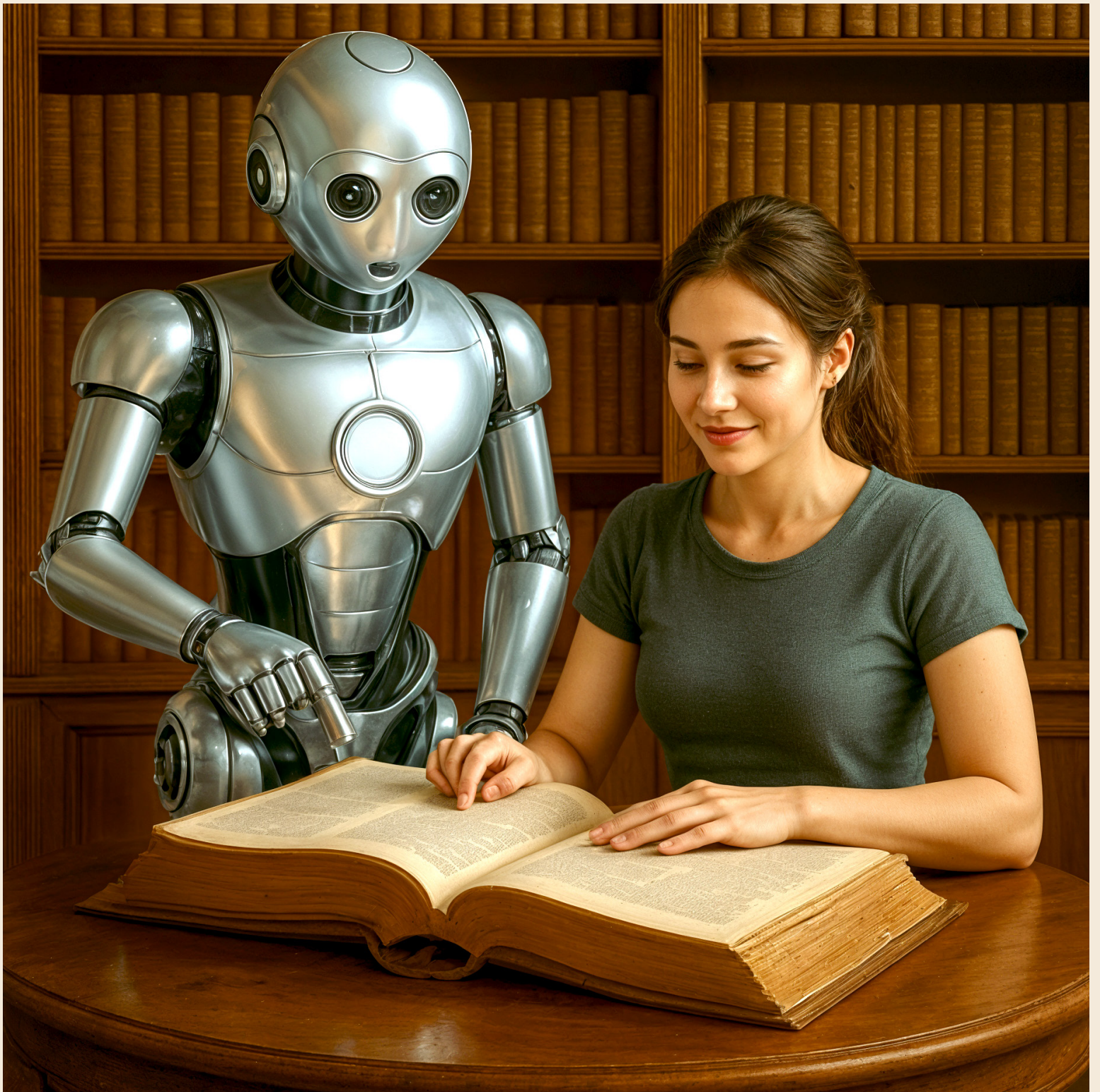


# ChatGPT for Genealogists

From first prompts to advanced workflows



ROBERT CAMERON WEIR

# **ChatGPT for Genealogists**

From first prompts to advanced workflows

Robert Cameron Weir

Antic Press  
Dover, NH

© 2026 Robert Cameron Weir

All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without prior written permission of the publisher.

Published by Antic Press, Dover, NH

Contact: [info@anticpress.com](mailto:info@anticpress.com)

[www.aiforgenealogists.com](http://www.aiforgenealogists.com)

ISBN: 978-1-7338971-5-0 (paperback)

LCCN: 2026905146

Publisher Cataloging-in-Publication Data:

Weir, Robert Cameron, 1969–

*ChatGPT for Genealogists : From first prompts to advanced workflows* / Robert Cameron Weir.

Dover, NH : Antic Press, 2026.

