Argentina. Cham Switzerland: Palgrave Macmillan.

O'Callaghan, Sean. 2000. To Hell or Barbados : The Ethnic Cleansing of Ireland. Dingle: Brandon.

O'Connor, Thomas; Mary Ann Lyons. 2003. Irish Migrants in Europe After Kinsale 1602-1820. Dublin: Four Courts Press.

O'Kane, Finola and Ciaran O'Neill. 2023. Ireland Slavery and the Caribbean: Interdisciplinary Perspectives. Manchester: Manchester University Press.

Onkey, Lauren. 2010. Blackness and Transatlantic Irish Identity: Celtic Soul Brothers. New York: Routledge.

O'Neill, Peter D. 2017. Famine Irish and the American Racial State. New York: Routledge.

Prados-Torreiram Teresa. 2005. Mambisas : Rebel Women in Nineteenth-Century Cuba. Gainesville Fla: University Press of Florida.

Putnam Lara. 2013. Radical Moves : Caribbean Migrants and the Politics of Race in the Jazz Age. Chapel Hill: University of North Carolina Press.

Recio Morales, Oscar. 2010. Ireland and the Spanish Empire 1600-1825. Dublin: Four Courts Press.

Roberts, Daniel Sanjiv, Jonathan Jeffrey Wright. 2019. Irelands Imperial Connections, 1775-1947. Palgrave MacMillan.

Rose, Sonya O. 1992. Limited Livelihoods : Gender and Class in Nineteenth-Century England. London: Routhledge.

Rose, Sonya O. 2010. What Is Gender History? Cambridge UK: Polity.

Scott, Joan Wallach. 1996. Feminism and History. Oxford: Oxford University Press.

Scott, Joan Wallach. 2018. Gender and the Politics of History (version 30th anniversary edition) 30Th anniversary ed. New York: Columbia University Press.

Scott, Joan Wallach. 2019. Knowledge Power and Academic Freedom. New York: Columbia University Press.

Scott, Joan Wallach. 2012. The Fantasy of Feminist History. Durham: Duke University Press.

Smith, Raymond T. 1988. Kinship and Class in the West Indies : A Genealogical Study of Jamaica and Guyana. Cambridge: Cambridge University Press.

Stradling, R. A. 1994. The Spanish Monarchy and Irish Mercenaries : The Wild Geese in Spain 1618-68. Blackrock Co. Dublin: Irish Academic Press.

Valente, Joseph. 2011. The Myth of Manliness in Irish National Culture 1880-1922. Urbana Ill: University of Illinois Press.