

ECON441B-Intro to ML Lab

Weiqi Xia

Github Repository Link:

```
In [1]: pip install openai wikipedia

Collecting openai
  Downloading openai-1.12.0-py3-none-any.whl (226 kB)
    226.7/226.7 kB 3.4 MB/s eta 0:00:00
Collecting wikipedia
  Downloading wikipedia-1.4.0.tar.gz (27 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.10/dist-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in /usr/lib/python3/dist-packages (from openai) (1.7.0)
Collecting httpx<1,>=0.23.0 (from openai)
  Downloading httpx-0.27.0-py3-none-any.whl (75 kB)
    75.6/75.6 kB 9.2 MB/s eta 0:00:00
Requirement already satisfied: pydantic<3,>=1.9.0 in /usr/local/lib/python3.10/dist-packages (from openai) (2.6.1)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-packages (from openai) (1.3.0)
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.10/dist-packages (from openai) (4.66.2)
Requirement already satisfied: typing-extensions<5,>=4.7 in /usr/local/lib/python3.10/dist-packages (from openai) (4.9.0)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from wikipedia) (4.12.3)
Requirement already satisfied: requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from wikipedia) (2.31.0)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai) (3.6)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai) (1.2.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0->openai) (2024.2.2)
Collecting httpcore==1.* (from httpx<1,>=0.23.0->openai)
  Downloading httpcore-1.0.4-py3-none-any.whl (77 kB)
    77.8/77.8 kB 4.9 MB/s eta 0:00:00
Collecting h11<0.15,>=0.13 (from httpcore==1.*->httpx<1,>=0.23.0->openai)
  Downloading h11-0.14.0-py3-none-any.whl (58 kB)
    58.3/58.3 kB 7.5 MB/s eta 0:00:00
Requirement already satisfied: annotated-types=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0->openai) (0.6.0)
Requirement already satisfied: pydantic-core=2.16.2 in /usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0->openai) (2.16.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.0.0->wikipedia) (3.3.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.0.0->wikipedia) (2.0.7)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->wikipedia) (2.5)
Building wheels for collected packages: wikipedia
  Building wheel for wikipedia (setup.py) ... done
  Created wheel for wikipedia: filename=wikipedia-1.4.0-py3-none-any.whl size=11678 sha256=2585acfe9be7da7725f4d2dd91fetc917d736fc40221949e6b897alb89391d48
  Stored in directory: /root/.cache/pip/wheels/5e/b6/c5/93f3dec388ae76edc830cb42901bb0232504dfc0df02fc50de
Successfully built wikipedia
Installing collected packages: h11, wikipedia, httpcore, httpx, openai
Successfully installed h11-0.14.0 httpcore-1.0.4 httpx-0.27.0 openai-1.12.0 wikipedia-1.4.0
```

```
In [2]: import openai
import os
import wikipedia
```

1.) Set up OpenAI and the enviornment

```
In [ ]: # DONE

In [3]: api_key = "*****"

In [4]: openai.api_key = api_key

In [5]: client = openai.OpenAI(
    api_key = openai.api_key
)
```

2.) Use the wikipedia api to get a function that pulls in the text of a wikipedia page

```
In [6]: page_titles = ["Artificial Intelligence", "UCLA"]

In [7]: page_title = page_titles[0]

In [8]: dir(wikipedia)

Out[8]: ['API_URL',
'BeautifulSoup',
'Decimal',
'DisambiguationError',
'HTTPTimeoutError',
'ODD_ERROR_MESSAGE',
'PageError',
'RATE_LIMIT',
'RATE_LIMIT_LAST_CALL',
'RATE_LIMIT_MIN_WAIT',
'RedirectError',
'USER_AGENT',
'WikipediaException',
'WikipediaPage',
'__builtins__',
'__cached__',
'__doc__',
'__file__',
'__loader__',
'__name__',
'__package__',
'__path__',
'__spec__',
'__version__',
'cache',
'datetime',
'debug',
'donate',
'exceptions',
'geosearch',
'languages',
'page',
'random',
're',
'requests',
'search',
'set_lang',
'set_rate_limiting',
'set_user_agent',
'stdout_encode',
'suggest',
'summary',
'sys',
'time',
'timedelta',
'unicode_literals',
'util',
'wikipedia']

In [9]: wikipedia.search(page_title)

Out[9]: ['Artificial intelligence',
'Generative artificial intelligence',
'Artificial general intelligence',
'A.I. Artificial Intelligence',
'Applications of artificial intelligence',
'Hallucination (artificial intelligence)',
'History of artificial intelligence',
'Ethics of artificial intelligence',
'Artificial intelligence art',
'Artificial intelligence in healthcare']

In [10]: search_results = wikipedia.search(page_title)

In [11]: page = wikipedia.page(search_results[0])

In [12]: # page.content
def get_wikipedia_content(page_title):
    search_results = wikipedia.search(page_title)
    page = wikipedia.page(search_results[0])
    return(page.content)

In [13]: content = get_wikipedia_content(page_title)
```

