3T1H Team Proposal

Topic: AI chatbox and try-on vitual

Dataset 1

VIETNAM DATATHON 2023

Table of contents

VIETNAM DATATHON 2023

Table of contents

VIETNAM DATATHON 2023

VIETNAM DATATHON 2023

01

Introduction

1. Introduction

Nowadays, shopping is a need for almost everyone, whether it's buying items according to a certain style, following trends, or simply dressing appropriately for the environment they live in. In the current retail market in Vietnam, it's impossible to count the number of brands, ranging from large to small. Many of these brands are looking to deploy chatbots to improve the interaction between customers and the brands.

1. Introduction

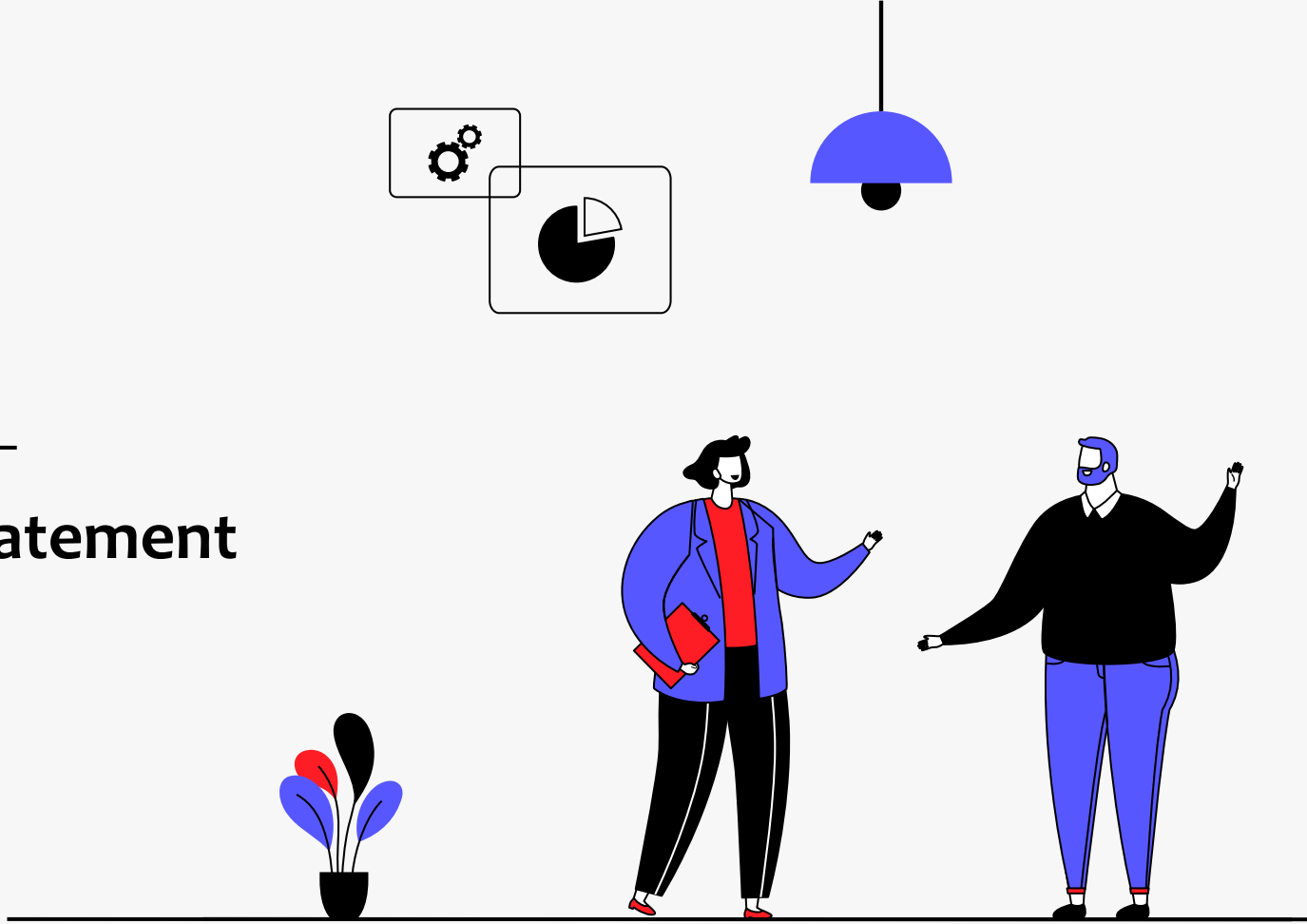
A chatbot with a fast response time can significantly enhance the customer’s experience by providing instant assistance and personalized recommendations. Moreover, it can also cut down the cost of hiring personnel, especially for large brands with a high volume of customers.

Our MVP is a chatbox and used to cater to almost all customers. In an era where e-commerce is booming, we expect this MVP to be able to reach almost all customers. Our MVP is a chatbot consultant that helps customers find their best match products based on their descriptions of what they want. We also want to implement a virtual try-on based on the input image and integrate it into the chatbot.

VIETNAM DATATHON 2023

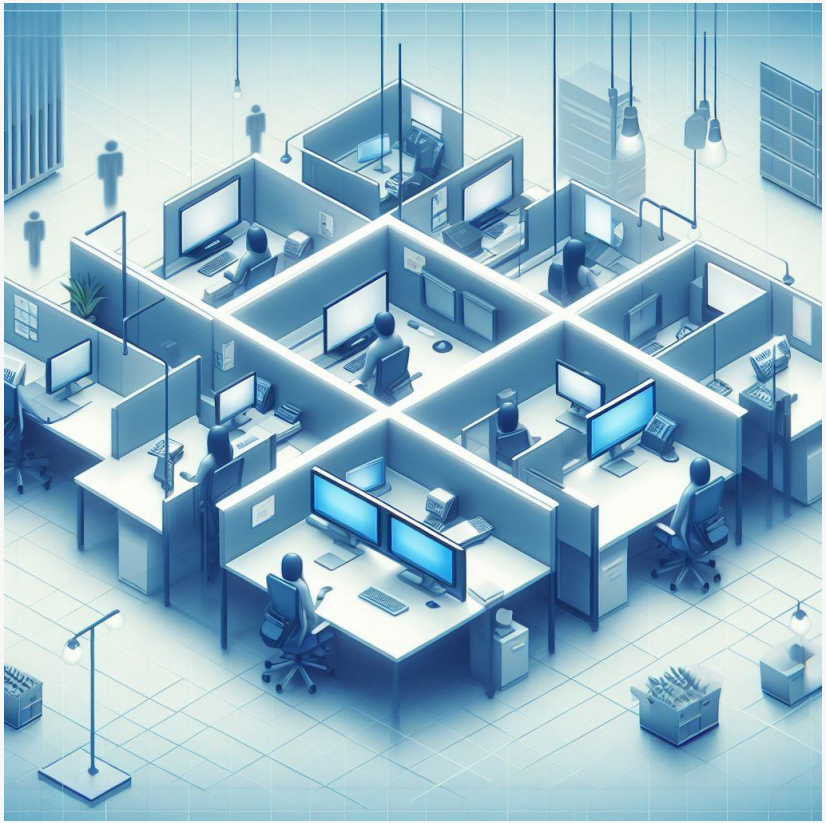
02

Problem Statement



2. Problem Statement

The problem that our MVP tries to solve is that we want to improve the customer’s experience when buying products from the company, cutting down the cost of hiring many employees.



2. Problem Statement

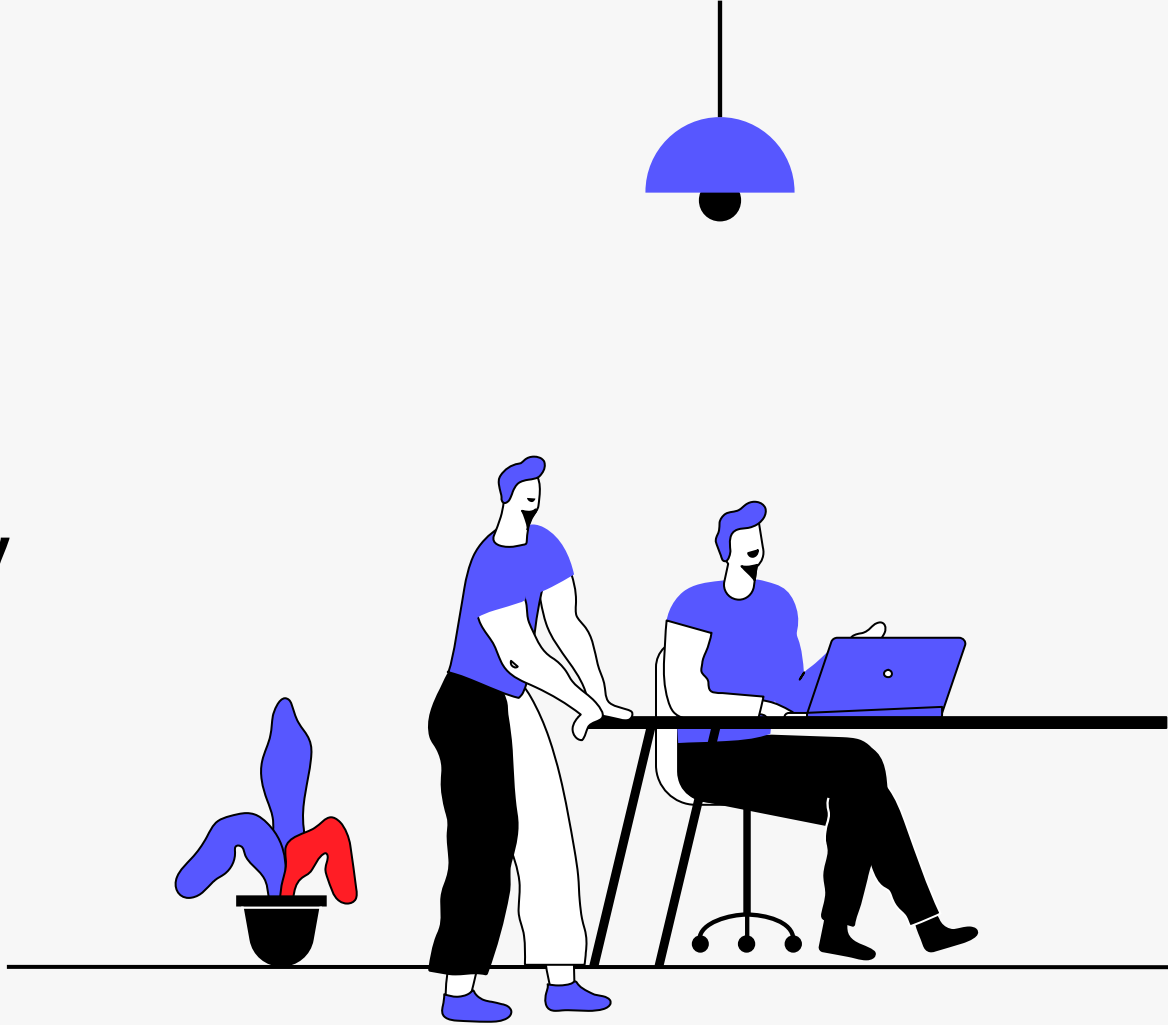
When it comes to retail, especially with large companies having many customers, consulting for a large number of customers will be extremely expensive if we don’t leverage AI solutions. Traditional consultants won't be able to handle this large number of customers, especially when the trend of buying online is increasing over time.

Many competitors now also leverage AI solutions, especially the LLMs to improve their customer’s experiences. This also cuts down a lot of cost for hiring personnel

VIETNAM DATATHON 2023

03

Solution Overview



3. Solution Overview

To answer the question: Present a high-level overview of the AI-based solution proposed in the MVP.

We intend to use a pre-trained Language Model (LLM) and then fine-tune it to serve as our chatbot consultant. This approach allows us to leverage the power of advanced NLP techniques without the need for extensive data or computational resources. The chatbot will be capable of understanding customer queries and providing a list of available products that best match the customer's demands.

3. Solution Overview

In addition to this, we plan to integrate a virtual try-on feature into the chatbot. This feature, powered by the Flow-Style-VTON model, will allow customers to visualize how different products would look on them, thereby enhancing their shopping experience and aiding in decision-making.

To ensure the chatbot can accurately interpret and respond to customer descriptions, we will incorporate other algorithms and models such as FashionCLIP for fashion-related queries and an entity recognition model for extracting key information from user inputs, which may help the FashionCLIP works better.

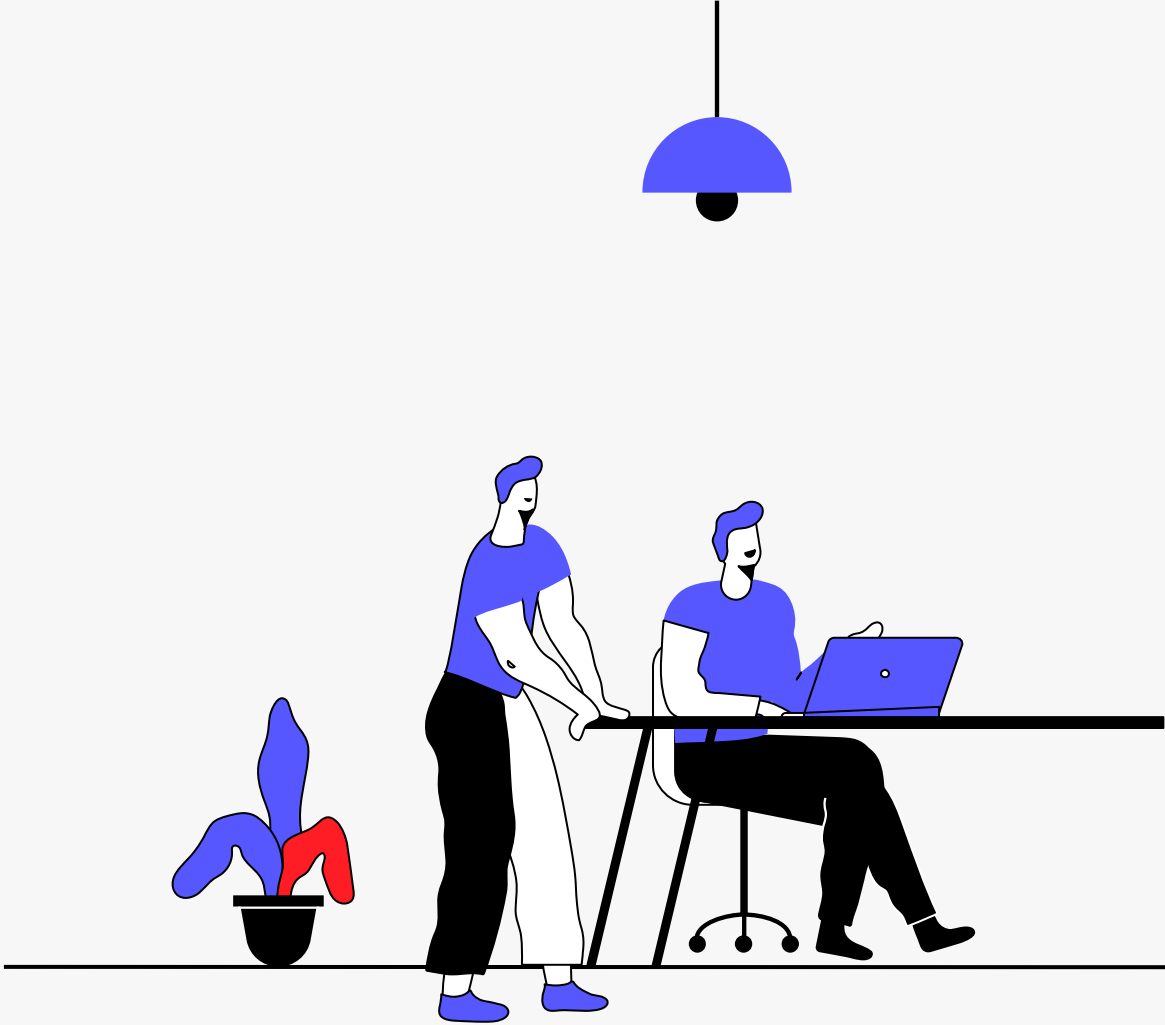
3. Solution Overview

Our solution is innovative since we will also integrate the virtual try-on to the chatbot using the Flow-Style-VTON model. To ensure the performance of consulting, we will use some other algorithms or models such as FashionCLIP, and entity recognition.

