day2 EXERCISE

January 6, 2025

1 Welcome to your first assignment!

Instructions are below. Please give this a try, and look in the solutions folder if you get stuck (or feel free to ask me!)

Just before we get to the assignment –

I thought I'd take a second to point you at this page of useful resources for the course. This includes links to all the slides. https://edwarddonner.com/2024/11/13/llm-engineering-resources/ Please keep this bookmarked, and I'll continue to add more useful links there over time.

2 HOMEWORK EXERCISE ASSIGNMENT

Upgrade the day 1 project to summarize a webpage to use an Open Source model running locally via Ollama rather than OpenAI

You'll be able to use this technique for all subsequent projects if you'd prefer not to use paid APIs.

Benefits: 1. No API charges - open-source 2. Data doesn't leave your box

Disadvantages: 1. Significantly less power than Frontier Model

2.1 Recap on installation of Ollama

Simply visit ollama.com and install!

Once complete, the ollama server should already be running locally.
If you visit:
http://localhost:11434/

You should see the message `Ollama is running`.

If not, bring up a new Terminal (Mac) or Powershell (Windows) and enter `ollama serve`
And in another Terminal (Mac) or Powershell (Windows), enter `ollama pull llama3.2`
Then try http://localhost:11434/ again.

If Ollama is slow on your machine, try using `llama3.2:1b` as an alternative. Run `ollama pull llama3.2:1b` from a Terminal or Powershell, and change the code below from `MODEL = "llama3.2"` to `MODEL = "llama3.2:1b"`

```
[1]: # imports
```

```python
import requests
from bs4 import BeautifulSoup
from IPython.display import Markdown, display
```

[2]:
```python
# Constants

OLLAMA_API = "http://localhost:11434/api/chat"
HEADERS = {"Content-Type": "application/json"}
MODEL = "llama3.2"
```

[3]:
```python
# Create a messages list using the same format that we used for OpenAI

messages = [
    {"role": "user", "content": "Describe some of the business applications of
    ↪Generative AI"}
]
```

[4]:
```python
payload = {
        "model": MODEL,
        "messages": messages,
        "stream": False
    }
```

[5]:
```python
# Let's just make sure the model is loaded

!ollama pull llama3.2
```

```
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling manifest   pulling manifest
pulling dde5aa3fc5ff… 100%              2.0 GB
pulling 966de95ca8a6… 100%              1.4 KB
pulling fcc5a6bec9da… 100%              7.7 KB
pulling a70ff7e570d9… 100%              6.0 KB
pulling 56bb8bd477a5… 100%                96 B
pulling 34bb5ab01051… 100%               561 B
verifying sha256 digest
writing manifest
success
```

[6]:
```python
# If this doesn't work for any reason, try the 2 versions in the following cells
# And double check the instructions in the 'Recap on installation of Ollama' at
    ↪the top of this lab
# And if none of that works - contact me!
```

```python
response = requests.post(OLLAMA_API, json=payload, headers=HEADERS)
print(response.json()['message']['content'])
```

Generative AI has numerous business applications across various industries, including:

1. **Content Creation**: Generative AI can be used to generate high-quality content such as articles, social media posts, product descriptions, and more. This can help businesses save time and resources while maintaining consistency in their content.
2. **Marketing Automation**: Generative AI can be used to automate marketing tasks such as email campaigns, ad targeting, and lead generation. It can also help personalize marketing messages based on customer data and behavior.
3. **Personalized Recommendations**: Generative AI can be used to generate personalized product recommendations for customers based on their browsing history, purchase behavior, and other factors.
4. **Chatbots and Virtual Assistants**: Generative AI can be used to create chatbots and virtual assistants that can provide 24/7 customer support and help businesses automate routine tasks.
5. **Data Analysis and Visualization**: Generative AI can be used to analyze large datasets and generate visualizations such as charts, graphs, and heatmaps to help businesses make data-driven decisions.
6. **Image and Video Generation**: Generative AI can be used to generate high-quality images and videos for marketing campaigns, product demos, and other business applications.
7. **Product Design and Development**: Generative AI can be used to design new products, prototypes, and concepts based on customer feedback, market trends, and other factors.
8. **Customer Service**: Generative AI can be used to generate responses to customer inquiries, resolve simple issues, and help businesses provide better customer service.
9. **Supply Chain Management**: Generative AI can be used to optimize supply chain operations, predict demand, and identify potential bottlenecks in the production process.
10. **Financial Analysis and Forecasting**: Generative AI can be used to analyze financial data, forecast revenue, and make predictions about market trends.