298 pages ; 28 cm

Subjects: LCSH: Genealogy—Data processing. | Genealogy—Research—Methodology. | Artificial intelligence. | ChatGPT.

Classification: LCC CS21 .W45 2026 | DDC 929.1—dc23

First edition, 2026

All product names, logos, and brands are property of their respective owners. All company, product, and service names used in this book are for identification purposes only. Use of these names does not imply endorsement.

This book is not affiliated with nor endorsed by OpenAI.

# Table of Contents

Dedication .....	iii
Foreword .....	vii
Preface .....	ix
Introduction.....	xi
Acknowledgments .....	xv

## Part I—Quick Start

Chapter 1: Quick Start .....	3
Chapter 2: Fundamentals of Prompting.....	11

## Part II—Under the Hood

Chapter 3: How We Got Here .....	17
Chapter 4: The Anatomy of an LLM.....	21
Chapter 5: Cautions .....	29
Chapter 6: Advanced Prompting .....	35
Chapter 7: Advanced Features .....	43

## Part III—Core Workflows

Chapter 8: Working with Text Documents.....	51
Chapter 9: Transcription .....	59
Chapter 10: Translation and Transliteration.....	69
Chapter 11: Reading Web Content.....	77
Chapter 12: Working with Genealogy Files.....	89
Chapter 13: AI as a Teacher .....	101
Chapter 14: Image Analysis and Interpretation .....	115
Chapter 15: Image Generation .....	127
Chapter 16: Image Modification.....	139
Chapter 17: Imagined Images .....	153
Chapter 18: Data Analysis.....	167
Chapter 19: DNA and Genetic Genealogy .....	183
Chapter 20: Genealogical Reasoning .....	185

## Part IV—Professional and Society Use

Chapter 21: Professional Genealogists.....	201
Chapter 22: Genealogical Societies.....	213

## Part V—Programming Assistant

Chapter 23: Code as Text .....	233
Chapter 24: Query Languages and Wildcards .....	235
Chapter 25: Date Conversion Macro .....	239
Chapter 26: Image Reviewer .....	243
Chapter 27: GEDCOM Parser .....	245
Chapter 28: Social Security Death Master File .....	257

## Part VI—Ethics and Futures

Chapter 29: AI, Genealogy, and Ethics .....	273
Chapter 30: The Perfect AI for Genealogy: A Wish List.....	275
Glossary.....	277
Further Resources .....	279

# Preface

Hello, you must be one of those people who actually read book prefaces. I'm so glad you stopped by. Most people skip right over the preface and sometimes even the introduction. It hardly seems worth the effort to write a preface anymore. But seeing that you're here...

Presumably you are here to hear a little about how this book came about. OK. Here's the skinny. There is certainly no shortage of information out there on AI. Some of it verges on apocalyptic with a heavy dose of Neo-Luddism, while from the other direction are those whose unhinged support of AI verges on idolatry. Even within our small world of genealogy, I've seen both poles of opinion.

My own view on AI and genealogy is more nuanced, but don't take that to mean that I hold middle-of-the-road, milquetoast views. As you'll see, I hold strong opinions. I think this comes from my background, both my scientific training (I started off studying astrophysics at Harvard) as well as my twenty-seven years at IBM as a software engineer, where I worked on AI-related technologies like information retrieval and machine learning. I understand how AI works on the inside. I developed my own AIs at home. I don't accept anyone's opinion on what these models can and cannot do. I try it, and then try it again differently. I conduct experiments. When I make a claim, I bring the receipts.

Combined with that scientific and technical background, I'm also an experienced genealogist. I actively do research, for clients and on my own ancestry. I lecture and teach classes on genealogy. And I write on the topic.

So, I think I have the right set of skills and knowledge and the right attitude to show my fellow genealogists the real power of AI today, its strengths and weaknesses, so you can use it effectively in your genealogical pursuits.

# Introduction

## Welcome

The Residenz, in Munich, was once the royal palace of the Wittelsbach dynasty, the kings of Bavaria. Among the many ornate rooms and hallways is the *Ahnengalerie* or ceremonial gallery of ancestors, in the rococo style, filled with portraits of illustrious ancestors, each framed in a gilded cartouche. On the facing wall, between the windows, is a giant family tree, “GENEALOGIA AUGUSTAE GENTIS WITTELSBACICAE BOICAE” (Genealogy of the August House of Wittelsbach of Bavaria), going over 35 generations deep, down to Theodo I, an 8th century Bavarian duke.

A question to ponder: Do you think that the Elector Karl Albrecht, the future Emperor Charles VII, when he created this gallery in 1726,<sup>1</sup> did the genealogical research himself? Probably not, no more so than he painted all the portraits himself. Of course, he had someone do the research for him.

Although there are exceptions, the wealthy and powerful generally rely on others to do their thinking as well as their labor. Then as now, brains and brawn are for sale. The rich can hire smart people to manage their money, to write and review contracts, to plan their vacations, to design their homes, to monitor their health, and even to research and write up stories of their illustrious forebears.

