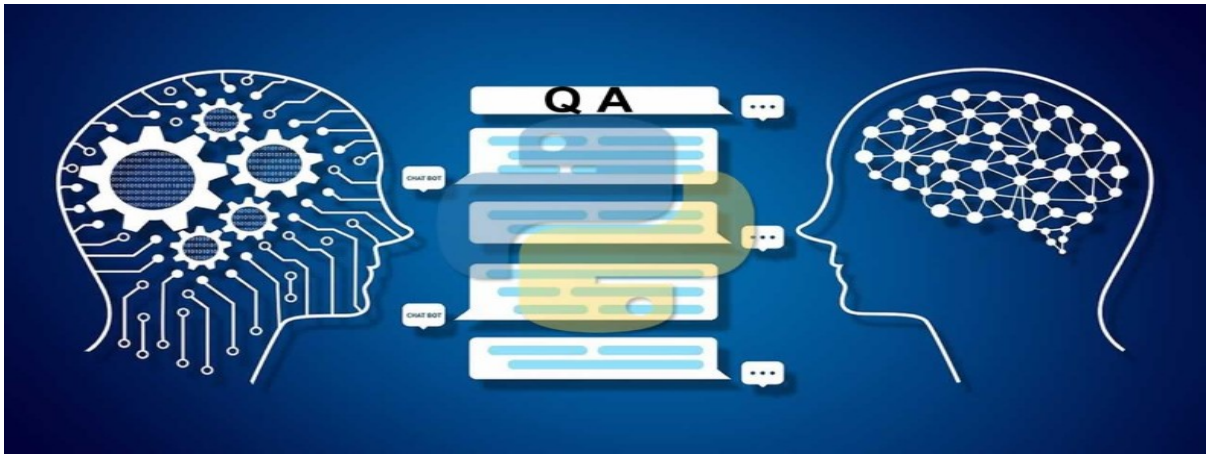
Natural Language Processing (BERT): Auto Model for Question Answering

The AutoModelForQuestionAnswering model with the configuration "bert-large-uncased-whole-word-masking-finetuned-squad" is a pre-trained language model specifically designed for question answering tasks.

In summary, this model is capable of:

- **Understanding the context of a given question and passage.**
- **Identifying the relevant information within the passage.**
- **Extracting the correct answer to the question.**

It is particularly well-suited for tasks that involve answering questions based on a given context, such as reading comprehension, document summarization, and chatbot applications.