Some specific business use cases for generative AI include:

* **Amazon**: Uses generative AI to generate product descriptions, reviews, and even entire product pages.
* **Netflix**: Uses generative AI to personalize recommendations for its users based on their viewing history and preferences.
* **Walmart**: Uses generative AI to generate personalized marketing messages and offers for its customers based on their purchase behavior and demographics.
* **Google**: Uses generative AI to generate high-quality images and videos for its search results, product demos, and other business applications.

These are just a few examples of the many business applications of generative AI. As the technology continues to evolve, we can expect to see even more innovative use cases emerge across various industries.

3 Introducing the ollama package

And now we'll do the same thing, but using the elegant ollama python package instead of a direct HTTP call.

Under the hood, it's making the same call as above to the ollama server running at localhost:11434

```python
import ollama

response = ollama.chat(model=MODEL, messages=messages)
print(response['message']['content'])
```

Generative AI has numerous business applications across various industries, including:

1. **Content Creation**: Generative AI can be used to create high-quality content such as text, images, videos, and music. This can be particularly useful for businesses that need to produce large amounts of content quickly, such as marketing agencies, e-learning platforms, and media companies.
2. **Product Design and Development**: Generative AI can be used to design new products, such as 3D models, product prototypes, and even entire product lines. This can help businesses reduce the time and cost associated with traditional product development methods.
3. **Customer Service Chatbots**: Generative AI can be used to create advanced chatbots that can understand natural language and respond accordingly. This can help businesses improve their customer service experience and provide 24/7 support to customers.
4. **Personalized Marketing**: Generative AI can be used to analyze customer data and behavior, and generate personalized marketing messages and offers. This can help businesses increase customer engagement and conversion rates.
5. **Financial Analysis and Modeling**: Generative AI can be used to analyze large datasets and identify patterns and trends that may not be visible to humans. This can help businesses make more informed investment decisions and predict market trends.
6. **Risk Management and Compliance**: Generative AI can be used to analyze financial data and identify potential risks and compliance issues. This can help businesses reduce the risk of non-compliance and improve their overall risk management strategy.
7. **Supply Chain Optimization**: Generative AI can be used to optimize supply chain operations, such as demand forecasting, inventory management, and logistics planning. This can help businesses reduce costs and improve delivery times.
8. **Predictive Maintenance**: Generative AI can be used to analyze equipment

data and predict when maintenance is required. This can help businesses reduce downtime and improve overall equipment effectiveness.
9. **Data Annotation and Labeling**: Generative AI can be used to annotate and label large datasets, which can help businesses prepare their data for machine learning models.
10. **Digital Twinning**: Generative AI can be used to create digital twins of physical assets, such as buildings and equipment. This can help businesses optimize their operations and reduce energy consumption.

Some specific examples of business applications of Generative AI include:

* **SAP's Digital Customer Service**: SAP is using generative AI to power its digital customer service platform, which allows customers to interact with the company through chatbots and other interfaces.
* **IBM's Watson Health**: IBM's Watson Health platform uses generative AI to analyze medical data and predict patient outcomes. This can help healthcare providers make more informed decisions about patient care.
* **Google's DeepMind Health**: Google's DeepMind Health platform uses generative AI to analyze medical images and diagnose diseases. This can help doctors and researchers improve their understanding of complex diseases.

These are just a few examples of the many business applications of Generative AI. As the technology continues to evolve, we can expect to see even more innovative applications across various industries.

3.1 Alternative approach - using OpenAI python library to connect to Ollama

```python
# There's actually an alternative approach that some people might prefer
# You can use the OpenAI client python library to call Ollama:

from openai import OpenAI
ollama_via_openai = OpenAI(base_url='http://localhost:11434/v1',
 api_key='ollama')

response = ollama_via_openai.chat.completions.create(
    model=MODEL,
    messages=messages
)

print(response.choices[0].message.content)
```

Generative AI has numerous business applications across various industries. Here are some examples:

1. **Content Creation**: Generative AI can automatically generate high-quality content, such as:
 * Product descriptions and product pages
 * Blog posts and articles