And the rest of us, those without palaces? Well, we do what we can with our own knowledge and experience, and this often is quite good. But what more could we do if we had access to the best minds in the world, and had their assistance whenever we wanted, 24×7? And what if this advice was free, or maybe only \$20/month? How much would that change things?

We’re not quite at that point yet, but we’re getting close. ChatGPT may not be the smartest mind in the world, but it is smarter than me, and perhaps you, at many things, and is rapidly getting better.

The purpose of this book is to orient you to the best uses of ChatGPT for genealogy, of using this state-of-the-art mind-for-hire to go deeper into your research, to open new avenues of investigation, and to break through brick walls.

In other words, I want to teach you how to live like a king.

## Using this Book

This book is organized into six parts:

- **Part I—Quick Start**, as the name suggests, provides the essential knowledge you need to get set up and productive with ChatGPT quickly.  
If you are already a ChatGPT user you should feel free to skim this part quickly, or skip it altogether.
- **Part II—Under the Hood** briefly covers the history of AI, and then provides an explanation of how ChatGPT was created, how it works, and what its limitations are. We also look at some of the more advanced features of ChatGPT like deep research and Custom GPTs.
- **Part III—Core Workflows** presents a series of chapters on specific genealogical workflows using ChatGPT, such as Transcription, Translation, AI as a Teacher, Data Analyst, Image Analysis and Interpretation.
- **Part IV—Professional and Society Use** explores additional workflows of particular interest to professional genealogists and to leaders and volunteers of genealogical societies.

---

1. <https://www.residenz-muenchen.de/englisch/museum/ahnen.htm>

- **Part V—Programming Assistant** presents super-powerful techniques available to ChatGPT users who also know a little about computer programming.

I appreciate that this is a very small portion of the you, but the opportunities here are so great that I don't want to ignore this area entirely. Even if you are not a programmer yourself, it is good to know what the “art of the possible” is in this area, especially if you have a niece or nephew or grandchild who is a programmer.

- **Part VI—Ethics and Futures** closes out the book with a few brief chapters.

The book then wraps up with a glossary and some further resources sources for learning more.

## How this Book is Formatted

The narrative of the book is formatted using a few conventions throughout.

The interactions with ChatGPT are formatted like this:

This is a prompt that I entered

This is the response returned by ChatGPT.

### Asides

Asides provide additional information to complement the topic under discussion. Read it or skip it as you wish.

## Never the Same Twice

The ChatGPT responses illustrated in this book are taken verbatim from the responses I received from the AI. I've only reformatted the response to be consistent with the typography used this book. But if you enter the exact same prompt that I used, you will likely get a different answer. That response might be just a rewording of the same information, but it might differ in more substantial ways.

What's going on here? There are a variety of things happening here.

First, ChatGPT's algorithms have a built-in element of randomness to them. This is intentional, done to make its answers appear more creative and spontaneous.

To illustrate this, I entered the following prompt three times and received three different responses:

Who was the 10th president of the United States?

The 10th president of the United States was John Tyler, who served from 1841 to 1845.

The 10th president of the United States was John Tyler, who served from 1841 to 1845. He assumed the presidency after the death of William Henry Harrison, becoming the first vice president to succeed to the office upon a president's death.

The 10th president of the United States was John Tyler, who served from 1841 to 1845. He assumed the presidency after the death of William Henry Harrison, becoming the first vice president to ascend to the office due to a president's death.

Note how the last two answers are nearly identical, differing only in a single word choice: "succeed" versus "ascend."

Second, OpenAI frequently updates their product, making small adjustments without notice, and making larger ones with announced version updates. The above responses were from ChatGPT 5.1.

If I give the same prompt to ChatGPT 5.0 I get a different response:

The 10th president of the United States was John Tyler.

He served from April 4, 1841, to March 4, 1845, succeeding William Henry Harrison, who died just one month into his term. Tyler was the first vice president to assume the presidency due to the death of his predecessor, setting a precedent for presidential succession later formalized in the 25th Amendment.

A few key facts:

- Political party: Whig (though later expelled from it during his presidency)
- Vice president: None (the office remained vacant during his term)
- Major events: Annexation of Texas (1845), Webster–Ashburton Treaty (1842), and conflicts over federal authority and banking.

Tyler was sometimes called "His Accidency" because many contemporaries viewed his succession as illegitimate, though he firmly established the principle that the vice president becomes president upon the death of the incumbent.

Third, users can set the "Base style and tone" of ChatGPT to one of several choices, including, Default, Professional, Friendly, Candid, Quirky, Efficient, Nerdy, or Cynical. I used Professional for my initial prompts. OpenAI defines this style as "polished and precise." If you set up your account with a different setting you're answers may be expressed differently. For example, when set to Cynical the response is:

John Tyler.

