RAG_and_LangChain_ConversationalRetrievalChain_round5

December 10, 2023

1 Import lib

```
[ ]: !pip install langchain
     !pip install pypdf
     !pip install openai
     !pip install chromadb
     !pip install tiktoken
```

```
[ ]: from langchain.vectorstores import Chroma
     from langchain.embeddings.openai import OpenAIEmbeddings
     from langchain.document_loaders import PyPDFLoader
     from langchain.text_splitter import CharacterTextSplitter
     from langchain.chat_models import ChatOpenAI
     from langchain.chains import RetrievalQA
```

```
[ ]: from google.colab import userdata
     openai_api_key = userdata.get('OPENAI_API_KEY')
```

2 Loading datasets and chunking

```
[ ]: from langchain.document_loaders import PyPDFLoader

     # I'm loading this summary of "Deep Work" book that you can find here:
     # https://briefer.com > books > deep-work > pdf
     pdf1 = path+"Deep_Work_summary.pdf"

     # and also RAG paper, to diversify source documents
     pdf2 = "https://arxiv.org/pdf/2005.11401.pdf"

     loaders = [
         # Duplicate documents on purpose - messy data
         PyPDFLoader(pdf1),
         PyPDFLoader(pdf1),
         PyPDFLoader(pdf2),
     ]
```

```python
docs = []
for i, loader in enumerate(loaders):
    pages = loader.load()
    print(f"For doc = {i}, number of pages: {len(pages)}")
    docs.extend(loader.load())

print(f" length of docs {len(docs)}")
```

```
For doc = 0, number of pages: 8
For doc = 1, number of pages: 8
For doc = 2, number of pages: 19
 length of docs 35
```

```python
text_splitter = CharacterTextSplitter(
    chunk_size = 1500,
    chunk_overlap = 150,
    separator='. '
)

chunks = text_splitter.split_documents(docs)
len(chunks)
```

```
[ ]: 80
```

3   Embedding and Vectorestores

```python
embedding = OpenAIEmbeddings(openai_api_key=openai_api_key)
```

```python
vectordb = Chroma.from_documents(
    documents=chunks,
    embedding=embedding,
)

print(vectordb._collection.count())
```

```
80
```

```python
llm = ChatOpenAI(model_name='gpt-3.5-turbo', openai_api_key=openai_api_key)
```

3.1   RetrievalQA

```python
qa_chain = RetrievalQA.from_chain_type(
    llm,
    retriever=vectordb.as_retriever(search_type = "mmr"),
    return_source_documents=True,
    chain_type="refine"
```

```
)

question = "What is a deep work"
result = qa_chain({"query": question})
print(f"Answer:\n {result['result']}")
```

Answer:
 Deep work refers to a state of focused and uninterrupted concentration on
cognitively demanding tasks, as opposed to shallow work. In today's digital age,
where we are constantly hyper-connected, it becomes increasingly challenging to
achieve deep work. The irony lies in the fact that our hyper-connectivity can
actually disconnect us from completing essential tasks. The constant
distractions and interruptions from digital devices and platforms can hinder our
ability to engage in deep work and fully immerse ourselves in meaningful,
valuable tasks. Therefore, practicing deep work requires consciously creating an
environment that minimizes distractions and allows for extended periods of
undisturbed concentration, enabling us to produce high-quality work and achieve
greater levels of productivity.

Here is another question without carrying out the previous LLM answer

```
[ ]: question = "How to put in place?"
     result = qa_chain({"query": question})
     print(f"Answer:\n {result['result']}")
```

Answer:
 In light of the additional context provided, the original answer remains
relevant and applicable. The importance of deep work and the need for focused
attention align with the idea that "focus, and not time, is what we all need
more of, to do the work that matters." The suggestions for creating a dedicated
workspace, minimizing distractions, and communicating availability still hold
true in order to prioritize deep work and improve overall focus.

4   ConversationBufferMemory

If we want to carry out this memory, we need to use **ConversationBufferMemory**.

ConversationBufferMemory is a memory that allows for storing messages and then extracts
the messages in a variable.

```
[ ]: from langchain.memory import ConversationBufferMemory
```

```
[ ]: memory = ConversationBufferMemory(
         memory_key="history",
         return_messages=True
     )

     #memory_key= history or chat_history ==> memory1 and memory2
```

3

```
memory1 = ConversationBufferMemory( memory_key="history")
memory1.save_context({"input": "hi"}, {"output": "whats up"})
print(f"memory1 ==> {memory1.load_memory_variables({})}")


memory2 = ConversationBufferMemory( memory_key="chat_history")
memory2.save_context({"input": "hi"}, {"output": "whats up"})
print(f"memory2 ==> {memory2.load_memory_variables({})}")
```

```
memory1 ==> {'history': 'Human: hi\nAI: whats up'}
memory2 ==> {'chat_history': 'Human: hi\nAI: whats up'}
```

```
[ ]: memory = ConversationBufferMemory(memory_key="chat_history",␣
       ↪return_messages=True)
     memory.save_context({"input": "hi"}, {"output": "whats up"})
     memory.save_context({"input": "tell me more"}, {"output": "more"})
     memory.load_memory_variables({})
```

```
[ ]: {'chat_history': [HumanMessage(content='hi'),
       AIMessage(content='whats up'),
       HumanMessage(content='tell me more'),
       AIMessage(content='more')]}
```

5   ConversationChain

ConversationChain with **ConversationBufferMemory** keeps in memory the previous LLMs answer. Combine all the previous question and LLM's answer in a new prompt and pass it again to the LLM for the current answer.

Here is how:

```
[ ]: from langchain.chains import ConversationChain
```

Using the ConversationBufferMemory in a chain, when calling an LLM. By specifying verbose = true, we will get the detailed prompt sent to the LLM. In this prompt, we can see that the results from the previous calls are included in it.

```
[ ]: conversation = ConversationChain(
         llm=llm,
         verbose=True, #To see the prompt
         memory=ConversationBufferMemory(memory_key="history")
     )

     # The prompt expects ['history', 'input']
     # Does not allow chat_history
```

```
[ ]: conversation.predict(input="Hi there!")
```

```
> Entering new ConversationChain chain…
Prompt after formatting:
The following is a friendly conversation between a human and an AI.

The AI is talkative and provides lots of specific details from its context. If

the AI does not know the answer to a question, it truthfully says it does not

know.

Current conversation:

Human: Hi there!

AI:

> Finished chain.
```

[ ]: 'Hello! How can I assist you today?'

[ ]: conversation.predict(input="I want to know how are you?")

```
> Entering new ConversationChain chain…
Prompt after formatting:
The following is a friendly conversation between a human and an AI.

The AI is talkative and provides lots of specific details from its context. If

the AI does not know the answer to a question, it truthfully says it does not

know.

Current conversation:

Human: Hi there!

AI: Hello! How can I assist you today?

Human: I want to know how are you?

AI:

> Finished chain.
```

[ ]: "Thank you for asking! As an AI, I don't have feelings or emotions, but I'm here
     and ready to help you with any questions or tasks you have. Is there something
     specific you'd like assistance with?"

[ ]: conversation.predict(input="Do you know 'deep work' book?")

```
> Entering new ConversationChain chain…
Prompt after formatting:
The following is a friendly conversation between a human and an AI.
The AI is talkative and provides lots of specific details from its context. If
the AI does not know the answer to a question, it truthfully says it does not
know.
Current conversation:
Human: Hi there!
AI: Hello! How can I assist you today?
Human: I want to know how are you?
AI: I'm an AI, so I don't have emotions or physical sensations like humans do.
However, I'm here and ready to help you with any questions or tasks you have.
How can I assist you today?
Human: Do you know 'deep work' book?
AI:

> Finished chain.
```

[ ]: 'Yes, I am familiar with the book "Deep Work" by Cal Newport. It was published
     in 2016 and explores the concept of deep work, which refers to the ability to
     focus without distraction on a cognitively demanding task. The book discusses
     the benefits of deep work, strategies for achieving it, and the impact of
     technology and distractions on our ability to engage in deep work. It has
     received positive reviews for its practical advice and insights into
     productivity in the modern age. Is there anything specific you would like to
     know about the book?'

Use **ConversationChain** with its own **PromptTemplate**

```python
from langchain.prompts.prompt import PromptTemplate

template = """You are a friendly AI, answering questions that you know. If you
  ↪don't know please say "I don't know".

Current conversation:
{history}
Human: {input}
AI:"""
PROMPT = PromptTemplate(input_variables=["history", "input"], template=template)
conversation = ConversationChain(
    prompt=PROMPT,
```

```
    llm=llm,
    verbose=True,
    memory=ConversationBufferMemory(human_prefix="Human"),
)
```

[ ]: `conversation.predict(input="Hi there!")`

> Entering new ConversationChain chain…
Prompt after formatting:
You are a friendly AI, answering questions that you know. If you

don't know please say "I don't know".

Current conversation:

Human: Hi there!

AI:

> Finished chain.

[ ]: 'Hello! How can I assist you today?'

[ ]: `conversation.predict(input="I want to know how are you?")`

> Entering new ConversationChain chain…
Prompt after formatting:
You are a friendly AI, answering questions that you know. If you

don't know please say "I don't know".

Current conversation:

Human: Hi there!

AI: Hello! How can I assist you today?

Human: I want to know how are you?

AI:

> Finished chain.

[ ]: "As an AI, I don't have feelings or emotions like humans do, but I'm here and
    ready to help. How can I assist you today?"

6 ConversationalRetrievalChain

With **ConversationBufferMemory** and **ConversationalRetrievalChain**, we keep in memory the LLM answer from the previous questions.

What is interesting with **ConversationalRetrievalChain** is that it combines the history' answers from the LLM with the newly asked question into one input. This input will be transformed to a standalone question which will be embedded and used in the vectorestore to retrieve the adequat documents.

It's also different from ConversationChain, which does not use any retrieval.

By default, in the retrieval part (RetrivalQA), the **chain_type** = '**stuff**', but you can modify it to '**refine**' or '**map_rerank**' or '**map_reduce**' as we already see in the previous notebook.

Also, the prompt template must contain: **chat_history** and **question** variables, like this example (source code):

condense_question_prompt = **PromptTemplate**(input_variables=['chat_history', 'question'], template='Given the following conversation and a follow up question, rephrase the follow up question to be a standalone question, in its original language.

```python
from langchain.chains import ConversationalRetrievalChain
from langchain.memory import ConversationBufferMemory
```

6.1 Default values

stuff in **chain_type** and **similarity** as **search_type** in the retrieval

```python
memory = ConversationBufferMemory(
    memory_key="chat_history",
    return_messages=True
)


retriever=vectordb.as_retriever()


qa = ConversationalRetrievalChain.from_llm(
    llm,
    retriever=retriever,
    memory=memory,
    verbose=True
)
```

6.1.1 Question1

```
question = "What is a deep work?"
result_1 = qa({"question": question})
```

```
result['answer']
```

```
'Deep work refers to a state of focused and concentrated work where an
individual can give their full attention to a demanding task without any
distractions or interruptions. It is characterized by a high level of
productivity and the ability to delve deeply into complex and meaningful work.
Deep work involves eliminating distractions, setting aside dedicated time for
concentrated work, and creating an environment conducive to deep focus. It is a
way to maximize productivity and produce high-quality work.'
```

```
result_1
```

```
{'question': 'What is a deep work?',
 'chat_history': [HumanMessage(content='What is a deep work?'),
  AIMessage(content='Deep work refers to a state of focused and concentrated
work where an individual can give their full attention to a demanding task
without any distractions or interruptions. It is characterized by a high level
of productivity and the ability to delve deeply into complex and meaningful
work. Deep work involves eliminating distractions, setting aside dedicated time
for concentrated work, and creating an environment conducive to deep focus. It
is a way to maximize productivity and produce high-quality work.')],
 'answer': 'Deep work refers to a state of focused and concentrated work where
an individual can give their full attention to a demanding task without any
distractions or interruptions. It is characterized by a high level of
productivity and the ability to delve deeply into complex and meaningful work.
Deep work involves eliminating distractions, setting aside dedicated time for
concentrated work, and creating an environment conducive to deep focus. It is a
way to maximize productivity and produce high-quality work.'}
```

6.1.2 Question2

Here is the second question in which I didn't explicitly talk about the Deep Work.

```
question = "How to put it in place?"
result_2 = qa({"question": question})
```

```
result_2
```

```
{'question': 'How to put it in place?',
 'chat_history': [HumanMessage(content='What is a deep work?'),
  AIMessage(content='Deep work refers to a state of focused and concentrated
work where an individual can give their full attention to a demanding task
without any distractions or interruptions. It is characterized by a high level
```

of productivity and the ability to delve deeply into complex and meaningful work. Deep work involves eliminating distractions, setting aside dedicated time for concentrated work, and creating an environment conducive to deep focus. It is a way to maximize productivity and produce high-quality work.'),
  HumanMessage(content='How to put it in place?'),
  AIMessage(content="Deep work can be implemented by following these steps:\n\n1. Make deep work part of your routine: Understand that willpower alone won't ensure commitment to deep work. Make deep work a regular part of your schedule and establish a routine around it.\n\n2. Schedule deep work: Look at your calendar and set aside dedicated time for deep work. Find a space where you won't be interrupted and choose a time when you feel most alert and productive. Start with scheduling just an hour and gradually increase the duration.\n\n3. Commit to the schedule ruthlessly: Stick to your deep work schedule and be disciplined about it. Use a scoreboard or a calendar to track and record the hours you spend doing deep work to stay motivated and on track.\n\n4. Eliminate distractions: Identify what steals your attention and eliminate or minimize those distractions. Practice digital minimalism, put up a do-not-disturb sign, use headphones, or consider working remotely if possible.\n\n5. Cultivate time awareness: Reflect on how much time you spend on shallow work instead of deep work. Conduct a time cost analysis and bring intention to what steals your attention. Be mindful of how you use your time and make adjustments accordingly.\n\n6. Divide your day into time blocks: Allocate specific tasks to time blocks of half-hour slots. Plan your day in advance and be disciplined about sticking to time limits. End each week with a review to evaluate what works and what doesn't, and make adjustments as needed.\n\nRemember, implementing deep work requires commitment and consistency. It's important to treat downtime as equally significant and make time for restorative rest to recharge and maintain focused attention.")],
  'answer': "Deep work can be implemented by following these steps:\n\n1. Make deep work part of your routine: Understand that willpower alone won't ensure commitment to deep work. Make deep work a regular part of your schedule and establish a routine around it.\n\n2. Schedule deep work: Look at your calendar and set aside dedicated time for deep work. Find a space where you won't be interrupted and choose a time when you feel most alert and productive. Start with scheduling just an hour and gradually increase the duration.\n\n3. Commit to the schedule ruthlessly: Stick to your deep work schedule and be disciplined about it. Use a scoreboard or a calendar to track and record the hours you spend doing deep work to stay motivated and on track.\n\n4. Eliminate distractions: Identify what steals your attention and eliminate or minimize those distractions. Practice digital minimalism, put up a do-not-disturb sign, use headphones, or consider working remotely if possible.\n\n5. Cultivate time awareness: Reflect on how much time you spend on shallow work instead of deep work. Conduct a time cost analysis and bring intention to what steals your attention. Be mindful of how you use your time and make adjustments accordingly.\n\n6. Divide your day into time blocks: Allocate specific tasks to time blocks of half-hour slots. Plan your day in advance and be disciplined about sticking to time limits. End each week with a review to evaluate what

works and what doesn't, and make adjustments as needed.\n\nRemember, implementing deep work requires commitment and consistency. It's important to treat downtime as equally significant and make time for restorative rest to recharge and maintain focused attention."}

```
[ ]: result_2['answer']
```

```
[ ]: "Deep work can be implemented by following these steps:\n\n1. Make deep work
     part of your routine: Understand that willpower alone won't ensure commitment to
     deep work. Make deep work a regular part of your schedule and establish a
     routine around it.\n\n2. Schedule deep work: Look at your calendar and set aside
     dedicated time for deep work. Find a space where you won't be interrupted and
     choose a time when you feel most alert and productive. Start with scheduling
     just an hour and gradually increase the duration.\n\n3. Commit to the schedule
     ruthlessly: Stick to your deep work schedule and be disciplined about it. Use a
     scoreboard or a calendar to track and record the hours you spend doing deep work
     to stay motivated and on track.\n\n4. Eliminate distractions: Identify what
     steals your attention and eliminate or minimize those distractions. Practice
     digital minimalism, put up a do-not-disturb sign, use headphones, or consider
     working remotely if possible.\n\n5. Cultivate time awareness: Reflect on how
     much time you spend on shallow work instead of deep work. Conduct a time cost
     analysis and bring intention to what steals your attention. Be mindful of how
     you use your time and make adjustments accordingly.\n\n6. Divide your day into
     time blocks: Allocate specific tasks to time blocks of half-hour slots. Plan
     your day in advance and be disciplined about sticking to time limits. End each
     week with a review to evaluate what works and what doesn't, and make adjustments
     as needed.\n\nRemember, implementing deep work requires commitment and
     consistency. It's important to treat downtime as equally significant and make
     time for restorative rest to recharge and maintain focused attention."
```

6.2 With MMR as search_type and Chain_type=Stuff

```
[ ]: memory = ConversationBufferMemory(
         memory_key="chat_history",
         return_messages=True
     )

     retriever=vectordb.as_retriever(search_type = "mmr", search_kwargs={"k": 3})

     qa = ConversationalRetrievalChain.from_llm(
         llm,
         retriever=retriever,
         memory=memory,
         verbose=True
     )
```

Here is a detailed schema of ConversationalRetrievalChain when using **stuff** chain_type for the

first and second question:

```
path_image1 = path_image + 'ConversationalRetrievalChain_stuff.png'
from IPython import display
display.Image(path_image1)
#source from LangChain documentation
```

[ ]:

ConversationalRetrievalChain

chain_type='stuff' , search_type = 'mmr' ==> Schema (1/1)

<question1>: "What is a deep work"

> Entering new StuffDocumentsChain chain...

    > Entering new LLMChain chain...

        Prompt after formatting:

    System: Use the following pieces of context to answer the user's question.

    If you don't know the answer, just say that you don't know, don't try to make up an answer.

    <doc1>

    <doc2>

    <doc3>

    Human: <question1>

    > Finished chain.

> Finished chain.

==> <**answer1**> ==> Correct: Definition and description of deep work

- With **MMR** you ensure to get diversified docs when retrieving from the vectorestore.
- With **Stuff**, you stuff the n retrieved documents in one context to be sent to LLM.

MMR *effect is not visible in the schema*

<question2>: "How to put it in place?"

> Entering new LLMChain chain...

    Prompt after formatting:

    Given the following conversation and a follow up question, rephrase the follow up question to be a standalone question, in its original language.

    Chat History:

    Human: <question1>

    Assistant: <answer1>

    Follow Up Input: <question2>

    Standalone question:

> Finished chain

> Entering new StuffDocumentsChain chain...

    > Entering new LLMChain chain...

        Prompt after formatting:

    System: Use the following pieces of context to answer the user's question.

    If you don't know the answer, just say that you don't know, don't try to make up an answer.

    <doc1>

    <doc2>

    <doc3> coming from RAG paper

    Human: <question2> but rephrased from history chat ==> How can deep work be implemented?

    > Finished chain.

> Finished chain.

==> <**answer2**>: correct answer: "Deep work can be implemented by following these steps: ...."

6.2.1 Question1

```
question = "What is a deep work"
result = qa({"question": question})
print(f"Answer:\n {result['answer']}")
```

> Entering new StuffDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:

All of the best, and most creative work, emerges from a state of clear focus and careful attention. So, perhaps deep work, along with restorative rest is just the antidote we need. Deep Work is a guide that shows us that we can take back control of our time, eliminate distractions, and improve our overall focus.

Focus, and not time, is what we all need more of, to do the work that matters.

Cal Newport, Associate Professor in computer science, popular author, and social media avoider, delves into the world of work, focus, and productivity. By distinguishing the two fundamental types of work, shallow and deep work, he offers practical advice on using your time more productively to produce better, more focused work.

This summary briefly looks at how to work smarter, not harder, how to avoid and negate distraction, and it busts the myth of multitasking. Furthermore, Newport advocates rest and rejuvenation and argues that downtime and relaxation are essential components to productivity.

Shallow Work Versus Deep Work

It's probably accurate to assume that your typical office day involves much of what Newport refers to as, "shallow work." This work involves activities such as checking emails, writing reports, and attending meetings. Shallow work leaves you feeling tired, and without the satisfaction of real accomplishment. A lot of the reason for this, is that shallow work is all of the tasks that we can perform when distracted, don't demand much thinking, and are easy to replicate. This kind of work means we don't create anything of value

This kind of work

means we don't create anything of value. So why is it that we gravitate towards shallow work?

The truth is that shallow work is easy, and deep work is difficult. Furthermore, shallow work is deceptive; it disguises itself as productivity.

```
> Finished chain.
```

```
> Finished chain.
Answer:
 Deep work refers to the state of focused, undistracted, and uninterrupted work
on a cognitively demanding task. It involves giving your full attention and
concentration to a single task, allowing you to produce high-quality work and
achieve a state of flow. Deep work requires sustained periods of concentration,
free from distractions such as social media, emails, or interruptions. It is the
opposite of shallow work, which consists of low-value tasks that can be easily
replicated and performed while distracted. Deep work is essential for producing
valuable and meaningful work.
```

6.2.2 Question2

```python
question = "How to put it in place?"
result_2 = qa({"question": question})
print(f"Answer:\n {result_2['answer']}")
```

```
> Entering new LLMChain chain…
Prompt after formatting:
Given the following conversation and a follow up question, rephrase

the follow up question to be a standalone question, in its original language.

Chat History:

Human: What is a deep work

Assistant: Deep work refers to the state of focused, undistracted, and

uninterrupted work on a cognitively demanding task. It involves giving your full

attention and concentration to a single task, allowing you to produce high-

quality work and achieve a state of flow. Deep work requires sustained periods

of concentration, free from distractions such as social media, emails, or

interruptions. It is the opposite of shallow work, which consists of low-value

tasks that can be easily replicated and performed while distracted. Deep work is

essential for producing valuable and meaningful work.

Follow Up Input: How to put it in place?

Standalone question:
```

```
> Finished chain.
```

```
> Entering new StuffDocumentsChain chain…


> Entering new LLMChain chain…
Prompt after formatting:
```

We've all heard the phrase, "work smarter, not harder." It's a big adjustment to make, because we've put so much value into working longer hours. Just because you're spending more time at the office, doesn't mean you're performing any better. You're probably just working harder at shallow work.

Newport encourages working smarter and developing your ability to delve into deep work. It's not easy - especially if you get distracted quickly. Deep Work offers some practical tips to help boost your focus fitness and, in turn, maximize your output.

Firstly, make deep work part of your routine. Understand that willpower alone, won't ensure you commit to doing deep work: you need to make it part of your routine. Think of the things and environment that might support you, like a nice cup of coffee to get you going.

You can schedule deep work, so look at your calendar and make time for it. It's best to find a dedicated space, where you know you won't be interrupted, and also when you feel you're most likely to be productive. Start slowly, by scheduling just an hour, and choose a time when you feel most alert. Once you've got the process started, try turning this into a daily habit that you can sustain. If you can only schedule time weekly, then be sure to set aside at least one full day.

If you work from home, find your space, put up the do-not-disturb sign, and practice digital minimalism

This kind of work

means we don't create anything of value. So why is it that we gravitate towards shallow work?

The truth is that shallow work is easy, and deep work is difficult. Furthermore, shallow work is deceptive; it disguises itself as productivity. Think about all of those little concrete tasks that make you look busy, but add very little value to your work output.

In contrast, deep work is difficult because it's hard to spend uninterrupted time routinely dedicated to your craft. Deep work isn't for

```
> Finished chain.

> Finished chain.
Answer:
 Deep work can be implemented by following these steps:

1. Make it a part of your routine: Willpower alone is not enough to commit to
deep work. You need to integrate it into your daily routine. Identify the things
and environment that support deep work, such as a quiet space or a cup of coffee
to help you focus.

2. Schedule dedicated time: Look at your calendar and set aside specific blocks
of time for deep work. Find a dedicated space where you won't be interrupted and
choose a time when you feel most alert and productive. Start with shorter
periods, like an hour, and gradually increase the duration.

3. Establish a daily habit: Once you've started scheduling deep work, aim to
make it a daily habit. Consistency is key. If you can only allocate time weekly,
make sure to set aside at least one full day for deep work.

4. Minimize distractions: If you work from home, create a designated workspace
and use a "do-not-disturb" sign to minimize interruptions. Practice digital
minimalism by reducing the use of distracting apps and notifications.

By implementing these strategies, you can cultivate the ability to engage in
deep work and maximize your productivity and output.
```

6.3   With MMR as search_type and Chain_type=Refine

You'll see in the following iterations with the refine chain_type the answers are not correct. Because, the refine method acts as it's continuing to refine answers from previous contexts. Also, if it finds a document context which is not related to your question (I did it in purpose by chunking RAG paper, alongside the deep work summary).

You need to explicitly ask in your question to get the needed steps.

```python
[ ]: memory = ConversationBufferMemory(
         memory_key="chat_history",
         return_messages=True
     )

     retriever=vectordb.as_retriever(search_type = "mmr", search_kwargs={"k": 3})

     qa_refine = ConversationalRetrievalChain.from_llm(
         llm,
         retriever=retriever,
         memory=memory,
```

```
    chain_type='refine',
    verbose=True

)
```

6.3.1 Question1

Here is a detailed schema of ConversationalRetrievalChain when using **refine** chain_type for the
first question:

```
[ ]: path_image1 = path_image + 'ConversationalRetrievalChain_refine1.png'
     # from IPython import display
     display.Image(path_image1)
     #source from LangChain documentation
```

[ ]:

ConversationalRetrievalChain

chain_type='refine' ==> Schema (1/2)

With **refine**:
- You ensure to send each document as a context to the LLM with the requested question,
- Get the answer
- Resend this answer with the next document and ask the LLM to refine the answer based on this new context

<question1>: "What is a deep work"

> Entering new RefineDocumentsChain chain...
    > Entering new LLMChain chain...
        Prompt after formatting:
        System: Context information is below.

        <doc1>

    Given the context information and not prior knowledge, answer any questions
        Human: <question1>
    > Finished chain.

    For doc in [<doc2>,<doc3>,...]:
        > Entering new LLMChain chain...
        Prompt after formatting:
        Human: <question1>
        AI: <answer from the previous call>
        Human: We have the opportunity to **refine** the existing answer (only if needed) with some more context below.

        <doc>

        Given the new context, **refine** the original answer to better answer the question. If the context isn't useful, return the original answer.

        > Finished chain.

> Finished chain.

==> **<answer1>** ==> Not that correct. It behaves like it's still in refining:
"The original answer remains relevant and provides a comprehensive explanation of deep work. Deep work refers to the ability to focus ...."

The answer of the first question is not that accurate. Let's see how refine will behave with the scond question →

```python
question = "What is a deep work"
result_refine1 = qa_refine({"question": question})
print(f"Answer:\n {result_refine1['answer']}")
```

> Entering new RefineDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:
System: Context information is below.

All of the best, and most creative work, emerges from a state of clear

focus and careful attention. So, perhaps deep work, along with restorative

rest is just the antidote we need. Deep Work is a guide that shows us that

we can take back control of our time, eliminate distractions, and improve

our overall focus.

Focus, and not time, is what we all need more of, to do the work that

matters.

Given the context information and not prior knowledge, answer any questions

Human: What is a deep work

> Finished chain.

> Entering new LLMChain chain…
Prompt after formatting:

Human: What is a deep work

AI: Deep work refers to the ability to focus without distraction on a cognitively demanding task. It is a state of intense concentration and productivity where one can produce high-quality work and make significant progress on complex tasks. Deep work requires eliminating or minimizing distractions and dedicating uninterrupted time to fully engage in the task at hand. It is a valuable skill in today's world of constant distractions and can lead to higher levels of productivity and creativity.

Human: We have the opportunity to refine the existing answer (only if needed) with some more context below.

Cal Newport, Associate Professor in computer science, popular author, and social media avoider, delves into the world of work, focus, and productivity. By distinguishing the two fundamental types of work, shallow and deep work, he offers practical advice on using your time more productively to produce better, more focused work.

This summary briefly looks at how to work smarter, not harder, how to avoid and negate distraction, and it busts the myth of multitasking. Furthermore, Newport advocates rest and rejuvenation and argues that downtime and relaxation are essential components to productivity.

Shallow Work Versus Deep Work

It's probably accurate to assume that your typical office day involves much of what Newport refers to as, "shallow work." This work involves activities such as checking emails, writing reports, and attending meetings. Shallow work leaves you feeling tired, and without the satisfaction of real accomplishment. A lot of the reason for this, is that shallow work is all of the tasks that we can perform when distracted, don't demand much thinking, and are easy to replicate. This kind of work means we don't create anything of value

----

Given the new context, refine the original answer to better answer the question. If the context isn't useful, return the original answer.

> Finished chain.

> Entering new LLMChain chain…
Prompt after formatting:
Human: What is a deep work

AI: Deep work refers to the ability to focus without distraction on a cognitively demanding task. It is a state of intense concentration and productivity where one can produce high-quality work and make significant progress on complex tasks. Deep work, as distinguished by Cal Newport, involves activities that require deep thinking, creativity, and problem-solving, such as research, writing, and strategic planning. In contrast, shallow work refers to tasks that are less demanding, can be easily replicated, and often performed while distracted, such as checking emails or attending meetings. By practicing deep work and minimizing shallow work, individuals can work smarter, avoid distractions, and produce more focused and valuable outcomes.
Human: We have the opportunity to refine the existing answer (only if needed) with some more context below.

This kind of work
means we don't create anything of value. So why is it that we gravitate towards shallow work?
The truth is that shallow work is easy, and deep work is difficult.
Furthermore, shallow work is deceptive; it disguises itself as productivity.
Think about all of those little concrete tasks that make you look busy, but add very little value to your work output.
In contrast, deep work is difficult because it's hard to spend uninterrupted time routinely dedicated to your craft. Deep work isn't for
----

Given the new context, refine the original answer to better answer the question.
If the context isn't useful, return the original answer.

> Finished chain.

```
> Finished chain.
```
Answer:
 The original answer remains relevant and provides a comprehensive explanation
of deep work. Deep work refers to the ability to focus without distraction on
cognitively demanding tasks and is characterized by intense concentration and
productivity. It involves activities that require deep thinking, creativity, and
problem-solving. Shallow work, on the other hand, refers to less demanding tasks
that can be easily replicated and often performed while distracted. While the
new context highlights the allure of shallow work and the challenges of
prioritizing deep work, the original answer already addresses these aspects by
explaining why people gravitate towards shallow work and the deceptive nature of
its perceived productivity. Therefore, the original answer adequately answers
the question.

6.3.2 Question 2

You'll see that the third document retrieved comes from the second pdf (RAG paper); which will
impact negatively the results of the LLM.

So I need to iterate several times, and ask the LLM to correct the answer. But this was possible
because with verbose=True, I saw why it wasn not working...But in production environnement,
you'll definitly noy going to do that...

Here is a detailed schema of ConversationalRetrievalChain when using **refine** chain_type for the
second question:

```python
path_image1 = path_image + 'ConversationalRetrievalChain_refine2.png'
# from IPython import displayx
display.Image(path_image1)
```

[ ]:

ConversationalRetrievalChain

chain_type='refine' ==> Schema (2/2)

<question2>: "How to put it in place?"

> Entering new LLMChain chain...
    Prompt after formatting:
    Given the following conversation and a follow up question, rephrase the follow up question to be a standalone question, in its original language.
    Chat History:
    Human: <question1>
    Assistant: <answer1>
    Follow Up Input: <question2>
    Standalone question:
> Finished chain

> Entering new RefineDocumentsChain chain...
    > Entering new LLMChain chain...
        Prompt after formatting:
        System: Context information is below.

        <doc1>

        Given the context information and not prior knowledge, answer any questions
        Human: : <question2> but rephrased from history chat ==> How can deep work be implemented?
    > Finished chain.

    For doc in [<doc2>,<doc3>,...]:
    > Entering new LLMChain chain...
        Prompt after formatting:
        Human: ***<question2> rephrased from step history chat (standalone question)***
        AI: <answer from the previous call>
            Human: We have the opportunity to **refine** the existing answer (only if needed) with some more context below.

        <doc>

        Given the new context, **refine** the original answer to better answer the question. If the context isn't useful, return the original answer.

    > Finished chain.

> Finished chain.

==> **<answer2>** ==> Not OK: "Apologies for the confusion, but the provided context does not seem to directly relate to the concept of "deep work" mentioned in the original question....."

```
question = "How to put it in place?"
result_refine2 = qa_refine({"question": question})
print(f"Answer:\n {result_refine2['answer']}")
```

> Entering new LLMChain chain…
Prompt after formatting:
Given the following conversation and a follow up question, rephrase

the follow up question to be a standalone question, in its original language.

Chat History:

Human: What is a deep work

Assistant: The original answer remains relevant and provides a comprehensive

explanation of deep work. Deep work refers to the ability to focus without

distraction on cognitively demanding tasks and is characterized by intense

concentration and productivity. It involves activities that require deep

thinking, creativity, and problem-solving. Shallow work, on the other hand,

refers to less demanding tasks that can be easily replicated and often performed

while distracted. While the new context highlights the allure of shallow work

and the challenges of prioritizing deep work, the original answer already

addresses these aspects by explaining why people gravitate towards shallow work

and the deceptive nature of its perceived productivity. Therefore, the original

answer adequately answers the question.
Follow Up Input: How to put it in place?
Standalone question:

> Finished chain.

> Entering new RefineDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:

System: Context information is below.

We've all heard the phrase, "work smarter, not harder." It's a big adjustment to make, because we've put so much value into working longer hours. Just because you're spending more time at the office, doesn't mean you're performing any better. You're probably just working harder at shallow work.

Newport encourages working smarter and developing your ability to delve into deep work. It's not easy - especially if you get distracted quickly. Deep Work offers some practical tips to help boost your focus fitness and, in turn, maximize your output.

Firstly, make deep work part of your routine. Understand that willpower alone, won't ensure you commit to doing deep work: you need to make it part of your routine. Think of the things and environment that might support you, like a nice cup of coffee to get you going.

You can schedule deep work, so look at your calendar and make time for it. It's best to find a dedicated space, where you know you won't be interrupted, and also when you feel you're most likely to be productive. Start slowly, by scheduling just an hour, and choose a time when you feel most alert. Once you've got the process started, try turning this into a daily habit that you can sustain. If you can only schedule time weekly, then be sure to set aside at least one full day.

If you work from home, find your space, put up the do-not-disturb sign, and practice digital minimalism

----

Given the context information and not prior knowledge, answer any questions
Human: How can deep work be implemented?

> Finished chain.

> Entering new LLMChain chain…
Prompt after formatting:

Human: How can deep work be implemented?

AI: Deep work can be implemented by following these steps:

1. Make deep work part of your routine: Don't rely solely on willpower to commit to deep work. Instead, make it a regular part of your schedule. Identify the factors and environment that support your focus, such as a quiet space or a favorite beverage.

2. Schedule dedicated time for deep work: Look at your calendar and allocate specific blocks of time for deep work. Start with shorter durations, like an hour, and choose a time when you feel most alert and productive. Gradually increase the duration and frequency of these sessions.

3. Find a dedicated space: Select a space where you can work without interruptions. It could be a designated office, a coffee shop, or any location that helps you concentrate. Make sure to communicate to others that you should not be disturbed during this time.

4. Start slowly and build a habit: Begin with small, manageable deep work sessions and gradually increase the duration and frequency. Aim to establish a daily habit of deep work that you can sustain. If daily deep work is not feasible, try to set aside at least one full day each week.

5. Practice digital minimalism: Minimize digital distractions during deep work sessions. Turn off notifications, put your phone on silent, and close unnecessary tabs and apps on your computer. This will help you maintain focus and avoid interruptions.

Remember, implementing deep work requires discipline and consistency. By making it a regular part of your routine and creating the right environment, you can enhance your focus and maximize your productivity.

Human: We have the opportunity to refine the existing answer (only if needed) with some more context below.

This kind of work

means we don't create anything of value. So why is it that we gravitate towards shallow work?

The truth is that shallow work is easy, and deep work is difficult.

Furthermore, shallow work is deceptive; it disguises itself as productivity. Think about all of those little concrete tasks that make you look busy, but add very little value to your work output.

In contrast, deep work is difficult because it's hard to spend

```
> Finished chain.


> Entering new LLMChain chain…
Prompt after formatting:
```

Human: How can deep work be implemented?

AI: In light of the additional context provided, the original answer remains relevant and applicable. Deep work can still be implemented by following the steps mentioned, regardless of the deceptive allure of shallow work. It is important to recognize the value and benefits of deep work, even if it requires more effort and discipline compared to shallow work. By incorporating the suggested strategies and maintaining a consistent deep work routine, individuals can overcome the challenges and experience the advantages of focused, meaningful productivity.

Human: We have the opportunity to refine the existing answer (only if needed) with some more context below.