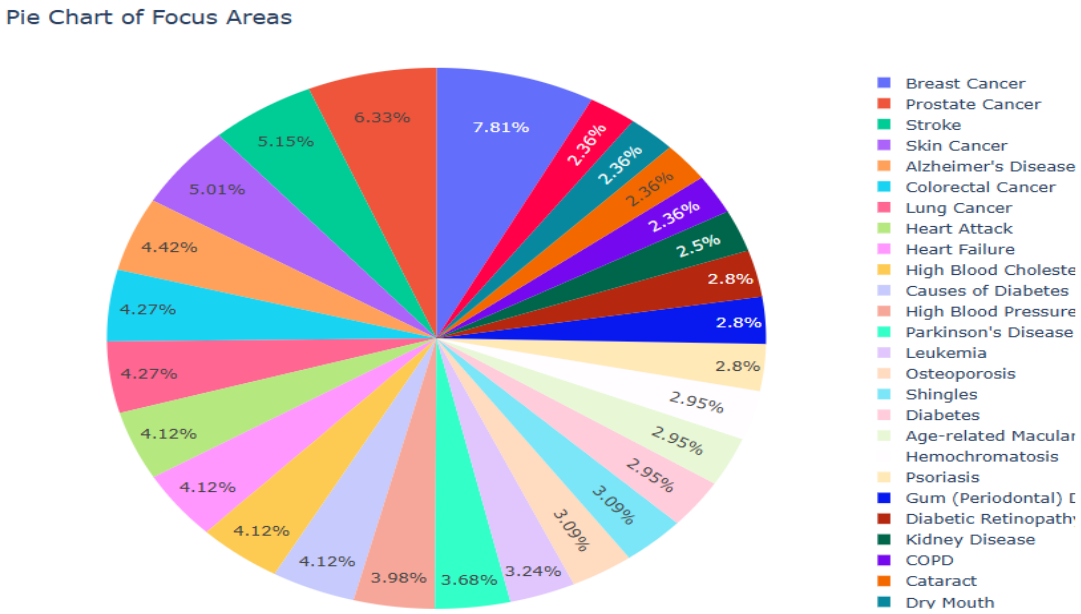

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv("/kaggle/input/medquad-medical-question-answer-for-ai-research/medquad.csv")
df = df.dropna()
df.head()
```

	question	answer	source	focus_area
0	What is (are) Glaucoma ?	Glaucoma is a group of diseases that can damag...	NIHSeniorHealth	Glaucoma
1	What causes Glaucoma ?	Nearly 2.7 million people have glaucoma, a lea...	NIHSeniorHealth	Glaucoma
2	What are the symptoms of Glaucoma ?	Symptoms of Glaucoma Glaucoma can develop in ...	NIHSeniorHealth	Glaucoma
3	What are the treatments for Glaucoma ?	Although open-angle glaucoma cannot be cured, ...	NIHSeniorHealth	Glaucoma
4	What is (are) Glaucoma ?	Glaucoma is a group of diseases that can damag...	NIHSeniorHealth	Glaucoma

Key components and their functionalities of AutoMLModel:

- ***BERT (Bidirectional Encoder Representations from Transformers):*** This is the underlying architecture of the model. BERT is a powerful language model that learns contextual representations of words by considering both left and right contexts.
- ***Large:*** This indicates that the model uses a larger version of the BERT architecture, typically with more parameters, which can lead to better performance on complex tasks.
- ***Uncased:*** This means the model is trained on a vocabulary without case distinctions, making it more robust to variations in capitalization.
- ***Whole-word-masking:*** During pre-training, the model is trained to predict masked whole words instead of individual characters. This helps it learn to understand and generate more coherent and meaningful text.
- ***Finetuned-squad:*** This indicates that the model has been fine-tuned on the SQuAD (Stanford Question Answering Dataset), a popular dataset for question answering tasks. This fine-tuning process adapts the model to the specific requirements of question answering, such as identifying the relevant context and extracting the correct answer.

```
import plotly.graph_objects as go
value_counts = df['focus_area'].value_counts()
result = value_counts[value_counts > 15]
fig = go.Figure(data=[go.Pie(labels=result.index, values=result.values)])
fig.update_layout(title='Pie Chart of Focus Areas', width=1200, heigh
t=800, font=dict(size=16))
fig.show()
```



```python
import random

# ANSI color codes
color_codes = {
    "blue": 34,
    "green": 32,
    "red": 31,
    "purple": 35,
    "orange": 33,
    "yellow": 33,
    "pink": 35,
    "brown": 33,
    "gray": 37
}

for i in range(0, len(df), 7):
    color = random.choice(list(color_codes.values()))
    print(f"\033[1;{color}mThe question is: {df['question'][i]}\033[0m\n\
033[1;{color}m The answer is: {df['answer'][i]}\033[0m\n")
    if i > 30:
        break
```

```
The question is: What is (are) Glaucoma ?
 The answer is: Glaucoma is a group of diseases that can damage the eye's optic nerve and result in vision loss and blindn
ess. While glaucoma can strike anyone, the risk is much greater for people over 60. How Glaucoma Develops  There are sever
al different types of glaucoma. Most of these involve the drainage system within the eye. At the front of the eye there is
a small space called the anterior chamber. A clear fluid flows through this chamber and bathes and nourishes the nearby ti
ssues. (Watch the video to learn more about glaucoma. To enlarge the video, click the brackets in the lower right-hand cor
ner. To reduce the video, press the Escape (Esc) button on your keyboard.) In glaucoma, for still unknown reasons, the flu
id drains too slowly out of the eye. As the fluid builds up, the pressure inside the eye rises. Unless this pressure is co
ntrolled, it may cause damage to the optic nerve and other parts of the eye and result in loss of vision. Open-angle Glauc
oma The most common type of glaucoma is called open-angle glaucoma. In the normal eye, the clear fluid leaves the anterior
chamber at the open angle where the cornea and iris meet. When fluid reaches the angle, it flows through a spongy meshwor
k, like a drain, and leaves the eye. Sometimes, when the fluid reaches the angle, it passes too slowly through the meshwor
k drain, causing the pressure inside the eye to build. If the pressure damages the optic nerve, open-angle glaucoma -- and
vision loss -- may result. There is no cure for glaucoma. Vision lost from the disease cannot be restored. However, there
are treatments that may save remaining vision. That is why early diagnosis is important.  See this graphic for a quick ove
rview of glaucoma,  including how many people it affects, whos at risk, what to do if you have it, and how to learn more.
See a glossary of glaucoma terms.