There, mystery solved. He's the guy who wound up in the job because William Henry Harrison couldn't stop giving the world's frostiest inaugural address and promptly died. Tyler then spent his term arguing with basically everyone, including the party that elected him.

But yes: **John Tyler.**

Finally, depending on how you have your privacy settings configured, ChatGPT may, over time, learn about you and your interests and what you are working on and try to "connect the dots" in the answers. For example, if I had previously mentioned that I was writing a school paper on the 25th Amendment it might have leaned into that aspect of the answer a bit more.

Note that none of the answers here were necessarily wrong. They differed mainly in length, style, and emphasis. But this was because I asked a question concerning an objective fact. If I had asked for an

opinion, like the “tell me the top 10 resources for reading old German handwriting” or asked for a creative response, like “Suggest 5 good names for a genealogy-related cookie” then we should expect wider and more substantive differences each time the prompt is entered. For example, here are two invocations of that cookie-naming prompt, with very different answers:

Below are five strong, brand-ready name options for a genealogy-themed cookie. Each is crafted to evoke ancestry, lineage, and tradition while still sounding like a treat someone would want to buy.

- Heritage Bites
- Ancestry Crunch
- Family Tree Cookies
- Legacy Crisp
- Roots & Crumbs

If you want names that lean more whimsical, more premium, or more historically themed, I can tailor additional options.

Here are five strong, marketable name concepts for a genealogy-themed cookie. Each option is designed to communicate heritage, lineage, and nostalgia while still sounding like a product you could put on a shelf or use for an event.

- Ancestry Bites Evokes family history in a light, approachable way.
- Heritage Crunch Cookies Connotes tradition and depth but still feels like a snack.
- Legacy Layers Plays on the idea of generations stacked over time.
- Family Tree Treats Friendly, easy to understand, and directly genealogy-linked.
- Lineage Shortbread Offers a premium, classic tone suitable for a specialty product.

If you want names that lean more whimsical, more formal, or themed around a specific culture or historical period, I can produce tailored options.

## Updates

Since the AI field is rapidly advancing, with model updates coming every few months. I am also maintaining a page on my website where I will list addenda as well as any necessary corrigenda to this book. You can find it at [www.AIForGenealogists.com](http://www.AIForGenealogists.com). If you want to be notified via email of updates related to this book you can sign up here: <https://aiforgenealogists.com/updates/>

## Chapter 9: Transcription

When we transcribe a document, we take it from one media, usually a older, original source, and create an accurate and literal copy of it. Traditionally, transcribing involved having the original document in front of you as you wrote the same text, letter for letter, word for word, line for line, into your notebook. In the computer era this task was accomplished by typing a literal copy directly into a word processor.

In all cases the goal is the same: creating a literal copy, preserving (to the extent possible) the idiosyncrasies of the original, including original spelling, capitalization, punctuation, superscripts, subscripts, insertions, marginal notes, cross-outs, line breaks, etc.

Since historical documents can be torn, stained, burned, wormed, or just suffer from really bad handwriting, scholars have adopted conventions for noting portions of the text that are missing or illegible, and indicating which readings are conjectural or uncertain. This is done by putting editorial additions within square brackets.

This approach, called semi-diplomatic transcription, is standard in genealogy and ChatGPT can help you do this. For example, you can upload a page from a scanned book, obituary, even a handwritten document like a letter or a deed, and ask ChatGPT to transcribe it for you.

The ability to take digital images and extract text from them is not new. If you've used Adobe Acrobat, Abbyy FineReader, or similar programs, you may have encountered Optical Character Recognition (OCR) before. ChatGPT can accomplish something similar, but it goes about it in a very different way. It's important that you understand this difference. So, let's compare how we as genealogists read text and compare that to traditional OCR and this newfangled AI.

### How We Read

By the time we get to mid to late elementary school most of us transition from deliberate decoding of text (sounding things out text letter-by-letter) to automatic word recognition. If the legibility of the text is good, we hardly even notice that we are decoding text. It is beneath the level of our conscious effort. It just happens.

But if the text is harder, maybe in a foreign language, or highly technical, or printed very small, or in an unfamiliar typeface or handwriting, then we revert back to a slower method of reading. In some especially difficult cases it is like we're back in the third grade, reading letter-by-letter.

As genealogists we have other tricks we can bring to bear on hard-to-read documents. In some cases, we know the context and the formulaic language involved and can orient ourselves to the text based on those expectations. Common record types like baptismal records or deeds are like this. In other cases we'll come across a letter we're uncertain of and then hunt around the page, or even on adjacent pages, to find other examples of that same set of strokes, hoping to identify it by finding it used in a word that we're more certain of. In a sense, we're not just reading, we're looking for clues while sleuthing around in the document

### Traditional OCR

Your classic, pre-LLM, OCR program starts by looking at the entire image to recognize the main text block. Is it a single block or is it multiple columns? Are there page headers and footers that are separate from the body of the page? Is the page crooked? If so, it may rotate the page to square things off. Is the contrast too low? If so, it might adjust the contrast.

Once it has the page structure figured out and the page optimized, the OCR identifies the lines and the line spacing and processes each line in turn, determining the character boundaries, and calculating for each character what the most likely letter is. Some programs will determine this purely on the shape of the letter. Other programs may also consult a dictionary and pick the letters that result in a sensible word.

Traditional OCR software works mostly line-by-line and word-by-word. They may use dictionaries and simple language models to nudge doubtful letters toward real words, but they do not do the kind of broader document sleuthing that we genealogists do, hunting around the page for a clearer example of the same letter shape.

## How an LLM Reads

Here it gets odd. Very odd. An LLM like ChatGPT does not do traditional OCR to extract text from an image. It doesn't run a separate preprocessing step to find text blocks, skew the page, or tweak contrast. It handles things like small rotations or lighting changes implicitly via the model's learned vision model, not by explicit rules. It does not slice the image into individual character boxes and recognize each letter the way classic OCR does. Instead, it looks at the whole image and then generates the transcription guided by what it "sees" and what it knows about spelling and language.

An LLM initially treats your document just as it would if you had uploaded a photo of your cat. It "embeds" the image, meaning it expresses, in detail, the image's visual appearance as a compact mathematical representation. The language model then uses that representation, together with your prompt, to generate text. The same internal representation of the image supports both low-level tasks ("what letters are printed here?") and high-level tasks ("what kind of animal is this?"). The model has learned, from training, how to decode those visual features into either a literal transcription or a conceptual description, depending on your prompt.

If the above explanation sounds intuitive and obvious to you, go back and read it again!

## Printed Type

As an example, I will upload this image from volume two of Jeremy Belknap's *The History of New Hampshire* (1784). It exhibits a range of complexities genealogists encounter with early texts, including marginal annotations, antique type, and the use of old letter forms, in this case the long-s (sometimes confused with the letter f) in initial and medial positions, e.g., fhadow or publifhed.

Create a semi-diplomatic transcription of this document.

Note that there is some discretion in how semi-diplomatic transcriptions are done. Although in all cases spelling, capitalization, punctuation, and line breaks are preserved, editors vary in how they treat things like obsolete letter forms and superscripted abbreviations. Some will write **publified** in the transcription, while others will normalize to **published**. Some will print s<sup>d</sup>, while others will lower the superscript to **sd**. None of these choices are wrong. All of these variations have been used in serious scholarly work. But if you have a very specific set of conventions in mind, by all means, write a longer prompt and specify your full requirements. In this case, however, I will just accept ChatGPT's default interpretation of semi-diplomatic transcriptions.

## NEW-HAMPSHIRE.

405

It ought ever to be remembered, that the declaration of our independence was made, at a point of time, when no royal Governor had even the shadow of authority in any of the Colonies; and when no British troops had any footing on this Continent. The country was then absolutely our own. A formidable force was indeed collected on our coasts, ready to invade us; and in the face of that armament this decisive step was taken. The declaration was received with joy by the American army then assembled at New-York. Within fourteen days it was published by beat of drum in all the shire towns of New-Hampshire. It relieved us from a state of embarrassment. We then knew the ground on which we stood, and from that time every thing assumed a new appearance. The jargon of distinctions between

1776.

July 4.

July 18.

## NEW-HAMPSHIRE. 405

It ought ever to be remembered, that [in margin: 1776. July 4.] the declaration of our independence was made, at a point of time, when no royal Governor had even the shadow of authority in any of the Colonies; and when no British troops had any footing on this Continent. 'The country was then absolutely our own.' A formidable force was indeed collected on our coasts, ready to invade us; and in the face of that armament this decisive step was taken. The declaration was received with joy by the American army then assembled at New-York.

Within fourteen days it was published by [in margin: July 18.] beat of drum in all the shire towns of New-Hampshire. It relieved us from a state of embarrassment." We then knew the ground on which we stood, and from that time every thing assumed a new appearance. The jargon of distinctions between

Generally, it did a good job here, preserving spelling, capitalization, punctuation, and line breaks. It normalized the long-s to its modern form, which is within allowable discretion.

The main errors were misreading some stray marks in the original as quotation marks, and inserting a paragraph break where one did not exist in the original.

In addition to semi-diplomatic, ChatGPT can produce a more normalized text which might be appropriate in other contexts, say for a school textbook.

Give a clean version of this, without the marginal notes, removing hyphens, and letting the text flow normally. Update to modern conventions for spelling, capitalization, and punctuation.

