Digital Sales Assistant

A Conversational Recommendation System

1st Eduardo Silva
Data Engineer
Critical TechWorks
Porto, Portugal
1180881@isep.ipp.pt

2nd Michel Silva
Engineering Lead
Farfetch
Porto, Portugal
1200169@isep.ipp.pt

3rd Miguel Costa
Software Engineer
Critical TechWorks
Porto, Portugal
1210064@isep.ipp.pt

4th Rui Oliveira
RPA Developer
Sonae MC
Porto, Portugal
1210067@isep.ipp.pt

***Abstract*—E-commerce has become, in recent times, the preferred shopping method for clients all around the world, as it allows them to purchase the products they need at any time, from the comfort of their homes.**

Despite this, when compared to physical retail, E-commerce still lacks the personalization provided by sales assistants, who accompany the client all the way from the product search and recommendation to checkout, with the benefit of being able to talk with the person and understand their preferences and needs.

As such, the team proposes the development of a digital sales assistant, supported by deep learning and natural language processing techniques, allowing the client to find the products they need, as if talking with a real sales assistant, through the continuous analysis of the client's needs, intentions and emotions.

As a proof of concept, the team also presents a solution implemented with state of the art techniques (e.g. Word2Vec) and models (GPT-3, LSTM, GRU), already capable of helping the client find the products they need and adapt its behaviour based on the client's feedback, while also showing flexibility and openness to new functionalities and further improvements to its dialog skills.

***Keywords*—Retail, E-commerce, Sales Assistant, Conversational Recommendation System, Artificial Intelligence, Deep Learning, Natural Language Processing, Word2Vec, Transfer Learning, GPT-3.**

I. Introduction

With E-commerce platforms becoming increasingly prevalent in retail, it is important to provide the clients with the best possible shopping experience, closing the gap or even surpassing the one found in traditional physical retail spaces.

In the past, the team had the opportunity to work on this very same topic, having implemented a **customer satisfaction diagnosis** tool, supported by an expert system that analysed client activity in an E-commerce platform and generated action plans to recover the confidence of less satisfied clients in the company's services.

Besides that, a **physical store layout planning** tool has also been implemented, that resorted to genetic algorithms and machine learning techniques to automatically come up with a familiar yet strategic store layout and product placement, aiming to improve the client's shopping experience, while also maximizing the company's sales and revenue.

In the present article, the team will present yet another solution that aims to close the aforementioned gap and produce the best and most personalized retail experience for every client around the world: a **digital sales assistant**.

When comparing physical retail with E-commerce, user experience has proven to be the biggest challenge for the latter, with personalized client service on online platforms proving to be more complex than in physical stores [1], since factors analysed by sales assistants while giving product recommendations, such as sentiment, emotion and body language, are not as easily accessible.

With a digital sales assistant, the team believes that clients will have an easier and more natural method to interact with their favorite online store, going as far as helping those who have avoided using E-commerce platforms because they are hard to navigate or to find the products they want, while also maximizing the conversion rate.

It's worth noting that, currently, the average conversion rate in an E-commerce platform is around 1%, which shows a lot of effort from the client's side, only to end the shopping session without a single purchase [2].

In a case study conducted by Satisfi Labs in 2019, the integration of a virtual assistant in Wicked, a official Broadway site, led to a 700% increase in its ROI and a 20% higher average ticket price since launch, having provided visitors with a more seamless ticket purchasing process [3] [4].

In the next few chapters, the reader will be presented with a state of the art analysis of the techniques (e.g. Word2Vec) and machine learning models (e.g. LSTM, GRU, GPT-3) applied in the proposed solution and how each of them contribute to the implementation of the digital sales assistant.

Afterwards, an overview of the proposed solution will be presented, along with its benefits and limitations when compared to other existing solutions.

Lastly, the various phases of the solution's development will be followed, loosely based the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, all the way from the business and data understanding to the implementation, evaluation and deployment phases, followed by some closing remarks by the team after this project's conclusion.

II. State of the Art

In order to produce a state of the art chat bot that is able to accompany and help a client in his shopping and find the products he needs, the application of state of the art techniques also becomes a requirement.

As such, in the following sections, the reader will be presented with a brief analysis of the current state of (i) deep and transfer learning, (ii) chat bots and (iii) conversational recommendation systems, in order to understand how each of these topics integrate with each other and relate to this project's subject, while also identifying what techniques are being applied by researchers and retail companies in the present day to tackle the problems involved in each of them.

