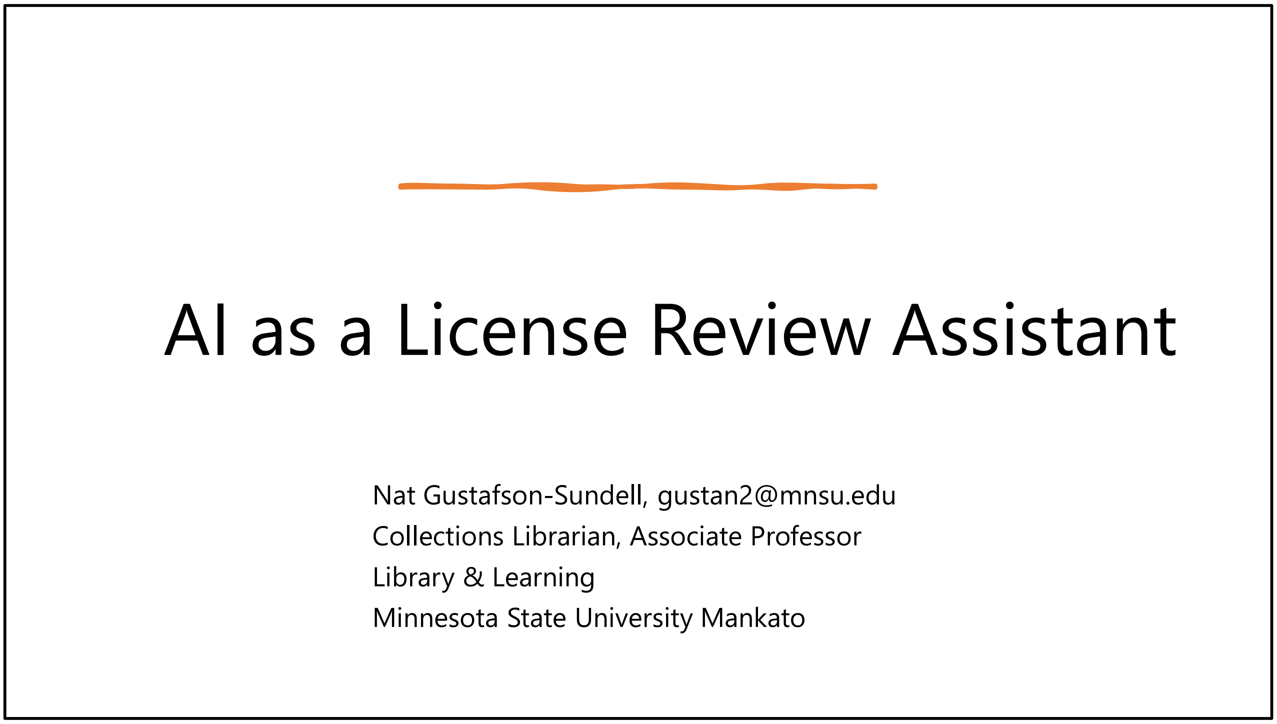


AI as a License Review Assistant

Nat Gustafson-Sundell, gustan2@mnsu.edu
Collections Librarian, Associate Professor
Library & Learning
Minnesota State University Mankato

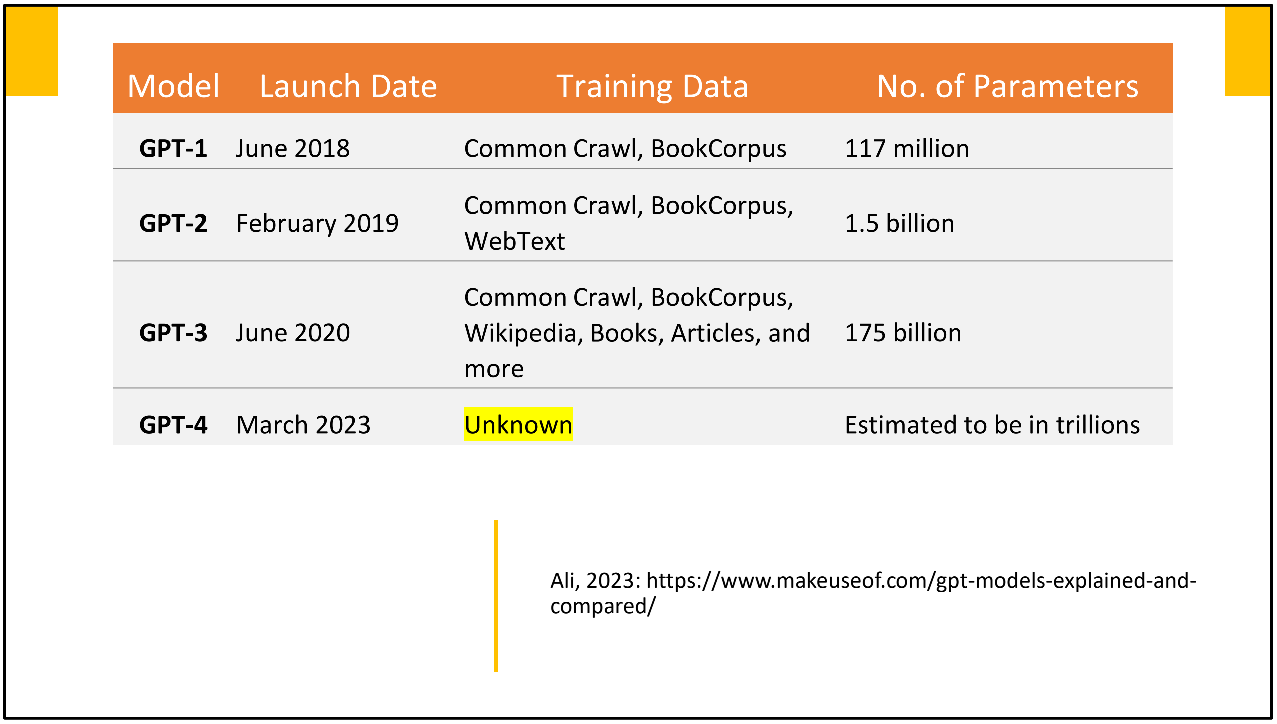
Hi, My name is Nat Gustafson-Sundell. I'm a Collections Librarian at Minnesota State University Mankato (MNSU), a regional, public, comprehensive university of about 14,000 students.

Today, I will present the steps we have taken to develop a prototype AI assistant for license review. I'll explain our criteria for the selection of an AI tool for this project. We reviewed ChatGPT, Claude 2, Bard, and PDF readers. My goal was to develop an initial prototype in a Jupyter Notebook environment so I could easily re-load context information, including a license checklist, but I'll explain why I revised this goal, instead to linger over license review interactions with ChatBots. I'll discuss early results, demonstrate example license review interactions, and outline my next steps.

This type of project would be easy for others to customize to meet their own needs, so, in this video, I won't dwell on licensing as such. Instead, I'll focus on the implementation issues.

link.mnsu.edu/ailicensedeck

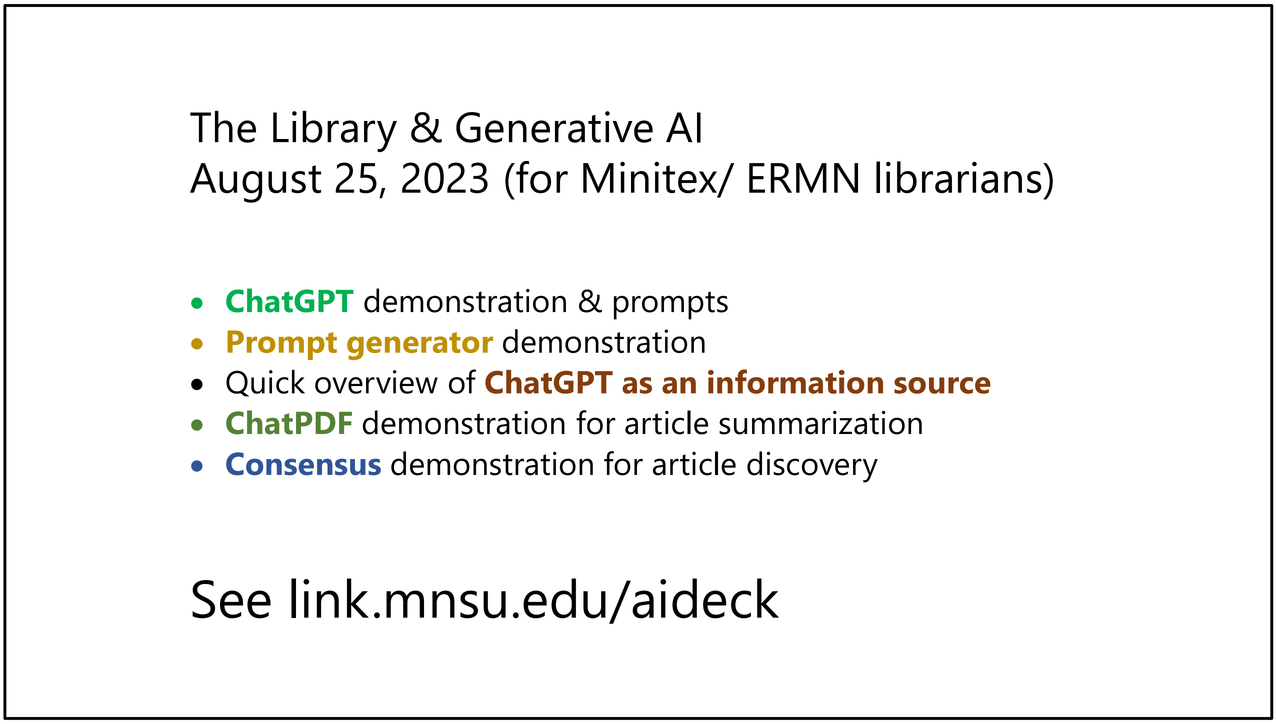
Sometimes it can be difficult to catch everything in a video. If you’d like to download my slides, as well as my notes, you can find them at “link.mnsu.edu/ailicensedeck”



Model	Launch Date	Training Data	No. of Parameters
GPT-1	June 2018	Common Crawl, BookCorpus	117 million
GPT-2	February 2019	Common Crawl, BookCorpus, WebText	1.5 billion
GPT-3	June 2020	Common Crawl, BookCorpus, Wikipedia, Books, Articles, and more	175 billion
GPT-4	March 2023	Unknown	Estimated to be in trillions

Ali, 2023: https://www.makeuseof.com/gpt-models-explained-and-compared/

I will describe the basic idea for my project, but first, I should provide some background. Hopefully, you are familiar with ChatBots, such as ChatGPT. At their most basic, ChatBots generate text responses to natural language queries. They have been trained on enormous quantities of text, including crawls of the internet over many years. ChatBots don't have any long-term memory, although the ChatBot interface might retain old conversations.

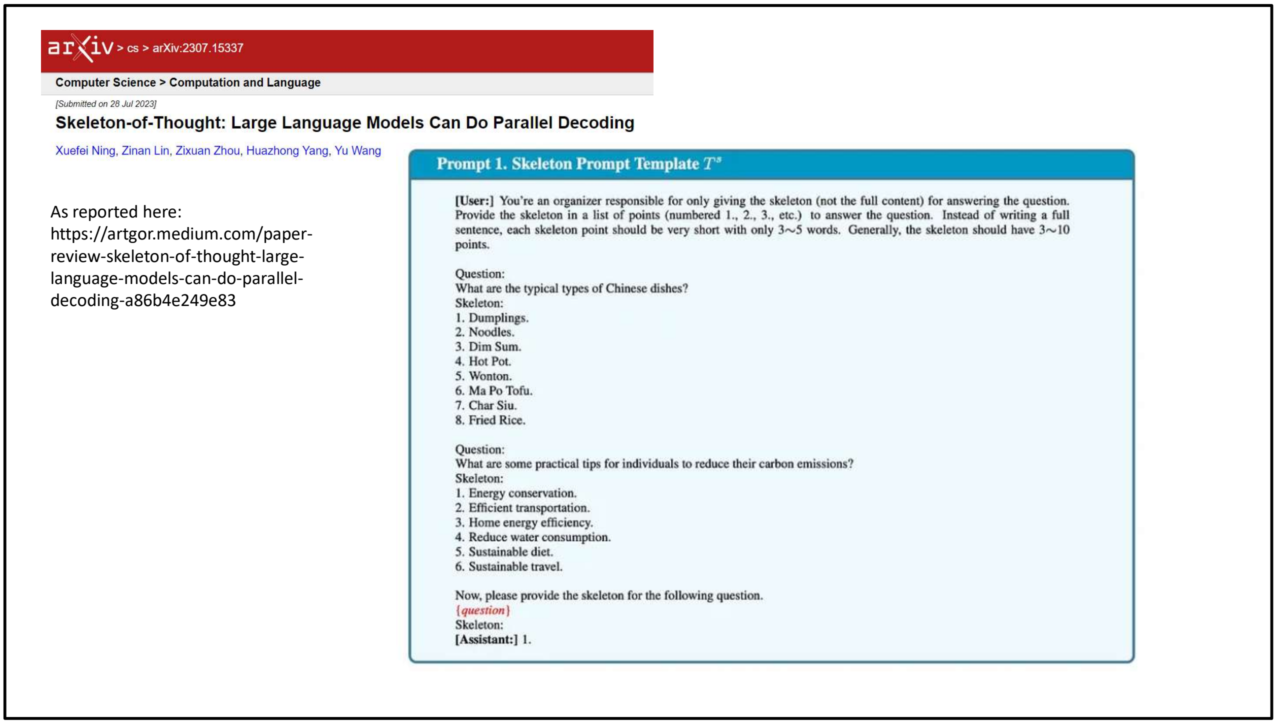


The Library & Generative AI
August 25, 2023 (for Minitex/ ERMN librarians)