3.) Build a chatgpt bot that will analyze the text given and try to locate any false info

```
In [20]: chat_completions = client.chat.completions.create(
    model = "gpt-4",
    messages = [
        {"role": "system", "content": "I will be giving you an article. I am looking for false information. I want to capture all potentially false info, if there is"},
        {"role": "user", "content": content[:8180]}]
)

In [21]: chat_completions.choices[0].message.content

Out[21]: '1. "Alan Turing was the first person to conduct substantial research in the field that he called machine intelligence." - This is potentially false, as even though Ala
n Turing made significant contributions to the field, it may not be accurate to call him the first person to conduct substantial research in the field, as other researc
hers were also making contributions around the same time.\n2. "Artificial intelligence was founded as an academic discipline in 1956." - This can be argued as the studi
es and theories that contributed to what we know today as AI have been developed years (even decades) before 1956.\n3. "Funding and interest vastly increased after 2012
when deep learning surpassed all previous AI techniques, and after 2017 with the transformer architecture." - This statement seems definitive and might not be entirely
correct as the interest and funding in AI could increase due to various other factors and not solely based on the advancements in deep learning and transformer architec
ture.\n4. "This led to the AI spring of the early 2020s, with companies, universities, and laboratories overwhelmingly based in the United States pioneering significant
advances in artificial intelligence." - This statement implies that the significant advances in AI are overwhelmingly based in the United States which may not represent
the global nature of AI advancements accurately.\n5. "The general problem of simulating (or creating) intelligence has been broken into sub-problems." - It might be an
oversimplification to suggest that the challenge of simulating or creating intelligence has been "broken into sub-problems". It may also imply an organized, globally-ac
cepted approach that may not be the case.\n6. "An "agent" is anything that perceives and takes actions in the world." - False, as in AI context, an agent is generally a
component that perceives its environment and takes actions that maximize its chance of success at some goal. It doesn't have to perceive and take actions in the 'worl
d' necessarily.\n7. "In classical planning, the agent knows exactly what the effect of any action will be." - Potentially false, as even in classical planning, there m
ay not be certainty about the effects of each action.'
```

```
In [22]: print(chat_completions.choices[0].message.content)

1. "Alan Turing was the first person to conduct substantial research in the field that he called machine intelligence." - This is potentially false, as even though Alan
Turing made significant contributions to the field, it may not be accurate to call him the first person to conduct substantial research in the field, as other researche
rs were also making contributions around the same time.
2. "Artificial intelligence was founded as an academic discipline in 1956." - This can be argued as the studies and theories that contributed to what we know today as A
I have been developed years (even decades) before 1956.
3. "Funding and interest vastly increased after 2012 when deep learning surpassed all previous AI techniques, and after 2017 with the transformer architecture." - This
statement seems definitive and might not be entirely correct as the interest and funding in AI could increase due to various other factors and not solely based on the a
dvancements in deep learning and transformer architecture.
4. "This led to the AI spring of the early 2020s, with companies, universities, and laboratories overwhelmingly based in the United States pioneering significant advanc
es in artificial intelligence." - This statement implies that the significant advances in AI are overwhelmingly based in the United States which may not represent the g
lobal nature of AI advancements accurately.
5. "The general problem of simulating (or creating) intelligence has been broken into sub-problems." - It might be an oversimplification to suggest that the challenge o
f simulating or creating intelligence has been "broken into sub-problems". It may also imply an organized, globally-accepted approach that may not be the case.
6. "An "agent" is anything that perceives and takes actions in the world." - False, as in AI context, an agent is generally a component that perceives its environment a
nd takes actions that maximize its chance of success at some goal. It doesn't have to perceive and take actions in the 'world' necessarily.
7. "In classical planning, the agent knows exactly what the effect of any action will be." - Potentially false, as even in classical planning, there may not be certain
y about the effects of each action.
```

```
In [32]: def chatgpt_error_correction(text):
    chat_completions = client.chat.completions.create(
        model = "gpt-4",
        messages = [
            {"role": "system", "content": "I will be giving you an article. I am looking for false information. I want to capture all potentially false info, if there is"},
            {"role": "user", "content": text[:8180]}]
        )
    print(chat_completions.choices[0].message.content)
```

4.) Make a for loop and check a few wikipedia pages and return a report of any potentially false info via wikipedia

```
In [33]: page_titles = ["Artificial Intelligence", "UCLA", "Rain"]

In [34]: for page_title in page_titles:
    try:
        print("_____"+page_title)
        content = get_wikipedia_content(page_title)
        chatgpt_error_correction(content)
    except:
        print("ERROR")
```

