Translating, Transcribing and Summarizing Documents Using AI

Thomas MacEntee, of Genealogy Bargains
https://genealogybargains.com hidefgen@gmail.com

The power of artificial intelligence can be used to assist genealogists in translating, transcribing, and summarizing a variety of genealogy records. A handwritten baptismal record from the 1800s? No problem. A faded newspaper obituary? No problem. Most AI platforms can help generate useful content for genealogical research.

What is Artificial Intelligence and How Does it Work?

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.

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 are incorporating AI into features provided 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.
- **Claude:** Claude (https://claude.ai/) is an artificial intelligence chatbot created by the company Anthropic that is designed to generate text content and engage in conversations with users using human-like responses.
- **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."

© 2025, copyright Thomas MacEntee. All rights reserved.

Summarization

AI document summarization typically uses special programs to focus on and pick out key sentences, phrases, or ideas from a text. These are chosen based on how often they appear and how relevant they are. The aim is to pull out the most important information while keeping the main message and context intact.

Summarize! ChatGPT

ChatGPT allows users to create and share customized programs that can be found by searching the **Explore ChatGPTs** section (https://chat.openai.com/gpts). **Summarize!** (https://chat.openai.com/g/g-Lj07Aq7sC-summarize) is an excellent tool for summarizing content that offers “In-depth Analysis, Summary, Action Plan and Further Reading Suggestions. Can be used also to summarize historical events, scientific subjects, news, etc…”

Transcription

If you’ve ever struggled to transcribe an older document, with the document propped up to the side of your monitor and you type word by word, then struggle no more. The transcribing abilities of AI platforms are outstanding, and even include the ability to create transcriptions of audio and video files as well as handwritten documents.

However, the resulting content is not always exact or perfect. Older handwriting such as English “secretary hand” or Latin used for German records can only be as good as the training data and examples used by the AI platform. Also review the resulting transcription before using the text in your research.

Translation

While the ability to translate documents and images containing text from one language to another is an amazing and useful feature for genealogists, this tool is not without its issues.

Review the text to be translated and determine when it was created. A Dutch language document from the 17th century may not produce adequate results compared to a more recent document in that language.

Consider using the services of a certified translator especially when translating documents for legal purposes. In addition, interview the translator or the translation service and make sure they can translate your document especially if it is older such as from the 17th century.

© 2025, copyright Thomas MacEntee. All rights reserved.

Improving Accuracy

Accuracy is dependent upon a variety of factors:

- **Document/Image Quality:** If the image or document is faded or damaged, consider using a photo editor to improve the quality.
- **Document/Image Type:** Not every platform will be able to handle PDF documents. Try using JPG, PNG, or other formats.
- **Document Length/Number of Images:** For large documents or a large number of images, try loading only a few at a time, in batches.
- **Print vs. Handwriting:** Printed text will always be converted faster and with more accuracy. Depending on the handwriting quality, you will see less accuracy. Also forms completed with handwriting (US World War I Draft Registration Cards) will be difficult due to the formatting of the form, especially the lines.
- **Platform:** Some of the AI platforms listed above will do a better job with better accuracy. Try different platforms for best results.

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.

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 translate a page from the book *Le troisième centenaire de l'Édit de Nantes en Amérique et en France*, here is the source citation I would use:

> "Translate to English" prompt using digital image of *Le troisième centenaire de l'Édit de Nantes en Amérique et en France*, page 3, published 1989. ChatGPT, ChatGPT 3.5 version, OpenAI, 12 March 2024, https://chat.openai.com/.

© 2025, copyright Thomas MacEntee. All rights reserved.

Tips and Tricks

General

- **Provide clear and concise instructions**. Make your prompt clear as to the desired result, such as “transcribe exactly” instead of “transcribe.” For translation use “translate to English” rather than “translate.”

- **Remember to provide feedback.** Most AI platforms provide a way to give feedback on the results generated. Remember these are “learning” platforms and this feedback is essential to better results in the future.

- **Break down the task into separate prompts.** You may find more success if you first ask the AI platform to transcribe a foreign language document and then follow up with the “translate to English” prompt.

- **Beware of platforms recommended by search engines.** There are many “specialized” AI platforms stating that they do a better job of transcribing or translating than “the bigger, better-known” platforms.

- **Review and test customized ChatGPTs.** Most of the offerings in the Explore ChatGPTs section are merely leads to external websites trying to sell an AI tool. Look for the number of chat conversations executed; the higher the number, the more popular the tool.

Summarization

- **Specify format.** Add modifiers to your prompt such as “bullet points.”

Transcription

- **Break down into smaller segments**. Especially when uploading a document or image containing handwritten text, use single pages or smaller segments for better accuracy.

- **Transcribe audio and video content**. Consider using AI to also upload video or audio files to create a transcription.

Translation

- **Craft specific prompts.** When uploading a document or image, use “translate to English” or other language instead of just “translate.”

- **Carefully review translated text.** The translated text should be similar in length to the original text. Some AI platforms may generate lengthy translations that are not useful.

- **Check the context for translated text.** While most AI platforms do a good job of translating text from one language to another, many results are literal rather than contextual.

© 2025, copyright Thomas MacEntee. All rights reserved.

Resources

- **ChatGPT**
  https://chat.openai.com/
- Claude
  https://claude.ai/
- **CoPilot**
  https://copilot.microsoft.com/
- **Explore ChatGPTs**
  https://chat.openai.com/gpts
- **Find Results with Full Text** - FamilySearch Labs
  https://www.familysearch.org/search/full-text
- **Gemini**
  https://gemini.google.com/
- **Genealogy and Artificial Intelligence (AI)** - Facebook group
  https://www.facebook.com/groups/1255245945084761
- **Summarize! ChatGPT**
  https://chat.openai.com/g/g-Lj07Aq7sC-summarize

© 2025, copyright Thomas MacEntee. All rights reserved.