VIETNAM DATATHON 2023

04

Methodologies

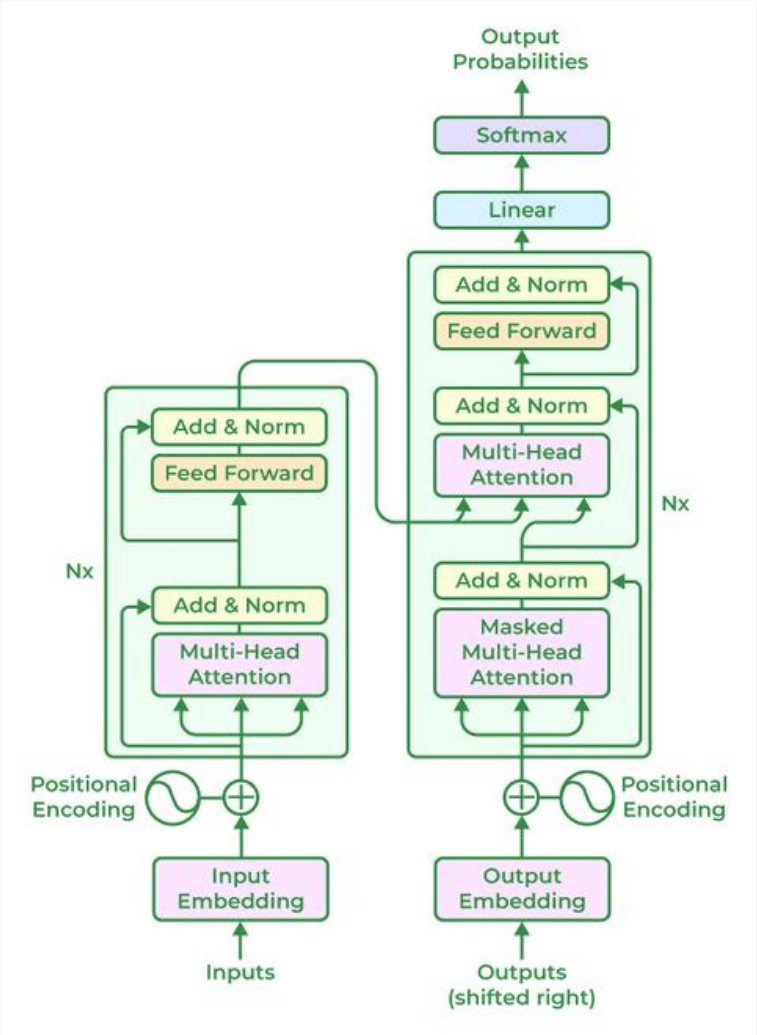


4. Methodologies

Our architecture will include 3 parts, which are the LLM, the consulting based on descriptions, and the virtual try-on. The main part is a pre-trained LLM, and integrated into this LLM is one or more than one model that will help to find the best match products for consulting. We intended to use FashionCLIP to calculate the similarity between the text description of a customer and our product's images. For better results, we might use a model that helps extract key information in the text, such as an entity recognition model (Which helps to extract the brand, color,...). We also want to use Flow-Style-VTON (a model that helps our virtual try-on part) and integrate it into our LLM

4.1: Methodologies: LLM:

A Large Language Model (LLM) is a type of artificial intelligence algorithm that applies neural network techniques with lots of parameters to process and understand human languages or text using self-supervised learning techniques.



4.1: Methodologies: LLM: keycomponents

1. **Input Embeddings**: Breaks down input text into smaller units and assigns each a **vector** representation to capture their meaning.

2. **Positional Encoding**: Adds information about the position of tokens in the sequence.

3. **Encoder**: Analyzes the input text and creates hidden states to preserve context and meaning.

4. **Self-Attention Mechanism**: Weighs the importance of different tokens in the input sequence.

4.1: Methodologies: LLM: keycomponents

5. **Feed-Forward Neural Network**: Captures complex interactions between tokens. the model can generate sequential outputs by attending to the previously generated tokens.

6. **Decoder Layers**: Generates sequential outputs by attending to previously generated tokens.

7. **Multi-Head Attention**: Allows the model to attend to various parts of the input sequence simultaneously.

8. **Layer Normalization**: Stabilizes the learning process and improves generalization.

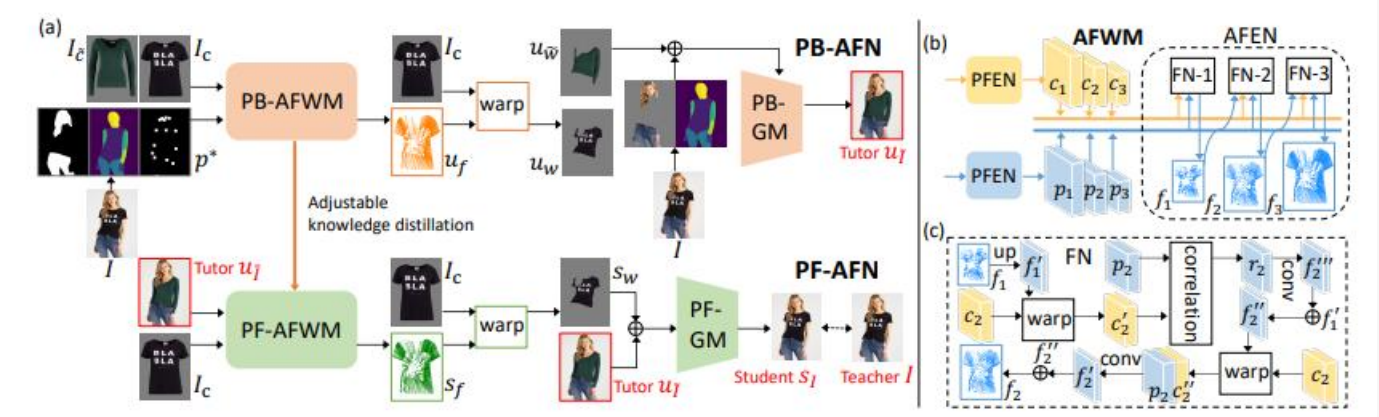
9. **Output Layers**: Varies depending on the task, often used to generate a probability distribution over the next token. Using softmax activation.

4.2: Methodologies: FashionCLIP:

The model uses a ViT-B/32 Transformer architecture as an image encoder and uses a masked self-attention Transformer as a text encoder. These encoders are trained, starting from a pre-trained checkpoint, to maximize the similarity of (image, text) pairs via a contrastive loss on a fashion dataset containing 800K products.

4.3: Methodologies: Flow-Style-VTON:

The Flow-Style-VTON model is a StyleGAN-based architecture that is used for appearance flow estimation1. It leverages a global style vector to encode a whole-image context, which helps to address various challenges.

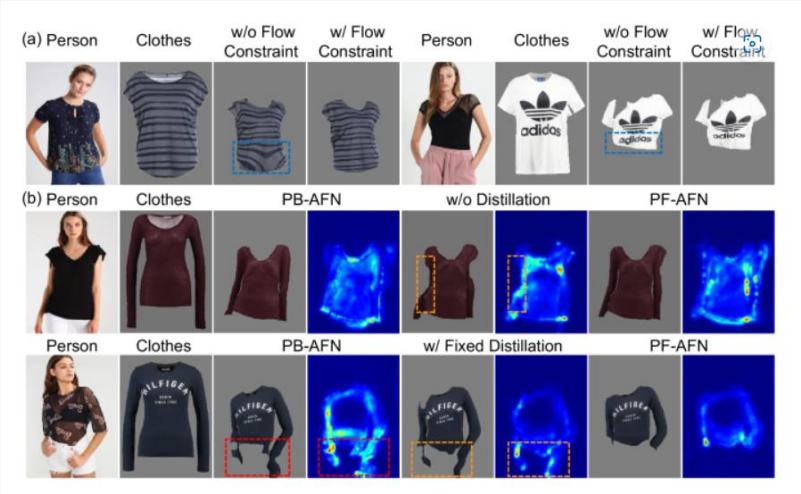


4.3: Methodologies: Flow-Style-VTON : key component

1. **Parser-Based Appearance Flow Style (PBAFS):** This is the first stage of the model. It uses paser-based to generate an appearance flow style such as segment different parts of the clothing or person might be identified and labeled.
2. **Parser-Based Generator (PBGen):** This is the second stage of the model. It uses the output from the PBAFS to generate the virtual try-on image.
3. **Parser-Free Appearance Flow Style (PFAFS):** This is the third stage of the model. It uses a parser-free approach to generate an appearance flow style by operates directly on the raw input data.
4. **Parser-Free Generator (PFGen):** This is the final stage of the model. It uses the output from the PFAFS to generate the final virtual try-on image.

4.3: Methodologies: Flow-Style-VTON : key component

Moreover: **Appearance Flow Warping Module (AFWM)**: This module is responsible for establishing accurate dense correspondences between the person image and the clothes image. It helps to fit the target clothes onto the person image. the AFWM would first fit the clothes onto the person image



$$\mathcal{L}_{sec} = \sum_{i=1}^{N} \sum_{t} \sum_{\pi \in \mathcal{N}_t} \mathcal{P}(f_i^{t-\pi} + f_i^{t+\pi} - 2f_i^t)$$

4.3: Methodologies: Flow-Style-VTON : key component

Moreover: **Generative Module (GM)**: Once the clothes have been accurately fitted onto the person image by the AFWM, the Generative Module takes over. It uses the warped clothes, the preserved regions on the person image, and human pose estimation as inputs to synthesize the final try-on image. the GM would use this output, along with other inputs, to generate a preliminary try-on image

$$\mathcal{L} = \lambda_l \mathcal{L}_l + \lambda_p \mathcal{L}_p + \lambda_{sec} \mathcal{L}_{sec}$$

where $\mathcal{L}_l$ is the pixel-wise L1 loss and $\mathcal{L}_p$ is the perceptual loss [14] to encourage the visual similarity between the try-on image (\iethe output s_I of the student network) and the real image I as below:

$$\mathcal{L}_l = ||s_I - I||_1$$

$$\mathcal{L}_p = \sum_m || \phi_m(s_I) - \phi_m(I) ||_1$$

4.3: Methodologies: Flow-Style-VTON : key component

Adjustable Knowledge Distillation: This technique is used to train a smaller, more efficient model (the student) to mimic the performance of a larger, more complex model (the teacher). In the context of virtual try-on systems, this could be used to train a model that can generate realistic try-on images without relying on human parsing. Adjustable Knowledge Distillation could be used to further refine this image, resulting in a final, highly realistic try-on image.