These successes leverage different retrieval-based architectures and optimization techniques to
achieve strong performance on a single task, while we show that a single retrieval-based architecture
can be ne-tuned for strong performance on a variety of tasks.
Memory-based Architectures Our document index can be seen as a large external memory for
neural networks to attend to, analogous to memory networks [64,55]. Concurrent work [14] learns
to retrieve a trained embedding for each entity in the input, rather than to retrieve raw text as in our
work. Other work improves the ability of dialog models to generate factual text by attending over
fact embeddings [15,13]. A key feature of our memory is that it is comprised of raw text rather
distributed representations, which makes the memory both (i) human-readable, lending a form of
interpretability to our model, and (ii) human-writable, enabling us to dynamically update the model's
memory by editing the document index. This approach has also been used in knowledge-intensive
dialog, where generators have been conditioned on retrieved text directly, albeit obtained via TF-IDF
rather than end-to-end learnt retrieval [9]

```
> Finished chain.
```

```
> Finished chain.
Answer:
 Apologies for the confusion, but the provided context does not seem to directly
relate to the concept of "deep work" mentioned in the original question. The
context appears to be discussing retrieval-based architectures and optimization
techniques in the field of natural language processing.

If there is any additional information or clarification you can provide
regarding the implementation of deep work, I would be happy to assist you
further.
```

6.3.3 Iteration 1 Question 2:

```python
question = """you gave non adequate answer. Your answer is based only on the
 ↪last document talking about RAG.
However, I need you to get answers from a relevant context. When I told you to
 ↪come upt with Deep work steps to put in place, you need to read only
 ↪documents related to this subject, and not RAG paper.
How to put it in place?"""
result_refine2 = qa_refine({"question": question})
print(f"Answer:\n {result_refine2['answer']}")
```

```
> Entering new LLMChain chain…
Prompt after formatting:
```

Given the following conversation and a follow up question, rephrase
the follow up question to be a standalone question, in its original language.
Chat History:
Human: What is a deep work
Assistant: The original answer remains relevant and provides a comprehensive
explanation of deep work. Deep work refers to the ability to focus without
distraction on cognitively demanding tasks and is characterized by intense
concentration and productivity. It involves activities that require deep
thinking, creativity, and problem-solving. Shallow work, on the other hand,
refers to less demanding tasks that can be easily replicated and often performed
while distracted. While the new context highlights the allure of shallow work
and the challenges of prioritizing deep work, the original answer already
addresses these aspects by explaining why people gravitate towards shallow work
and the deceptive nature of its perceived productivity. Therefore, the original
answer adequately answers the question.
Human: How to put it in place?
Assistant: Apologies for the confusion, but the provided context does not seem
to directly relate to the concept of "deep work" mentioned in the original
question. The context appears to be discussing retrieval-based architectures and
optimization techniques in the field of natural language processing.
If there is any additional information or clarification you can provide
regarding the implementation of deep work, I would be happy to assist you
further.
Follow Up Input: you gave non adequate answer. Your answer is based only on the
last document talking about RAG.
However, I need you to get answers from a relevant context. When I told you to
come upt with Deep work steps to put in place, you need to read only documents
related to this subject, and not RAG paper.
How to put it in place?
Standalone question:

> Finished chain.

> Entering new RefineDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:
System: Context information is below.

We've all heard the phrase, "work smarter, not harder." It's a big
adjustment to make, because we've put so much value into working
longer hours. Just because you're spending more time at the office,
doesn't mean you're performing any better. You're probably just working
harder at shallow work.

Newport encourages working smarter and developing your ability to
delve into deep work. It's not easy - especially if you get distracted
quickly. Deep Work offers some practical tips to help boost your focus
fitness and, in turn, maximize your output.

Firstly, make deep work part of your routine. Understand that willpower
alone, won't ensure you commit to doing deep work: you need to make it
part of your routine. Think of the things and environment that might
support you, like a nice cup of coffee to get you going.

You can schedule deep work, so look at your calendar and make time for
it. It's best to find a dedicated space, where you know you won't be
interrupted, and also when you feel you're most likely to be productive.
Start slowly, by scheduling just an hour, and choose a time when you feel
most alert. Once you've got the process started, try turning this into a
daily habit that you can sustain. If you can only schedule time weekly,
then be sure to set aside at least one full day.

If you work from home, find your space, put up the do-not-disturb sign,
and practice digital minimalism

Given the context information and not prior knowledge, answer any questions
Human: How can the steps to implement deep work be identified and applied?

```
> Finished chain.


> Entering new LLMChain chain…
Prompt after formatting:
```

Human: How can the steps to implement deep work be identified and applied?

AI: To implement deep work effectively, you can follow these steps:

1. Recognize the importance: Understand the value and benefits of deep work. Recognize that it leads to higher productivity, improved focus, and better quality work.

2. Make it a priority: Treat deep work as a priority and allocate dedicated time for it. Set aside specific blocks of time in your schedule solely for deep work.

3. Create a routine: Make deep work a regular part of your routine. Choose a consistent time and place for deep work sessions. This helps train your brain to focus during those specific times.

4. Eliminate distractions: Minimize or eliminate distractions during deep work sessions. Find a quiet space where you won't be interrupted. Put your phone on silent or in another room, close unnecessary tabs or apps on your computer, and use tools like website blockers to avoid distractions.

5. Start small: Begin with shorter deep work sessions, such as an hour, and gradually increase the duration as you build your focus stamina. This helps prevent burnout and allows you to ease into the practice.

6. Find your optimal time: Identify the time of day when you feel most alert and focused. Schedule your deep work sessions during this period to optimize your productivity.

7. Develop rituals: Create rituals or habits that signal your brain it's time for deep work. It could be as simple as making a cup of coffee, listening to a specific playlist, or doing a short meditation before diving into deep work.

8. Set goals and track progress: Define clear goals for each deep work session and track your progress. This helps you stay motivated and provides a sense of accomplishment.

9. Practice discipline and consistency: Commit to following your deep work routine consistently. It may require discipline to resist the temptation of distractions or the urge to multitask. Over time, it will become easier and more natural.

10. Adapt and iterate: Continuously evaluate your deep work routine and make adjustments as needed. Experiment with different strategies, environments, and techniques to find what works best for you.

By following these steps, you can gradually incorporate deep work into your daily routine and experience the benefits of increased focus and productivity.

```
> Finished chain.


> Entering new LLMChain chain…
Prompt after formatting:
```

Human: How can the steps to implement deep work be identified and applied?

AI: The steps to implement deep work remain relevant in the given context. However, we can incorporate the provided context to refine the answer as follows:

To implement deep work effectively and reclaim control over your time and focus, follow these steps:

1. Recognize the value: Understand that deep work is the key to producing your best and most creative work. It allows you to focus without distractions and achieve a state of flow.

2. Make it a priority: Prioritize deep work and acknowledge that it is crucial for accomplishing meaningful tasks. Set aside dedicated time for deep work sessions in your schedule.

3. Create a routine: Establish a regular routine for deep work. Designate specific times and a conducive environment that promote focused and uninterrupted work. Consistency is key.

4. Minimize distractions: Identify and eliminate potential distractions during your deep work sessions. Put away your phone, close unnecessary tabs, and communicate to others that you should not be interrupted.

5. Start with small steps: Begin with shorter deep work sessions to build your focus stamina. Gradually increase the duration as you become more comfortable with sustained concentration.

6. Find your optimal focus time: Determine the time of day when you are most alert and focused. Schedule your deep work sessions during this period to leverage your peak productivity.

7. Establish rituals: Develop rituals or pre-work habits that signal your brain it's time for deep work. These rituals could include preparing a cup of coffee, listening to a specific playlist, or doing a short mindfulness exercise.

8. Set goals and track progress: Define clear goals for each deep work session and track your progress. Having specific objectives helps maintain motivation and provides a sense of accomplishment.

9. Cultivate discipline and consistency: Commit to following your deep work routine consistently. Train yourself to resist distractions and the temptation to multitask. Over time, it will become easier to maintain focus.

10. Adapt and refine: Continuously evaluate your deep work routine and make adjustments as needed. Experiment with different techniques, environments, or

> Finished chain.

> Finished chain.
Answer:
 The existing answer remains relevant and helpful in the given context. The
steps provided can still be applied to counter the distractions and
disconnection caused by hyper-connectivity in the digital age. No further
refinements are needed.