Artificial Intelligence

- "AI technology is widely used throughout industry, government, and science." - This is partly false as it depends on the definition of AI technology. While some forms of AI are used throughout these sectors, not all forms of AI are used widely across all industries, governments, or branches of science.
- "Alan Turing was the first person to conduct substantial research in the field that he called machine intelligence." - This is potentially false information as, although Alan Turing was a pioneer in the field of AI, it could be claimed that others like John McCarthy, Allen Newell, or Herbert Simon might have conducted substantial research before or simultaneously.
- "Funding and interest vastly increased after 2012 when deep learning surpassed all previous AI techniques, and after 2017 with the transformer architecture." - The information that deep learning surpassed all previous AI techniques in 2012 can be considered as false as it's more of an opinion rather than verifiable fact. Different AI techniques may excel in various areas and it's not universally agreed upon that deep learning is superior to all other techniques.
- "The field's long-term goals ... ability to complete any task performable by a human." - This is potentially false information as it may not be a universal goal among all AI researchers or for the entire field. AI research goals vary greatly.
- "Knowledge bases need to represent things such as ... default reasoning" - This claim is not entirely accurate as some knowledge bases may not hold such broad information or be created to represent such things depending on their application.
- "In some problems, the agent's preferences may be uncertain, especially if there are other agents or humans involved." - This statement might be false as it depends on the design and objectives of the agent. Some AI agents are designed with specific, fixed preferences.
- "A Markov decision process ... reward function that supplies the utility of each state and the cost of each action." - This information can be false as not all Markov decision processes employ a 'cost' related to each action, it might be more accurate to talk about a 'reward' instead.
- "Machine learning is the study of programs that can improve their performance on a given task automatically. It has been a part of AI from the beginning." - Potentially false statement. Machine learning has been a part of AI development but wasn't necessarily included from the very beginning.
- "Unsupervised learning analyzes a stream of data and finds patterns and makes predictions without any other guidance." - This sentence is partially false, as unsupervised learning can find patterns, but making predictions typically falls under the realm of supervised learning.

UCLA

- UCLA's academic roots being established in 1881 as a normal school known as the southern branch of the California State Normal School. It actually became an independent entity known as the Los Angeles State Normal School in 1887 before eventually merging with UC in 1919.
- UCLA being the second-oldest campus in the UC system after UC Berkeley. Upon these two campuses' establishment, Cal was already functioning as a university, while the UCLA campus was a teacher training school.
- UCLA enrolling about 31,600 undergraduate and 14,300 graduate and professional students annually. The actual number is approximately 44,000 undergraduates and 28,000 graduates annually.
- UCLA's three graduate-level professional health science schools are Medicine, Dentistry, and Public Health. The university also has a school of Nursing, which offers advanced degrees.
- UCLA's athletic teams have won 121 NCAA team championships. As of 2021, they have won 118.
- UCLA has been represented in every Olympics since the university's founding except in 1924. The university was founded in 1919, while the Summer Olympics were held in 1920 and 1924.
- UCLA having a gold medalist in every Olympics in which the U.S. has participated since 1932. While the university does have a consistent history of Olympic success, it is highly unlikely to have had an affiliated gold medalist in every single Olympics.
- UCLA being considered a Public Ivy. The term "Public Ivy" wasn't coined until 1985 by Richard Moll.
- 27 Nobel laureates, five Turing Award winners, two Chief Scientists of the U.S. Air Force and one Fields Medalist have been affiliated with UCLA as faculty, researchers, and alumni. Numbers for Nobel laureates and other award winners being affiliated can vary depending on the affiliations considered.
- The new "Beverly Site" for UCLA being announced in 1925. The site of the currently existing campus in Westwood was announced in 1927.
- UCLA's original Vermont Avenue campus outgrowing its 25-acre location by the mid-1920s. The Vermont Avenue campus was approximately 20 acres, not 25, and the new Westwood campus wasn't purchased until 1926.
- The first undergraduate classes on the new campus being held in 1929 with 5,500 students. The actual number of students at the time of the campus' opening was around 4,000.
- Franklin David Murphy being appointed as Chancellor in 1960, sparking an era of tremendous growth for UCLA. Murphy didn't become Chancellor until 1961.
- In February 2022, Matthew Harris, a former lecturer and postdoc at UCLA, was arrested after allegedly making threats of violence against students and faculty members of UCLA's Philosophy Department. The threats happened in February 2022, but Harris was not arrested until March 2022.
- In 2022, UCLA signed an agreement to partner with the Tongva for caretaking and landscaping various areas of the campus. As of now, no such agreement is publicly known or announced.

Rain

- The article states that the steam locomotive was invented in the United Kingdom in 1802, but the first full-scale working railway steam locomotive was built in the UK by George Stephenson in 1814.
- The article states that the first ever steam train was powered by a steam locomotive built by Richard Trevithick in 1804. While Trevithick did build the first railway steam locomotive in 1804, the first full-scale working railway steam locomotive was built by George Stephenson in 1814.
- The article incorrectly states that the rapid transit and trams were built in the late 1800s, while the first full-scale working railway steam locomotives for public passenger transit were built in the early 1800s.
- It is mentioned that trains entered service in South America, Africa, and Asia through construction by imperial powers starting in the 1840s. Yet, the first railway in Africa was the Alexandria-Cairo line, which opened in 1856, later than stated.
- The article incorrectly states that steam locomotives were used to haul mainline trains as late as 2005 in Inner Mongolia, China. There is no evidence to support the claim that steam locomotives were in mainstream operation during this time.