$$\mathcal{L}_{hint} = \psi \sum_{i=1}^{N} ||u_{p_i} - s_{p_i}||_2$$

$$\mathcal{L}_{pred} = \psi \sum_{i=1}^{N} ||\sqrt{(u_{f_i} - s_{f_i})^2}||_1$$

$$\psi = \begin{cases} 1, \text{if} \quad ||u_I - I||_1 < ||s_I - I||_1 0, \text{otherwise} \end{cases}$$

$$\mathcal{L}_{kd} = \lambda_{hint}\mathcal{L}_{hint} + \lambda_{pred}\mathcal{L}_{pred}$$

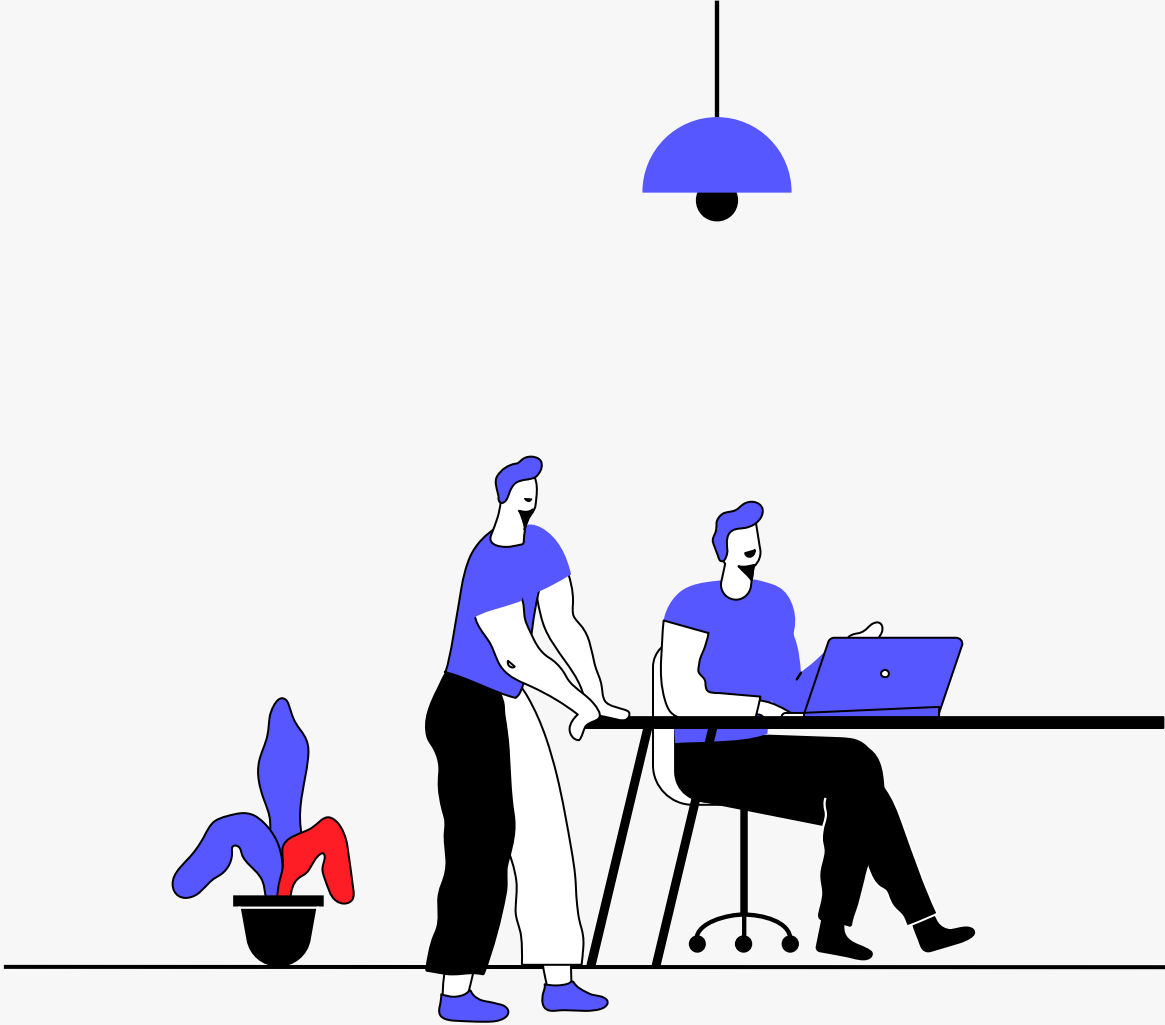
4.3: Methodologies: Technologies

1. **PyTorch** for Deep Learning framework
2. **Torchvision** for handling image data
3. **TensorboardX** for visualization
4. **OpenCV2** for processing task
5. **Human Parser, OpenPose, Dense Pose:** these tool are used for preparing the training data.

VIETNAM DATATHON 2023

05

Core Functionality



5: Core Functionality: Chatbox

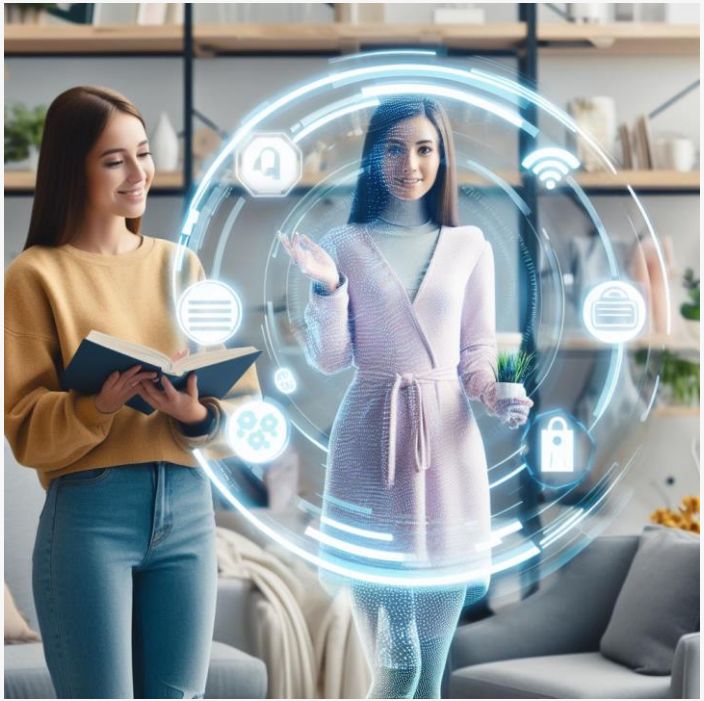
Chatbot Consultant: The MVP features a chatbot consultant that can interact with customers in real time. It can understand and respond to customer queries, helping them find the best match products based on their descriptions.

5: Core Functionality: Product Recommendation

Product Recommendation: Based on the customer's description of what they want, the chatbot can recommend products that best match the customer's needs. This feature leverages advanced natural language processing techniques to understand the customer's preferences and suggest relevant products.

5: Core Functionality: Virtual Try-On

Virtual Try-On: The MVP also includes a virtual try-on feature. Customers can upload an image of themselves, and the system will use this image to create a virtual model of the customer. The customer can then “try on” different products virtually, allowing them to see how the products would look on them before making a purchase.