6.3.4   Iteration 2 question 2

```
question = """You didn't give the steps in your final answers. Provide them␣
 ↪please"""
result_refine3 = qa_refine({"question": question})
print(f"Answer:\n {result_refine3['answer']}")
```

> Entering new LLMChain chain…
Prompt after formatting:

Given the following conversation and a follow up question, rephrase
the follow up question to be a standalone question, in its original language.
Chat History:
Human: What is a deep work
Assistant: The original answer remains relevant and provides a comprehensive
explanation of deep work. Deep work refers to the ability to focus without
distraction on cognitively demanding tasks and is characterized by intense
concentration and productivity. It involves activities that require deep
thinking, creativity, and problem-solving. Shallow work, on the other hand,
refers to less demanding tasks that can be easily replicated and often performed
while distracted. While the new context highlights the allure of shallow work
and the challenges of prioritizing deep work, the original answer already
addresses these aspects by explaining why people gravitate towards shallow work
and the deceptive nature of its perceived productivity. Therefore, the original
answer adequately answers the question.
Human: How to put it in place?
Assistant: Apologies for the confusion, but the provided context does not seem
to directly relate to the concept of "deep work" mentioned in the original
question. The context appears to be discussing retrieval-based architectures and
optimization techniques in the field of natural language processing.
If there is any additional information or clarification you can provide
regarding the implementation of deep work, I would be happy to assist you
further.
Human: you gave non adequate answer. Your answer is based only on the last
document talking about RAG.
However, I need you to get answers from a relevant context. When I told you to
come upt with Deep work steps to put in place, you need to read only documents
related to this subject, and not RAG paper.
How to put it in place?
Assistant: The existing answer remains relevant and helpful in the given
context. The steps provided can still be applied to counter the distractions and
disconnection caused by hyper-connectivity in the digital age. No further
refinements are needed.
Follow Up Input: You didn't give the steps in your final answers. Provide them
please
Standalone question:

```
> Finished chain.


> Entering new RefineDocumentsChain chain…


> Entering new LLMChain chain…
Prompt after formatting:
```

System: Context information is below.

We've all heard the phrase, "work smarter, not harder." It's a big adjustment to make, because we've put so much value into working longer hours. Just because you're spending more time at the office, doesn't mean you're performing any better. You're probably just working harder at shallow work.

Newport encourages working smarter and developing your ability to delve into deep work. It's not easy – especially if you get distracted quickly. Deep Work offers some practical tips to help boost your focus fitness and, in turn, maximize your output.

Firstly, make deep work part of your routine. Understand that willpower alone, won't ensure you commit to doing deep work: you need to make it part of your routine. Think of the things and environment that might support you, like a nice cup of coffee to get you going.

You can schedule deep work, so look at your calendar and make time for it. It's best to find a dedicated space, where you know you won't be interrupted, and also when you feel you're most likely to be productive. Start slowly, by scheduling just an hour, and choose a time when you feel most alert. Once you've got the process started, try turning this into a daily habit that you can sustain. If you can only schedule time weekly, then be sure to set aside at least one full day.

If you work from home, find your space, put up the do-not-disturb sign, and practice digital minimalism

----

Given the context information and not prior knowledge, answer any questions
Human: What are the steps to put deep work in place?

> Finished chain.

> Entering new LLMChain chain…
Prompt after formatting:

Human: What are the steps to put deep work in place?

AI: To put deep work in place, the following steps can be followed:

1. Make deep work part of your routine: Recognize that relying solely on willpower is not enough to commit to doing deep work consistently. Make it a regular part of your routine by incorporating it into your schedule.

2. Schedule dedicated time for deep work: Look at your calendar and allocate specific time slots for deep work. Choose a dedicated space where you won't be easily interrupted and select a time when you feel most productive and alert.

3. Start slowly and gradually increase: Begin by scheduling just one hour of deep work and gradually increase the duration as you become more comfortable and proficient. This allows you to build up your focus and concentration over time.

4. Establish a daily habit: Aim to make deep work a daily habit that you can sustain in the long term. If daily deep work is not feasible, make sure to set aside at least one full day dedicated to deep work.

5. Create a conducive environment: If you work from home, find a space where you can minimize distractions and interruptions. Put up a "do-not-disturb" sign to signal to others that you need uninterrupted focus. Additionally, practice digital minimalism by reducing unnecessary digital distractions.

By following these steps, you can incorporate deep work into your routine and maximize your productivity and output.

Human: We have the opportunity to refine the existing answer (only if needed) with some more context below.

All of the best, and most creative work, emerges from a state of clear focus and careful attention. So, perhaps deep work, along with restorative rest is just the antidote we need. Deep Work is a guide that shows us that we can take back control of our time, eliminate distractions, and improve our overall focus.

Focus, and not time, is what we all need more of, to do the work that matters.

Given the new context, refine the original answer to better answer the question. If the context isn't useful, return the original answer.

> Finished chain.

> Entering new LLMChain chain…
Prompt after formatting:
Human: What are the steps to put deep work in place?

AI: The original answer remains relevant and applicable in the given context.

The steps outlined in the original answer still align with the idea of deep work

as a means to regain control of our time, eliminate distractions, and improve

focus. Therefore, no further refinement is needed in this case.

Human: We have the opportunity to refine the existing answer (only if needed)

with some more context below.

Switch off your phone or set it to airplane

mode, shut down all web browsers, and, if you can, disconnect from the

internet altogether.

If you're in an office environment, it's a good idea to use noise-canceling

----

Given the new context, refine the original answer to better answer the question.

If the context isn't useful, return the original answer.

> Finished chain.

> Finished chain.
Answer:
 To put deep work in place, you can follow these steps:

1. Eliminate distractions: Switch off your phone or set it to airplane mode,
close all unnecessary web browsers, and, if possible, disconnect from the
internet completely. This helps create an environment conducive to deep work.

2. Create a designated workspace: If you're in an office environment, consider
using noise-canceling headphones or finding a quiet area where you can focus
without interruptions. If you work from home, designate a specific area as your
workspace to minimize distractions.

3. Set clear intentions and goals: Before starting deep work, define what you
aim to accomplish during that session. Set specific goals and prioritize tasks
to maintain focus and direction.

4. Establish a routine or schedule: Allocate dedicated time blocks for deep work in your daily or weekly schedule. Consistency is key, so try to establish a routine that allows for regular deep work sessions without interruptions.

5. Practice time blocking: Use time blocking techniques to allocate specific time slots solely for deep work. This involves segmenting your schedule into focused work periods and scheduling other tasks and meetings around them.

6. Communicate your availability: If you work in a collaborative environment, communicate your deep work schedule to your colleagues or team. Let them know when you'll be unavailable or less responsive, ensuring they understand and respect your dedicated deep work time.

7. Start with shorter sessions: If you're new to deep work or find it challenging to sustain focus for extended periods, start with shorter sessions and gradually increase their duration. Building up your deep work stamina over time will help you improve your ability to concentrate for longer periods.

8. Prioritize self-care: Deep work requires mental energy and focus, so it's crucial to prioritize self-care practices. Get enough sleep, eat well, exercise regularly, and take breaks during deep work sessions to recharge and maintain productivity.

Remember, implementing deep work takes practice and discipline. Start by incorporating these steps gradually into your routine and adjust them as needed to find what works best for you.