A. Deep and Transfer Learning

Despite being a field with significant breakthroughs throughout the years, Natural Language Processing (NLP) has benefited greatly from the advancements made in the field of Deep Learning in the past decade, with some of the most successful applications of architectures such as Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) being in the field of NLP (e.g. machine translation, text generation).

Despite the success of these architectures, as NLP problems grew more complex, the need for more data and more complex models grew with it.

In 2018, one of the biggest breakthroughs of the field was made, with the proposal of the BERT deep learning model, which stands for Bidirectional Encoder Representations from Transformers [5].

With the introduction of BERT, the field of NLP started experiencing the benefit of Transfer Learning, a technique that had already been popular for many years in the field of Computer Vision, which consists in pre-training a complex model with a large dataset in order to solve a general set of problems (e.g. generate text) and only need some fine-tuning, with a smaller dataset, to obtain a state of the art model for a more specific problem (generate chat bot responses) [5] [6].

Since then, many have been the models proposed by the scientific community to further improve the benefits provided by Transfer Learning in this field. One such case has been the GPT models, developed by OpenAI, which is currently in its 3rd iteration: GPT-3, released in 2020 [7].

GPT-3 is described by OpenAI as an autoregressive language model with 175 billion parameters, 10 times more than any previous non-sparse language model [7], as shown in figure 1.

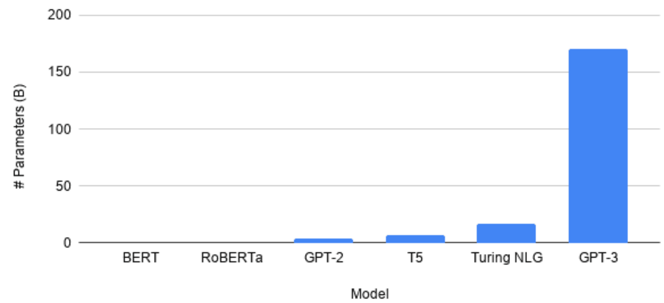

Fig. 1. Comparison of no. of parameters between models

In the GPT-3 proposal article, the writers test the model under 3 different circumstances known as zero-shot, one-shot and few-shot, which means that the model is given a description of the problem and no examples, one example or many examples, respectively [7].

As expected, since the model has been pre-trained to solve general tasks, it performed relatively well, even when given only a description of the problem, although it shows significant improvements when provided a set of examples [7].

The graph in figure 2 displays an accuracy comparison between 3 different models (i.e. GPT-2, Turing NLG and GPT-3, in ascending order of parameters), under the aforementioned circumstances, with the GPT-3 model emerging as a significant upgrade when compared to its previous version, GPT-2, which only had around 1.3 billion parameters, compared to the 175 billion that comprise GPT-3.

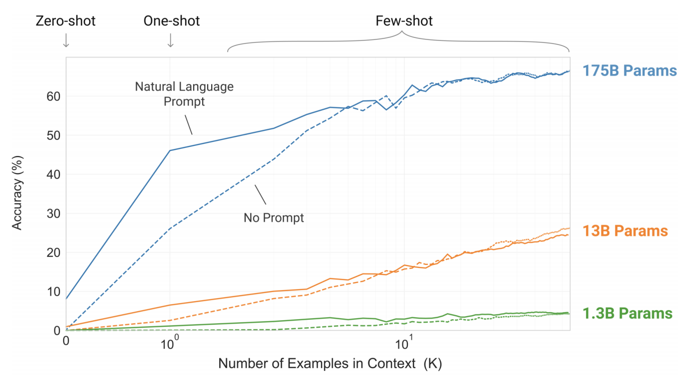

Fig. 2. Zero-shot, One-shot and Few-shot model comparison

The impressive performance displayed by GPT-3 has not gone unnoticed by the scientific community, with many projects being built around it, all the way from source code generators (e.g. GitHub CoPilot [8]) to conversational systems (e.g. chat bots [9]).

The performance shown under fine-tuning with a relatively small dataset and its proven effectiveness in the implementation of chat bots, gives the team confidence to use GPT-3 in this project's implementation, while also seeing it as a good opportunity to apply state of the art techniques to obtain the best possible solution.

B. Chat Bots

Chat bots, defined in the lexicon as a ”computer program designed to simulate conversation with human users, especially over the internet”, have risen in popularity over the last decade, with users finding in them a quick and convenient productivity tool to interact with [10].

In the article ”An Overview of Chatbot Technology”, it is stated that when designing the architecture of a chat bot, the requirements include a language understanding module, a dialogue management component, an accurate knowledge representation and, finally, an answer generation strategy [10].

