```
def DOCUMENTATION_WRITER_SOP(
    task: str,
    module: str,
):
    documentation = f"""
    Create multi-page long and explicit professional pytorch-like documentation for the {module} code below follow the outline for the {module} library,
    provide many examples and teach the user about the code, provide examples for every function, make the documentation 10,000 words,
    provide many usage examples and note this is markdown docs, create the documentation for the code to document,
    put the arguments and methods in a table in markdown to make it visually seamless

     Now make the professional documentation for this code, provide the architecture and how the class works and why it works that way,
    it's purpose, provide args, their types, 3 ways of usage examples, in examples show all the code like imports main example etc

    BE VERY EXPLICIT AND THOROUGH, MAKE IT DEEP AND USEFUL

    ######## INSTRUCTIONS ########
    Step 1: Understand the purpose and functionality of the module or framework

     Read and analyze the description provided in the documentation to understand the purpose and functionality of the module or framework.
    Identify the key features, parameters, and operations performed by the module or framework.
```

Step 2: Provide an overview and introduction

Start the documentation by providing a brief overview and introduction to the module or framework.

Explain the importance and relevance of the module or framework in the context of the problem it solves.

Highlight any key concepts or terminology that will be used throughout the documentation.

Step 3: Provide a class or function definition

Provide the class or function definition for the module or framework.

Include the parameters that need to be passed to the class or function and provide a brief description of each parameter.

Specify the data types and default values for each parameter.

Step 4: Explain the functionality and usage

Provide a detailed explanation of how the module or framework works and what it does.

Describe the steps involved in using the module or framework, including any specific requirements or considerations.

Provide code examples to demonstrate the usage of the module or framework.

Explain the expected inputs and outputs for each operation or function.

Step 5: Provide additional information and tips

Provide any additional information or tips that may be useful for using the module or framework effectively.

Address any common issues or challenges that developers may encounter and provide recommendations or workarounds.

Step 6: Include references and resources

Include references to any external resources or research papers that provide further information or background on the module or framework.

Provide links to relevant documentation or websites for further exploration.

Example Template for the given documentation:

EXAMPLE

Module/Function Name: MultiheadAttention

class torch.nn.MultiheadAttention(embed_dim, num_heads, dropout=0.0, bias=True, add_bias_kv=False, add_zero_attn=False, kdim=None, vdim=None, batch_first=False, device=None, dtype=None):

```

Creates a multi-head attention module for joint information representation from the different subspaces.

Parameters:

- embed_dim (int): Total dimension of the model.

- num_heads (int): Number of parallel attention heads. The embed_dim will be split across num_heads.

- dropout (float): Dropout probability on attn_output_weights. Default: 0.0 (no dropout).

- bias (bool): If specified, adds bias to input/output projection layers. Default: True.

- add_bias_kv (bool): If specified, adds bias to the key and value sequences at dim=0. Default: False.
```

- add_zero_attn (bool): If specified, adds a new batch of zeros to the key and value sequences at dim=1. Default: False.

- kdim (int): Total number of features for keys. Default: None (uses kdim=embed_dim).

- vdim (int): Total number of features for values. Default: None (uses vdim=embed_dim).

- batch_first (bool): If True, the input and output tensors are provided as (batch, seq, feature). Default: False.

- device (torch.device): If specified, the tensors will be moved to the specified device.

- dtype (torch.dtype): If specified, the tensors will have the specified dtype.

```

def forward(query, key, value, key_padding_mask=None, need_weights=True, attn_mask=None, average_attn_weights=True, is_causal=False):

```

Forward pass of the multi-head attention module.

Parameters:

- query (Tensor): Query embeddings of shape (L, E_q) for unbatched input, (L, N, E_q) when batch_first=False, or (N, L, E_q) when batch_first=True.

- key (Tensor): Key embeddings of shape (S, E_k) for unbatched input, (S, N, E_k) when batch_first=False, or (N, S, E_k) when batch_first=True.

- value (Tensor): Value embeddings of shape (S, E_v) for unbatched input, (S, N, E_v) when batch_first=False, or (N, S, E_v) when batch_first=True.

- key_padding_mask (Optional[Tensor]): If specified, a mask indicating elements to be ignored in key for attention computation.

- need_weights (bool): If specified, returns attention weights in addition to attention outputs. Default: True.

- attn_mask (Optional[Tensor]): If specified, a mask preventing attention to certain positions.

- average_attn_weights (bool): If true, returns averaged attention weights per head. Otherwise, returns attention weights separately per head. Note that this flag only has an effect when need_weights=True. Default: True.

- is_causal (bool): If specified, applies a causal mask as the attention mask. Default: False.

Returns:

Tuple[Tensor, Optional[Tensor]]:

- attn_output (Tensor): Attention outputs of shape (L, E) for unbatched input, (L, N, E) when batch_first=False, or (N, L, E) when batch_first=True.

- attn_output_weights (Optional[Tensor]): Attention weights of shape (L, S) when unbatched or (N, L, S) when batched. Optional, only returned when need_weights=True.

```

# Implementation of the forward pass of the attention module goes here

return attn_output, attn_output_weights

```

Usage example:

multihead_attn = nn.MultiheadAttention(embed_dim, num_heads)

attn_output, attn_output_weights = multihead_attn(query, key, value)

Note:

The above template includes the class or function definition, parameters, description, and

usage example.

To replicate the documentation for any other module or framework, follow the same structure and provide the specific details for that module or framework.

```
    ############## DOCUMENT THE FOLLOWING CODE ########
    {task}
    """
  return documentation
```