* Social media posts and engaging content
 * Images and videos (using text-to-image models)
2. **Marketing Automation**: Generative AI can help automate marketing workflows by generating:
 * Personalized marketing messages and emails
 * Ad copy and ads (with high accuracy)
 * Product recommendations and upselling/cross-selling suggestions
3. **Customer Service Chatbots**: Generative AI-powered chatbots can:
 * Respond to customer inquiries and provide support
 * Offer personalized solutions and recommendations
 * Automate routine tasks, freeing up human customer service agents for more complex issues
4. **Data Analysis and Visualization**: Generative AI can help analyze and visualize large datasets by:
 * Identifying trends and patterns in data
 * Generating visualizations (e.g., charts, graphs) to present insights
 * Predicting future trends and outcomes
5. **Financial Modeling and Forecasting**: Generative AI can aid in financial modeling by generating:
 * Predictive models for sales, revenue, and expenses
 * Scenario-based forecasting and cash flow analysis
 * Risk assessment and sensitivity analysis
6. **Design and Engineering**: Generative AI can assist design teams with tasks such as:
 * Conceptual designs and product development
 * 3D modeling and visualization
 * Material selection and optimization
7. **Risk Management**: Generative AI can support risk management by generating:
 * Automated alerts for potential risks and threats
 * Scenario-based risk assessments and mitigation strategies
 * Predictive models for identifying high-risk events
8. **Cybersecurity**: Generative AI can aid in cybersecurity by:
 * Analyzing patterns of malware and detecting new threats
 * Generating custom security protocols and configurations
 * Predicting potential vulnerabilities and weaknesses
9. **Supply Chain Optimization**: Generative AI can help optimize supply chain operations by generating:
 * Route optimization for delivery vehicles
 * Demand forecasts and inventory management
 * Predictive models for material availability and lead times
10. **Document Review and Management**: Generative AI can assist with document review and management tasks such as:
 * Automated document summarization and abstracting
 * Document translation (voice-to-text) and transcription
 * Redaction and removal of sensitive information

These are just a few examples of the many business applications of Generative AI. As this technology continues to evolve, we're likely to see even more

```
innovative uses across industries.
```

4 NOW the exercise for you

Take the code from day1 and incorporate it here, to build a website summarizer that uses Llama 3.2 running locally instead of OpenAI; use either of the above approaches.

```python
[9]:  # A class to represent a Webpage
      # If you're not familiar with Classes, check out the "Intermediate Python"
       ↪notebook

      # Some websites need you to use proper headers when fetching them:
      headers = {
       "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
       ↪(KHTML, like Gecko) Chrome/117.0.0.0 Safari/537.36"
      }

      class Website:

          def __init__(self, url):
              """
              Create this Website object from the given url using the BeautifulSoup
       ↪library
              """
              self.url = url
              response = requests.get(url, headers=headers)
              soup = BeautifulSoup(response.content, 'html.parser')
              self.title = soup.title.string if soup.title else "No title found"
              for irrelevant in soup.body(["script", "style", "img", "input"]):
                  irrelevant.decompose()
              self.text = soup.body.get_text(separator="\n", strip=True)
```

```python
[10]: # Define our system prompt - you can experiment with this later, changing the
       ↪last sentence to 'Respond in markdown in Spanish."

      system_prompt = "You are an assistant that analyzes the contents of a website \
      and provides a short summary, ignoring text that might be navigation related. \
      Respond in markdown."
```

```python
[11]: # A function that writes a User Prompt that asks for summaries of websites:

      def user_prompt_for(website):
          user_prompt = f"You are looking at a website titled {website.title}"
          user_prompt += "\nThe contents of this website is as follows; \
      please provide a short summary of this website in markdown. \
      If it includes news or announcements, then summarize these too.\n\n"
          user_prompt += website.text
```

```
        return user_prompt
```

[12]:
```python
# See how this function creates exactly the format above

def messages_for(website):
    return [
        {"role": "system", "content": system_prompt},
        {"role": "user", "content": user_prompt_for(website)}
    ]
```

[13]:
```python
# And now: call the OpenAI API. You will get very familiar with this!

def summarize(url):
    website = Website(url)
    response = ollama.chat(model=MODEL, messages=messages_for(website))
    return response['message']['content']
```

[14]:
```python
# A function to display this nicely in the Jupyter output, using markdown

def display_summary(url):
    summary = summarize(url)
    display(Markdown(summary))
```

[15]:
```python
display_summary("https://edwarddonner.com")
```

5 Website Summary: Edward Donner

5.1 Overview

The website is hosted by Ed, the co-founder and CTO of Nebula.io. The platform applies AI to help people discover their potential and pursue their reason for being.

5.2 News/Announcements

- **December 21, 2024**: Welcome message for SuperDataScientists.
- **November 13, 2024**: Announcement of "Mastering AI and LLM Engineering – Resources".
- **October 16, 2024**: Announcement of "From Software Engineer to AI Data Scientist – resources".
- **August 6, 2024**: Introduction to the "Outsmart LLM Arena" - a battle of diplomacy and deviousness between LLMs.

5.3 Features

- The Outsmart arena pits LLMs against each other in a competitive environment.
- Nebula.io's platform applies AI to help people discover their potential and pursue their reason for being.

[ ]: