

# Final Assignment

January 11, 2025

## Thành viên nhóm

- 22280003 - Phạm Bá Hoàng Anh
- 22280004 - Trương Bình Ba
- 22280013 - Phạm Lê Hồng Đức
- 22280016 - Bạch Ngọc Lê Duy

```
[51]: #!pip install -q -U google-generativeai  
#!pip install langdetect  
#!pip install pycountry  
#!pip install beautifulsoup4  
#!pip install sentence-transformers  
#!pip install faiss-cpu  
#!pip install nltk
```

```
[52]: import requests  
from bs4 import BeautifulSoup  
  
import time  
import random  
import numpy as np  
  
from langdetect import detect  
import pycountry  
from sentence_transformers import SentenceTransformer  
from sklearn.metrics.pairwise import cosine_similarity  
import google.generativeai as genai  
import re  
import nltk  
from nltk.corpus import stopwords  
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to  
[nltk_data] C:\Users\ThinkPad\AppData\Roaming\nltk_data...  
[nltk_data] Package stopwords is already up-to-date!
```

## Question 1: LLM integration (Score: 30%)

The task involves building an AI capable of language translation.

- **1.1 Single Text Translation: (Score: 15%)**

You are asked to write a Python code using the OpenAI API to translate a given text into Vietnamese.

(You should check the text if it's already the destination language).

For example, translating “Hello” into Vietnamese should return “Xin chào”, but “Xin chào” should return the same.

- **1.2 Multiple Texts Translation: (Score: 15%)** Similar to 2.1, but the input is a list of texts.

The Python code should accept a list of strings

and return their translations in the specified language.

For instance, translating [“Hello”, “I am John”, “Tôi là sinh viên”] into Vietnamese should return [“Xin chào”, “Tôi tên là John”, “Tôi là sinh viên”].

```
[53]: # Hàm dịch văn bản đơn lẻ
def translate_single_text(json_input):
    text = json_input['text']
    dest_language = json_input['dest_language']

    # Detect language (kiểm tra text đã là ngôn ngữ đích hay chưa thông qua
    ⇨ thư viện langdetect)
    current_language = detect(text)
    if current_language == dest_language:
        return text

    # Pycountry (đưa dest_language từ dạng ISO 639-1 về dạng ngôn ngữ tự nhiên
    ⇨ để request được rõ ràng)
    dest_language_name = pycountry.languages.get(alpha_2=dest_language).name
    genai.configure(api_key='')
    model = genai.GenerativeModel("gemini-1.5-flash")
    response = model.generate_content(f"Dịch '{text}' sang ngôn ngữ
    ⇨ {dest_language_name} mà không thêm thông tin khác kể cả dấu chấm")
    return response.text.strip() # strip(): Xóa khoảng trắng thừa ở đầu và cuối.
```

```
[54]: # Hàm dịch danh sách văn bản
def translate_multiple_texts(json_input):
    texts = json_input['text']
    dest_language = json_input['dest_language']
    translations = []

    for text in texts:
        multiple_texts = translate_single_text({'text': text, 'dest_language':
        ⇨ dest_language})
        translations.append(multiple_texts)

    return translations
```

```
[55]: # Dữ liệu đầu vào
json_1 = {
    'text': 'Hello',
    'dest_language': 'vi'
}
json_2 = {
    'text': ['Hello', 'I am John', 'Tôi là sinh viên'],
    'dest_language': 'vi'
}
```

```
[56]: # Kiểm tra kết quả
translate_single_text(json_1)
```

```
[56]: 'Xin chào'
```

```
[57]: translate_multiple_texts(json_2)
```

```
[57]: ['Xin chào', 'Tôi là John', 'Tôi là sinh viên']
```

## Question 2: Chatbot Development (Score: 70%)

Assignment Test: Chatbot Development from Website Data. The data is at <https://www.presight.io/privacy-policy.html>

- **2.1 Data Access and Indexing (Score: 40%)**

Tasked with creating a chatbot, begin by web crawling the specified website to gather relevant data, then preprocess and structure this data into a searchable index, ready for query retrieval. (Short version: crawling then embedding data, you can use selenium or requests)

- **2.2 Chatbot Development (Score: 30%)**

Develop a chatbot that employs natural language processing to comprehend user questions, searches the indexed data from 2.1 for the best match, and delivers the most accurate response drawn from the website's information. (Use any distance/similarity metrics to get the best match paragraph then feed to LLM to get answer)

```
[58]: class Chatbot:
    def __init__(self, api_key, url):
        self.url = url
        self.data = []
        self.crawl_data()
        self.sentence_transformer_model =
↳ SentenceTransformer('all-mpnet-base-v2')
        self.embeddings = self.sentence_transformer_model.
↳ encode([item["cleaned"] for item in self.data])
        genai.configure(api_key=api_key)
        self.gemini_model = genai.GenerativeModel("gemini-1.5-flash")

    def crawl_data(self):
```

```

response = requests.get(self.url)
soup = BeautifulSoup(response.content, "html.parser")

# Danh sách các thẻ cần trích xuất
tags_to_extract = ["h2", "i", "p", "li"]
segments = []
current_segment = []

for element in soup.find_all(tags_to_extract):
    if element.name in ["h2", "i"]:
        if current_segment:
            segments.append(current_segment)
            current_segment = [element.text.strip()]
        elif element.name == "li":
            current_segment.append(element.text.strip().rstrip(".")) + ';')
        else:
            current_segment.append(element.text.strip())
# Thêm đoạn cuối cùng nếu còn nội dung
if current_segment:
    segments.append(current_segment)

segments[0][0] = segments[1][0] + " - " + segments[2][0]
segments[1][0], segments[2][0] = segments[2][1].split(" ")
segments[5][0] = segments[4][0] + " - " + segments[5][0]
segments[6][0] = segments[4][0] + " - " + segments[6][0]
segments[10][0] = segments[9][0] + " - " + segments[10][0]
segments[11][0] = segments[9][0] + " - " + segments[11][0]
del segments[23][2:]
del segments[9]
del segments[4]
del segments[2][1]
del segments[0][1]

# Hiển thị nội dung trích xuất
for idx, segment in enumerate(segments):
    if (idx > 2):
        segments[idx][0] = segments[idx][0].upper() + ":"
        segments[idx] = " ".join(segment).strip()

    self.data.append({
        "original": segments[idx],
        "cleaned": self.clean(segments[idx])
    })
    print(f"Segment {idx}:")
    print("- " + "\n- ".join(segment))

```

```

        print("Original:", self.data[-1]["original"])
        print("Cleaned :", self.data[-1]["cleaned"])
        print("-" * 50)

def clean(self, text):
    """
    Làm sạch đoạn văn bản trước khi embedding.
    Args:
        text (str): Văn bản cần làm sạch.
    Returns:
        str: Đoạn văn bản đã được làm sạch.
    """
    # text = re.sub(r'http\S+/www\S+', '', text) # Loại bỏ URL
    # text = re.sub(r'\d+', '', text) # Loại bỏ các số
    text = text.lower() # Chuyển văn bản về chữ thường
    text = re.sub(r'\S+@\S+\.\S+', 'email', text) # Thay email bằng email
    text = re.sub(r'[^w\s]', '', text) # Loại bỏ ký tự đặc biệt và dấu câu
    ↪ (trừ ký tự khoảng trắng)
    # Loại bỏ stop words
    stop_words = set(stopwords.words('english'))
    # Danh sách các đại từ cần giữ lại
    'pronouns_to_keep' = {
        "i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you",
        ↪ "you're", "you've", "you'll", "you'd", "your",
        "yours", "yourself", "yourselves", "he", "him", "his", "himself",
        ↪ "she", "she's", "her", "hers", "herself", "it",
        "it's", "its", "itself", "they", "them", "their", "theirs",
        ↪ "themselves"}

    # Loại bỏ các đại từ ra khỏi danh sách stop words
    stop_words = stop_words - pronouns_to_keep
    text = ' '.join(word for word in text.split() if word not in stop_words)
    text = re.sub(r'\s+', ' ', text).strip() # Loại bỏ khoảng trắng thừa

    return text

def search_query(self, query, top_k):
    """
    Tìm kiếm các đoạn văn bản liên quan dựa trên câu hỏi.

    Args:
        query (str): Câu hỏi của người dùng.
        top_k (int): Số lượng kết quả trả về.

    Returns:
        contexts: Các đoạn văn bản gốc liên quan nhất.
        scores: Điểm tương đồng của các đoạn văn bản tương ứng.
    """
    # Làm sạch câu hỏi

```

```

query_cleaned = self.clean(query)

# Tạo embedding cho câu hỏi đã làm sạch
query_embedding = self.sentence_transformer_model.
↳ encode([query_cleaned])

# Tính cosine similarity giữa câu hỏi và các embeddings
similarities = cosine_similarity(query_embedding, self.embeddings)

# Lấy chỉ số của các đoạn văn bản có điểm tương đồng cao nhất
top_indices = np.argsort(similarities[0])[:, :-1][:top_k]

# Lấy các văn bản gốc và điểm tương đồng tương ứng
contexts = [self.data[idx]["original"] for idx in top_indices]
scores = [similarities[0][idx] for idx in top_indices]
return contexts, scores

# Tích hợp Gemini Flash 1.5
def generate_answer_with_gemini(self, context, question):
    prompt = (
        f"You are a highly knowledgeable and detail-oriented assistant. "
        f"Begin your answer by addressing the general concept or definition,
↳ of the question directly. "
        f"Then, if necessary, use the provided context to elaborate or,
↳ refine your answer. "
        f"Ensure your response is accurate, detailed, and concise. "
        f"Do not include irrelevant information or assumptions not,
↳ supported by the context.\n\n"
        f"Question:\n{question}\n\n"
        f"Context (if relevant):\n{context}\n\n"
        f"Answer:"
    )

    try:
        response = self.gemini_model.generate_content(prompt)
        return response.text
    except Exception as e:
        return f"Error: {e}"

# Tạo request
def make_request(self, question):
    # Tìm kiếm đoạn văn liên quan
    start = time.time()
    results, scores = self.search_query(question, 3)
    context = "\n".join([f"\t- (Score: {score:.4f}) {item}" for item, score
↳ in zip(results, scores)])
    answer = self.generate_answer_with_gemini(context, question)

```

```

end = time.time()
# Debug: In context và question
print(f"Context : \n{context}")
print(f"Chatbot : {answer.rstrip()}")
print(f"Running time: {end - start:.2f} seconds")
# return answer

```

```

[60]: url = "https://www.presight.io/privacy-policy.html"
      api_key = ''
      chatbot = Chatbot(api_key, url)

```

Segment 0:

- PRIVACY POLICY - Last updated 15 Sep 2023  
Original: PRIVACY POLICY - Last updated 15 Sep 2023  
Cleaned : privacy policy last updated 15 sep 2023  
-----

Segment 1:

- At Presight, we are committed to protecting the privacy of our customers and visitors to our website  
Original: At Presight, we are committed to protecting the privacy of our customers and visitors to our website  
Cleaned : presight committed protecting privacy customers visitors website  
-----

Segment 2:

- This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.  
Original: This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.  
Cleaned : privacy policy explains collect use disclose information customers visitors  
-----

Segment 3:

- INFORMATION COLLECTION AND USE:  
- We collect several different types of information for various purposes to provide and improve our Service to you.  
Original: INFORMATION COLLECTION AND USE: We collect several different types of information for various purposes to provide and improve our Service to you.  
Cleaned : information collection use collect several different types information various purposes provide improve service  
-----

Segment 4:

- TYPES OF DATA COLLECTED - PERSONAL DATA:  
- While using our Service, we may ask you to provide us with certain personally identifiable information that can be used to contact or identify you ("Personal Data"). Personally identifiable information may include, but is not limited to:  
- Email address;  
- First name and last name;

- Phone number;
- Address, State, Province, ZIP/Postal code, City;
- Cookies and Usage Data;

Original: TYPES OF DATA COLLECTED - PERSONAL DATA: While using our Service, we may ask you to provide us with certain personally identifiable information that can be used to contact or identify you ("Personal Data"). Personally identifiable information may include, but is not limited to: Email address; First name and last name; Phone number; Address, State, Province, ZIP/Postal code, City; Cookies and Usage Data;

Cleaned : types data collected personal data using service may ask provide us certain personally identifiable information used contact identify personal data personally identifiable information may include limited email address first name last name phone number address state province zippostal code city cookies usage data

-----

Segment 5:

- TYPES OF DATA COLLECTED - USAGE DATA:

- We may also collect information that your browser sends whenever you visit our Service or when you access the Service by or through a mobile device ("Usage Data"). This Usage Data may include information such as your computer's Internet Protocol address (e.g. IP address), browser type, browser version, the pages of our Service that you visit, the time and date of your visit, the time spent on those pages, unique device identifiers, and other diagnostic data.

Original: TYPES OF DATA COLLECTED - USAGE DATA: We may also collect information that your browser sends whenever you visit our Service or when you access the Service by or through a mobile device ("Usage Data"). This Usage Data may include information such as your computer's Internet Protocol address (e.g. IP address), browser type, browser version, the pages of our Service that you visit, the time and date of your visit, the time spent on those pages, unique device identifiers, and other diagnostic data.

Cleaned : types data collected usage data may also collect information browser sends whenever visit service access service mobile device usage data usage data may include information computers internet protocol address eg ip address browser type browser version pages service visit time date visit time spent pages unique device identifiers diagnostic data

-----

Segment 6:

- USE OF DATA:

- Presight uses the collected data for various purposes:

- To provide and maintain our Service;
- To notify you about changes to our Service;
- To allow you to participate in interactive features of our Service when you choose to do so;
- To provide customer support;
- To gather analysis or valuable information so that we can improve our Service;
- To monitor the usage of our Service;
- To detect, prevent and address technical issues;

Original: USE OF DATA: Presight uses the collected data for various purposes: To



provide and maintain our Service; To notify you about changes to our Service; To allow you to participate in interactive features of our Service when you choose to do so; To provide customer support; To gather analysis or valuable information so that we can improve our Service; To monitor the usage of our Service; To detect, prevent and address technical issues;

Cleaned : use data presight uses collected data various purposes provide maintain service notify changes service allow participate interactive features service choose provide customer support gather analysis valuable information improve service monitor usage service detect prevent address technical issues

-----

Segment 7:

- CONSENT:

- As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

Original: CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

Cleaned : consent personal information collected asked confirm information correct prior submitting presight

-----

Segment 8:

- ACCESS TO PERSONAL INFORMATION - ACCESSING YOUR PERSONAL INFORMATION:

- You have the right to access all of your personal information that we hold.

Through the application, you can correct, amend, or append your personal information by logging into the application and navigating to your settings and profile.

Original: ACCESS TO PERSONAL INFORMATION - ACCESSING YOUR PERSONAL INFORMATION:

You have the right to access all of your personal information that we hold.

Through the application, you can correct, amend, or append your personal information by logging into the application and navigating to your settings and profile.

Cleaned : access personal information accessing personal information right access personal information hold application correct amend append personal information logging application navigating settings profile

-----

Segment 9:

- ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS:

- Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

Original: ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

Cleaned : access personal information automated edit checks presight employs automated edit checks ensure data entry fields completed properly collecting personal information edit checks help maintain data integrity accuracy

encouraged provide complete valid information ensure smooth processing personal data

-----  
Segment 10:

- DISCLOSURE OF INFORMATION:

- We may disclose your application data to third-party service providers who help us provide our services such as Datadog, AWS, Google Cloud and Google Workspace. We may also disclose your information in response to a legal request, such as a subpoena or court order, or to protect our rights or the rights of others.

Original: DISCLOSURE OF INFORMATION: We may disclose your application data to third-party service providers who help us provide our services such as Datadog, AWS, Google Cloud and Google Workspace. We may also disclose your information in response to a legal request, such as a subpoena or court order, or to protect our rights or the rights of others.

Cleaned : disclosure information may disclose application data thirdparty service providers help us provide services datadog aws google cloud google workspace may also disclose information response legal request subpoena court order protect rights rights others

-----  
Segment 11:

- SHARING OF PERSONAL DATA:

- Your personal data will not be subject to sharing, transfer, rental or exchange for the benefit of third parties, including AI models.

Original: SHARING OF PERSONAL DATA: Your personal data will not be subject to sharing, transfer, rental or exchange for the benefit of third parties, including AI models.

Cleaned : sharing personal data personal data subject sharing transfer rental exchange benefit third parties including ai models

-----  
Segment 12:

- GOOGLE USER DATA AND GOOGLE WORKSPACE APIS:

- In all cases when users authenticate the platform to Google Workspace, the following applies:

- We do not retain or use Google User Data to develop, improve, or train generalized/non-personalized AI and/or ML models;

- We do not use Google Workspace APIs to develop, improve, or train generalized/non-personalized AI and/or ML models;

- We do not transfer Google User Data to third-party AI tools for the purpose of developing, improving, or training generalized or non-personalized AI and/or ML models;

Original: GOOGLE USER DATA AND GOOGLE WORKSPACE APIS: In all cases when users authenticate the platform to Google Workspace, the following applies: We do not retain or use Google User Data to develop, improve, or train generalized/non-personalized AI and/or ML models; We do not use Google Workspace APIs to develop, improve, or train generalized/non-personalized AI and/or ML models; We do not transfer Google User Data to third-party AI tools for the purpose of developing, improving, or training generalized or non-personalized AI and/or ML

models;

Cleaned : google user data google workspace apis cases users authenticate platform google workspace following applies retain use google user data develop improve train generalizednonpersonalized ai andor ml models use google workspace apis develop improve train generalizednonpersonalized ai andor ml models transfer google user data thirdparty ai tools purpose developing improving training generalized nonpersonalized ai andor ml models

-----  
Segment 13:

- DATA SECURITY:

- All data is encrypted both in transit and at rest, using industry-standard encryption methods;
- We regularly perform security audits and vulnerability assessments to ensure the safety of our platform and the data stored within it;
- Our employees are trained on best practices for data security, and access to customer data is restricted on a need-to-know basis;

Original: DATA SECURITY: All data is encrypted both in transit and at rest, using industry-standard encryption methods; We regularly perform security audits and vulnerability assessments to ensure the safety of our platform and the data stored within it; Our employees are trained on best practices for data security, and access to customer data is restricted on a need-to-know basis;

Cleaned : data security data encrypted transit rest using industrystandard encryption methods regularly perform security audits vulnerability assessments ensure safety platform data stored within employees trained best practices data security access customer data restricted needtoknow basis

-----  
Segment 14:

- DATA RETENTION & DISPOSAL:

- Customer data is retained for as long as the account is in active status. Data enters an "expired" state when the account is voluntarily closed. Expired account data will be retained for 60 days. After this period, the account and related data will be removed.

Original: DATA RETENTION & DISPOSAL: Customer data is retained for as long as the account is in active status. Data enters an "expired" state when the account is voluntarily closed. Expired account data will be retained for 60 days. After this period, the account and related data will be removed.

Cleaned : data retention disposal customer data retained long account active status data enters expired state account voluntarily closed expired account data retained 60 days period account related data removed

-----  
Segment 15:

- QUALITY, INCLUDING DATA SUBJECTS' RESPONSIBILITIES FOR QUALITY:

- We are committed to maintaining the quality and accuracy of the personal information we collect and process;
- We rely on data subjects to provide accurate and up-to-date information;
- Data subjects have the responsibility to inform us of any changes or inaccuracies in their personal data;
- If you believe that any information we hold about you is inaccurate,

incomplete, or outdated, please contact us promptly to rectify the information;  
Original: QUALITY, INCLUDING DATA SUBJECTS' RESPONSIBILITIES FOR QUALITY: We are committed to maintaining the quality and accuracy of the personal information we collect and process; We rely on data subjects to provide accurate and up-to-date information; Data subjects have the responsibility to inform us of any changes or inaccuracies in their personal data; If you believe that any information we hold about you is inaccurate, incomplete, or outdated, please contact us promptly to rectify the information;

Cleaned : quality including data subjects responsibilities quality committed maintaining quality accuracy personal information collect process rely data subjects provide accurate uptodate information data subjects responsibility inform us changes inaccuracies personal data believe information hold inaccurate incomplete outdated please contact us promptly rectify information

-----  
Segment 16:

- MONITORING AND ENFORCEMENT:

- We regularly monitor its data processing activities to ensure compliance with this privacy policy and applicable data protection laws;  
- In the event of a data breach or any unauthorized access to your personal information, we will notify you and the appropriate authorities as required by law;

- We committed to cooperating with data protection authorities and complying with their advice and decisions regarding data protection and privacy matters;

Original: MONITORING AND ENFORCEMENT: We regularly monitor its data processing activities to ensure compliance with this privacy policy and applicable data protection laws; In the event of a data breach or any unauthorized access to your personal information, we will notify you and the appropriate authorities as required by law; We committed to cooperating with data protection authorities and complying with their advice and decisions regarding data protection and privacy matters;

Cleaned : monitoring enforcement regularly monitor data processing activities ensure compliance privacy policy applicable data protection laws event data breach unauthorized access personal information notify appropriate authorities required law committed cooperating data protection authorities complying advice decisions regarding data protection privacy matters

-----  
Segment 17:

- COOKIES:

- We use cookies to enhance your experience on our website. You can control the use of cookies through your web browser settings.

Original: COOKIES: We use cookies to enhance your experience on our website. You can control the use of cookies through your web browser settings.

Cleaned : cookies use cookies enhance experience website control use cookies web browser settings

-----  
Segment 18:

- THIRD-PARTY WEBSITES:

- Our website may contain links to third-party websites. We are not responsible

for the privacy practices or content of those websites.

Original: THIRD-PARTY WEBSITES: Our website may contain links to third-party websites. We are not responsible for the privacy practices or content of those websites.

Cleaned : thirdparty websites website may contain links thirdparty websites responsible privacy practices content websites

---

Segment 19:

- CHANGES TO PRIVACY POLICY:

- We may update this Privacy Policy from time to time. The updated Privacy Policy will be posted on our website.

