

AI and Genealogy: Trouble Ahead?

Thomas MacEntee, of Genealogy Bargains

<https://genealogybargains.com>

hidefgen@gmail.com

While AI (artificial intelligence) might be the current “hot” buzz word, the fact is that many genealogy vendors and even genealogists have already been using this technology for years. The AI industry is at a crossroads and within the next five years, it will permeate almost every aspect of business and society. Learn how AI is currently being used to improve the genealogy experience, and whether or not you should seek out other uses of artificial intelligence for your own genealogy research.

What is Artificial Intelligence and Why Should I Be Concerned?

AI stands for Artificial Intelligence and represents computer-based systems that can “mimic” human intelligence. The goal is to have these systems perform human tasks.

A task could be as simple as entering a customer service-related question on a company’s website and having AI generate a response. It might appear simple - a looks up of a response posted to the chat panel. However, AI-based systems might prompt you with more questions in order to generate the most helpful answer. The system could capture your questions and “learn” more about the way you use the product in order to better respond in the future.

Another example: you could ask **ChatGPT** - one of more popular AI platforms - to compose a poem about genealogy. And taking it one step further, you could add “in the style of” Shakespeare, Keats, Robert Frost, Maya Angelou, etc.

The most discussed features of artificial intelligence are “deep learning” and “generative AI.” Deep learning mimics the human brain in that it looks for patterns using vast amounts of information to interpret photos, audio, and text. Generative AI actually “generates” new photos, audio, and text, based on information provided by the user, and again, uses its own database of “training data” to understand patterns and generate output that matches the user’s query.

AI Platforms

While genealogy vendors such as MyHeritage are incorporating artificial intelligence into the features they provide to users, there are some popular AI platforms open to the public that you might want to consider using.

- **ChatGPT:** Meaning “Chat Generative Pre-trained Transformer,” ChatGPT (<https://chat.openai.com/>) is the most popular publicly-accessible artificial intelligence platform.
- **Copilot:** Copilot (<https://copilot.microsoft.com>) is an AI-powered intelligent assistant that helps you get answers and inspirations from across the web, supports creativity and collaboration, and helps you focus on the task at hand.

- **Gemini:** Developed by Google, **Gemini** (<https://gemini.google.com/>) formerly known as Bard, describes itself as “a family of AI models developed by Google's AI research labs DeepMind and Google Research. Gemini is Google's largest and most flexible AI model, able to run on data centers and mobile devices.”
- **Perplexity:** A relative newcomer in the world of AI platforms, Perplexity (<https://www.perplexity.ai/>) presents a curated list of sources when answering user queries.

AI and Genealogy

Current uses of AI by Genealogy Vendors and others

Believe it or not, genealogists have already benefited from artificial intelligence whether it is just spelling and grammar check in Microsoft Word when writing a family story or genealogical report ... to ... searching for family in the 1950 United States Census which was indexed using artificial intelligence’s ability to decipher handwriting.

- **Family photos:** MyHeritage has been offering a variety of photo enhancement tools over the past three years including ways to colorize images and make them clearer. In addition there are tools that can “animate” an ancestor based on a photo and even help determine the date of an image based on characteristics such as fashion styles, hair styles, and more. The new Reimagine tool announced recently offers all these tools in an easy-to-use app.
- **Transcription:** The National Archives and Records Administration (NARA) in conjunction with Ancestry and FamilySearch used artificial intelligence to index the 1950 US Census population schedules released in April 2022. Entries made by enumerators were scanned and transcribed then released for use at a much faster rate than what was accomplished with manual indexing performed for the 1940 US Census release in 2012. For the 1950 US Census, users were encouraged to review the transcriptions and submit corrections as part of a community effort by genealogists and other researchers.
- **Searching for and suggesting records:** Ancestry and other genealogy platforms have been listing “related” or “suggested” records in the sidebar of the webpage when a user is viewing a record as part of a search. In addition “hints” will often pop up suggesting records and family trees that a researcher might want to review due to similarities in data.
- **DNA matches:** With over 30 million people having used personal DNA testing kits, 23andMe, AncestryDNA, FamilyTreeDNA, and MyHeritage all leverage AI to find connections between testers based on shared DNA data. Given the sheer amount of information involved, these match results are only possible with artificial intelligence.