It ought always to be remembered that the Declaration of Independence was made at a point in time when no royal governor had even the shadow of authority in any of the colonies, and when no British troops had any footing on this continent. "The country was then absolutely our own." A formidable force was indeed collected on our coasts, ready to invade us; and in the face of that armament, this decisive step was taken. The declaration was received with joy by the American army then assembled at New York.

Within fourteen days, it was published by beat of drum in all the shire towns of New Hampshire. It relieved us from a state of embarrassment. We then knew the ground on which we stood, and from that time everything assumed a new appearance. The jargon of distinctions between...

Note that the errors in the original transcription, the added quote marks and new paragraph, carried over into this version.

In the above example, and the one that we'll look at next, are cases where the original document is clean and legible. A genealogist experienced in the writing of this period would have little difficulty making an accurate transcription. ChatGPT, on the other hand, introduced errors in its transcription.

So, what value is an AI that makes mistakes? What use is a human assistant that makes mistakes? I think the cases are similar. **I recommend using ChatGPT as a time saver.** It does not transcribe better than me, but it does transcribe far faster than me. I'll still need to carefully check its transcription against the original. Occasionally I will find errors. But even accounting for time spent verifying, my total effort is less than if I had done the transcription entirely on my own.

## Manuscripts

Now let's look at a handwritten example, from the metes and bounds from a nineteenth century deed:<sup>13</sup>

8d bounded as follows - viz. beginning at a maple tree standing on the borders of the Mill pond thence south 64° west on land formerly Samuel Stone's three rods & one fourth to a Beach tree thence south 7½° east fifteen rods & a half to a stake & stones by a corner of the wall on land of said Luther Stone thence on land of said Stone south 32° east by the wall & fence twelve rods to a stake & stones thence on land of said Stone south 62° east twenty seven rods & a half by the fence to an Elm tree near the bridge thence bounding on the French river - up stream fifty one rods to the bounds first mentioned - Also the dwelling

Transcribe this land description. Give a literal, word for word, line for line transcription, preserving the period spelling and punctuation and line breaks of the original. If a letter or word is unclear, put your guess in square brackets.

8d bounded as follows.- viz. beginning at a maple tree standing on the borders of the Mill pond thence south 64° west on land formerly Samuel Stone's three rods & one fourth to a Beach tree thence south 7½° east fifteen rods & a half to a stake & stones by a corner of the wall on land of said Luther Stone thence on land of said Stone south 32° east by the wall & fence twelve rods to a stake & stones thence on land of said Stone south 62° east twenty seven rods & a half by the fence to an Elm tree near the bridge thence bounding on the French river- up stream fifty one rods to the bounds first mentioned- Also the dwelling

This also did well. The only error I see is that it got confused by the very first character, which looks to be an ampersand with a loop of another letter descending from text in the line above.

But what about trickier documents, the kind that even an experienced genealogist will struggle with? How about a messy example like the following?<sup>14</sup>

13. Stone to Rice, Worcester Co., Mass. Deeds, 495:59 (1852).

14. "City of Dover, NH Town Records 1657-1753," page 1.

Dover, N.H.      Pay      Anno Domini 1717

Memorandum. It is this day ordered, constituted, constituted, and agreed upon, that the Inhabitants of Dover, New Hampshire, shall and shall cause to be made and set into a forme of levying of Rates and Assessments, for raising of publique charges according to an order of Court made and held at Boston. dated: 13. of the 9. m: the following: viz.

1. That the select men and one man more chosen by the Inhabitants aforesaid, shall have full power and authority to make Rates, and to follow this forme as followeth.

2. That they take and make, an exact and perfect list of all the Male Inhabitants, wite in this Townshipp, from the Age of sixteen yeared and upwards, according to a true valuation, and estimate of all Estates, as well real, as personal, of houses, and lands of all sorts, unbroken as others. Except, such as shall be given for free feeding of Cattle to the use of the Inhabitants in general, whether belonging to Towns, or particular persons, viz. Mills, Shoppes, merchantable goods, Cranes, wharves, and all other sorts of Cattle and other valuable goods, or Estates either at Sea or Land. The aforesaid lands and Estates are by the said select men, and Commissioners to be assessed and rated, as here after followeth: viz.

For real estate, as aforesaid: Magistrates only excepted, Two shillings and six pence per acre. And all Estates both real & personal, at One penny for every Twenty shillings, and

Transcribe this passage from a 17th century town record book from Dover, NH. Give a literal, word for word, line for line transcription, preserving the period spelling and punctuation and line breaks of the original. If a letter or word is unclear, put your guess in square brackets.

For example, if you think it says "land" but are not certain, write it as "[land]". If a character is so uncertain that you cannot guess, put a period as a placeholder for each uncertain letter, e.g., "F[...]d" for three consecutive uncertain characters within a word.

You may recognize here the use of few-shot prompting, which we discussed back on page 35.