Original: CHANGES TO PRIVACY POLICY: We may update this Privacy Policy from time to time. The updated Privacy Policy will be posted on our website.

Cleaned : changes privacy policy may update privacy policy time time updated privacy policy posted website

---

Segment 20:

- CONTACT US:

- If you have any questions about this Privacy Policy, please contact us through the customer portal or by email at [presight@presight.io](mailto:presight@presight.io).

Original: CONTACT US: If you have any questions about this Privacy Policy, please contact us through the customer portal or by email at [presight@presight.io](mailto:presight@presight.io).

Cleaned : contact us questions privacy policy please contact us customer portal email email

---

Segment 21:

- PURPOSEFUL USE ONLY:

- We commit to only use personal information for the purposes identified in the entity's privacy policy.

Original: PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

Cleaned : purposeful use commit use personal information purposes identified entitys privacy policy

---

```
[64]: questions = [
    {"section": "URELATED QUESTION",
     "question": "What is policy?"},

    {"section": "PRIVACY POLICY - Last updated 15 Sep 2023",
     "question": "When was the privacy policy last updated?"},

    {"section": "Commitment to Privacy",
     "question": "What does Presight commit to protecting for its customers and
↵visitors?"},
```

```

{"section": "Privacy Policy Explanation",
 "question": "What does this Privacy Policy explain regarding customers and
↳visitors?"},

{"section": "INFORMATION COLLECTION AND USE",
 "question": "Why does Presight collect different types of information?"},

{"section": "TYPES OF DATA COLLECTED - PERSONAL DATA",
 "question": "What types of personally identifiable information does
↳Presight collect?"},

{"section": "TYPES OF DATA COLLECTED - USAGE DATA",
 "question": "What is included in the usage data collected by Presight?"},

{"section": "USE OF DATA",
 "question": "For what purposes does Presight use the collected data?"},

{"section": "CONSENT",
 "question": "How does Presight ensure that personal information submitted
↳is correct?"},

{"section": "ACCESS TO PERSONAL INFORMATION - ACCESSING YOUR PERSONAL
↳INFORMATION",
 "question": "How can users access and update their personal information
↳held by Presight?"},

{"section": "ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS",
 "question": "What is the purpose of automated edit checks in collecting
↳personal information?"},

{"section": "DISCLOSURE OF INFORMATION",
 "question": "Under what circumstances might Presight disclose application
↳data to third parties?"},

{"section": "SHARING OF PERSONAL DATA",
 "question": "Does Presight share personal data with third parties or AI
↳models?"},

{"section": "GOOGLE USER DATA AND GOOGLE WORKSPACE APIS",
 "question": "What restrictions does Presight place on the use of Google
↳User Data and Google Workspace APIs?"},

{"section": "DATA SECURITY",
 "question": "What encryption and security measures does Presight use to
↳protect customer data?"},

```

```

{"section": "DATA RETENTION & DISPOSAL",
 "question": "How long does Presight retain customer data after account_
↪closure?"},

{"section": "QUALITY, INCLUDING DATA SUBJECTS' RESPONSIBILITIES FOR_
↪QUALITY",
 "question": "What responsibilities do users have in maintaining the_
↪accuracy of their personal data?"},

{"section": "MONITORING AND ENFORCEMENT",
 "question": "What actions does Presight take to monitor data compliance_
↪and handle data breaches?"},

{"section": "COOKIES",
 "question": "How can users manage cookies on Presight's website?"},

{"section": "THIRD-PARTY WEBSITES",
 "question": "What is Presight's responsibility regarding third-party_
↪websites?"},

{"section": "CHANGES TO PRIVACY POLICY",
 "question": "Where will updates to the Privacy Policy be posted?"},

{"section": "CONTACT US",
 "question": "What is the email address provided for contacting Presight if_
↪I have questions about the Privacy Policy?"},

{"section": "PURPOSEFUL USE ONLY",
 "question": "For what purposes does Presight commit to using personal_
↪information?"}
]

# === Vòng lặp tự động hỏi ===
for item in questions:
    delay = random.uniform(0, 2)
    time.sleep(delay)
    print("\n==== Debug Information =====")
    print(f"Section: {item['section']}")
    print(f"Question: {item['question']}")
    chatbot.make_request(item['question'])

```

==== Debug Information =====

Section: URELATED QUESTION

Question: What is policy?

Context :

- (Score: 0.5053) PRIVACY POLICY - Last updated 15 Sep 2023

- (Score: 0.3356) CHANGES TO PRIVACY POLICY: We may update this Privacy Policy from time to time. The updated Privacy Policy will be posted on our website.

- (Score: 0.3030) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.  
Chatbot : Policy is a course or principle of action adopted or proposed by a government, party, business, or individual. The provided text exemplifies this definition through the mention of a "PRIVACY POLICY," outlining principles for the handling of personal information. The context further clarifies that a policy can be updated and that adherence to its stated purposes is crucial.

Running time: 0.96 seconds

==== Debug Information =====

Section: PRIVACY POLICY - Last updated 15 Sep 2023

Question: When was the privacy policy last updated?

Context :

- (Score: 0.8726) PRIVACY POLICY - Last updated 15 Sep 2023

- (Score: 0.7766) CHANGES TO PRIVACY POLICY: We may update this Privacy Policy from time to time. The updated Privacy Policy will be posted on our website.

- (Score: 0.5045) This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.

Chatbot : The privacy policy was last updated on September 15, 2023.

Running time: 0.62 seconds

==== Debug Information =====

Section: Commitment to Privacy

Question: What does Presight commit to protecting for its customers and visitors?