6.4 With MMR as search_type and Chain_type=map_rerank

```python
memory = ConversationBufferMemory(
    memory_key="chat_history",
    return_messages=True
)

retriever=vectordb.as_retriever(search_type = "mmr")

qa_rerank = ConversationalRetrievalChain.from_llm(
    llm,
    retriever=retriever,
    memory=memory,
    chain_type='map_rerank',
    verbose=True

)
```

6.4.1 Question1

Here is a detailed schema of ConversationalRetrievalChain when using **refine** chain_type for the first question:

```python
path_image1 = path_image + 'ConversationalRetrievalChain_map_rerank1.png'
# from IPython import display
display.Image(path_image1)
#source from LangChain documentation
```

[ ]:

ConversationalRetrievalChain

chain_type='map_rerank' ==> Schema (1/2)

<question1>: "What is a deep work"

> Entering new MapRerankDocumentsChain chain...
>   > Entering new LLMChain chain...
>   For doc in [<doc1>,<doc2>,<doc3>,<doc4>]
>   *Prompt after formatting*:
>   *Use the following pieces of context to answer the question at the end. If you don't know the answer, just say that you don't know, don't try to make up an answer.*
>   *In addition to giving an answer, also return a score of how fully it answered the user's question. This should be in the following format:*
>   *Question: [question here]*
>   *Helpful Answer: [answer here]*
>   *Score: [score between 0 and 100]*

>   *How to determine the score:*
>   *– Higher is a better answer*
>   *– Better responds fully to the asked question, with sufficient level of detail*
>   *– If you do not know the answer based on the context, that should be a score of 0*
>   *– Don't be overconfident!*

>   *Example #1*
>   *Context:*
>   *---------*
>   *Apples are red*
>   *---------*
>   *Question: what color are apples?*
>   *Helpful Answer: red*
>   *Score: 100*

>   *<.... Example #2 and Example #3 ==> **few-shot learning to score the answer**>*

>   **Begin**!

>   Context:
>   ---------
>   <doc>
>   ---------
>   Question: <question 1>
>   Helpful Answer:
>   > Finished chain.
> Finished chain.

==> <**answer1**> ==> Correct answer ==> "Deep Work is a concept and a guide that emphasizes the importance ...."

With **rerank**:
- You ensure to send each document as a context to the LLM with the question,
- Get the answer and score it
- Get as final answer, the one with the highest score

The answer of the first question is accurate. Let's see how rerank will behave with the scond question
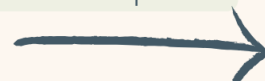
```
question = "What is a deep work"
result_rerank1 = qa_rerank({"question": question})
print(f"Answer:\n {result_rerank1['answer']}")
```

> Entering new MapRerankDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.
In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:
Question: [question here]
Helpful Answer: [answer here]
Score: [score between 0 and 100]
How to determine the score:
- Higher is a better answer
- Better responds fully to the asked question, with sufficient level of detail
- If you do not know the answer based on the context, that should be a score of
0
- Don't be overconfident!
Example #1
Context:

Apples are red

Question: what color are apples?
Helpful Answer: red
Score: 100
Example #2
Context:
----
it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?
Helpful Answer: a sports car or an suv
Score: 60
Example #3
Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.
In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:
Question: [question here]
Helpful Answer: [answer here]
Score: [score between 0 and 100]
How to determine the score:
- Higher is a better answer
- Better responds fully to the asked question, with sufficient level of detail
- If you do not know the answer based on the context, that should be a score of
0
- Don't be overconfident!
Example #1
Context:

Apples are red

Question: what color are apples?
Helpful Answer: red
Score: 100
Example #2
Context:
----
it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?
Helpful Answer: a sports car or an suv
Score: 60
Example #3
Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

```
/usr/local/lib/python3.10/dist-packages/langchain/chains/llm.py:344:
UserWarning: The apply_and_parse method is deprecated, instead pass an output
parser directly to LLMChain.
  warnings.warn(
```

> Finished chain.

> Finished chain.
Answer:
 Deep Work is a concept and a guide that emphasizes the importance of clear
focus and careful attention in order to produce the best and most creative work.
It teaches us how to take back control of our time, eliminate distractions, and
improve our overall focus. Deep Work argues that it is the quality of focus,
rather than the amount of time spent, that is crucial for doing meaningful work.
In essence, it is a practice that allows individuals to fully engage in their
work and achieve higher levels of productivity and satisfaction.

6.4.2 Question2

Here is a detailed schema of ConversationalRetrievalChain when using **rerank** chain_type for the second question:

```python
path_image1 = path_image + 'ConversationalRetrievalChain_map_rerank2.png'
# from IPython import display
display.Image(path_image1)
#source from LangChain documentation
```

[ ]:

ConversationalRetrievalChain

chain_type='map_rerank' ==> Schema (2/2)

<question2>: "How to put it in place?"

> Entering new LLMChain chain...
    Prompt after formatting:
    Given the following conversation and a follow up question, rephrase the follow up question to be a standalone question, in its original language.

    Chat History:

    Human: <question1>
    Assistant: <answer1>
    Follow Up Input: <question2>
    Standalone question:

> Finished chain.

> Entering new MapRerankDocumentsChain chain...
    > Entering new LLMChain chain...
        For doc in [<doc1>,<doc2>,<doc3>,<doc4>]:
            Prompt after formatting:
            Use the following pieces of context to answer the question at the end. If you don't know the answer, just say that you don't know, don't try to make up an answer.
                In addition to giving an answer, also return a score of how fully it answered the user's question. This should be in the following format:
                Question: [question here]

                ....
                <same prompt than before ==> **few-shot learning to score the answer>**

            Begin!

            Context:

            <doc>

            Question: <question2> **rephrased from step history chat (standalone question)**
            Helpful Answer:
        > Finished chain.
> Finished chain.

==> **<answer2>** ==> Correct one : "To implement Deep Work, you can start by making it part of your...."

```
question = "How to put it in place?"
result_rerank2 = qa_rerank({"question": question})
print(f"Answer:\n {result_rerank2['answer']}")
```

> Entering new LLMChain chain…
Prompt after formatting:
Given the following conversation and a follow up question, rephrase

the follow up question to be a standalone question, in its original language.

Chat History:

Human: What is a deep work

Assistant: Deep Work is a concept and a guide that emphasizes the importance of

clear focus and careful attention in order to produce the best and most creative

work. It teaches us how to take back control of our time, eliminate

distractions, and improve our overall focus. Deep Work argues that it is the

quality of focus, rather than the amount of time spent, that is crucial for

doing meaningful work. In essence, it is a practice that allows individuals to

fully engage in their work and achieve higher levels of productivity and

satisfaction.

Follow Up Input: How to put it in place?

Standalone question:

> Finished chain.

> Entering new MapRerankDocumentsChain chain…

> Entering new LLMChain chain…
Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

Prompt after formatting:

Use the following pieces of context to answer the question at the
end. If you don't know the answer, just say that you don't know, don't try to
make up an answer.

In addition to giving an answer, also return a score of how fully it answered
the user's question. This should be in the following format:

Question: [question here]

Helpful Answer: [answer here]

Score: [score between 0 and 100]

How to determine the score:

- Higher is a better answer

- Better responds fully to the asked question, with sufficient level of detail

- If you do not know the answer based on the context, that should be a score of
0

- Don't be overconfident!

Example #1

Context:

Apples are red

Question: what color are apples?

Helpful Answer: red

Score: 100

Example #2

Context:

----

it was night and the witness forgot his glasses. he was not sure if it was a
sports car or an suv

Question: what type was the car?

Helpful Answer: a sports car or an suv

Score: 60

Example #3

Context:

Pears are either red or orange

```
/usr/local/lib/python3.10/dist-packages/langchain/chains/llm.py:344:
UserWarning: The apply_and_parse method is deprecated, instead pass an output
parser directly to LLMChain.
  warnings.warn(
```

> Finished chain.

> Finished chain.
Answer:
 To implement Deep Work, you can start by making it part of your routine and
schedule dedicated time for it. Find a space where you won't be interrupted and
choose a time when you feel most alert. Start with scheduling just an hour and
gradually turn it into a daily habit. If you work from home, create a designated
workspace, use a do-not-disturb sign, and practice digital minimalism to
minimize distractions.