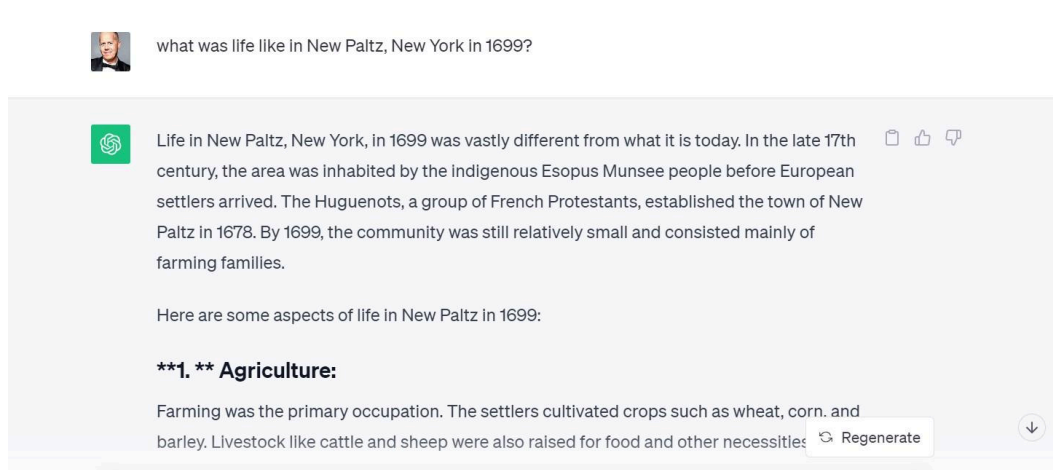
Examples

Here are a few examples of how you can use the major AI platforms when searching for ancestors.

Social history

Hugo Freer, my 9th great-grandfather, settled in New Paltz, New York along with other Huguenot settlers about 1675. His house - the Freer-Low House - built in 1699 is still standing. Wanting to know more about how Freer lived, I used this query at ChatGPT: **What was life like in New Paltz, New York in 1699?**

The results:



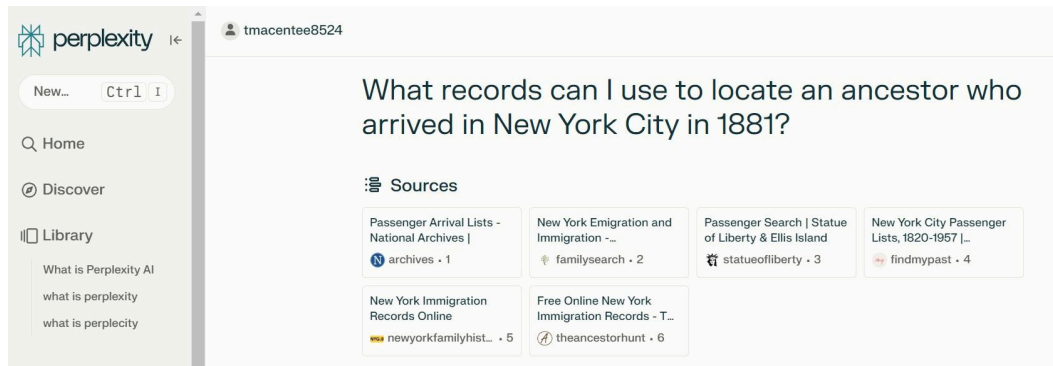
Source: "What was life like in New Paltz, New York in 1699?" prompt. ChatGPT, ChatGPT 3.5 version, OpenAI, 1 October 2023, <https://chat.openai.com/>.

The information provided is extensive but also rather generic for any location on the East Coast of the United States. In addition, the section on housing states the use of wooden buildings with thatched roofs, when in fact New Paltz was known for its early homes built using stones excavated from the fields.

Record sets

My 3rd great-grandfather Gustave Henneberg arrived in New York from Germany about 1881. I posed the following query at Perplexity in order to determine which records I should use for research purposes: **What records can I use to locate an ancestor who arrived in New York City in 1881?**

The results:



Source: "What records can I use to locate an ancestor who arrived in New York City in 1881?" prompt. Perplexity, version 1.0.21, Perplexity, 13 October 2023, <https://www.perplexity.ai/>.

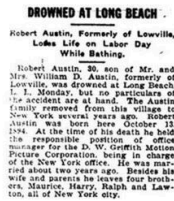
Perplexity bills itself as different from ChatGPT and Gemini in that it curates sources which are presented at the top of the generated content.

Transcription

Robert Austin was the brother of my great-grandfather John Ralph Austin (1896-1976). Robert drowned in 1924 at Long Beach, New York while trying to rescue a child. I uploaded a newspaper article about the incident to Gemini and in the query prompt entered: **Transcribe**

The results:

Gemini **Advanced** ▾



Show drafts ▾ 🔊



Sure, here is the verbatim transcription of the newspaper article in the image:

DROWNED AT LONG BEACH

Robert Austin, Formerly of Lowville,

Source: Uploaded article "Drowned at Long Beach" with the query "Transcribe verbatim". Gemini, Gemini Advanced version, Google, 8 Feb 2024, <https://gemini.google.com/>.

Gemini Advanced did a great job despite the article image having some clarity issues!

Relationships

During my genealogy research, I've located a person to whom I'm related via a maternal great-grandmother. The document I am using for research states that the person was my great-grandmother's niece. Using Gemini Advanced, I want to determine how I am related to this person. **How am I related to my great-grandmother's niece?**

The results:

Gemini Advanced ▾



how am I related to my great-grandmother's niece?

Show drafts ▾



Your great-grandmother's niece is your first cousin, twice removed. Here's the breakdown:

- **Your great-grandmother's niece:** This person shares your great-grandparents as their grandparents.
- **First cousin:** This indicates you are in the same generation, descended from the same great-grandparents.
- **Twice removed:** This means there are two generations of difference between you. You are descended from your great-grandparents, while your great-grandmother's niece is descended from their sibling.

Source: "How am I related to my great-grandmother's niece?". Gemini, Gemini Advanced version, Google, 8 February 2024, <https://gemini.google.com/>.

Copyright and AI

There are several issues involving artificial intelligence and intellectual property that should concern you. Some of these issues have already been discussed and decided by agencies and courts. Many of them, however, have not been resolved.

The two main issues are:

- **AI-generated content:** Can content that is created by artificial intelligence based on your query be copyrighted? What if you ask Gemini to generate an image of what your 5th great-grandfather who fought in the Revolutionary War might look like? And the query was based on your research information as to his physical description taken from letters or diaries? Who owns the resulting image?

Currently, lower courts have stated that AI-generated content cannot be copyrighted since there is no human author. Much like the case of the "Macaque monkey selfie" where a monkey took a selfie photograph using equipment set up by a British photographer, there is no "consent" involved. Animals cannot give consent or enter into a legal agreement so it was determined that the resulting image was copyright free. The courts are using the same method to determine who owns that ancestor photo you generated using artificial intelligence.

- **Source or reference content:** Which leads to the next copyright issue of how AI platforms are gathering their reference information used to generate content. When asking ChatGPT to generate a sonnet about genealogy in the style of Shakespeare, the algorithm must have Shakespeare's sonnets in order to understand his writing style and create the genealogy sonnet.

In this case, all of Shakespeare's works are in the public domain according to United States copyright laws. But what about an author such as Tom Clancy or Stephen King whose works are still under copyright? And what about AI-generated images or even recordings based on a celebrity's image and voice? Most platforms are not transparent as to what reference content is being used and how it was acquired. This becomes an ethical issue and only furthers general fears about artificial intelligence.

AI and Source Citations

Those new to genealogy and family history soon learn the importance of source citations in proving relationships as well as facts about an ancestor. Usually source citations document how we find and use records such as census population schedules, death certificates, and even letters or diaries.

For the most part, you won't find records when making queries on an AI platform. But you may find information that serves as a clue for further research or, more likely, as social history about how an ancestor lived. In these situations, a method of citing AI-generated content is needed.

Citing sources need not be intimidating or time consuming. Stick to the basics: the information found, how it was found, information about where it was found, and locator data so another researcher can find the information.

For artificial intelligence content, here's the formula you might consider using as proposed by the Modern Language Association of America (MLA):

“[QUERY]” prompt. [NAME OF AI PLATFORM], [DATE OR VERSION OF PLATFORM], [NAME OF AI COMPANY], [DATE OF QUERY], [PLATFORM URL]

So, if I asked ChatGPT to determine the value of my great-grandfather's home in the 1930 US Census listed as \$80,000 in 2024 dollars, here is the source citation I would use:

“Value of home in the 1930 US Census listed as \$80,000 in 2024 dollars” prompt. ChatGPT, ChatGPT 3.5 version, OpenAI, 1 October 2023, <https://chat.openai.com/>.

Future Uses of AI for Genealogy

The concern over artificial intelligence in general, and specifically in family history research, is similar to the concern over social media almost 15 years ago. Remember when genealogists were worried about Facebook and X (formerly known as Twitter)?

We are experiencing the “First Phase” of using artificial intelligence when it comes to genealogy and family history research. Five years from now we should be in “Second Phase” mode. What does this mean?

Remember when Netscape was THE BROWSER everyone used when the Internet became popular in the early 1990s? And MYSPACE was THE SOCIAL MEDIA PLATFORM? Well, the second phase apps - Google and Facebook respectively - are now the most popular. The second phase of a new technology usually brings vast improvements in terms of functionality, ease-of-use, and value.

Here’s a short list of what you can expect to see in the next five years:

- DNA triangulation tools that will quickly determine relationships on family trees.
- Conversion of handwriting into searchable text including older forms of English and German handwriting.
- Creation of source citations for a variety of records using specific formats such as MLA, Evidence Explained, and AP style.
- Discovering connections between F.A.N. club members using digitized historical newspapers content and other records.
- Identification of ancestors in old family photos based on “traits” such as facial features including connecting family members based on similar traits.
- Suggesting records for expanding genealogical searches including those records not yet digitized.

AI: The Good and the Bad

With all the “hype” about artificial intelligence, it can be difficult to figure out not only what is true about this technology, but also the benefits and drawbacks. Here is an explanation of the elements of AI that have impacted or will impact how we search for ancestors.

The Good

There seem to be endless possibilities for using artificial intelligence when searching for ancestors. This makes for an exciting time to be doing genealogy.

- **Analyzing vast amounts of data:** Yes there is a lot of information available online for genealogy research, but humans cannot possibly analyze that information as quickly as artificial intelligence. This allows for discovering new connections between data points and better understanding migration patterns

and motivations, F.A.N. club relationships, the impact of social history on our ancestors, and more. What is not obvious immediately to our human minds can be quickly determined by using artificial intelligence.

- **Block chaining:** I've long been an advocate of using block chaining for genealogy data, especially DNA data. Block chaining involves tagging data with specific information including ownership, and tracking its use by others. The chain of use is kept in a public "ledger" and the owner can better understand who is using that data and why.
- **Translation and transcription:** As already demonstrated with the release of the 1950 US Census images, AI promises to make the transcription and translation of record images faster and easier. I recently uploaded a newspaper clipping from a historical newspaper that has not yet been digitized, and the AI platform did an amazing job in transcribing the content.
- **Timelines and mapping:** For those genealogists who want to fill in the "dash" between an ancestor's birth date and death date, artificial intelligence can help build complex timelines as well as "map" event dates to locations for a better understanding of how our ancestors lived.

The Bad

While many see artificial intelligence as a panacea that can cure many problems that come with genealogical research, AI can also be a Pandora's box filled with its own set of problems.

- **Lack of transparency:** One of the biggest issues for users of AI is the inability of the user to determine the source of the reference material used when generating content. Another issue: recognizing AI-generated content. Most users do not add source citations to AI-generated content or watermarks to AI-generated images.
- **Bias:** Studies have proven that many AI platforms can be biased, especially since content used as reference material is supplied by humans. The same biases we see in terms of race, gender, and age are easily replicated by artificial intelligence. Recent examples have included a bias towards generating white or caucasian faces rather than people of color when creating certain images.
- **Copyright:** Many copyright and intellectual property issues related to AI have popped up in the past year. US courts have ruled that content created by artificial intelligence cannot be copyrighted. In addition, several content creators including authors and performers have sued major AI platforms such as ChatGPT and Gemini for scraping copyright protected content from the internet to help create AI-generated content.

- **False information:** As platforms using AI gather information, who or what is determining what is true and what is false? A recent example of a law firm submitting a legal filing created by artificial intelligence - resulting in a list of fictitious court cases to support legal arguments - demonstrates the problem. This is another reason that “human review” is often required before relying upon AI-generated content.
- **Privacy violations:** Artificial intelligence can quickly collect data entered at genealogy platforms when performing research and creating family trees. In addition, users are tracked as to searches performed and this data is analyzed to create new features and products. More importantly, DNA data is captured and despite privacy policies that ensure the use of only metadata, recent computer hacks at vendors such as 23andMe have caused a steep decline in the number of people using personal DNA test kits.
- **High costs:** While not often discussed, deploying artificial intelligence can be expensive for vendors, resulting in higher prices for the genealogy consumer. The machines and servers used for AI processes require more powerful chips as well as simply just more power to run. Besides an increase in costs, there are environmental and climate impact costs through the need for more energy to power AI computers.

Conclusion

Artificial intelligence is seen as powerful but also as something to be feared. It isn't much different than how our earliest ancestors reacted to the discovery of fire. That new technology had great benefits and advanced progress in many areas of human life. But fire also brought new dangers and uses that might not have been anticipated.

The best way to cut through the current hype and misinformation around AI is to stay informed. Learn from other genealogists how they are using artificial intelligence to improve their genealogy research.

Whether you decide to take a full plunge or just dip your toe in the AI pond, you'll discover amazing possibilities and ways to take your search for your roots to the next level.

Resources

- **AI & Genealogy: Harnessing the Power of Artificial Intelligence for Family History Research** - MyHeritage Knowledgebase
<https://education.myheritage.com/article/ai-genealogy-harnessing-the-power-of-artificial-intelligence-for-family-history-research/?lcpt=article>
- **AI Record Finder** - MyHeritage
<https://www.myheritage.com/research/ai-record-finder/>
- **BanyanDNA**
<https://www.banyandna.com>
- **ChatGPT**
<https://chat.openai.com/>
- **CoPilot**
<https://copilot.microsoft.com/>
- **Disclosing Use of AI for Writing Assistance in Genealogy** - Family Locket
<https://familyloket.com/disclosing-use-of-ai-for-writing-assistance-in-genealogy/>
- **Gemini**
<https://gemini.google.com/>
- **Genealogy and Artificial Intelligence (AI)** - Facebook group
<https://www.facebook.com/groups/1255245945084761>
- **Genealogy Eyes** - ChatGPT
<https://chat.openai.com/g/g-gmlAn5mh6-genealogy-eyes>
- **How do I cite generative AI in MLA style?** - Modern Language Association of America
<https://style.mla.org/citing-generative-ai/>
- **Intersection of AI & Copyright** - Copyright Clearance Center
<https://www.copyright.com/resource-library/insights/intersection-ai-copyright/>
- **Introducing AI Biographer™: Create a Wikipedia-like Biography for Any Ancestor Using AI, Enriched with Historical Context** - MyHeritage
<https://blog.myheritage.com/2023/12/introducing-ai-biographer-create-a-wikipedia-like-biography-for-any-ancestor-using-ai-enriched-with-historical-context/>
- **MyHeritage Photo Features: What They Are & How to Use Them** - MyHeritage
<https://education.myheritage.com/article/myheritage-photo-features/>
- **Perplexity**
<https://www.perplexity.ai/>