I'll explain why in a moment, but I took the exact same manuscript image and the exact same prompt and had ChatGPT create the transcription twice, in two different chat sessions.

Dovr. Pay

1

[diretimo] [Juno] [17] Anno Dm: [16]47

memorand. It is this day ordered concluded and agreed upon, that the Inhabitants of Dover doe and shall [consent] and assist unto the forme of leuiting of Rates and Assessments, for raising of publique charge according to the Order of Court made to be held at Boston. dated 13 of the 9: mo: the following: vizt.

That the Select men and one man more chosen by the Inhabitants aforesaid, Shall have full power and authority to make Rates, and to follow this forme as followeth.

That they take and make an exact and true list of all the Male Inhabitants within this Township, from the Age of sixteen yeares and upward, according to a true valuation: and likewise of all Estates of what reall, and personall, of houses and lands, and Cattle, such as horses, oxen, cows, young cattell, [sheepe] [swine] as also of [Corne] and other [moveable] goods, [Cranes] [Wharves] [Woods] of Mills, Shippes, merchantable goods, [wares] [Marchandize] and all sorts of Cattle and other visible goods, or Estates within the said Towne, the persons and Estates are by the said Select men, and [Comissioners] to be assessed and rated as here after followeth.

[Rate two pence] upon [said] Magistrates only excepted: Two [shillings] and [six pence] in head and all Estates both reall & personall, per pound, at one penny for the twenty shillings after to the rate of Cattle hereafter mentioned. [vizt.]

Dover.

pag 1

Anno Dni. [1647]

Memorand. It is this day ordered consented and agreed upon, that the Inhabitants of Dover for and shall [raise] the publique charges according to order of Court made and held at Boston. dated 13 of the 9th mo Anno [16..] followeth viz:

That the select men and one man more chosen by the Inhabitants aforesaid, shall have full power and authority to make Rates, and to follow this forme as followeth.

That they take and make an Exact and per-fect list, of all the Male Inhabitants with in this Towneship, from the Age of sixteen yeares and upward, Attording to a true Valuation, and Estimation of all Estates, of what nature and kind soever, as in Lands and houses, Cattell, [..] [..] [..] as also [..] [..] with other belonging to Townes or particular persons, vizt Mills, shippes, merchantable goods, [wares] moneys and debts, or other visible goods, or Estates within the said

[towne], and [dominions] to be assessed and Rated as herein after followeth: viz:

That no person aforesaid Magistrates only excepted, Two shillings and six pence in the [pound] and all Estates both Real & personall, at One penny for each Twenty shillings, Attording to the rate of Cattell hereafter mentioned. viz:

Clearly, the AI is clearly struggling here with this transcription. There are some scribal conventions and letterforms that it simply does not understand.

Whereas traditional OCR would produce nonsense, random characters, or symbols when it erred, here everything looks reasonable, at least at first glance. That is the nature and the particular danger of hallucinations in an LLM. **They often look right while being wrong.**

The expression of uncertainty, using the square brackets for conjectured readings, is also inconsistent. There are passages that differ between the two transcriptions even though neither one was marked as uncertain, and there are passages marked as uncertain in one transcription that the other transcription did not mark as uncertain.

I'd like to proposal a rule of thumb: **If a factual question is given to an LLM twice and it gives two different answers, you should trust neither of them.**

That said, let's see if ChatGPT can correlate these two transcriptions and help us make sense of the discrepancies:

Here is another transcription, made by you earlier, of the same image, using the same prompt. Compare this transcription with the one you just returned. Determine a best reading and put in bold the text where the two transcriptions are in agreement and in italics where they differ.

<paste in early transcription>

*Dovr. Pay*

1

*[diretimo] [Juno] [17] Anno Dm: [16]47*

**memorand. It is this day ordered concluded and agreed upon, that the Inhabitants of Dover doe and shall [consent] and assist unto the forme of leuiting of Rates and Assessments, for raising of publique charge according to the Order of Court made to be held at Boston. dated 13 of the 9: mo: the following: vizt.**

**That the Select men and one man more chosen by the Inhabitants aforesaid, Shall have full power and authority to make Rates, and to follow this forme as followeth.**

**That they take and make an exact and true list of all the Male Inhabitants within this Township, from the Age of sixteen yeares and upward, according to a true valuation: and likewise of all Estates of what reall, and personall, of houses and lands, and Cattle, such as horses, oxen, cows, young cattell, [sheepe] [swine] as also of [Corne] and other [moveable] goods, [Cranes] [Wharves] [Woods] of**

**Mills, Shipps, merchantable goods, [wares] *[Marchandize]* and all sorts of Cattle and other visible goods, or Estates within the said Towne, the persons and Estates are by the said Select men, and *[Comissioners]* to be assessed and rated as here after followeth.**

*[Rate two pence] upon [said] Magistrates only excepted: Two [shillings] and [six pence] in head and all Estates both reall & personall, per pound, at one penny for the twenty shillings after to the rate of Cattle hereafter mentioned. [vizt.]*

In theory, we could take that and go back to the original manuscript image, confirm whether the bold parts were transcribed correctly and give deeper attention to the parts in italics, where the two transcriptions differed, and to the parts in square brackets.