- **ChatGPT** demonstration & prompts
- **Prompt generator** demonstration
- Quick overview of **ChatGPT as an information source**
- **ChatPDF** demonstration for article summarization
- **Consensus** demonstration for article discovery

See link.mnsu.edu/aideck

By the way, if you haven’t had much time yet to investigate ChatBots, you might want to take a look at a presentation I gave to librarians in my region. The slides are available at “link.mnsu.edu/aideck.” It was an introductory presentation covering basic prompting and a couple of other AI tools. I also incorporated a section on the limits of ChatGPT as an information source which I prepared for a professional development session to returning faculty through our University’s Center for Excellence in Teaching & Learning.



arXiv > cs > arXiv:2307.15337

Computer Science > Computation and Language

[Submitted on 28 Jul 2023]

Skeleton-of-Thought: Large Language Models Can Do Parallel Decoding

Xuefei Ning, Zinan Lin, Zixuan Zhou, Huazhong Yang, Yu Wang

As reported here: https://artgor.medium.com/paper-review-skeleton-of-thought-large-language-models-can-do-parallel-decoding-a86b4e249e83

Prompt 1. Skeleton Prompt Template T^s

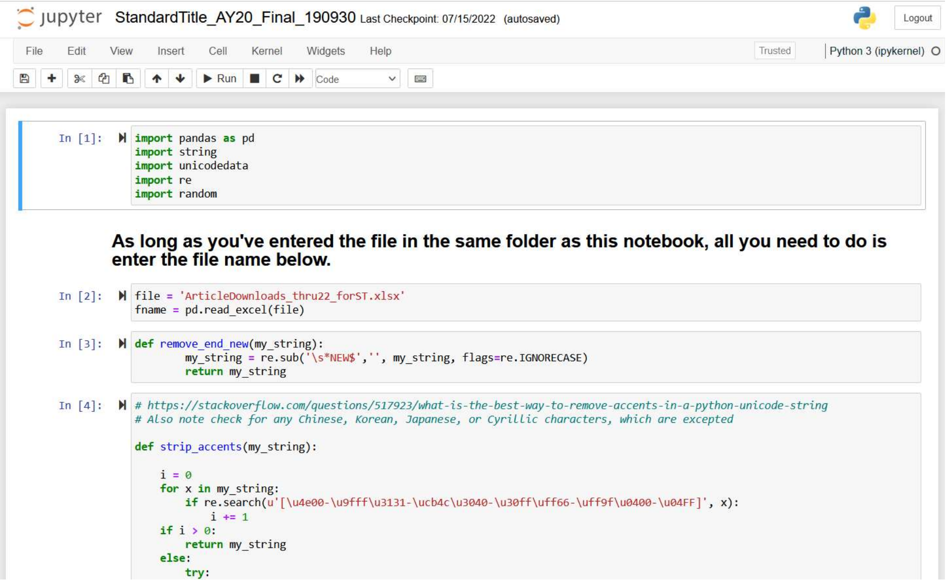
[User:] You're an organizer responsible for only giving the skeleton (not the full content) for answering the question. Provide the skeleton in a list of points (numbered 1., 2., 3., etc.) to answer the question. Instead of writing a full sentence, each skeleton point should be very short with only 3∼5 words. Generally, the skeleton should have 3∼10 points.

Question:
What are the typical types of Chinese dishes?
Skeleton:
1. Dumplings.
2. Noodles.
3. Dim Sum.
4. Hot Pot.
5. Wonton.
6. Ma Po Tofu.
7. Char Siu.
8. Fried Rice.

Question:
What are some practical tips for individuals to reduce their carbon emissions?
Skeleton:
1. Energy conservation.
2. Efficient transportation.
3. Home energy efficiency.
4. Reduce water consumption.
5. Sustainable diet.
6. Sustainable travel.

Now, please provide the skeleton for the following question.
{*question*}
Skeleton:
[Assistant:] 1.

When prompting ChatBots, it's possible to engineer very specific kinds of responses in an array of formats. If you've spent anytime studying prompt engineering, you've probably run across approaches such as chain-of-thought and skeleton-of-thought prompts, preference maps, and so on. The point of these approaches is to define clearly what is expected of the ChatBot, so that the ChatBot responds with the right kind of information, in the right format, with the right voice, etc.


jupyter StandardTitle_AY20_Final_190930 Last Checkpoint: 07/15/2022 (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code

```
In [1]:  import pandas as pd
         import string
         import unicodedata
         import re
         import random
```

As long as you've entered the file in the same folder as this notebook, all you need to do is enter the file name below.

```
In [2]:  file = 'ArticleDownloads_thru22_forST.xlsx'
         fname = pd.read_excel(file)
```

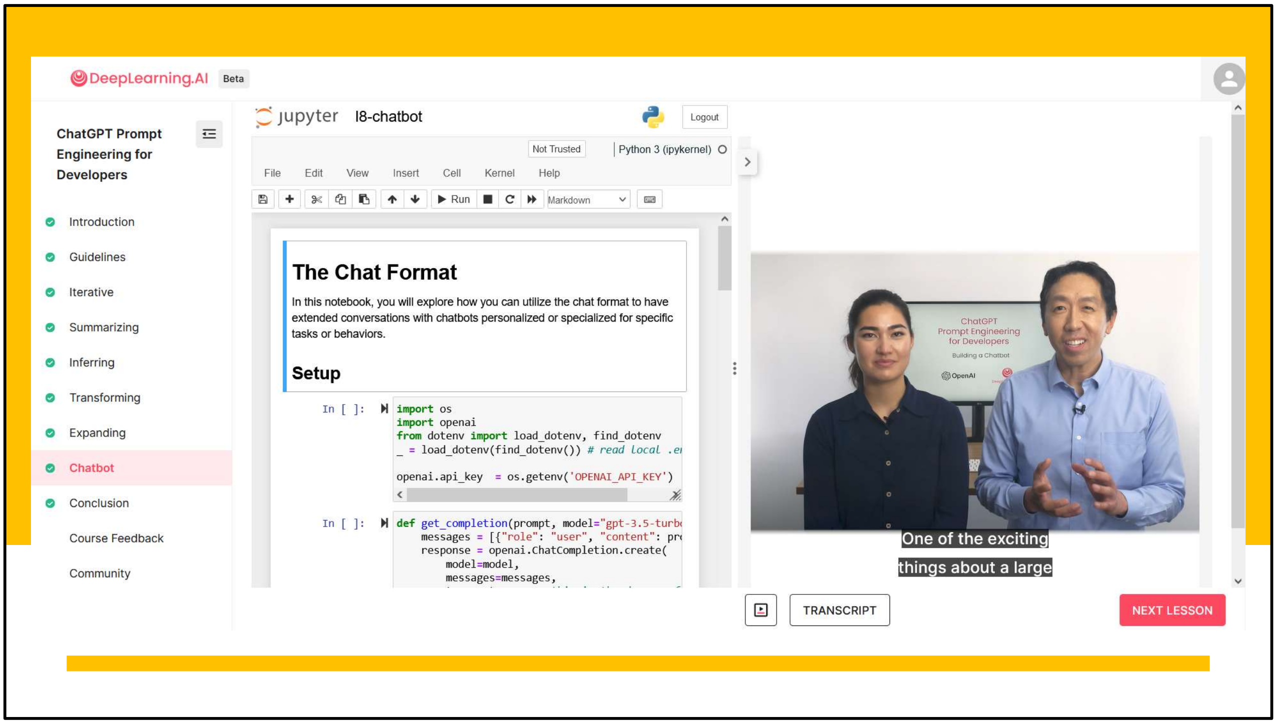
```
In [3]:  def remove_end_new(my_string):
                 my_string = re.sub('\s*NEW$','', my_string, flags=re.IGNORECASE)
                 return my_string
```

```
In [4]:  # https://stackoverflow.com/questions/517923/what-is-the-best-way-to-remove-accents-in-a-python-unicode-string
         # Also note check for any Chinese, Korean, Japanese, or Cyrillic characters, which are excepted

         def strip_accents(my_string):

             i = 0
             for x in my_string:
                 if re.search(u'[\u4e00-\u9fff\u3131-\ucb4c\u3040-\u30ff\uff66-\uff9f\u0400-\u04FF]', x):
                     i += 1
             if i > 0:
                 return my_string
             else:
                 try:
```

It is also possible to interact with ChatBots through a Jupyter Notebook. A Jupyter Notebook is an environment for programming, initially developed, I think, so that code and data could be shared for the replication of research. On the screen, you can see the first few cells of a Jupyter Notebook I developed several years ago for data processing. A Jupyter Notebook can be a repository for consistent inputs to a ChatBot. So, for example, you could create a preference map to load every time you use the notebook. A preference map is basically just a list of settings and expectations.



The Chat Format

In this notebook, you will explore how you can utilize the chat format to have extended conversations with chatbots personalized or specialized for specific tasks or behaviors.

Setup

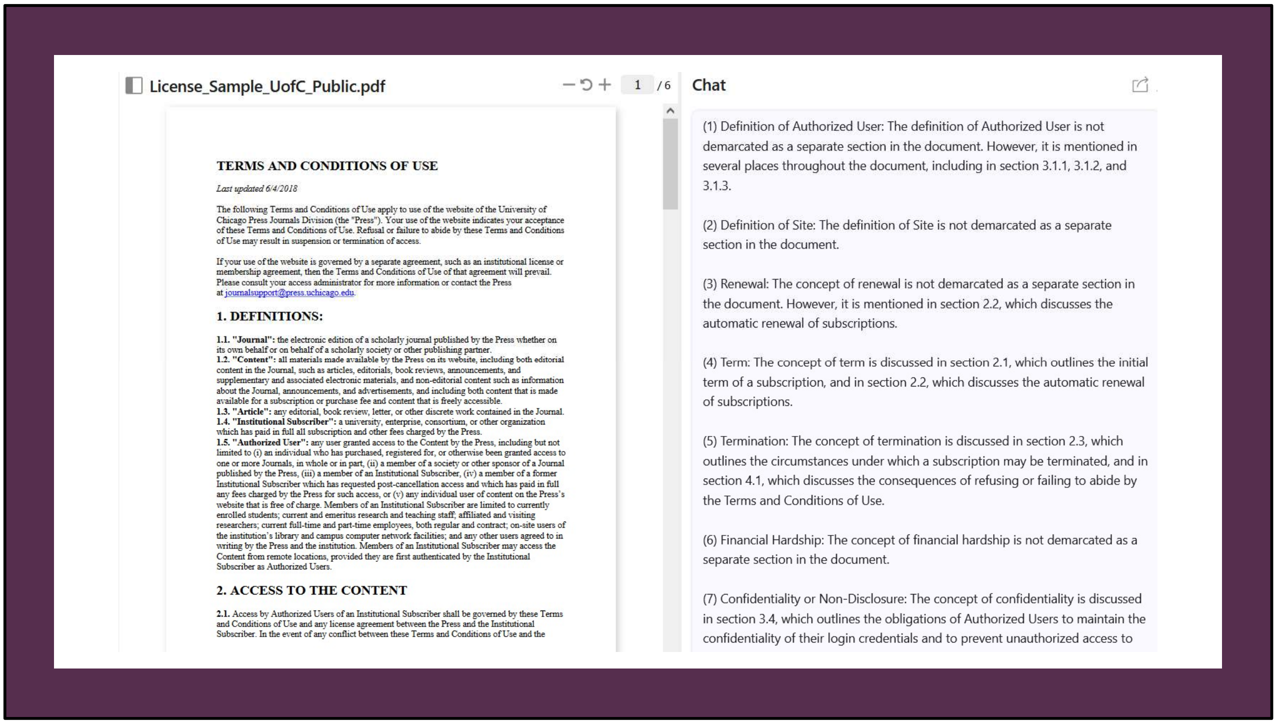
```
In [ ]:  import os
         import openai
         from dotenv import load_dotenv, find_dotenv
         _ = load_dotenv(find_dotenv()) # read local .e

         openai.api_key  = os.getenv('OPENAI_API_KEY')
```

```
In [ ]:  def get_completion(prompt, model="gpt-3.5-turb
             messages = [{"role": "user", "content": pr
             response = openai.ChatCompletion.create(
                 model=model,
                 messages=messages,
```

One of the exciting
things about a large

My inspiration to use Jupyter Notebook to interact with a ChatBot came from an online class I took over the summer, entitled “ChatGPT Prompt Engineering for Developers.” I learned that it’s possible to interact with ChatGPT from within Jupyter Notebook using an API key. An API key is just a programmatic means of access to the ChatBot.



TERMS AND CONDITIONS OF USE

Last updated 6/4/2018

The following Terms and Conditions of Use apply to use of the website of the University of Chicago Press Journals Division (the "Press"). Your use of the website indicates your acceptance of these Terms and Conditions of Use. Refusal or failure to abide by these Terms and Conditions of Use may result in suspension or termination of access.

If your use of the website is governed by a separate agreement, such as an institutional license or membership agreement, then the Terms and Conditions of Use of that agreement will prevail. Please consult your access administrator for more information or contact the Press at journalsupport@press.uchicago.edu.

1. DEFINITIONS:

1.1. "Journal": the electronic edition of a scholarly journal published by the Press whether on its own behalf or on behalf of a scholarly society or other publishing partner.
1.2. "Content": all materials made available by the Press on its website, including both editorial content in the Journal, such as articles, editorials, book reviews, announcements, and supplementary and associated electronic materials, and non-editorial content such as information about the Journal, announcements, and advertisements, and including both content that is made available for a subscription or purchase fee and content that is freely accessible.
1.3. "Article": any editorial, book review, letter, or other discrete work contained in the Journal.
1.4. "Institutional Subscriber": a university, enterprise, consortium, or other organization which has paid in full all subscription and other fees charged by the Press.
1.5. "Authorized User": any user granted access to the Content by the Press, including but not limited to (i) an individual who has purchased, registered for, or otherwise been granted access to one or more Journals, in whole or in part, (ii) a member of a society or other sponsor of a Journal published by the Press, (iii) a member of an Institutional Subscriber, (iv) a member of a former Institutional Subscriber which has requested post-cancellation access and which has paid in full any fees charged by the Press for such access, or (v) any individual user of content on the Press's website that is free of charge. Members of an Institutional Subscriber are limited to currently enrolled students; current and emeritus research and teaching staff; affiliated and visiting researchers; current full-time and part-time employees, both regular and contract; on-site users of the institution's library and campus computer network facilities; and any other users agreed to in writing by the Press and the institution. Members of an Institutional Subscriber may access the Content from remote locations, provided they are first authenticated by the Institutional Subscriber as Authorized Users.

2. ACCESS TO THE CONTENT

2.1. Access by Authorized Users of an Institutional Subscriber shall be governed by these Terms and Conditions of Use and any license agreement between the Press and the Institutional Subscriber. In the event of any conflict between these Terms and Conditions of Use and the

(1) Definition of Authorized User: The definition of Authorized User is not demarcated as a separate section in the document. However, it is mentioned in several places throughout the document, including in section 3.1.1, 3.1.2, and 3.1.3.

(2) Definition of Site: The definition of Site is not demarcated as a separate section in the document.

(3) Renewal: The concept of renewal is not demarcated as a separate section in the document. However, it is mentioned in section 2.2, which discusses the automatic renewal of subscriptions.

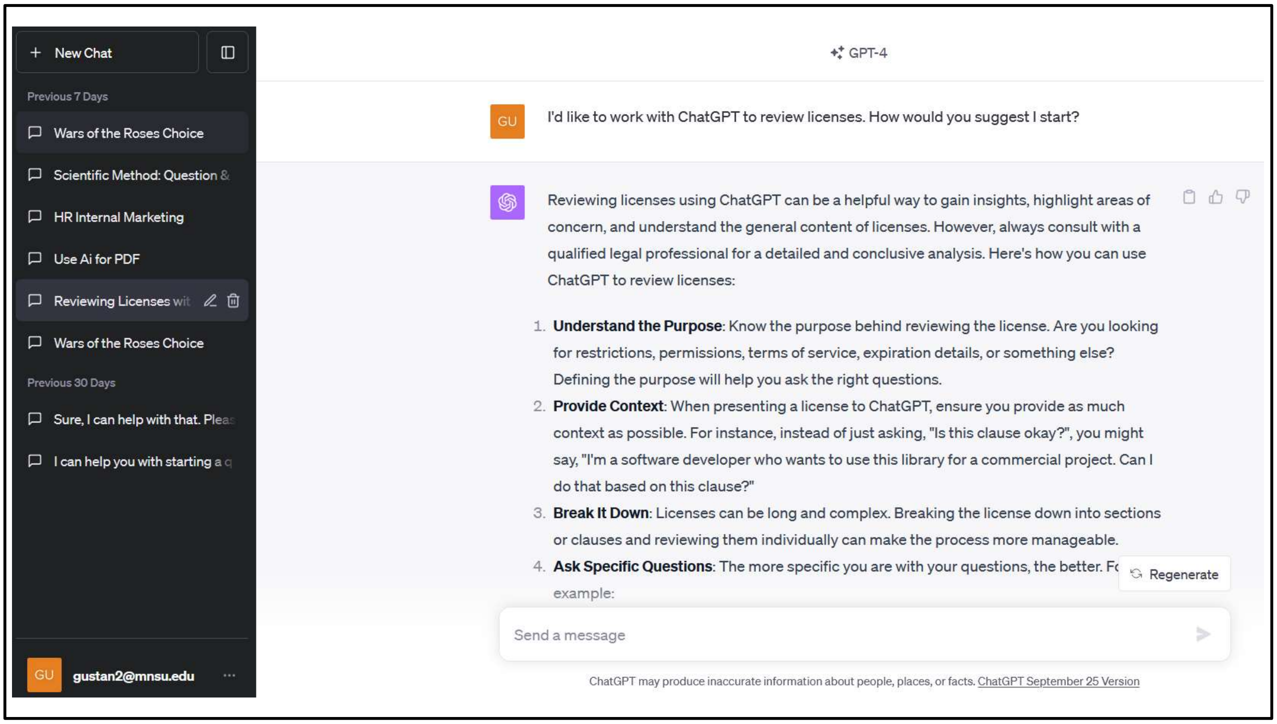
(4) Term: The concept of term is discussed in section 2.1, which outlines the initial term of a subscription, and in section 2.2, which discusses the automatic renewal of subscriptions.

(5) Termination: The concept of termination is discussed in section 2.3, which outlines the circumstances under which a subscription may be terminated, and in section 4.1, which discusses the consequences of refusing or failing to abide by the Terms and Conditions of Use.

(6) Financial Hardship: The concept of financial hardship is not demarcated as a separate section in the document.

(7) Confidentiality or Non-Disclosure: The concept of confidentiality is discussed in section 3.4, which outlines the obligations of Authorized Users to maintain the confidentiality of their login credentials and to prevent unauthorized access to

So then, my initial idea was pretty straightforward. I would create a Jupyter Notebook to interact with a ChatBot to review licenses. I would adapt our licensing toolkit as a checklist of terms which I could re-load every time I used the notebook. The Jupyter Notebook would only require a few pieces, first the API key, second a checklist of terms, third a preference map or set of instructions for the responses. For any given session, I'd also load a license. After loading the license, I'd simply interact with the ChatBot to review terms. Over time, I could refine the checklist and preferences. This process would not replace my own review of the license, but it could complement that review. The new process might help me perform more consistent, careful reviews.



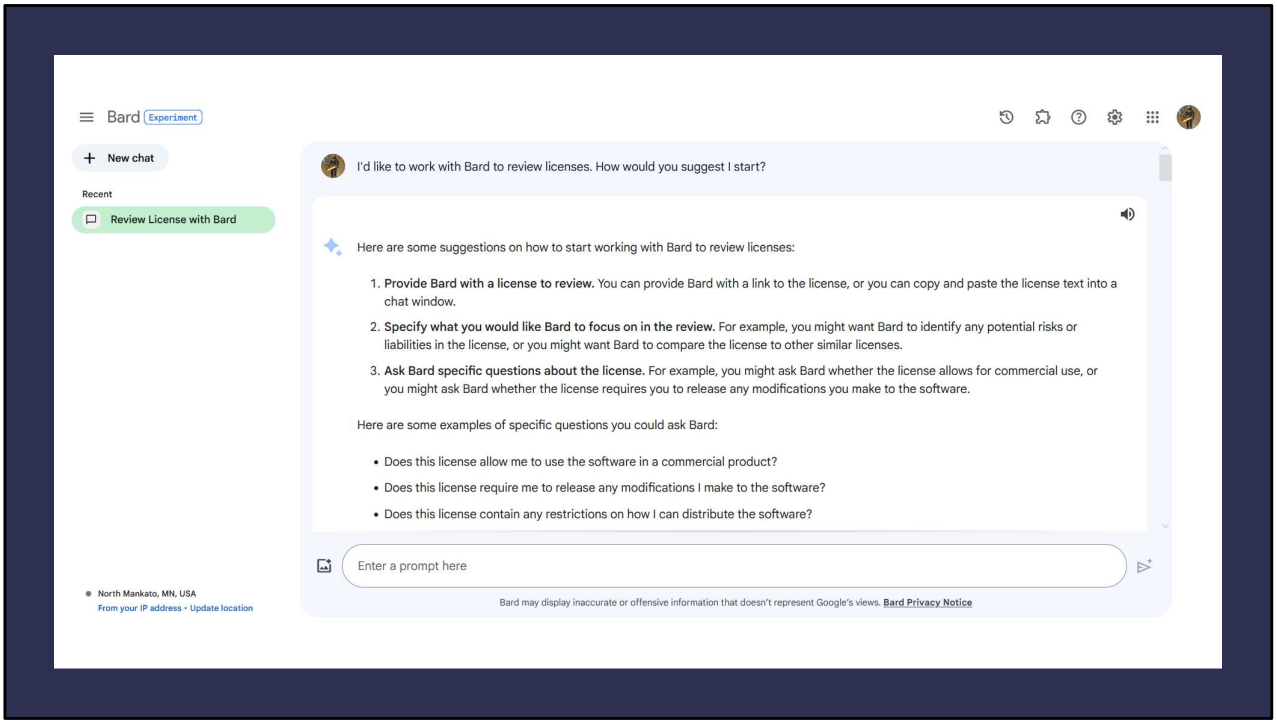
I'd like to work with ChatGPT to review licenses. How would you suggest I start?

Reviewing licenses using ChatGPT can be a helpful way to gain insights, highlight areas of concern, and understand the general content of licenses. However, always consult with a qualified legal professional for a detailed and conclusive analysis. Here's how you can use ChatGPT to review licenses:

1. **Understand the Purpose**: Know the purpose behind reviewing the license. Are you looking for restrictions, permissions, terms of service, expiration details, or something else? Defining the purpose will help you ask the right questions.
2. **Provide Context**: When presenting a license to ChatGPT, ensure you provide as much context as possible. For instance, instead of just asking, "Is this clause okay?", you might say, "I'm a software developer who wants to use this library for a commercial project. Can I do that based on this clause?"
3. **Break It Down**: Licenses can be long and complex. Breaking the license down into sections or clauses and reviewing them individually can make the process more manageable.
4. **Ask Specific Questions**: The more specific you are with your questions, the better. Fo example:

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT September 25 Version

I began the project by investigating which of the leading ChatBots I should use to implement the project. This investigation was very simple. I just asked ChatGPT, Claude, and then Bard how I would go about working with them. On the screen, you can see the start of my conversation with ChatGPT.



I'd like to work with Bard to review licenses. How would you suggest I start?

Here are some suggestions on how to start working with Bard to review licenses:

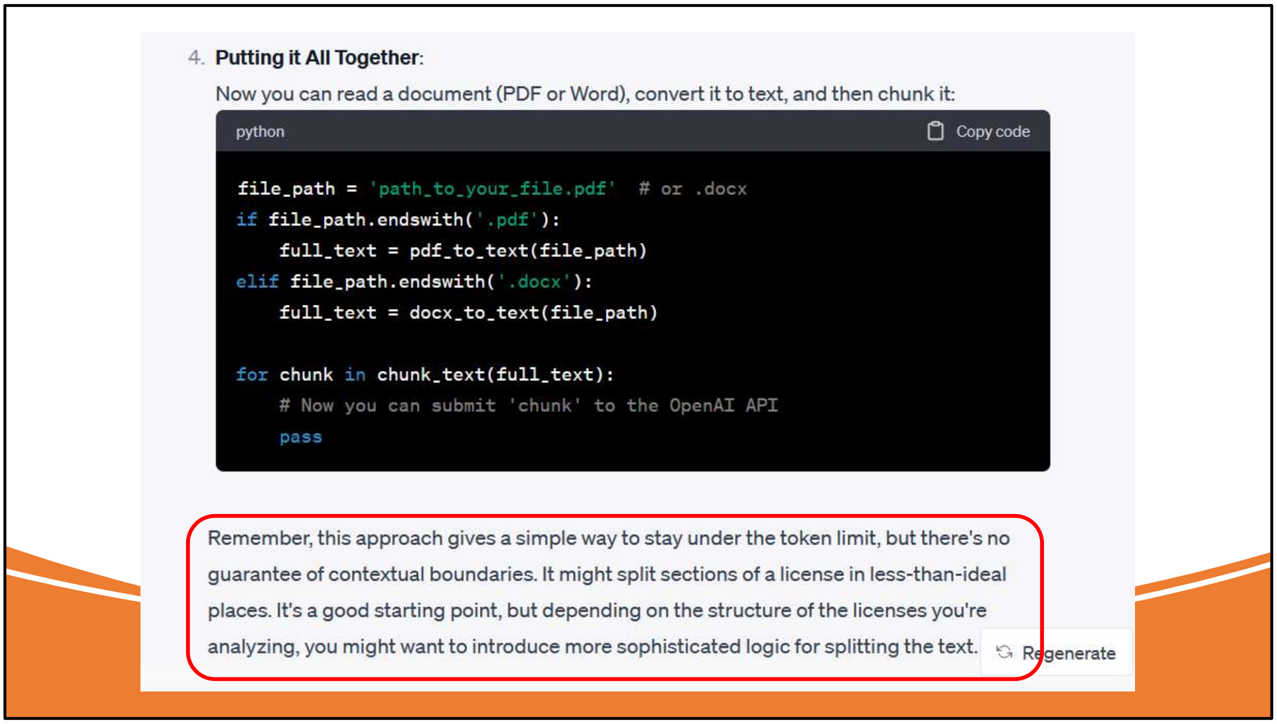
1. **Provide Bard with a license to review.** You can provide Bard with a link to the license, or you can copy and paste the license text into a chat window.
2. **Specify what you would like Bard to focus on in the review.** For example, you might want Bard to identify any potential risks or liabilities in the license, or you might want Bard to compare the license to other similar licenses.
3. **Ask Bard specific questions about the license.** For example, you might ask Bard whether the license allows for commercial use, or you might ask Bard whether the license requires you to release any modifications you make to the software.

Here are some examples of specific questions you could ask Bard:

- Does this license allow me to use the software in a commercial product?
- Does this license require me to release any modifications I make to the software?
- Does this license contain any restrictions on how I can distribute the software?

Bard may display inaccurate or offensive information that doesn't represent Google's views. **Bard Privacy Notice**

Basically, they each said they'd love to help me and they each provided information about how to proceed, including suggested code to set up Jupyter Notebook. Here you can see the same start to the conversation with Bard.



4. **Putting it All Together**:

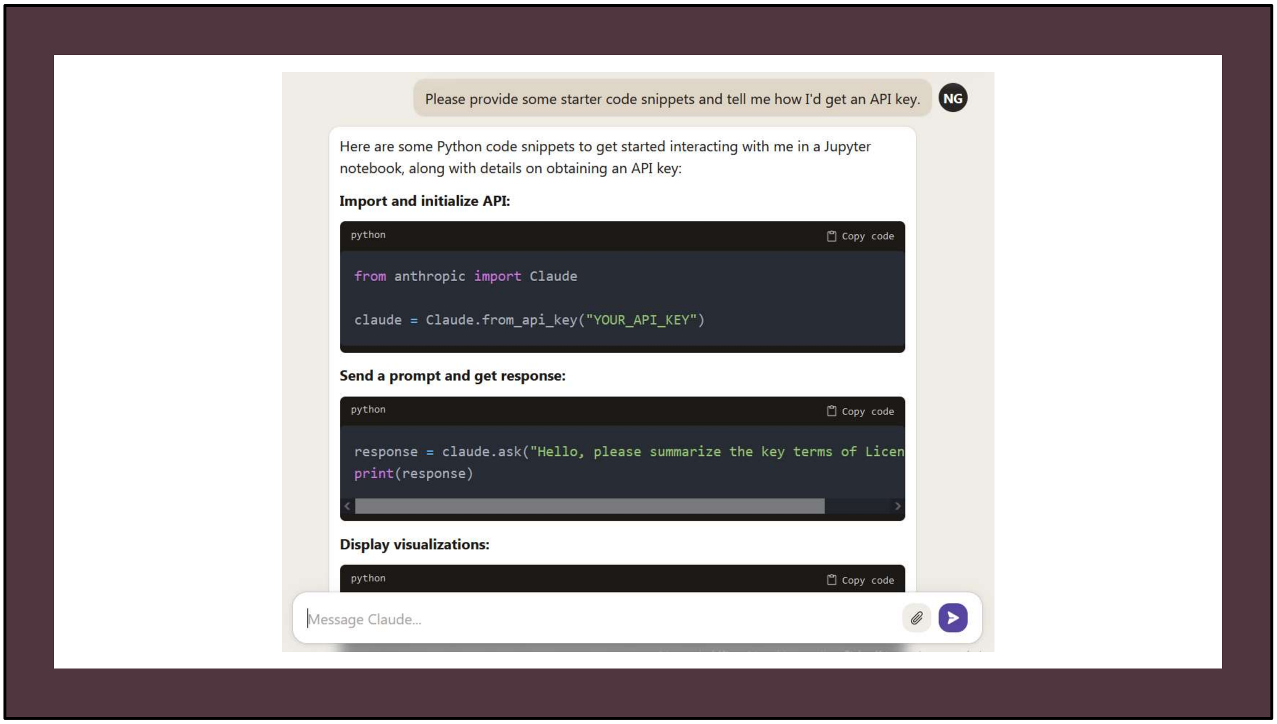
Now you can read a document (PDF or Word), convert it to text, and then chunk it:

```python
file_path = 'path_to_your_file.pdf'  # or .docx
if file_path.endswith('.pdf'):
    full_text = pdf_to_text(file_path)
elif file_path.endswith('.docx'):
    full_text = docx_to_text(file_path)

for chunk in chunk_text(full_text):
    # Now you can submit 'chunk' to the OpenAI API
    pass
```

Remember, this approach gives a simple way to stay under the token limit, but there's no guarantee of contextual boundaries. It might split sections of a license in less-than-ideal places. It's a good starting point, but depending on the structure of the licenses you're analyzing, you might want to introduce more sophisticated logic for splitting the text.

These 3 ChatBots are definitely not all the same. They have distinct advantages and disadvantages for a project like mine. My first impression was that ChatGPT would not be the best option for my project. At this time, it is only possible to input PDFs into the subscription version of ChatGPT using plug-ins and then only through a URL. Text inputs are limited in size, less than 2000 words. Although ChatGPT offered a work-around, the work-around was not optimal and ChatGPT admitted it, as you can see on the screen. Basically, I would have needed to break text inputs into chunks to enter them into ChatGPT.



Please provide some starter code snippets and tell me how I'd get an API key.

Here are some Python code snippets to get started interacting with me in a Jupyter notebook, along with details on obtaining an API key:

Import and initialize API:

```python
from anthropic import Claude

claude = Claude.from_api_key("YOUR_API_KEY")
```

Send a prompt and get response:

```python
response = claude.ask("Hello, please summarize the key terms of Licen
print(response)
```

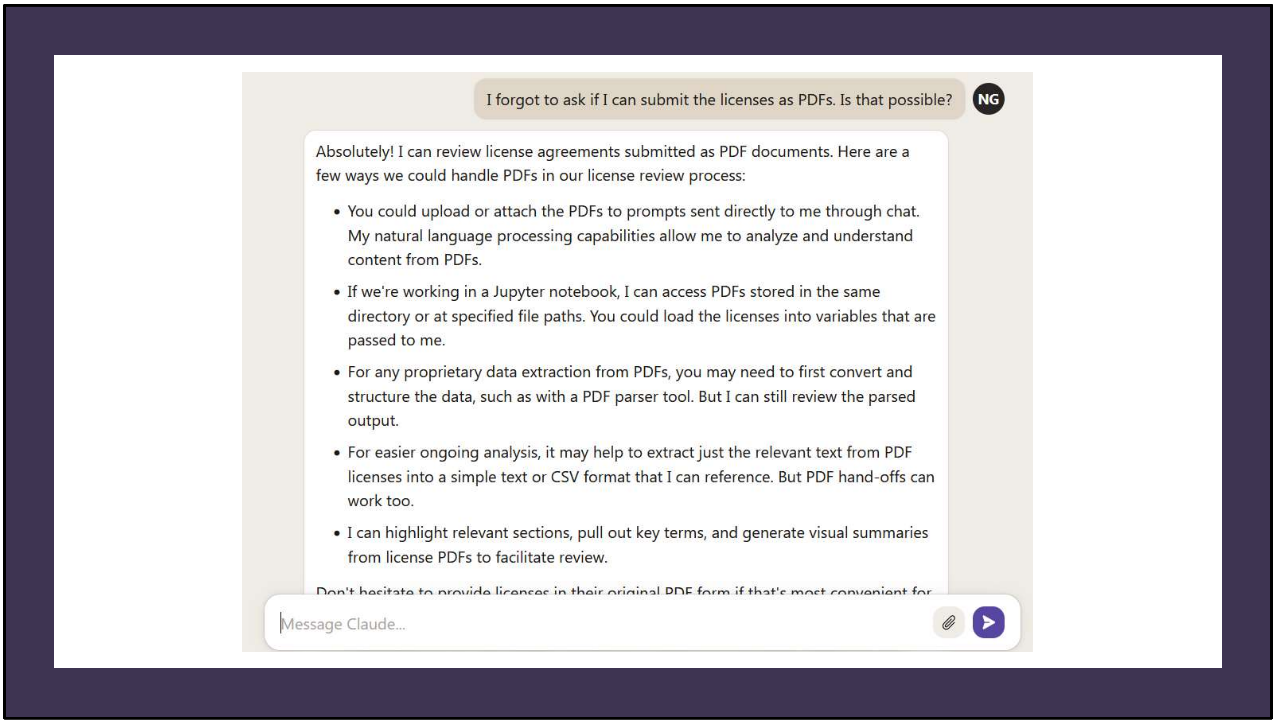
Display visualizations:

All of the ChatBots offered tips about how to implement the project. In all 3 cases, the ChatBots said I could access them via Jupyter Notebook using an API key. I don’t have time now to go through each of these conversations in detail, even though they were each pretty brief, but if you’re interested in reading the transcripts, I’ll provide the links on the next screen.

link.mnsu.edu/licensebard

link.mnsu.edu/licenseclaude

link.mnsu.edu/licensechatgpt



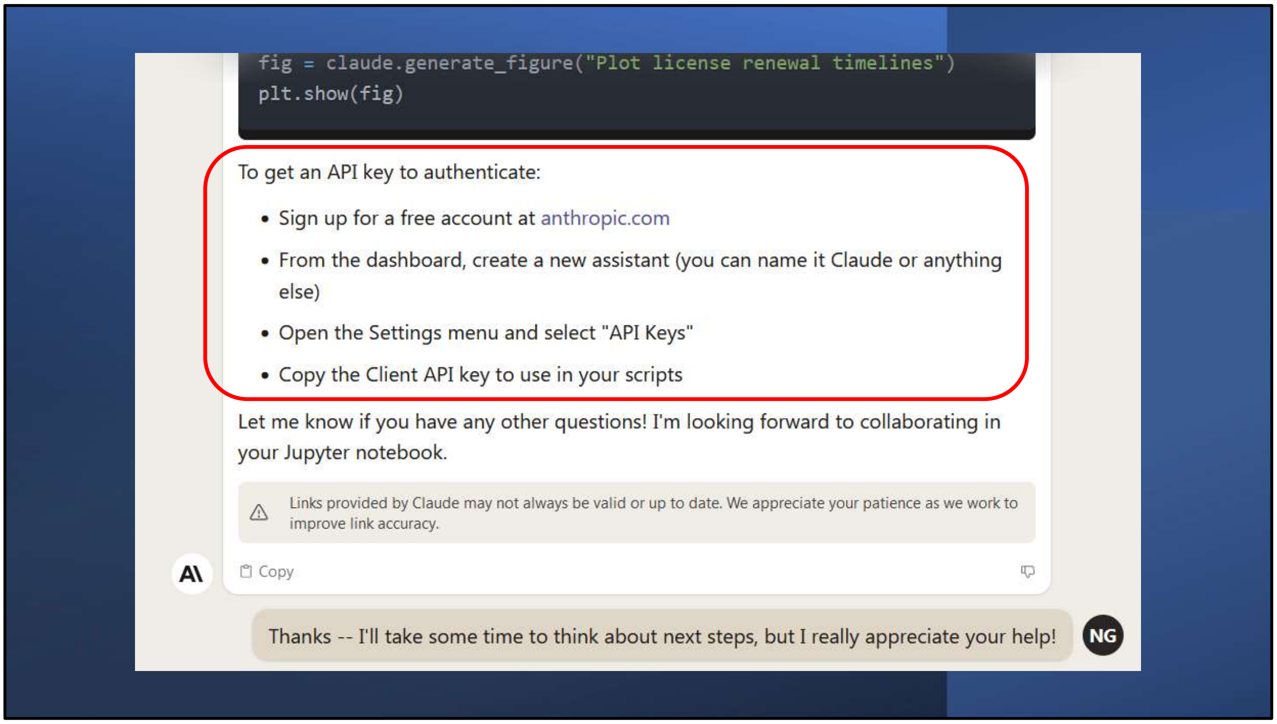
I forgot to ask if I can submit the licenses as PDFs. Is that possible?

Absolutely! I can review license agreements submitted as PDF documents. Here are a few ways we could handle PDFs in our license review process:

- You could upload or attach the PDFs to prompts sent directly to me through chat. My natural language processing capabilities allow me to analyze and understand content from PDFs.
- If we're working in a Jupyter notebook, I can access PDFs stored in the same directory or at specified file paths. You could load the licenses into variables that are passed to me.
- For any proprietary data extraction from PDFs, you may need to first convert and structure the data, such as with a PDF parser tool. But I can still review the parsed output.
- For easier ongoing analysis, it may help to extract just the relevant text from PDF licenses into a simple text or CSV format that I can reference. But PDF hand-offs can work too.
- I can highlight relevant sections, pull out key terms, and generate visual summaries from license PDFs to facilitate review.

Don't hesitate to provide licenses in their original PDF form if that's most convenient for

Claude seemed to me to be the best ChatBot to use. Claude makes it easy to upload PDFs. Claude's instructions for implementation were very clear. Claude also was really enthusiastic about the project. I've worked the most with ChatGPT in the past, and I like it, but Claude just seemed best for the purposes of this project.



```
fig = claude.generate_figure("Plot license renewal timelines")
plt.show(fig)
```

To get an API key to authenticate:

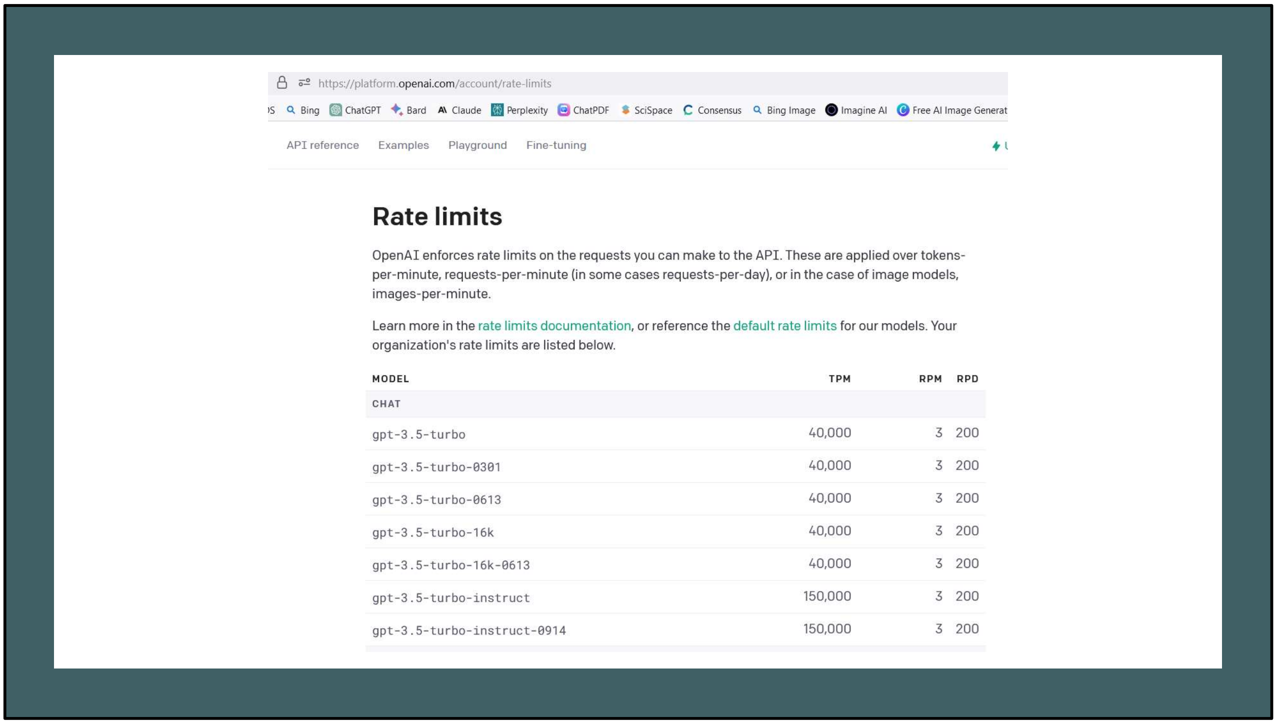
- Sign up for a free account at anthropic.com
- From the dashboard, create a new assistant (you can name it Claude or anything else)
- Open the Settings menu and select "API Keys"
- Copy the Client API key to use in your scripts

Let me know if you have any other questions! I'm looking forward to collaborating in your Jupyter notebook.

⚠ Links provided by Claude may not always be valid or up to date. We appreciate your patience as we work to improve link accuracy.

Thanks -- I'll take some time to think about next steps, but I really appreciate your help!

After getting a sense of how each of the ChatBots could work, I proceeded to practical steps, but here I ran into a problem. Although Claude had made it sound easy to get an API key, it turned out I couldn't. In fact, for both Claude and Bard, I have had to apply for an API key, which is a beta feature right now for each of these ChatBots. I haven't yet received a reply from either Anthropic or Google, the companies behind Claude and Bard respectively.



Rate limits

OpenAI enforces rate limits on the requests you can make to the API. These are applied over tokens-per-minute, requests-per-minute (in some cases requests-per-day), or in the case of image models, images-per-minute.

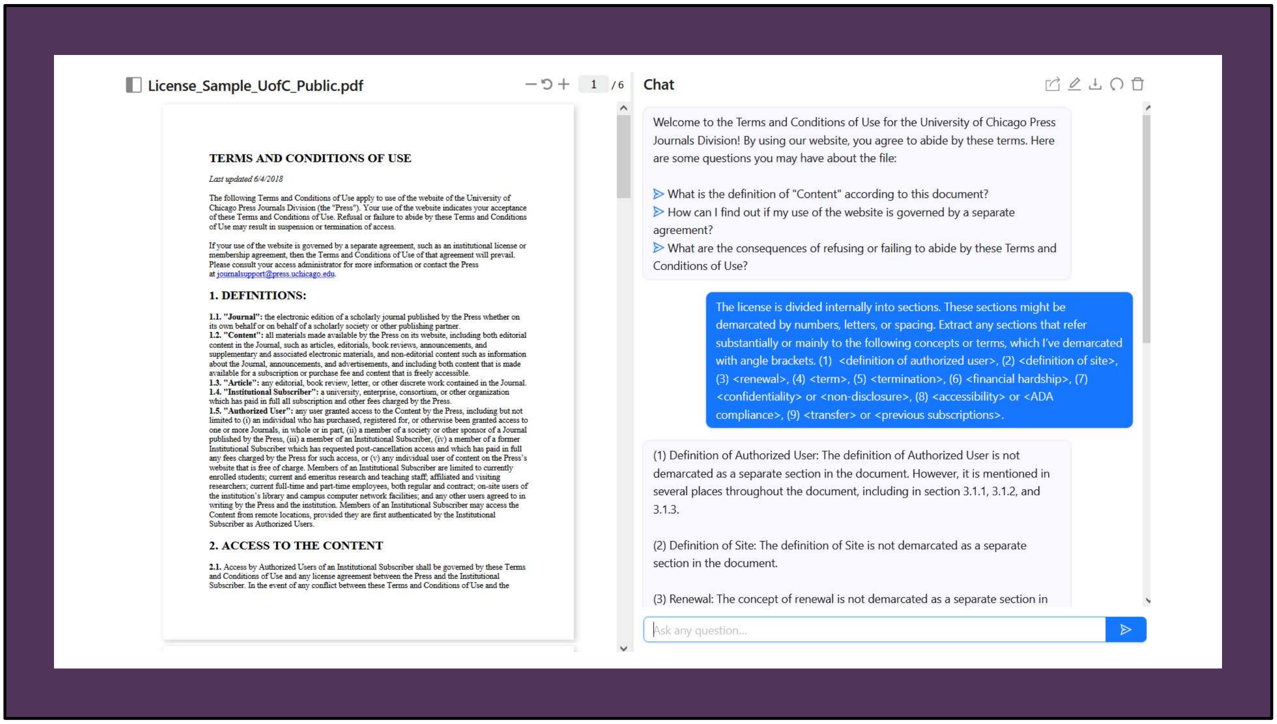
Learn more in the rate limits documentation, or reference the default rate limits for our models. Your organization's rate limits are listed below.

MODEL	TPM	RPM	RPD
CHAT			
gpt-3.5-turbo	40,000	3	200
gpt-3.5-turbo-0301	40,000	3	200
gpt-3.5-turbo-0613	40,000	3	200
gpt-3.5-turbo-16k	40,000	3	200
gpt-3.5-turbo-16k-0613	40,000	3	200
gpt-3.5-turbo-instruct	150,000	3	200
gpt-3.5-turbo-instruct-0914	150,000	3	200

It is possible for me to get an API key to ChatGPT, but only by paying more money. I'm currently a subscriber to ChatGPT+ at $20 per month. This version does not provide an API key. It's necessary to add a pay-as-you-go plan to another version. Before doing this, I wondered if there were any other options I hadn't tried?

Because I’ve been very busy on several projects, my thinking has been compartmentalized. It took me a little while to realize that I didn’t need to proceed to the full Jupyter Notebook implementation to make progress on my project. I could instead just interact with a ChatBot over a license. And actually, as I started doing so, I realized this probably should have been part of the plan all along – it’s important to test and refine one’s prompts. Later on, when I can proceed with the Jupyter Notebook implementation, I’ll be better prepared to design my inputs.

By the way, the image on the screen is meant to illustrate the notion of compartmentalized thinking. I prompted Bing Image Creator with the phrase “numerous small furry monsters in compartments.” I don’t know why I naturally thought of my thoughts as small furry monsters, but now that I see them on the screen, I know this is definitely what my thoughts look like…



TERMS AND CONDITIONS OF USE

Last updated 6/4/2018

The following Terms and Conditions of Use apply to use of the website of the University of Chicago Press Journals Division (the "Press"). Your use of the website indicates your acceptance of these Terms and Conditions of Use. Refusal or failure to abide by these Terms and Conditions of Use may result in suspension or termination of access.

If your use of the website is governed by a separate agreement, such as an institutional license or membership agreement, then the Terms and Conditions of Use of that agreement will prevail. Please consult your access administrator for more information or contact the Press at journalsupport@press.uchicago.edu.

1. DEFINITIONS:

1.1. "Journal": the electronic edition of a scholarly journal published by the Press whether on its own behalf or on behalf of a scholarly society or other publishing partner.
1.2. "Content": all materials made available by the Press on its website, including both editorial content in the Journal, such as articles, editorials, book reviews, announcements, and supplementary and associated electronic materials, and non-editorial content such as information about the Journal, announcements, and advertisements, and including both content that is made available for a subscription or purchase fee and content that is freely accessible.
1.3. "Article": any editorial, book review, letter, or other discrete work contained in the Journal.
1.4. "Institutional Subscriber": a university, enterprise, consortium, or other organization which has paid in full all subscription and other fees charged by the Press.
1.5. "Authorized User": any user granted access to the Content by the Press, including but not limited to (i) an individual who has purchased, registered for, or otherwise been granted access to one or more Journals, in whole or in part, (ii) a member of a society or other sponsor of a Journal published by the Press, (iii) a member of an Institutional Subscriber, (iv) a member of a former Institutional Subscriber which has requested post-cancellation access and which has paid in full any fees charged by the Press for such access, or (v) any individual user of content on the Press's website that is free of charge. Members of an Institutional Subscriber are limited to currently enrolled students; current and emeritus research and teaching staff; affiliated and visiting researchers; current full-time and part-time employees, both regular and contract; on-site users of the institution's library and campus computer network facilities; and any other users agreed to in writing by the Press and the institution. Members of an Institutional Subscriber may access the Content from remote locations, provided they are first authenticated by the Institutional Subscriber as Authorized Users.

2. ACCESS TO THE CONTENT

2.1. Access by Authorized Users of an Institutional Subscriber shall be governed by these Terms and Conditions of Use and any license agreement between the Press and the Institutional Subscriber. In the event of any conflict between these Terms and Conditions of Use and the

Welcome to the Terms and Conditions of Use for the University of Chicago Press Journals Division! By using our website, you agree to abide by these terms. Here are some questions you may have about the file:

▷ What is the definition of "Content" according to this document?
▷ How can I find out if my use of the website is governed by a separate agreement?
▷ What are the consequences of refusing or failing to abide by these Terms and Conditions of Use?