Context :

- (Score: 0.7479) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.4817) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

- (Score: 0.4231) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT

CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

Chatbot : Presight commits to protecting the privacy of its customers and visitors to its website. This commitment includes using personal information only for the purposes specified in its privacy policy.

Running time: 0.74 seconds

==== Debug Information =====

Section: Privacy Policy Explanation

Question: What does this Privacy Policy explain regarding customers and



visitors?

Context :

- (Score: 0.7975) This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.

- (Score: 0.6648) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.5800) PRIVACY POLICY - Last updated 15 Sep 2023

Chatbot : This Privacy Policy explains how the organization collects, uses, and discloses information about its customers and website visitors.

Running time: 0.68 seconds

==== Debug Information =====

Section: INFORMATION COLLECTION AND USE

Question: Why does Presight collect different types of information?

Context :

- (Score: 0.4486) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT

CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.4075) INFORMATION COLLECTION AND USE: We collect several different types of information for various purposes to provide and improve our Service to you.

- (Score: 0.3951) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

Chatbot : Presight collects different types of information to provide and improve its service. The provided text highlights this as the primary reason, emphasizing the use of this information to facilitate smooth processing of personal data and maintain data integrity and accuracy.

Running time: 1.11 seconds

==== Debug Information =====

Section: TYPES OF DATA COLLECTED - PERSONAL DATA

Question: What types of personally identifiable information does Presight collect?

Context :

- (Score: 0.5352) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.5266) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT

CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.5106) TYPES OF DATA COLLECTED - PERSONAL DATA: While using our Service, we may ask you to provide us with certain personally identifiable information that can be used to contact or identify you ("Personal Data").

Personally identifiable information may include, but is not limited to: Email address; First name and last name; Phone number; Address, State, Province, ZIP/Postal code, City; Cookies and Usage Data;  
Chatbot : Presight collects personally identifiable information (PII) that can be used to contact or identify an individual. This includes, but is not limited to: email address; first name and last name; phone number; address (including state, province, ZIP/postal code, and city); and cookies and usage data.  
Running time: 0.94 seconds

==== Debug Information =====

Section: TYPES OF DATA COLLECTED - USAGE DATA

Question: What is included in the usage data collected by Presight?

Context :

- (Score: 0.5410) TYPES OF DATA COLLECTED - USAGE DATA: We may also collect information that your browser sends whenever you visit our Service or when you access the Service by or through a mobile device ("Usage Data"). This Usage Data may include information such as your computer's Internet Protocol address (e.g. IP address), browser type, browser version, the pages of our Service that you visit, the time and date of your visit, the time spent on those pages, unique device identifiers, and other diagnostic data.

- (Score: 0.4673) USE OF DATA: Presight uses the collected data for various purposes: To provide and maintain our Service; To notify you about changes to our Service; To allow you to participate in interactive features of our Service when you choose to do so; To provide customer support; To gather analysis or valuable information so that we can improve our Service; To monitor the usage of our Service; To detect, prevent and address technical issues;

- (Score: 0.4561) TYPES OF DATA COLLECTED - PERSONAL DATA: While using our Service, we may ask you to provide us with certain personally identifiable information that can be used to contact or identify you ("Personal Data").

Personally identifiable information may include, but is not limited to: Email address; First name and last name; Phone number; Address, State, Province, ZIP/Postal code, City; Cookies and Usage Data;

Chatbot : Presight's usage data includes information sent by a user's browser or mobile device when accessing the service. This encompasses the user's IP address, browser type and version, visited pages, timestamps of visits, time spent on pages, unique device identifiers, and other diagnostic data.

Running time: 0.91 seconds

==== Debug Information =====

Section: USE OF DATA

Question: For what purposes does Presight use the collected data?

Context :

- (Score: 0.5950) USE OF DATA: Presight uses the collected data for various purposes: To provide and maintain our Service; To notify you about changes to our Service; To allow you to participate in interactive features of our Service when you choose to do so; To provide customer support; To gather analysis or valuable information so that we can improve our Service; To monitor the usage of our Service; To detect, prevent and address technical issues;

- (Score: 0.4689) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT  
CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.4669) At Presight, we are committed to protecting the privacy of our customers and visitors to our website  
Chatbot : Presight uses collected data to provide and maintain its service, notify users of service changes, enable participation in interactive features (when chosen by the user), provide customer support, gather analysis for service improvement, monitor service usage, and detect, prevent, and address technical issues.

Running time: 0.88 seconds

==== Debug Information =====

Section: CONSENT

Question: How does Presight ensure that personal information submitted is correct?

Context :

- (Score: 0.7380) CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

- (Score: 0.6962) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT  
CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.5694) At Presight, we are committed to protecting the privacy of our customers and visitors to our website  
Chatbot : Presight employs two primary methods to ensure the accuracy of submitted personal information: first, users are prompted to confirm the correctness of their information before submission; second, automated edit checks are implemented to validate data entry during the collection process. These checks help maintain data integrity and accuracy.

Running time: 0.93 seconds

==== Debug Information =====

Section: ACCESS TO PERSONAL INFORMATION - ACCESSING YOUR PERSONAL INFORMATION

Question: How can users access and update their personal information held by Presight?

Context :

- (Score: 0.5717) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT  
CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal

data.

- (Score: 0.5442) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.5279) ACCESS TO PERSONAL INFORMATION - ACCESSING YOUR PERSONAL INFORMATION: You have the right to access all of your personal information that we hold. Through the application, you can correct, amend, or append your personal information by logging into the application and navigating to your settings and profile.

Chatbot : Presight users can access and update their personal information held by the company through the Presight application. To do so, users must log in to the application and then navigate to their settings and profile section. Within this section, they can correct, amend, or append their personal information.