But at this point we need to ask ourselves: Is ChatGPT actually saving us time in preparing this transcription? Did it do a good enough job? Or are we going to spend more time cleaning this up than it would take us to transcribe ourselves? It is not clear to me that, with tough documents like this, using ChatGPT is a net benefit.

On the other hand, suppose we were not necessarily seeking a full and accurate transcription of this manuscript. We just wanted a quick sense of what it was about, a summary or abstract, allowing us to judge whether the material was at all relevant to our research. If it is, then we'd knuckle down and create a full and accurate transcription.

So, let's try this, in a new chat session, using the original image:

Summarize this passage from a 17th century town record book from Dover, NH, preserving any dates, places, and persons mentioned, but eliminating formulaic language.

In Dover, the town agreed to follow a standardized method for raising public charges by rates and assessments, based on an order of the Court held at Boston, dated the 13th of the 9th month (year written as Anno Dom. [16]47).

- Authority to assess: The selectmen, plus one additional man chosen by the inhabitants, were given full authority to make the rates.
- Who and what must be listed: They were to compile an exact list of all male inhabitants aged sixteen and upward, and make a “true valuation” of all estates, both real and personal—including houses and lands, cattle (and other livestock), and other taxable property such as mills, ships, and merchantable/vendible goods.
- How it would be taxed (as stated here): a poll (head) rate of two shillings and six pence per head, with magistrates excepted; and an estate rate stated as one penny per twenty shillings (with further livestock rating rules indicated as following after this passage).

I believe that a summary like this would give a good sense to the reader whether or not the manuscript was relevant to their research and worth investing time for a deeper reading.

## The Danger Zone

Note that in everything we've done here, we've used ChatGPT to produce an initial draft of a transcription, which we then manually reviewed and corrected. Where the original document was clear and easy to read, the AI did a great job and we had very few corrections to make. Where the original was harder to read, we had a great deal to clean up.

Also note that when ChatGPT made a mistake the error did not stand out in an obvious way. If it misreads "Corne" it will not print "C0me" or something that. It may just print "Potato" or some other plausible word that fits the context. If you are not doing a careful word-for-word review of the transcription, you could entirely miss the error.

The danger to avoid here is attempting to use ChatGPT to transcribe text that you yourself cannot accurately read. How could you possibly be an effective reviewer of a transcription if you are not adept at reading the text yourself? The danger is compounded if the manuscript is in a foreign language or even uses a non-Latin script, as we'll look at in the next chapter.

The best practice here is to use ChatGPT as a time saver, a faster path to a first draft for texts that you could, given the time, transcribe accurately yourself. Avoid putting yourself in a situation where you are trying to use an AI-generated transcription that you have no way of verifying. That's just asking for trouble.

## Meet Your New Research Assistant

ChatGPT for Genealogists is a practical, real-world guide to using generative AI responsibly and effectively in family history research. Written for genealogists, it takes you from your first prompts to advanced workflows used by serious researchers and professionals.

Printed in full color so the screenshots, graphics, charts, and shaded prompt-and-response callouts are easy to follow.

You will learn how to use ChatGPT to transcribe and translate historical documents, extract key details from common genealogical sources, analyze data (including GEDCOM files), date photographs, search websites more effectively, and even write Microsoft Office macros.

Just as importantly, the book is candid about the limits and risks of AI. You will learn why large language models sometimes invent facts, how to verify every claim, and how to use ChatGPT in ways that align with genealogical standards. Clear guidance is included on ethics, and appropriate handling of sensitive information. Two chapters address the unique needs of professional genealogists and genealogical societies.

Whether you are a hobbyist exploring your roots or a professional seeking a more efficient workflow, ChatGPT for Genealogists shows you how to use AI as a capable collaborator.

### About the Author



Robert Cameron Weir is a professional genealogist based in Dover, New Hampshire. A 12th generation New Englander with more than twenty years of research experience, he specializes in solving complex family history problems through careful analysis, sound methodology, and the effective use of modern research tools. His primary research focus is New England from 1620 to the present.

He holds a Certificate in Genealogical Research from Boston University's Center for Professional Education and is a member of the Association of Professional Genealogists. He is a graduate of Harvard College (1991).

In addition to genealogical research, Weir teaches genealogists how to use artificial intelligence tools, especially large language models such as ChatGPT, as a practical research assistant for tasks like document interpretation, translation, analysis, and writing. His approach emphasizes verification, clear citations, and responsible use with attention to privacy, ethics, and reliability.