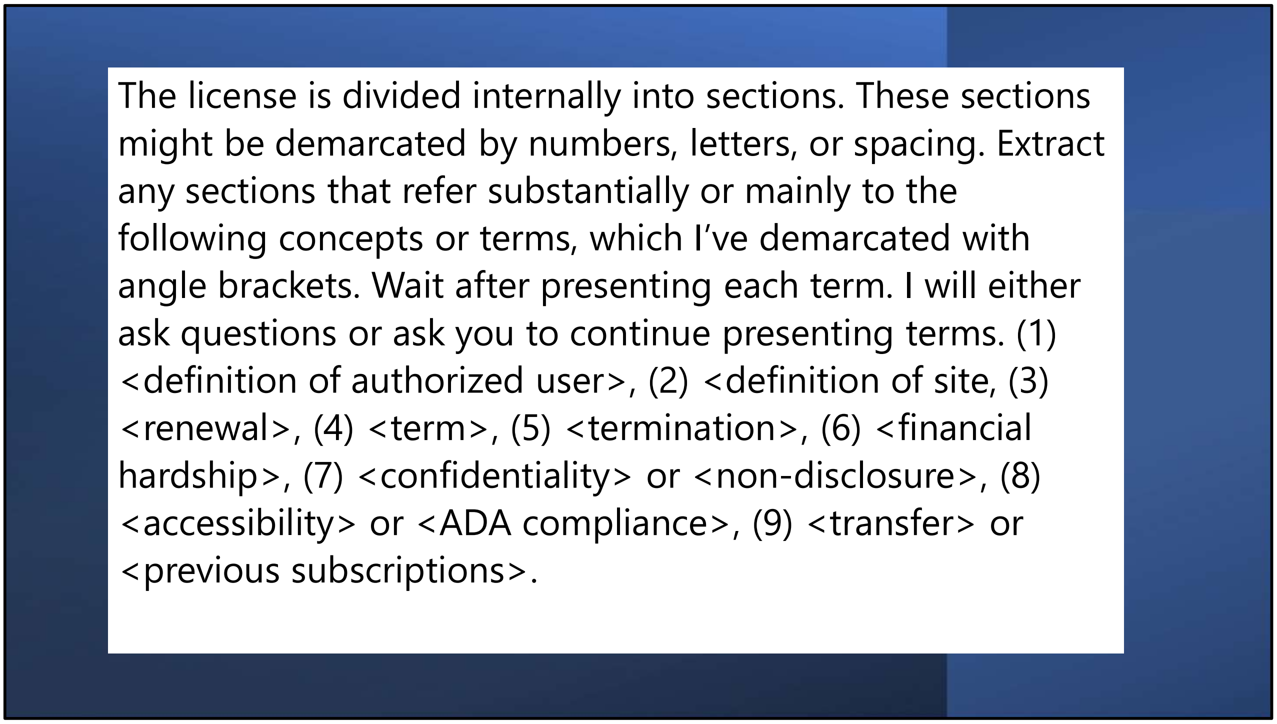
The license is divided internally into sections. These sections might be demarcated by numbers, letters, or spacing. Extract any sections that refer substantially or mainly to the following concepts or terms, which I've demarcated with angle brackets. (1) <definition of authorized user>, (2) <definition of site>, (3) <renewal>, (4) <term>, (5) <termination>, (6) <financial hardship>, (7) <confidentiality> or <non-disclosure>, (8) <accessibility> or <ADA compliance>, (9) <transfer> or <previous subscriptions>.

(1) Definition of Authorized User: The definition of Authorized User is not demarcated as a separate section in the document. However, it is mentioned in several places throughout the document, including in section 3.1.1, 3.1.2, and 3.1.3.

(2) Definition of Site: The definition of Site is not demarcated as a separate section in the document.

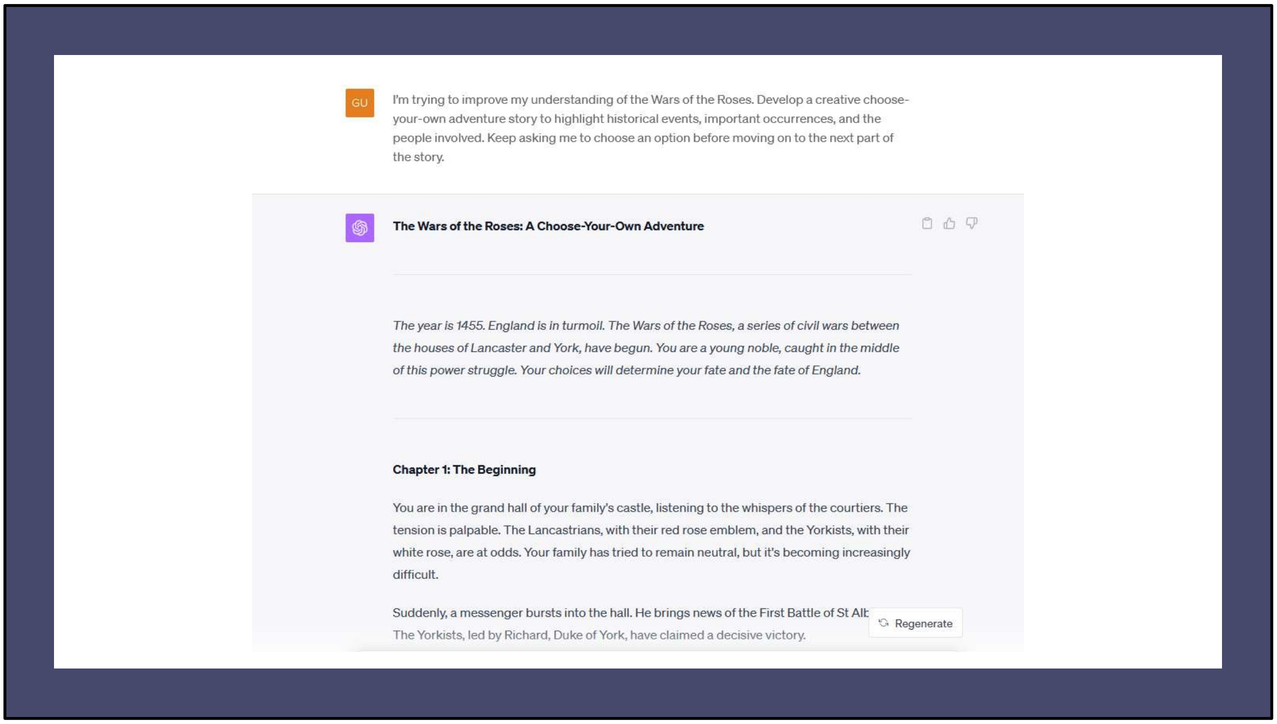
(3) Renewal: The concept of renewal is not demarcated as a separate section in

I started testing with ChatPDF. I was already familiar with this tool because it is popular for journal article summarization, an issue I addressed in a presentation to regional librarians, which I previously mentioned. Although ChatPDF is optimized for PDFs, it's actually not a great tool for my license review project. ChatPDF is limited to prompts of about 90 words or less. The limit on words made the prompts quite lean.



The license is divided internally into sections. These sections might be demarcated by numbers, letters, or spacing. Extract any sections that refer substantially or mainly to the following concepts or terms, which I've demarcated with angle brackets. Wait after presenting each term. I will either ask questions or ask you to continue presenting terms. (1) <definition of authorized user>, (2) <definition of site, (3) <renewal>, (4) <term>, (5) <termination>, (6) <financial hardship>, (7) <confidentiality> or <non-disclosure>, (8) <accessibility> or <ADA compliance>, (9) <transfer> or <previous subscriptions>.

On the screen, you can see the first draft prompt I developed when I was still working with ChatPDF. The prompt incorporates some preferences for the format and delivery of the response, but the approach is far simpler than I initially envisioned because I wrote it to conform with the limits of ChatPDF. In Jupyter Notebook, I probably would have started with something much more verbose, so I'm glad I had to rethink the project because I might have frontloaded a lot of effort trying to get ideal responses, but the reality might not have been better than I can get with simple interactions.



I'm trying to improve my understanding of the Wars of the Roses. Develop a creative choose-your-own adventure story to highlight historical events, important occurrences, and the people involved. Keep asking me to choose an option before moving on to the next part of the story.

The Wars of the Roses: A Choose-Your-Own Adventure

The year is 1455. England is in turmoil. The Wars of the Roses, a series of civil wars between the houses of Lancaster and York, have begun. You are a young noble, caught in the middle of this power struggle. Your choices will determine your fate and the fate of England.

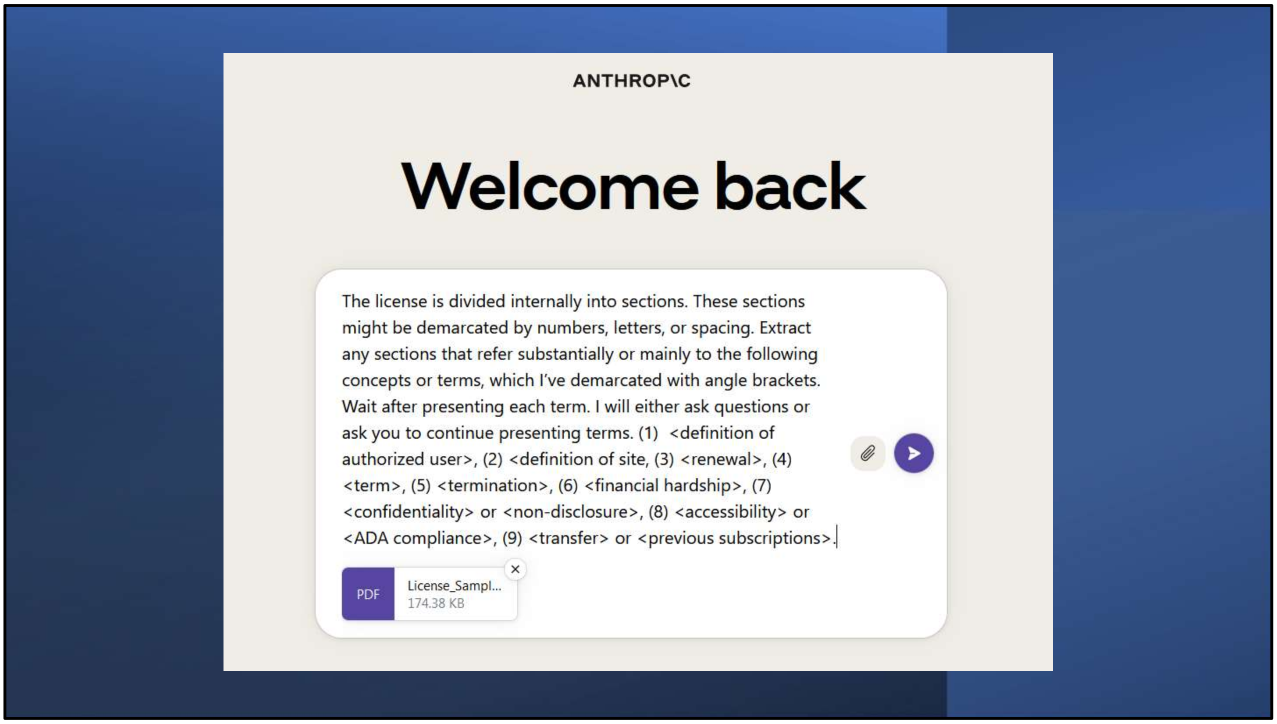
Chapter 1: The Beginning

You are in the grand hall of your family's castle, listening to the whispers of the courtiers. The tension is palpable. The Lancastrians, with their red rose emblem, and the Yorkists, with their white rose, are at odds. Your family has tried to remain neutral, but it's becoming increasingly difficult.

Suddenly, a messenger bursts into the hall. He brings news of the First Battle of St Alb

The Yorkists, led by Richard, Duke of York, have claimed a decisive victory.

After my initial tests with ChatPDF, I paused the project to prepare an internal presentation for my colleagues on a variety of AI topics. I wanted to share what I was learning from this project, as well as other AI projects – one of which was a 5-hour workshop for high school students, which was way more fun, as you can see on the screen. One reason for the pause was that I wanted to solicit feedback and hear concerns about the license review project before getting too far ahead of myself. I should mention in this context that I have been careful to test only with licenses that are publicly available on the internet.