Running time: 0.93 seconds

==== Debug Information =====

Section: ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS

Question: What is the purpose of automated edit checks in collecting personal information?

Context :

- (Score: 0.6318) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.4621) CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

- (Score: 0.4437) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

Chatbot : Automated edit checks, in the context of collecting personal information, serve to ensure the accuracy and integrity of the data entered. They verify that data entry fields are completed correctly, preventing errors and inconsistencies in the collected personal information.

Running time: 0.85 seconds

==== Debug Information =====

Section: DISCLOSURE OF INFORMATION

Question: Under what circumstances might Presight disclose application data to third parties?

Context :

- (Score: 0.6371) DISCLOSURE OF INFORMATION: We may disclose your application data to third-party service providers who help us provide our services such as Datadog, AWS, Google Cloud and Google Workspace. We may also disclose your information in response to a legal request, such as a subpoena or court order, or to protect our rights or the rights of others.

- (Score: 0.5875) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.5602) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.  
Chatbot : Presight may disclose application data to third-party service providers (such as Datadog, AWS, Google Cloud, and Google Workspace) that assist in providing their services. Additionally, Presight may disclose information in response to legal requests like subpoenas or court orders, or to protect its own rights or the rights of others.  
Running time: 0.97 seconds

==== Debug Information =====

Section: SHARING OF PERSONAL DATA

Question: Does Presight share personal data with third parties or AI models?

Context :

- (Score: 0.6699) SHARING OF PERSONAL DATA: Your personal data will not be subject to sharing, transfer, rental or exchange for the benefit of third parties, including AI models.

- (Score: 0.5658) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.5439) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

Chatbot : Presight explicitly states that personal data will not be shared, transferred, rented, or exchanged with third parties, including AI models.

Running time: 0.77 seconds

==== Debug Information =====

Section: GOOGLE USER DATA AND GOOGLE WORKSPACE APIS

Question: What restrictions does Presight place on the use of Google User Data and Google Workspace APIs?

Context :

- (Score: 0.5955) GOOGLE USER DATA AND GOOGLE WORKSPACE APIS: In all cases when users authenticate the platform to Google Workspace, the following applies: We do not retain or use Google User Data to develop, improve, or train generalized/non-personalized AI and/or ML models; We do not use Google Workspace APIs to develop, improve, or train generalized/non-personalized AI and/or ML models; We do not transfer Google User Data to third-party AI tools for the purpose of developing, improving, or training generalized or non-personalized AI and/or ML models;

- (Score: 0.4118) DISCLOSURE OF INFORMATION: We may disclose your application data to third-party service providers who help us provide our services such as Datadog, AWS, Google Cloud and Google Workspace. We may also disclose your information in response to a legal request, such as a subpoena or court order, or to protect our rights or the rights of others.

- (Score: 0.3917) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

Chatbot : Presight restricts the use of Google User Data and Google Workspace APIs to explicitly prohibit their use in the development, improvement, or training of generalized or non-personalized AI and/or ML models. This restriction applies to Presight's own use of this data and also prevents the transfer of Google User Data to third-party AI tools for the same purpose.  
Running time: 1.00 seconds

==== Debug Information =====

Section: DATA SECURITY

Question: What encryption and security measures does Presight use to protect customer data?

Context :

- (Score: 0.6461) At Presight, we are committed to protecting the privacy of our customers and visitors to our website
- (Score: 0.6209) DATA SECURITY: All data is encrypted both in transit and at rest, using industry-standard encryption methods; We regularly perform security audits and vulnerability assessments to ensure the safety of our platform and the data stored within it; Our employees are trained on best practices for data security, and access to customer data is restricted on a need-to-know basis;

- (Score: 0.4956) MONITORING AND ENFORCEMENT: We regularly monitor its data processing activities to ensure compliance with this privacy policy and applicable data protection laws; In the event of a data breach or any unauthorized access to your personal information, we will notify you and the appropriate authorities as required by law; We committed to cooperating with data protection authorities and complying with their advice and decisions regarding data protection and privacy matters;

Chatbot : Presight employs industry-standard encryption methods to protect customer data both in transit and at rest. They also conduct regular security audits and vulnerability assessments, restrict data access to a need-to-know basis through employee training on data security best practices. In the event of a data breach, they are committed to notifying affected customers and relevant authorities as legally required.

Running time: 1.11 seconds

==== Debug Information =====

Section: DATA RETENTION & DISPOSAL

Question: How long does Presight retain customer data after account closure?

Context :

- (Score: 0.6463) DATA RETENTION & DISPOSAL: Customer data is retained for as long as the account is in active status. Data enters an "expired" state when the account is voluntarily closed. Expired account data will be retained for 60 days. After this period, the account and related data will be removed.

- (Score: 0.4315) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.3422) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

Chatbot : Presight retains customer data for 60 days after account closure.

After this 60-day period, the account and associated data are removed.

Running time: 0.73 seconds

==== Debug Information =====

Section: QUALITY, INCLUDING DATA SUBJECTS' RESPONSIBILITIES FOR QUALITY

Question: What responsibilities do users have in maintaining the accuracy of their personal data?