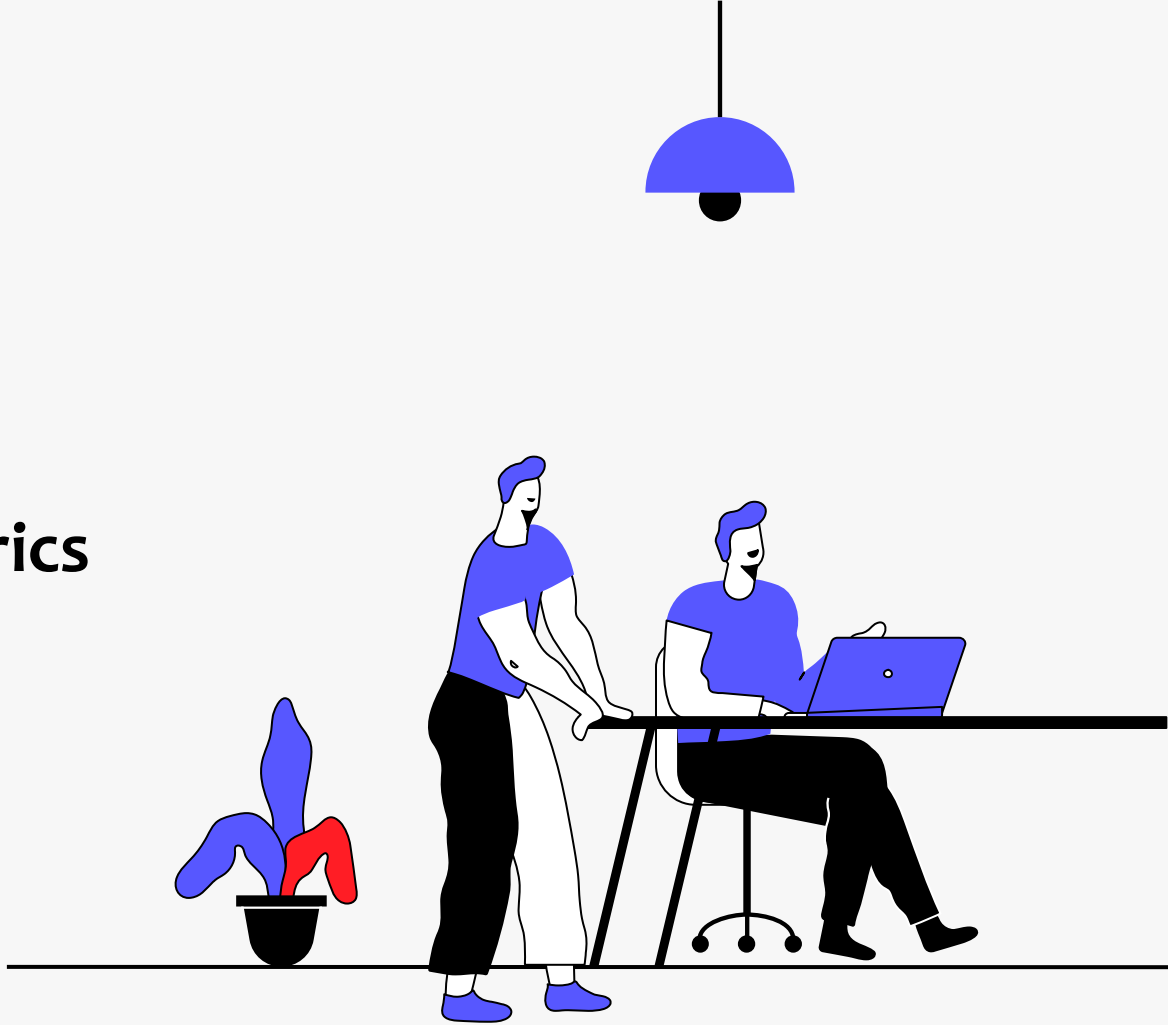
5: Core Functionality: Multi-Language Support

Multi-Language Support: To cater to a global customer base, the chatbot supports multiple languages. This ensures that customers from different regions can effectively interact with the chatbot.

VIETNAM DATATHON 2023

06

Performance Metrics



6: Performance Metrics

FashionCLIP:

The key performance metrics is the latency, the accuracy of FashionCLIP.

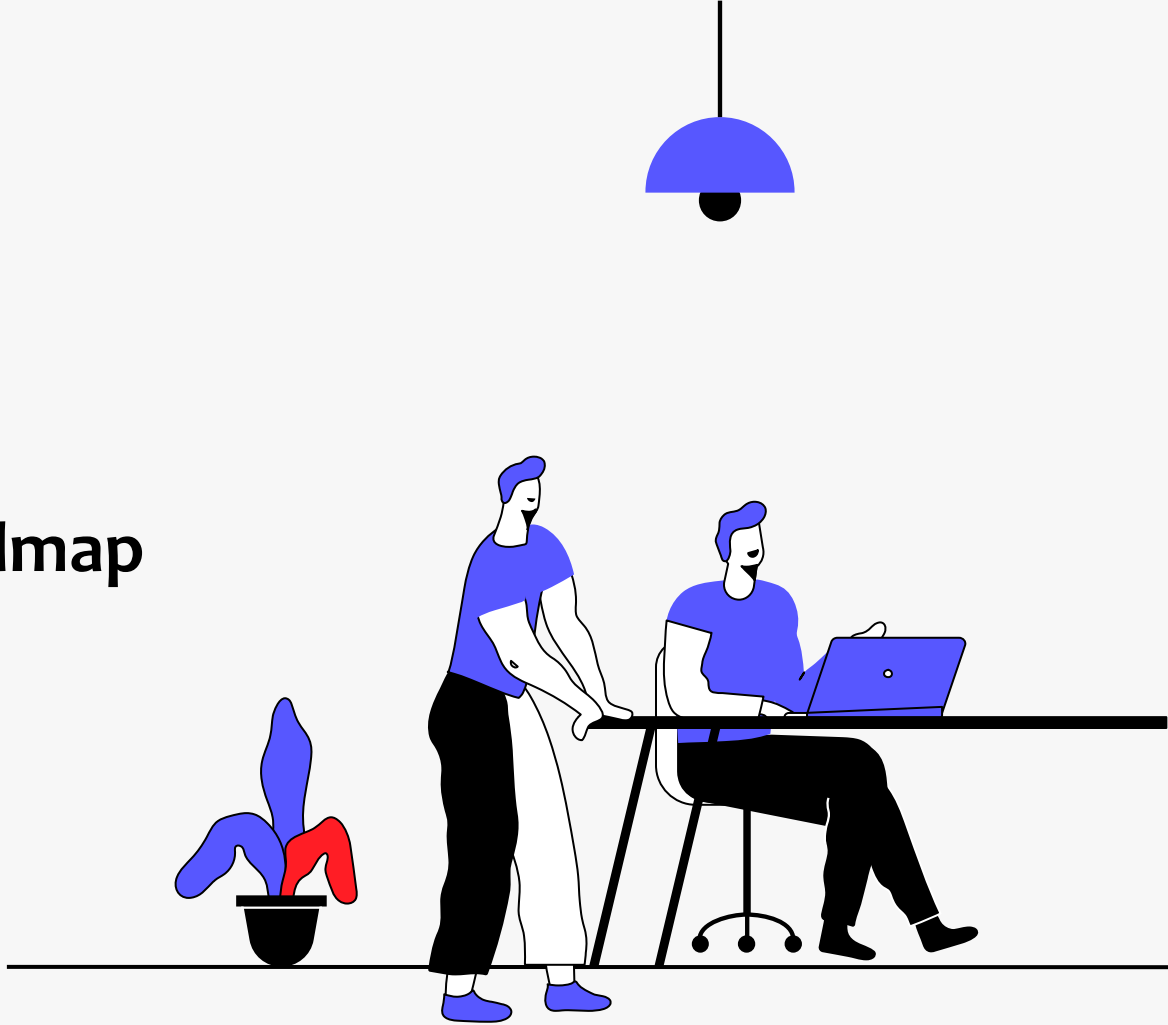
Flow-Style-VTON:

Frechet Inception Distance (FID): FID measures the similarity between two datasets of images. It was shown to correlate well with human judgement of visual quality and is most commonly used to evaluate the quality of samples of Generative Adversarial Networks (GAN). FID is calculated by computing the Frechet distance between two Gaussians fitted to feature representations of the Inception network.

VIETNAM DATATHON 2023

07

Timeline and Roadmap



7: Timeline and Roadmap

→ Chosing and finetuning a LLM
→ Integrate FashionCLIP (and an entity recognition model) into the LLM
→ Integrate Flow-Style-VTON model for virtual try-on
→ Deploy to Web

VIETNAM DATATHON 2023

08

User Interface (UI) or Interaction

8: User Interface and Interaction

Chat Interface: The primary interface is a chat window in the website where users can communicate with the AI chatbot. Users can type in their queries or requirements, and the chatbot responds in real-time. The chat interface is designed to be intuitive and easy to use, even for users who are not tech-savvy.

Image Upload: For the virtual try-on feature, users can upload an image of themselves using a simple file upload button. The interface provides clear instructions on how to take and upload a suitable photo.

8: User Interface and Interaction

Product Display: Once the chatbot has understood the user's requirements, it displays a list of recommended products. Each product is displayed with an image, a brief description, and a 'try-on' button.

Virtual Try-On: When a user clicks the 'try-on' button, the system uses the uploaded image to create a virtual model of the user wearing the selected product. This image is displayed in a separate window, allowing the user to see how the product would look on them.