The question is: Who is at risk for Glaucoma? ?
 The answer is: Anyone can develop glaucoma. Some people are at higher risk than others. They include - African-Americans
over age 40  - everyone over age 60, especially Hispanics/Latinos  - people with a family history of glaucoma. African-Ame
ricans over age 40 everyone over age 60, especially Hispanics/Latinos people with a family history of glaucoma.  See this
graphic for a quick overview of glaucoma, including how many people it affects, whos at risk, what to do if you have it, a
nd how to learn more.


The question is: How to prevent High Blood Pressure ?
 The answer is: Two key measures are used to determine if someone is overweight or obese. These are body mass index, or BM
I, and waist circumference. Body mass index (BMI) is a measure of weight in relation to height, and provides an estimate o
f your total body fat. As your BMI goes up, so do your chances of getting high blood pressure, heart disease, and other he
alth problems. A BMI -  below 18.5 is a sign that you are underweight.  - between 18.5 and 24.9 is in the healthy range.
- between 25 and 29.9 is considered overweight.  - of 30 or more is considered obese.  below 18.5 is a sign that you are u
nderweight. between 18.5 and 24.9 is in the healthy range. between 25 and 29.9 is considered overweight. of 30 or more is
considered obese. See the Body Mass Index Table, available from the National Heart, Lung, and Blood Institute (NHLBI). Bod
y mass index (BMI) applies to both men and women, but it does have some limits. - It may overestimate body fat in in someo
ne who is very muscular or who has swelling from fluid retention (called edema)  - It may underestimate body fat in older
persons and others who have lost muscle mass. It may overestimate body fat in in someone who is very muscular or who has s
welling from fluid retention (called edema) It may underestimate body fat in older persons and others who have lost muscle
mass. Thats why waist measurement is often checked as well. Another reason is that too much body fat in the stomach area a
lso increases disease risk. A waist measurement of more than 35 inches in women and more than 40 inches in men is consider
ed high.
```

```
unique_focus_area = df['focus_area'][:500].unique()
print(f"\033[34m\033[1m{unique_focus_area}\033[0m")
```

```
['Glaucoma' 'High Blood Pressure' "Paget's Disease of Bone"
 'Urinary Tract Infections' 'Alcohol Use and Older Adults'
 'Osteoarthritis' 'Problems with Taste' 'Anxiety Disorders' 'Diabetes'
 'Medicare and Continuing Care' 'Knee Replacement' 'Balance Problems'
 'Quitting Smoking for Older Adults' 'Prostate Cancer' 'Dry Mouth'
 'Osteoporosis' 'Kidney Disease' "Alzheimer's Disease"
 'Rheumatoid Arthritis' 'Hearing Loss' 'Low Vision' 'COPD'
 'Age-related Macular Degeneration' 'Diabetic Retinopathy' 'Depression'
 'Problems with Smell' 'Breast Cancer' 'Colorectal Cancer'
 "Parkinson's Disease" 'Leukemia' 'Lung Cancer' 'Urinary Incontinence']
```

```
import shutil

def get_terminal_width():
        return shutil.get_terminal_size().columns

def wrap_text(text, width):
    words = text.split()
    lines = []
    current_line = []
    current_length = 0
    for word in words:
        if current_length + len(word) + 1 > width:
            lines.append(' '.join(current_line))
            current_line = [word]
            current_length = len(word)
        else:
            current_line.append(word)
            current_length += len(word) + 1
    lines.append(' '.join(current_line))
    return lines

def display_full_width_row(df, index):
    """Display a row from the DataFrame with full-width answer and alternate coloring."""
    question = df['question'][index]
    answer = df['answer'][index]
    focus_area = df['focus_area'][index]
```