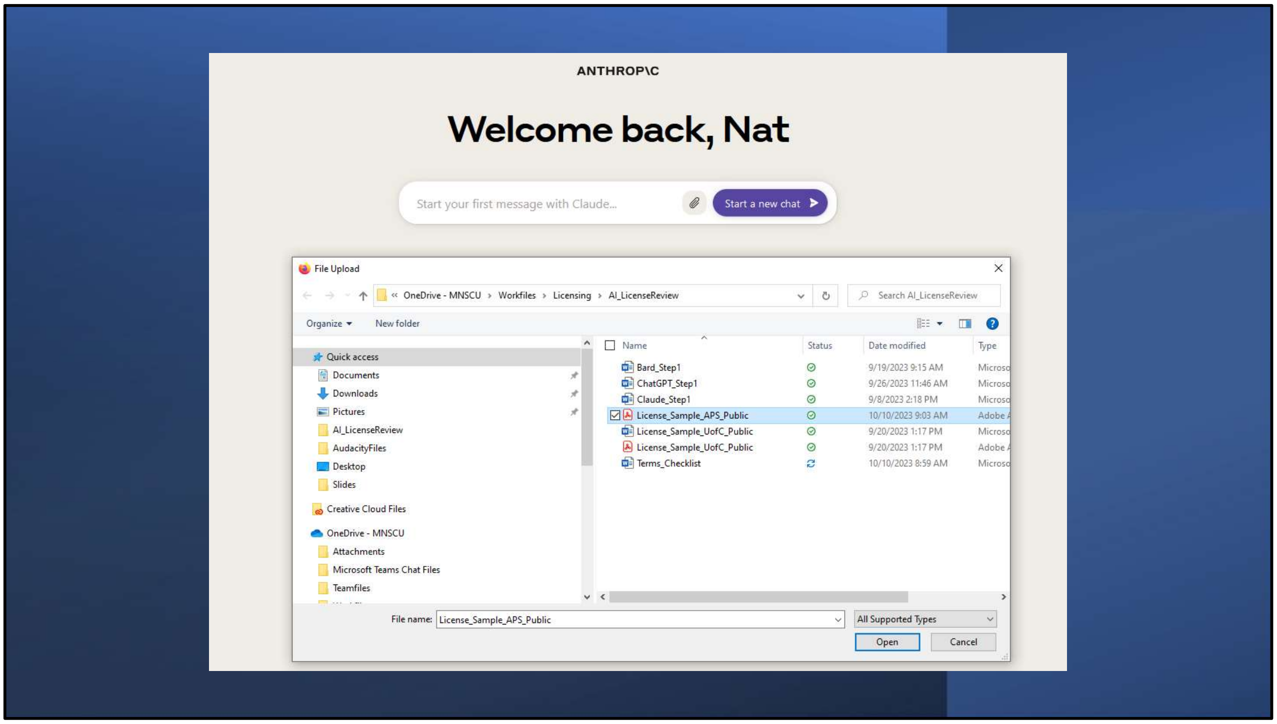
Welcome back

The license is divided internally into sections. These sections might be demarcated by numbers, letters, or spacing. Extract any sections that refer substantially or mainly to the following concepts or terms, which I've demarcated with angle brackets. Wait after presenting each term. I will either ask questions or ask you to continue presenting terms. (1) <definition of authorized user>, (2) <definition of site, (3) <renewal>, (4) <term>, (5) <termination>, (6) <financial hardship>, (7) <confidentiality> or <non-disclosure>, (8) <accessibility> or <ADA compliance>, (9) <transfer> or <previous subscriptions>.

The pause was helpful because when I came back to the project, I saw that I had been compartmentalizing again. I had thought to use ChatPDF because of my previous familiarity with the tool, but, of course, Claude can also accept PDF inputs, as I previously mentioned. I knew this in one part of my mind, but when I ran into the problem with API keys, I started looking in different directions altogether.

Working with Claude, I've continued building on what I learned from ChatPDF. I haven't done anything fancy yet, but I have learned there's value simply interacting with Claude over a license. In a very basic sense, this gamifies the work, if only because the interaction makes the work a little more enjoyable and because it causes me to see the licenses somewhat differently. Let's move onto a live-action demo of Claude so you can see how this works.

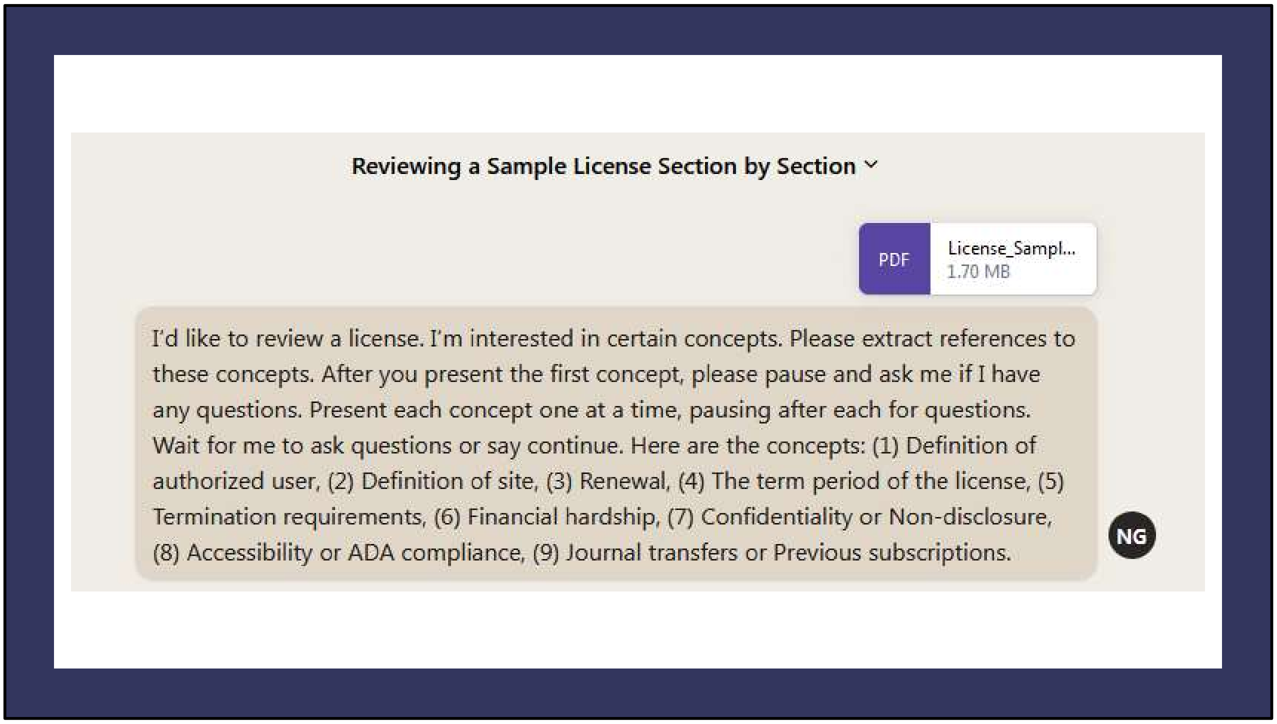
(The following slides are screenshots from Claude. In the video I demonstrate Claude in live-action, but I also wanted to provide my notes in these slides, in case they can be helpful to anyone.)



Name	Status	Date modified	Type
Bard_Step1	✓	9/19/2023 9:15 AM	Microso
ChatGPT_Step1	✓	9/26/2023 11:46 AM	Microso
Claude_Step1	✓	9/8/2023 2:18 PM	Microso
☑ License_Sample_APS_Public	✓	10/10/2023 9:03 AM	Adobe A
License_Sample_UofC_Public	✓	9/20/2023 1:17 PM	Microso
License_Sample_UofC_Public	✓	9/20/2023 1:17 PM	Adobe A
Terms_Checklist	⟳	10/10/2023 8:59 AM	Microso

File name: License_Sample_APS_Public

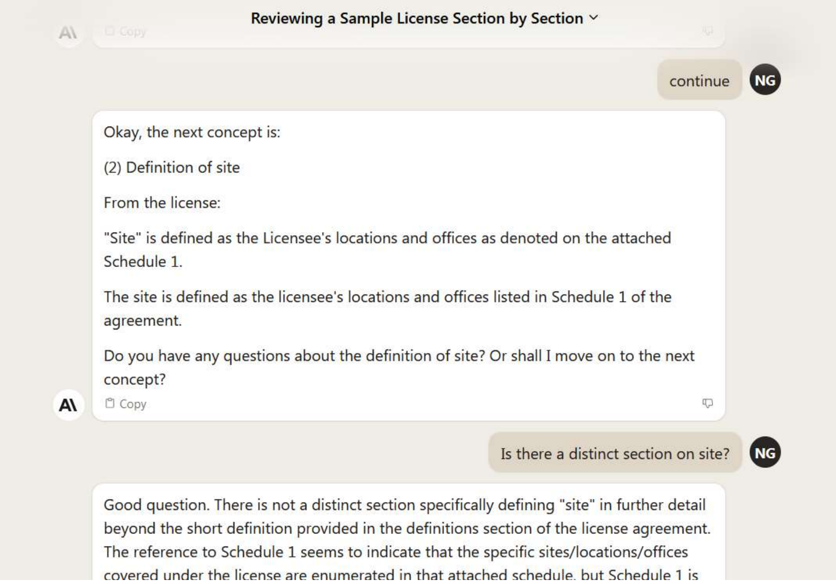
This demonstration will be very brief and basic because I’m limited for time, but it should provide a good enough sense of how anyone can get started. First, I’ll start by inputting a license into Claude. I’ll then enter one of the current versions of my starting prompt.



Reviewing a Sample License Section by Section

I'd like to review a license. I'm interested in certain concepts. Please extract references to these concepts. After you present the first concept, please pause and ask me if I have any questions. Present each concept one at a time, pausing after each for questions. Wait for me to ask questions or say continue. Here are the concepts: (1) Definition of authorized user, (2) Definition of site, (3) Renewal, (4) The term period of the license, (5) Termination requirements, (6) Financial hardship, (7) Confidentiality or Non-disclosure, (8) Accessibility or ADA compliance, (9) Journal transfers or Previous subscriptions.

This prompt is still based on the lean version I developed for ChatPDF. Basically, I've revised the prompt so that the mechanics of my interaction are correct. With earlier versions, Claude sometimes presented term references incorrectly or presented all terms at once. My goal was for Claude to pause over each of my terms of interest, which it now does. Curiously, the responses have improved as I've simplified the prompts so far, but I'm glad to know I'm not limited to 90 words as I was in ChatPDF, because, in the future, I will probably expand the checklist and associated context.



continue

Okay, the next concept is:

(2) Definition of site

From the license:

"Site" is defined as the Licensee's locations and offices as denoted on the attached Schedule 1.

The site is defined as the licensee's locations and offices listed in Schedule 1 of the agreement.

Do you have any questions about the definition of site? Or shall I move on to the next concept?

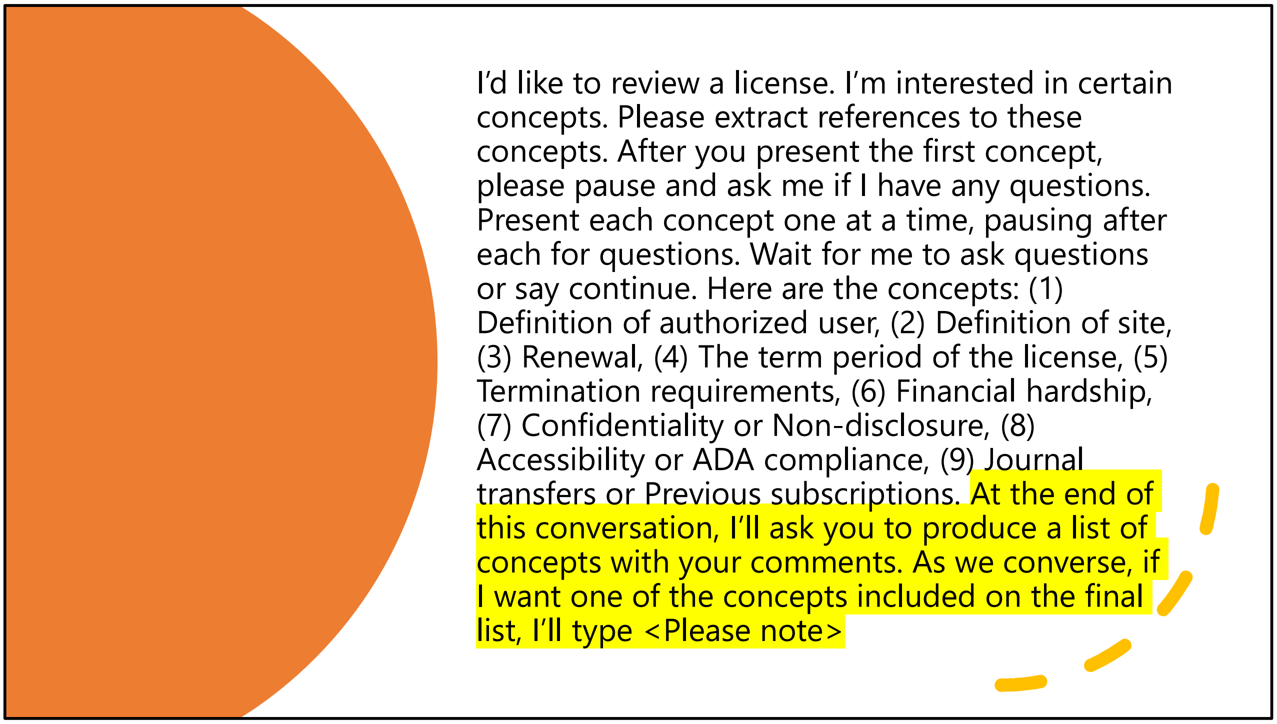
Is there a distinct section on site?

Good question. There is not a distinct section specifically defining "site" in further detail beyond the short definition provided in the definitions section of the license agreement. The reference to Schedule 1 seems to indicate that the specific sites/locations/offices covered under the license are enumerated in that attached schedule, but Schedule 1 is

I've gone back and forth deciding whether I prefer Claude to pause over each term on my checklist or whether I want Claude to provide a full list. I can see virtues in both approaches. For now, however, I'm pausing over each term, so that I can ask for follow-up – this helps me learn about the license and it also helps me think about how much I trust Claude to catch everything related to the terms in question.