8: User Interface and Interaction

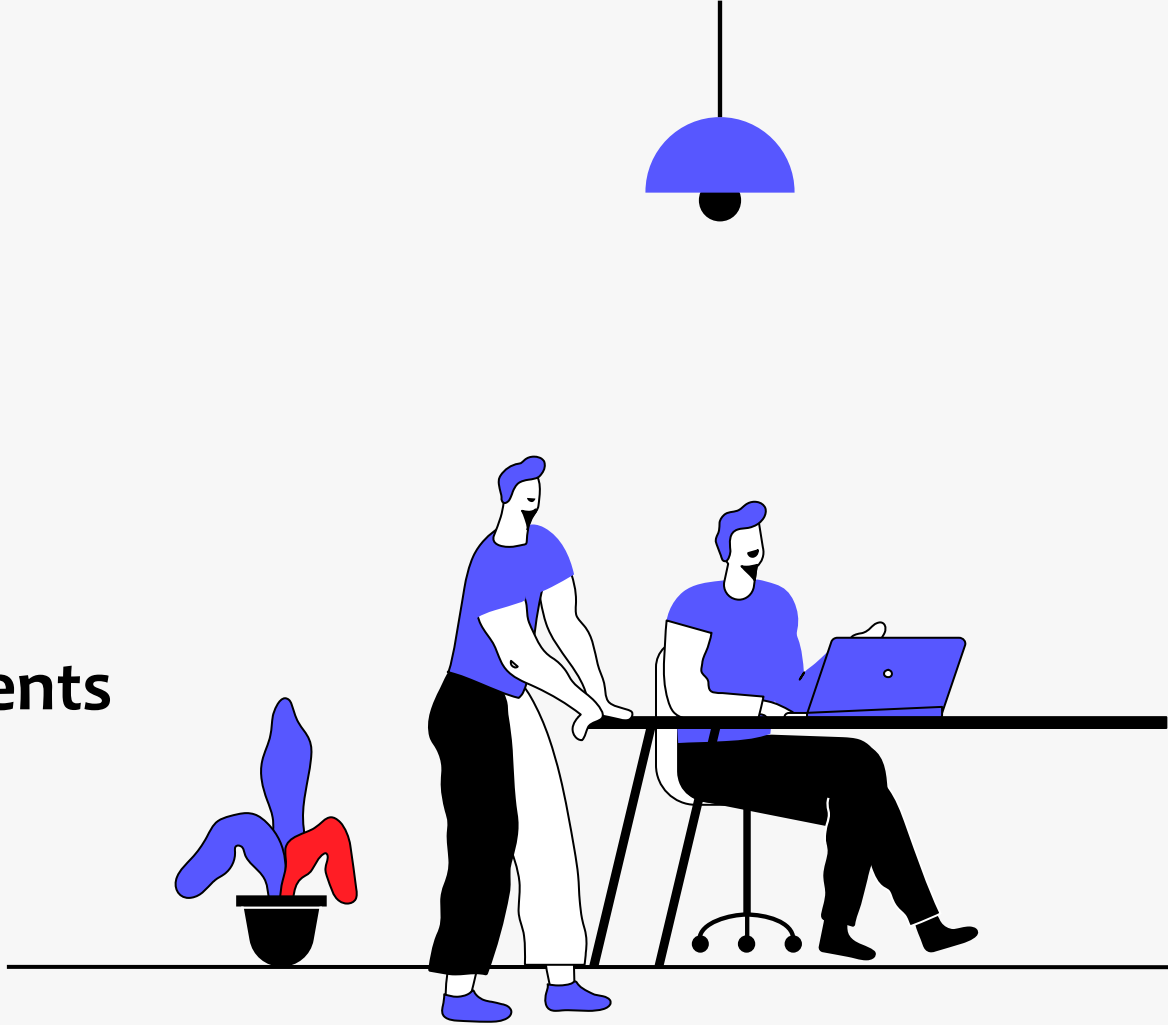
Feedback and Rating: After the virtual try-on, users can provide feedback or rate the recommended products. This helps the system learn and improve its recommendations over time.

Checkout Integration: If a user decides to purchase a product, they can do so directly through the chat interface. The system is integrated with the e-commerce platform’s checkout process, making it easy for users to complete their purchase without leaving the chat.

VIETNAM DATATHON 2023

09

Limitations and Future Enhancements



9: Limitations:

Performance: Our MVP depends a lot on the performance of LLM (to provide natural response) and FashionCLIP (calculate the similarity between description text and image of product).

Dependence on Quality of User Images: The effectiveness of the virtual try-on feature relies heavily on the quality of the images uploaded by the users. Poor lighting, low resolution, or images taken from unusual angles could affect the accuracy of the virtual try-on.

Lack of Personalization: The MVP might not offer personalized recommendations based on user's past behavior or preferences as it primarily relies on the user's current description.

9: Enhancements:

Performance: More data, computation, and training time can improve the performance of LLM. FashionCLIP may be more performant in longer queries than shorter ones.

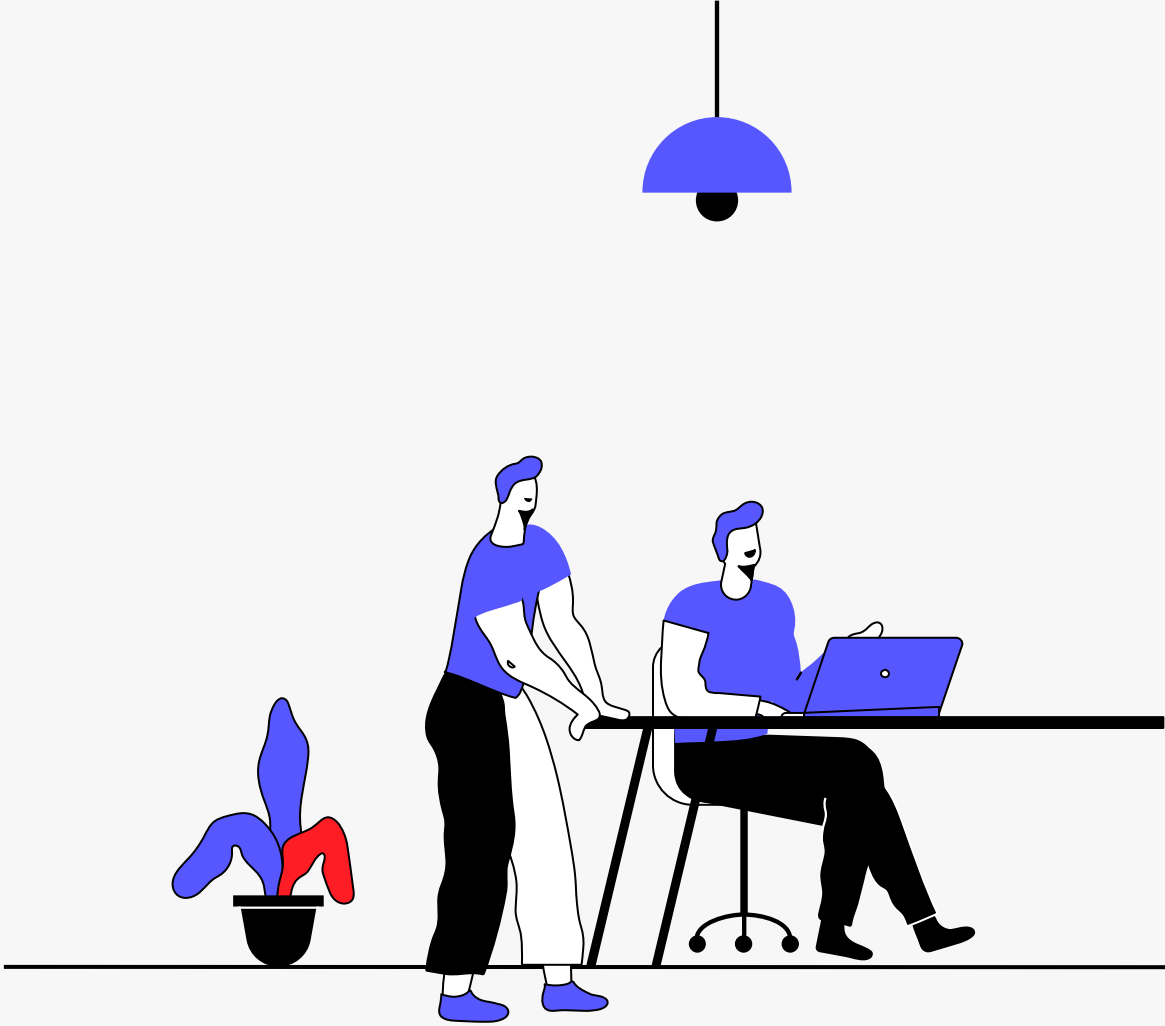
Dependence on Quality of User Images: guideline for user, check image quality and using techniques to improve the quality of image

Lack of Personalization: Allow users to create a profiles where they can save their preferences and past behavior. This information can be used to offer personalized recommendations. Implement a feedback system where users can rate or review items. This information can be used to improve the personalization of the recommendations.

VIETNAM DATATHON 2023

10

Conclusion



10: Conclusion:

First, this MVP: key points:

1. **Broad Customer Reach:** The MVP aims to reach almost all customers, capitalizing on the growing trend of e-commerce.

2. **Chatbot Consultant:** The MVP is a chatbot that helps customers find their best match products based on their descriptions.

3. **Virtual Try-On:** The MVP plans to implement a virtual try-on feature based on input images, further enhancing the user experience.

10: Conclusion:

Second, this MVP: Value proposition:

1. **Personalized Shopping Experience:** By understanding customer descriptions, the chatbot can provide a personalized shopping experience, helping customers find products that best match their needs and preferences.

2. **Convenience:** The virtual try-on feature allows customers to visualize how a product would look on them, providing a convenient and engaging shopping experience from the comfort of their homes.