```
    width = get_terminal_width()
    question_width = width - 12  # Adjust for "Question: " prefix
    answer_width = width - 16    # Adjust for "Answer: " prefix
    focus_width = width - 14     # Adjust for "Focus Area: " prefix

    # Wrap text to fit within the calculated widths
    wrapped_question = wrap_text(question, question_width)
    wrapped_answer = wrap_text(answer, answer_width)
    wrapped_focus = wrap_text(focus_area, focus_width)

    # Print with alternate coloring using background colors
    print("\033[45m" + "Question:".ljust(width) + "\033[0m")  # Magent
a background
    for line in wrapped_question:
        print(line)
    print()

    print("\033[46m" + "Answer:".ljust(width) + "\033[0m")    # Cyan ba
ckground
    for line in wrapped_answer:
        print(line)
    print()

    print("\033[43m" + "Focus Area:".ljust(width) + "\033[0m")  # Yello
w background
    for line in wrapped_focus:
        print(line)
    print()
display_full_width_row(df, 100)
```

```
Question:
What are the treatments for Anxiety Disorders ?

Answer:
Most insurance plans, including health maintenance
organizations (HMOs), will cover treatment for anxiety
disorders. Check with your insurance company and find out. If
you dont have insurance, the Health and Human Services division
of your county government may offer mental health care at a
public mental health center that charges people according to how
much they are able to pay. If you are on public assistance, you
may be able to get care through your state Medicaid plan. To
learn about more mental health resources, see Help for Mental
Illness, from the National Institute of Mental Health at NIH.

Focus Area:
Anxiety Disorders
```

```
display_full_width_row(df, 1)
```

```
Question:
What causes Glaucoma ?

Answer:
Nearly 2.7 million people have glaucoma, a leading cause of
blindness in the United States. Although anyone can get
glaucoma, some people are at higher risk. They include -
African-Americans over age 40 - everyone over age 60, especially
Hispanics/Latinos - people with a family history of glaucoma.
African-Americans over age 40 everyone over age 60, especially
Hispanics/Latinos people with a family history of glaucoma. In
addition to age, eye pressure is a risk factor. Whether you
develop glaucoma depends on the level of pressure your optic
nerve can tolerate without being damaged. This level is
different for each person. Thats why a comprehensive dilated eye
exam is very important. It can help your eye care professional
determine what level of eye pressure is normal for you. Another
risk factor for optic nerve damage relates to blood pressure.
Thus, it is important to also make sure that your blood pressure
is at a proper level for your body by working with your medical
doctor. (Watch the animated video to learn more about the causes
of glaucoma. To enlarge the video, click the brackets in the
lower right-hand corner. To reduce the video, press the Escape
(Esc) button on your keyboard.)

Focus Area:
Glaucoma
```

```
%%time

from transformers import AutoTokenizer, AutoModelForQuestionAnswering
import torch

# ANSI escape codes for text color
RED = '\033[91m'
RESET = '\033[0m'

# Load the dataset
print("Dataset loaded successfully.")

# Initialize the model and tokenizer
```

```
print("Initializing the model and tokenizer...")
model_name = "bert-large-uncased-whole-word-masking-finetuned-squad"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForQuestionAnswering.from_pretrained(model_name)
print("Model and tokenizer initialized.")

def answer_question(question, context):
    # Tokenize the input
    inputs = tokenizer.encode_plus(question, context, return_tensors="pt", max_length=512, truncation=True)

    # Get the model's answer
    with torch.no_grad():
        outputs = model(**inputs)

    # Process the output
    answer_start = torch.argmax(outputs.start_logits)
    answer_end = torch.argmax(outputs.end_logits) + 1