Context :

- (Score: 0.6616) QUALITY, INCLUDING DATA SUBJECTS' RESPONSIBILITIES FOR QUALITY: We are committed to maintaining the quality and accuracy of the personal information we collect and process; We rely on data subjects to provide accurate and up-to-date information; Data subjects have the responsibility to inform us of any changes or inaccuracies in their personal data; If you believe that any information we hold about you is inaccurate, incomplete, or outdated, please contact us promptly to rectify the information;

- (Score: 0.5953) MONITORING AND ENFORCEMENT: We regularly monitor its data processing activities to ensure compliance with this privacy policy and applicable data protection laws; In the event of a data breach or any unauthorized access to your personal information, we will notify you and the appropriate authorities as required by law; We committed to cooperating with data protection authorities and complying with their advice and decisions regarding data protection and privacy matters;

- (Score: 0.5708) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

Chatbot : Users have a responsibility to ensure the accuracy of their personal data by providing accurate and up-to-date information initially and promptly informing the data controller of any subsequent changes or inaccuracies. This includes notifying the controller if any held information is believed to be inaccurate, incomplete, or outdated.

Running time: 0.90 seconds

==== Debug Information =====

Section: MONITORING AND ENFORCEMENT

Question: What actions does Presight take to monitor data compliance and handle data breaches?

Context :

- (Score: 0.6573) MONITORING AND ENFORCEMENT: We regularly monitor its data processing activities to ensure compliance with this privacy policy and applicable data protection laws; In the event of a data breach or any unauthorized access to your personal information, we will notify you and the appropriate authorities as required by law; We committed to cooperating with data protection authorities and complying with their advice and decisions regarding data protection and privacy matters;

- (Score: 0.4914) ACCESS TO PERSONAL INFORMATION - AUTOMATED EDIT

CHECKS: Presight employs automated edit checks to ensure that data entry fields are completed properly when collecting personal information. These edit checks help maintain data integrity and accuracy. You are encouraged to provide complete and valid information to ensure the smooth processing of their personal data.

- (Score: 0.4884) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

Chatbot : Presight monitors its data processing activities to ensure compliance with its privacy policy and applicable data protection laws. In the event of a data breach or unauthorized access to personal information, Presight will notify affected individuals and the appropriate authorities as legally required. Furthermore, Presight is committed to cooperating with data protection authorities and adhering to their guidance on data protection and privacy matters.

Running time: 1.00 seconds

==== Debug Information =====

Section: COOKIES

Question: How can users manage cookies on Presight's website?

Context :

- (Score: 0.6196) COOKIES: We use cookies to enhance your experience on our website. You can control the use of cookies through your web browser settings.

- (Score: 0.6077) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.3810) This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.

Chatbot : Users can manage cookies on Presight's website through their web browser settings. The provided text explicitly states this as a method of controlling cookie use.

Running time: 0.73 seconds

==== Debug Information =====

Section: THIRD-PARTY WEBSITES

Question: What is Presight's responsibility regarding third-party websites?

Context :

- (Score: 0.7294) THIRD-PARTY WEBSITES: Our website may contain links to third-party websites. We are not responsible for the privacy practices or content of those websites.

- (Score: 0.5256) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

- (Score: 0.4115) CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

Chatbot : Presight is not responsible for the privacy practices or content of third-party websites linked from its own website.

Running time: 0.66 seconds



===== Debug Information =====

Section: CHANGES TO PRIVACY POLICY

Question: Where will updates to the Privacy Policy be posted?

Context :

- (Score: 0.8170) PRIVACY POLICY - Last updated 15 Sep 2023

- (Score: 0.7549) CHANGES TO PRIVACY POLICY: We may update this Privacy Policy from time to time. The updated Privacy Policy will be posted on our website.

- (Score: 0.5836) This Privacy Policy explains how we collect, use, and disclose information about our customers and visitors.

Chatbot : Updates to the Privacy Policy will be posted on the website.

Running time: 0.63 seconds

===== Debug Information =====

Section: CONTACT US

Question: What is the email address provided for contacting Presight if I have questions about the Privacy Policy?

Context :

- (Score: 0.6253) CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

- (Score: 0.5991) PRIVACY POLICY - Last updated 15 Sep 2023

- (Score: 0.5401) CONTACT US: If you have any questions about this Privacy Policy, please contact us through the customer portal or by email at [presight@presight.io](mailto:presight@presight.io).

Chatbot : The provided text states that to contact Presight with questions about the Privacy Policy, you can use the email address [presight@presight.io](mailto:presight@presight.io).

Running time: 0.69 seconds

===== Debug Information =====

Section: PURPOSEFUL USE ONLY

Question: For what purposes does Presight commit to using personal information?

Context :

- (Score: 0.7081) PURPOSEFUL USE ONLY: We commit to only use personal information for the purposes identified in the entity's privacy policy.

- (Score: 0.6420) CONSENT: As personal information is collected, you will be asked to confirm that your information is correct prior to submitting it to Presight.

- (Score: 0.6183) At Presight, we are committed to protecting the privacy of our customers and visitors to our website

Chatbot : Presight commits to using personal information only for the purposes specified in its privacy policy.

Running time: 0.63 seconds

**Đánh giá chatbot:** Chatbot có khả năng trả lời chính xác các câu hỏi được đưa ra. Khi câu hỏi chứa nhiều từ khóa liên quan đến section tương ứng, khả năng chatbot cung cấp câu trả lời đúng và phù hợp sẽ cao hơn.

```
[50]: # Chạy Chatbot
      '''print("Chatbot is ready! Type 'exit' or 'quit' to end the session.")
      while True:
          user_input = input("You: ")
          if user_input.lower() in ["exit", "quit"]:
              print("Chatbot: Goodbye!")
              break
          chatbot(user_input)'''
```

```
[50]: 'print("Chatbot is ready! Type \'exit\' or \'quit\' to end the session.")\nwhile
True:\n    user_input = input("You: ")\n    if user_input.lower() in ["exit",
"quit"]:\n        print("Chatbot: Goodbye!")\n        break\nchatbot(user_input)'
```