10: Conclusion:

Third, this MVP: Potential Impact and Benefits of the AI Solution:

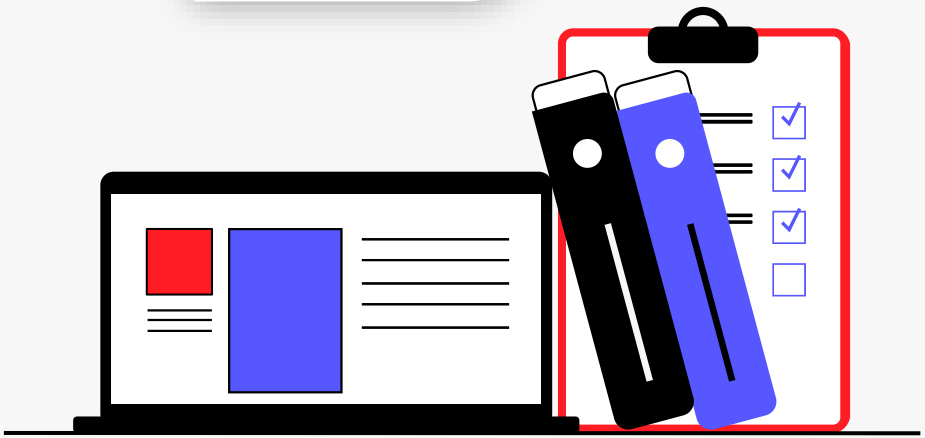
1. **Increased Customer Satisfaction:** By providing personalized recommendations and a virtual try-on feature, the AI solution can significantly enhance the shopping experience, leading to increased customer satisfaction.

2. **Increased Sales:** A more satisfying shopping experience can lead to increased sales, as customers are more likely to purchase products that they can visualize on themselves and that match their described preferences.

3. **Competitive Advantage:** In the competitive e-commerce industry, offering a unique and engaging shopping experience can give

VIETNAM DATATHON 2023

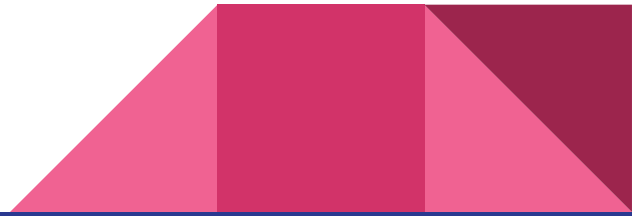
Thanks!



Vietnam Datathon 2023

Team 3T1H

Contents



Introduction

Customer Experience Statistic 2023

Introduction - Consumer Experience

Investment in CX is critical to bottom-line growth

- *80 percent of leaders plan to increase customer service budgets over the next year. (Zendesk CX Trends Report 2023)*
- *73 percent of customers now say CX is the number one thing they consider when deciding whether to purchase from a company (PwC).*

Customer expectations are on the rise

- *Customers are 2.4 times more likely to stick with a brand when their problems are solved quickly. (Forrester)*
- *72 percent of customers want immediate service. (Zendesk CX Trends Report 2023)*

Introduction - Customer Experience

The promise of AI has yet to be fully realized

- 80 percent of executives have reported demonstrable improvements in customer satisfaction, delivery of service, and overall contact center performance as a result of implementing conversational AI. (MIT Technology Review)
- 57 percent of business leaders feel that conversational chatbots deliver a large ROI on minimal investment. (Accenture)

Personalization drives loyalty

- 88 percent of online shoppers are more likely to continue shopping on a retailer website that offers a personalized experience, including 96 percent of Gen Zers and 97 percent of Millennials. (Elastic)
- 60 percent of consumers report that they will become repeat buyers after a personalized purchasing experience. (Twilio)

Introduction to MVP

Our MVP is conversation chatbot to enhance customer experience, helping customers answer questions about return, repair and warranty policies as quickly and effectively as possible.

Problem Statement

Problem Statement

- Customers often encounter product-related problems such as damage, manufacturer errors, and difficulty installing the product.
- **BUT** the company's customer service department is often overloaded and cannot resolve the entire issue immediately.
- **AND** Competitors are willing to pay a substantial amount of money for advanced AI technologies in products that help provide the best customer experience.

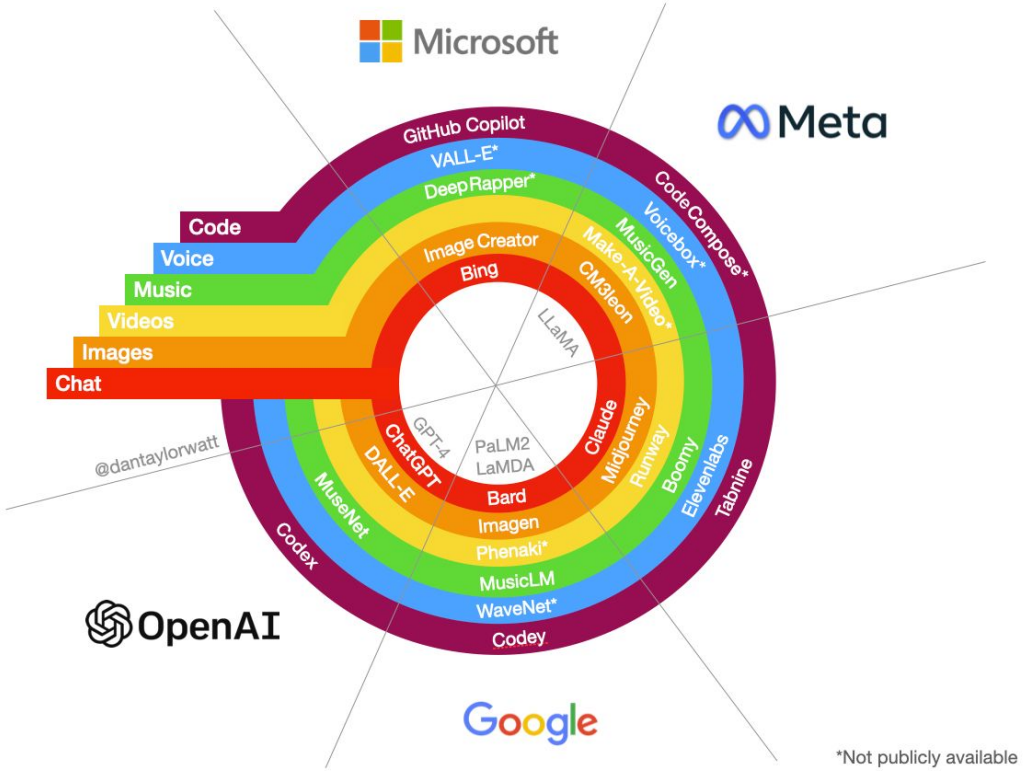
Problem Statement

- The problem that needs to be solved by MVP is to improve customer experience, helping customers answer questions about return, repair and warranty policies as quickly and effectively as possible.
- Reduce pressure on the customer consulting department and reduce costs for the business.

Solution Overview

Solution Overview

We plan to use an open source pretrained large language model (Document Retrieval & Question-Answering Models) and then fine-tune it on the company's policy dataset to produce better responses based on domain-knowledge. Moreover, AI can able to learn from customer feedback to increase effectiveness.



Solution Overview

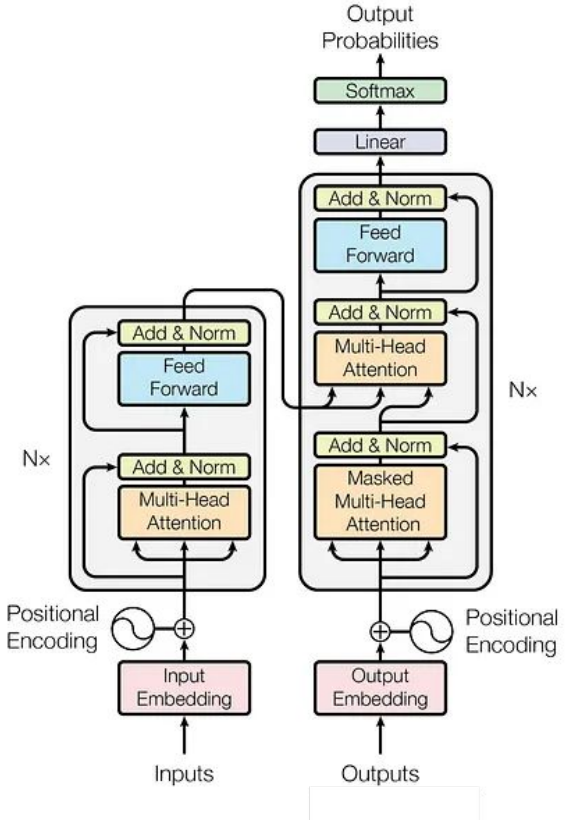
- Our solution increases accuracy, reliability and saves costs by making use of pretrained models from technology giants.
- The innovation of this solution is that AI can automatically learn more from customer feedback and become more and more understanding of customers.
- This solution helps users be more satisfied with the company's purchasing experience, reduce the burden on customer services. Thereby bringing profits to the business.

Methodologies

Methodologies - Transformers-base Models

BERT

Encoder



GPT

Decoder

Methodologies - Transformers-base Models

Transformer: Encoder-Decoder

The transformer encoder-decoder architecture is used for tasks like language translation, where the model must take in a sentence in one language and output a sentence in another language. The encoder takes in the input sentence and produces a fixed-size vector representation of it, which is then fed into the decoder to generate the output sentence. The decoder uses both self-attention and cross-attention, where the attention mechanism is applied to the output of the encoder and the input of the decoder.