    # Get the tokens for the entire input
    all_tokens = tokenizer.convert_ids_to_tokens(inputs["input_ids"][0])

    # Extract the answer tokens
    answer_tokens = all_tokens[answer_start:answer_end]

    # Convert answer tokens back to text
    answer = tokenizer.convert_tokens_to_string(answer_tokens)

    # Clean up the answer
    answer = answer.strip()

    # If the answer is empty or just punctuation, return a message
    if not answer or answer.strip('.,;:!?') == '':
        return "I'm sorry, I couldn't find a specific answer in the given context."

    return answer
```

```
def print_colored_answer(answer):
    print(f"{RED}Answer:{RESET}")
    print(f"{RED}{'=' * 50}{RESET}")
    print(f"{RED}{answer}{RESET}")
    print(f"{RED}{'=' * 50}{RESET}")

print("\nProcessing the first question from the dataset as an exampl
e:")
# Example usage with the first row of the dataset
first_row = df.iloc[0]
question = first_row['question']
context = first_row['answer']

print(f"Question: {question}")
print(f"Context: {context}")  # Print the full context

answer = answer_question(question, context)
print_colored_answer(answer)
print()

print("Now entering interactive mode. You can ask your own medica
l questions.")
# Interactive loop for user questions
while True:
    user_question = input("Enter your medical question (or 'quit' to exi
t): ")
    if user_question.lower() == 'quit':
        break

    # Find the most relevant context (simple approach: exact match)
    relevant_rows = df[df['question'].str.contains(user_question, case
=False, na=False)]

    if not relevant_rows.empty:
        for _, row in relevant_rows.iterrows():
            context = row['answer']
            print(f"\nFound relevant context: {context[:100]}...")  # Print fi
rst 100 characters of context
            print("Full context:")
            print(context)
            answer = answer_question(user_question, context)
```

```
            print_colored_answer(answer)

            user_input = input("Is this answer helpful? (yes/no): ").lower()
            if user_input == 'yes':
                break
        else:
            print("Sorry, none of the answers were helpful.")
    else:
        print("Sorry, I couldn't find a relevant context for your question.
")
    print()

print(f"{RESET}Thank you for using the MedQuAD Question Answerin
g System!")
```

```
Question: What is (are) Glaucoma ?
Context: Glaucoma is a group of diseases that can damage the eye's optic nerve and result in vision loss and blindness. W
hile glaucoma can strike anyone, the risk is much greater for people over 60. How Glaucoma Develops  There are several di
fferent types of glaucoma. Most of these involve the drainage system within the eye. At the front of the eye there is a s
mall space called the anterior chamber. A clear fluid flows through this chamber and bathes and nourishes the nearby tiss
ues. (Watch the video to learn more about glaucoma. To enlarge the video, click the brackets in the lower right-hand corn
er. To reduce the video, press the Escape (Esc) button on your keyboard.) In glaucoma, for still unknown reasons, the flu
id drains too slowly out of the eye. As the fluid builds up, the pressure inside the eye rises. Unless this pressure is c
ontrolled, it may cause damage to the optic nerve and other parts of the eye and result in loss of vision. Open-angle Gla
ucoma The most common type of glaucoma is called open-angle glaucoma. In the normal eye, the clear fluid leaves the anter
ior chamber at the open angle where the cornea and iris meet. When fluid reaches the angle, it flows through a spongy mes
hwork, like a drain, and leaves the eye. Sometimes, when the fluid reaches the angle, it passes too slowly through the me
shwork drain, causing the pressure inside the eye to build. If the pressure damages the optic nerve, open-angle glaucoma
-- and vision loss -- may result. There is no cure for glaucoma. Vision lost from the disease cannot be restored. Howeve
r, there are treatments that may save remaining vision. That is why early diagnosis is important.  See this graphic for a
quick overview of glaucoma,  including how many people it affects, whos at risk, what to do if you have it, and how to le
arn more.  See a glossary of glaucoma terms.
Answer:
================================================
a group of diseases that can damage the eye ' s optic nerve and result in vision loss and blindness
================================================
```

Prepared by: Syed Afroz Ali

Data Scientist (Kaggle Grandmaster)

Kaggle: https://www.kaggle.com/pythonafroz

LinkedIn : https://www.linkedin.com/in/syed-afroz-70939914/