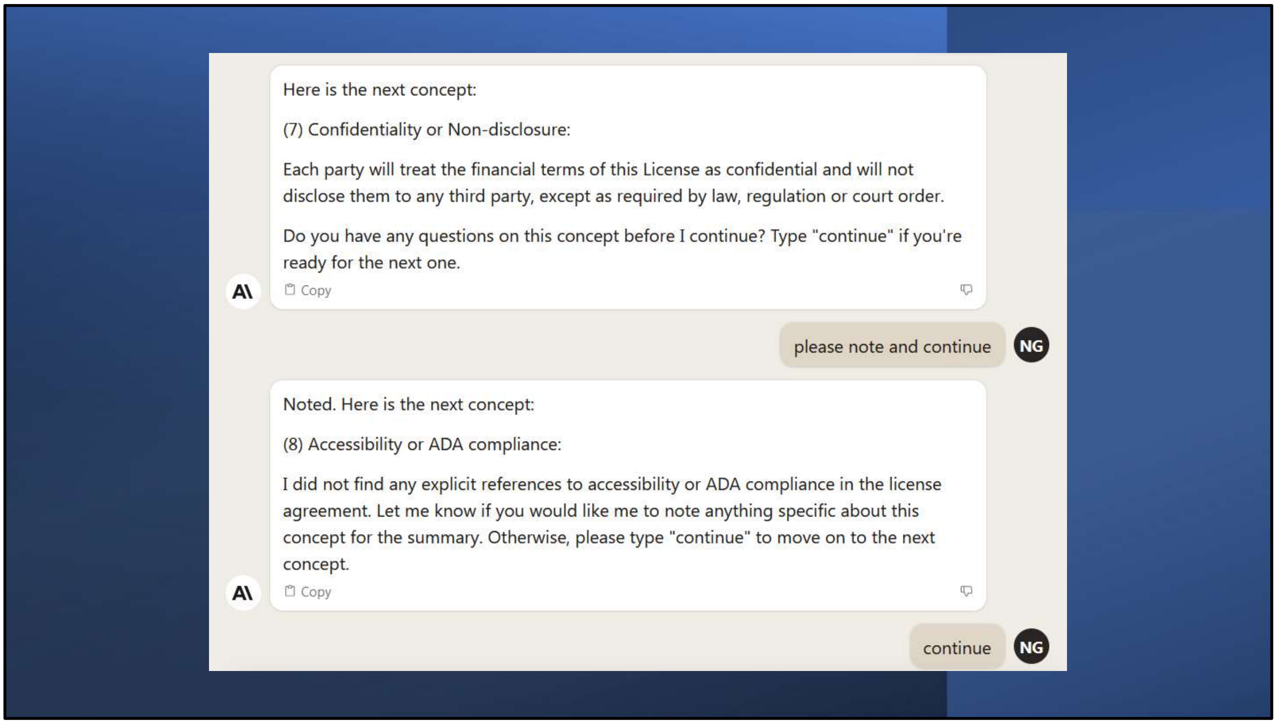
I’d like to review a license. I’m interested in certain concepts. Please extract references to these concepts. After you present the first concept, please pause and ask me if I have any questions. Present each concept one at a time, pausing after each for questions. Wait for me to ask questions or say continue. Here are the concepts: (1) Definition of authorized user, (2) Definition of site, (3) Renewal, (4) The term period of the license, (5) Termination requirements, (6) Financial hardship, (7) Confidentiality or Non-disclosure, (8) Accessibility or ADA compliance, (9) Journal transfers or Previous subscriptions.

To this point, I have only tested and tweaked my basic prompt and associated follow-up. So far, I think it’s helpful to interact with Claude to read licenses because it helps me slow down and check for the terms that typically need extra attention. I definitely could not replace my own careful reading of the license at this time, but the interaction is enjoyable and sometimes enlightening. These interactions are also leading me to think about how I can improve my checklist. An additional next step will be to work on a formatted output. If Claude outputs a report of terms requiring follow-up, I could revise and augment the report based on my own careful reading.



I'd like to review a license. I'm interested in certain concepts. Please extract references to these concepts. After you present the first concept, please pause and ask me if I have any questions. Present each concept one at a time, pausing after each for questions. Wait for me to ask questions or say continue. Here are the concepts: (1) Definition of authorized user, (2) Definition of site, (3) Renewal, (4) The term period of the license, (5) Termination requirements, (6) Financial hardship, (7) Confidentiality or Non-disclosure, (8) Accessibility or ADA compliance, (9) Journal transfers or Previous subscriptions. At the end of this conversation, I'll ask you to produce a list of concepts with your comments. As we converse, if I want one of the concepts included on the final list, I'll type <Please note>

And, actually, as I was working on this video, I realized how easy it would be to have Claude produce a list of concerns, at least as a first step toward a report. I added two sentences to the prompt…



Here is the next concept:

(7) Confidentiality or Non-disclosure:

Each party will treat the financial terms of this License as confidential and will not disclose them to any third party, except as required by law, regulation or court order.

Do you have any questions on this concept before I continue? Type "continue" if you're ready for the next one.

please note and continue

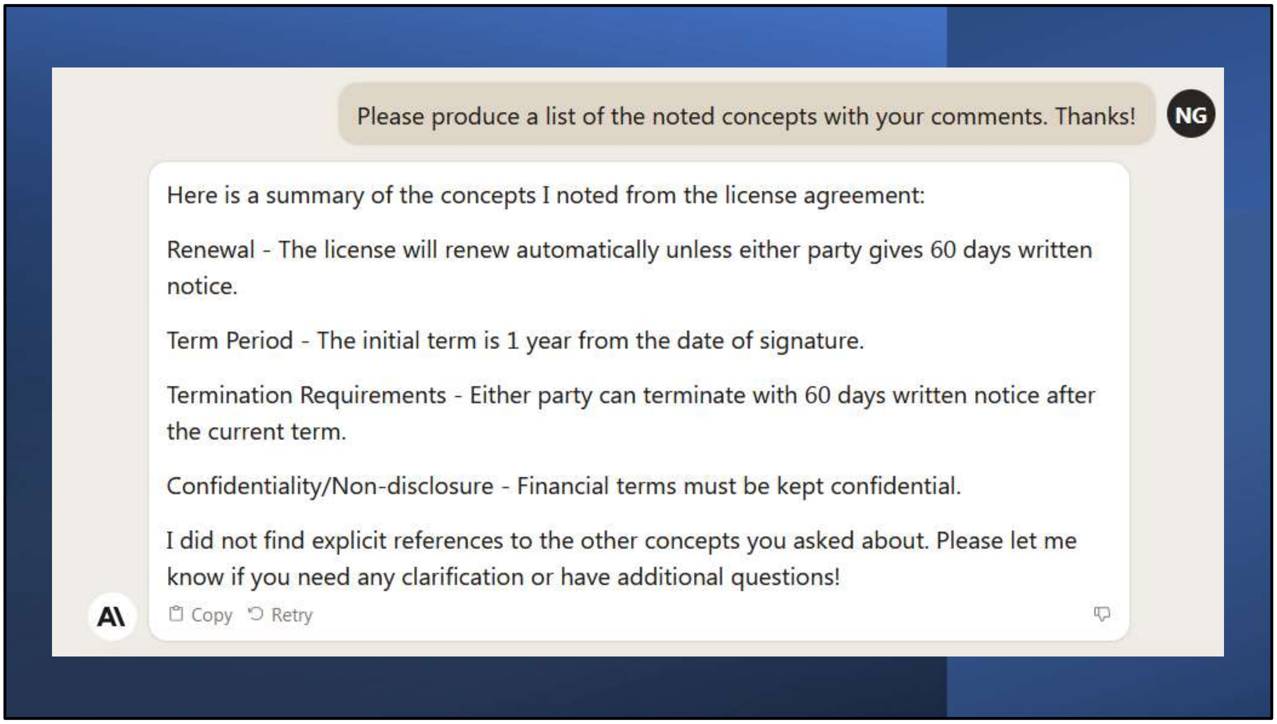
Noted. Here is the next concept:

(8) Accessibility or ADA compliance:

I did not find any explicit references to accessibility or ADA compliance in the license agreement. Let me know if you would like me to note anything specific about this concept for the summary. Otherwise, please type "continue" to move on to the next concept.

continue

As you can see on the screen, I simply needed to ask Claude to note concepts as we proceeded.



Please produce a list of the noted concepts with your comments. Thanks!

Here is a summary of the concepts I noted from the license agreement:

Renewal - The license will renew automatically unless either party gives 60 days written notice.

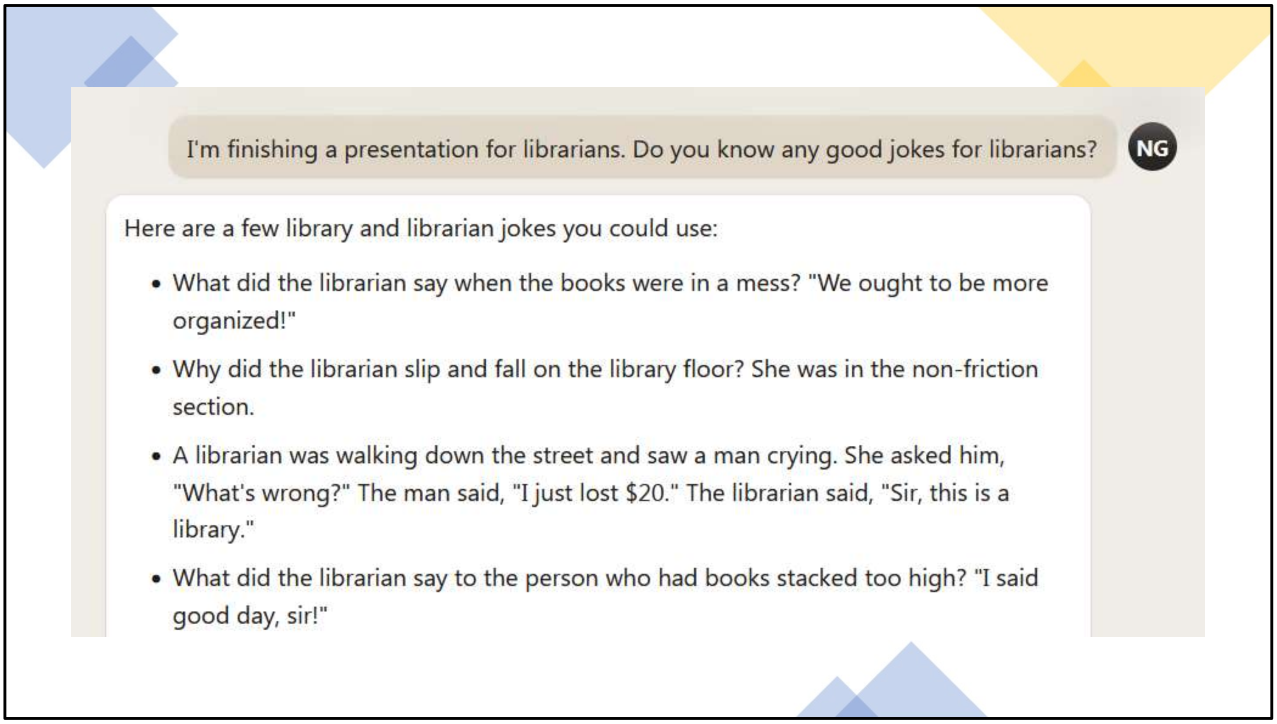
Term Period - The initial term is 1 year from the date of signature.

Termination Requirements - Either party can terminate with 60 days written notice after the current term.

Confidentiality/Non-disclosure - Financial terms must be kept confidential.

I did not find explicit references to the other concepts you asked about. Please let me know if you need any clarification or have additional questions!

At the end of the conversation, I just asked for a list and Claude provided it. I’ll probably play around with this a little more, but I like where it’s going so far…



I'm finishing a presentation for librarians. Do you know any good jokes for librarians?

Here are a few library and librarian jokes you could use:

- What did the librarian say when the books were in a mess? "We ought to be more organized!"
- Why did the librarian slip and fall on the library floor? She was in the non-friction section.
- A librarian was walking down the street and saw a man crying. She asked him, "What's wrong?" The man said, "I just lost $20." The librarian said, "Sir, this is a library."
- What did the librarian say to the person who had books stacked too high? "I said good day, sir!"

I still don't know when or if I'll return to the idea of using a Jupyter Notebook for license review specifically, although I do think Jupyter Notebook would be a good environment for some other AI projects. This has been a sort of hobby project which I have advanced in little bits and pieces when I've had time, but right now, I do feel confident utilizing Claude in my real-life license review process. I also think others could easily implement Claude for their own purposes, but maybe not help with jokes to end presentations. As you can see on the screen, Claude isn't quite ready for stand-up.

Thank you!

gustan2@mnsu.edu

link.mnsu.edu/ailicensevideo
link.mnsu.edu/ailicensedeck

Anyway, thank you for bearing with me… and I want to thank NASIG for the opportunity to present what I’ve done so far.

If you’d like to talk, my email address is on the screen.