Methodologies - Transformers-base Models

Transformer: Encoder

The transformer encoder architecture is used for tasks like text classification, where the model must classify a piece of text into one of several predefined categories, such as sentiment analysis, topic classification, or spam detection. The encoder takes in a sequence of tokens and produces a fixed-size vector representation of the entire sequence, which can then be used for classification. It applies self-attention mechanism to the input tokens, allowing it to focus on the most relevant parts of the input for the given task.

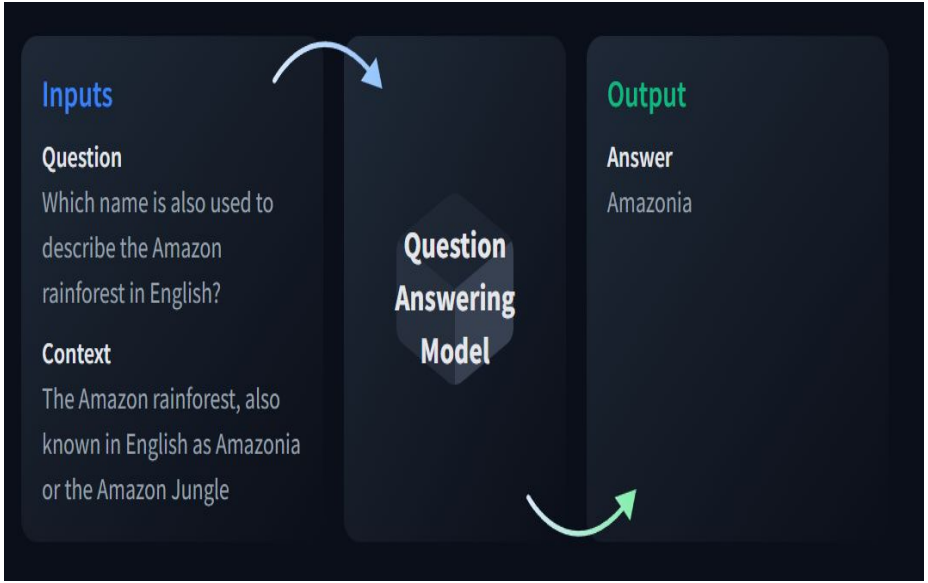
Methodologies - Transformers-base Models

Transformer: Decoder

The transformer decoder architecture is used for tasks like language generation, where the model must generate a sequence of words based on an input prompt or context. The decoder takes in a fixed-size vector representation of the context and uses it to generate a sequence of words one at a time, with each word being conditioned on the previously generated words. The transformer decoder architecture introduces a technique called triangle masking for attention, which ensures that the attention mechanism only looks at tokens to the left of the current token being generated. This prevents the model from "cheating" by looking at tokens that it hasn't generated yet.

Methodologies - Core LLMs models

- Use Question-Answering pre-trained models like from Hugging Face.
- Using English-Vietnamese, Vietnamese-English language translation models from VinAI translation models.
- Apply Semantic Analysis model to understand customer emotions and behavior.
-



Methodology - Techniques

- Crawl documents related to the company's return, warranty, and product recall policies from IKEA's official website (Using Selenium)

Methodology - Techniques

- Store documents in the Chroma database.
- Use Langchain and Prompting Engineering to manipulate LLM models.

Methodology - Techniques

- Deploy the chatbot using frameworks such as Streamlit, and Gradio to create a user interface for communication between the chatbot and customers.

Core Functionality

Core Functionality

Responding to customer inquiries related to the company's policies such as:

+ Returns
+ Warranty
+ Repairs

Recalling defective products

Providing additional information and usage tips to ensure product durability and longevity

Suggesting the purchase of replacement parts for damaged components.

Creating special policies for loyal and valued customers

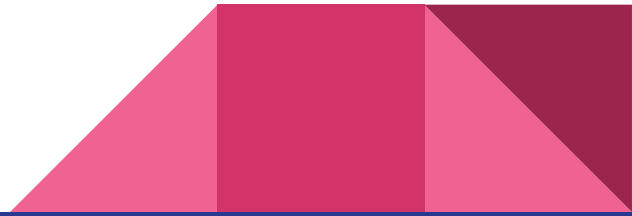
Personalizing the customer experience, sending thank-you emails,...

Offering promotions on special days,...

Performance Metrics

Performance Metrics

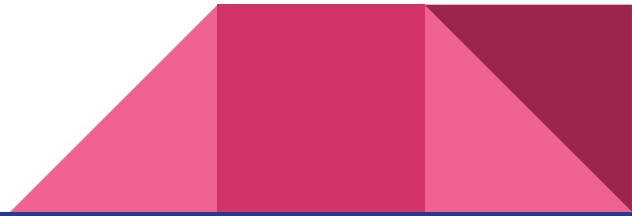
- ROUGE metric
- Customer satisfaction from feedback



Performance Metrics

- **Exact Match** is a metric based on the strict character match of the predicted answer and the right answer. For answers predicted correctly, the Exact Match will be 1. Even if only one character is different, Exact Match will be 0

- The **F1-Score** metric is useful if we value both false positives and false negatives equally. The F1-Score is calculated on each word in the predicted sequence against the correct answer



User Interface (UI) / Interaction (optional)

User Interface (UI) / Interaction (optional)

- Description of the user interface or interactive components of the MVP
- Explain how users interact with the powerful AI feature.

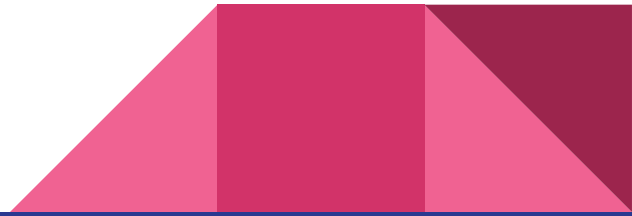
User Interface (UI)

The website features a Frequently Asked Questions (FAQ) section, a text input box for users to engage in a chat, and displays answers from the system. An emoji bar is included to indicate user satisfaction with the responses.

Limitations & Future Enhancement (optional)

Limitations & Future Enhancements (optional)

- Acknowledge any limitations and constraints of the MVP.
- Discuss potential improvements in the future or features that you may integrate.

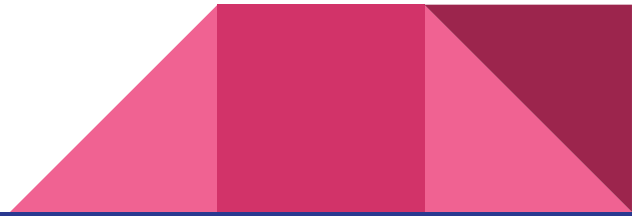


Future Enhancements

Add features to support customers in assembling products.

Introduce tailored promotional programs for each customer to encourage additional purchases and maintain loyal customer engagement.

Possibly incorporate voice support and image search functionality.



Conclusion

Conclusion

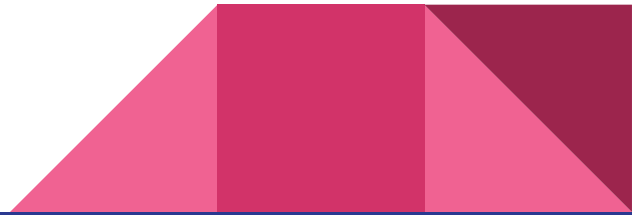
- Summarize the key points and reiterate the value proposition of the MVP.
- Highlight the potential impact and benefits of the Al solution.

The MVP Objectives:

+ Create an AI-powered chatbot to assist customers in understanding IKEA's return and warranty policies.

AI-Based Solution:

+ Utilize a data scraping model to extract information from IKEA's PDF documents, storing the data in the ChromaDB database.
+ Select a powerful machine learning language model capable of fine-tuning to effectively address user queries through the chatbot.
+ Employ Langchain for flexible interaction with language models.



The Value of the MVP:

Automation of Information: The solution automates the process of gathering and providing information on return and warranty policies, alleviating the burden on customer support resources.

Performance and Accuracy: Fine-tuning the Large Language Model (LLM) enhances the performance and accuracy of the chatbot, ensuring precise and comprehensive responses to user queries.

Natural Interaction: The chatbot interacts through natural language, creating a smooth and efficient user experience.

Resource Optimization: Minimizing manual tasks and enhancing customer self-service capabilities, the solution optimizes business resources

Highlight the potential impact and benefits of the AI solution.

+ Swift and Efficient Enhancement: The chatbot enables customers to swiftly access accurate information, reducing response time and increasing efficiency.
+ Intelligent Feedback: Gathering data from interactions with the chatbot to provide intelligent feedback helps businesses understand customer needs and concerns comprehensively.
+ Improved Customer Experience: Offering a modern and user-friendly approach, creating a positive customer experience, and enhancing their loyalty.
+ Resource Savings: Alleviating pressure on customer support departments, the chatbot aids businesses in resource savings and strengthens customer service capabilities