These requirements result in the general architecture represented in figure 3.

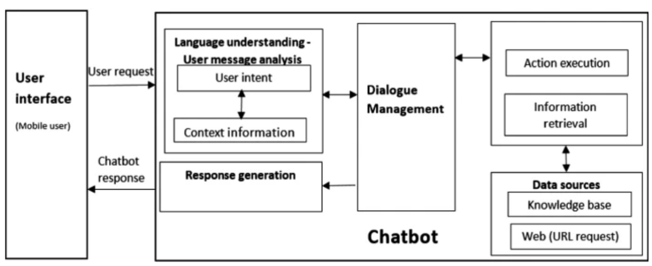

Fig. 3. General chat bot architecture

With the rise in popularity of E-commerce platforms in the past few years and chat bots becoming increasingly more effective and natural, due to the advancements in the field of NLP, the demand for digital assistants in these platforms has risen, such as the one proposed in this article, with many possible end goals, ranging from shopping assistance to customer service.

Examples of such chat bots are the aforementioned solution implemented by Satisfi Labs [3] [4], which helped clients with the ticket purchasing process, and SuperAgent, a customer service chat bot for E-commerce websites, that answers frequently asked questions from users regarding products in the platform [11] (e.g. what is the screen resolution, does it have a HDMI port).

It is worth noting that the lack of empathy and friendliness of a chat bot remains one of the biggest challenges in the field, as it impacts the customers' trust [12]. As such, the team will make it a priority to implement a chat bot that feels natural and friendly to have a dialogue with, while also adapting to the emotions and sentiment displayed by the user.

C. Conversational Recommendation Systems

Recommendation Systems have become an indispensable tool for information seeking on E-commerce platforms and content sharing platforms, as they aim to personalize the user experience and come up with the set of information that best suits the user's needs [13].

Traditional recommendation systems typically resort to the user's activity in the platform to base its recommendations on, usually with good accuracy and response time, but they are still based on an assumption established on limited knowledge about the user.

Conversational Recommendation Systems (CRS) have emerged in recent times as an alternative solution, that instead of coming up with recommendations established on top of assumptions, they are instead based on the direct input provided by the user, through a conversation, similar to how a sales assistant helps a client find the clothing item they need in a physical store, resulting in a better shopping experience and higher conversion rates [13] [14].

Despite the apparent benefits of this approach, the field of CRS is still an emerging one, with a great effort being shown by the scientific community to better understand how these systems can be implemented and what clients are looking for, despite retail companies still seeming a bit hesitant on investing their resources in it [13].

III. Proposal

In the following chapter, a brief overview of the solution and its main characteristics will be presented, complemented with an analysis of its benefits and limitations when compared to other existing solutions.

A. Architecture

Based on the state of the art techniques identified in the previous chapter, the team proposes the development of a chat bot that closely follows the architecture represented in figure 3, with a few modifications that the team felt were necessary to better accommodate the domain at hand.

The chat bot architecture proposed by the team is represented in figure 4.

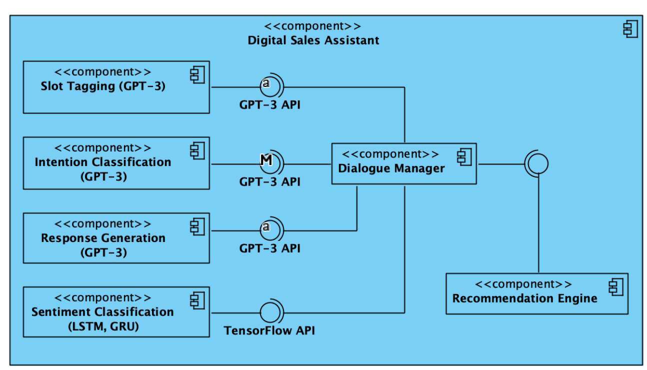

Fig. 4. Proposed chat bot architecture

At the core of the chat bot, there's a Dialogue Manager module, which is responsible for maintaining a state of the conversation and help the bot adapt its behaviour to the user (i.e. by adjusting its expressiveness, based on the sentiment shown by the user), while also interacting with the machine learning models and the Recommendation Engine, the module responsible for formulating the set of products to be shown to the user, based on the store's product catalog.

As suggested in the general chat bot architecture presented in figure 3, language understanding and generation are the two main topics to be covered by the aforementioned machine learning models in this solution, with further subtopics also being identified and distributed to different deep learning architectures as follows:

- Language understanding
  - Slot tagging (GPT-3)
  - Intention classification (GPT-3)
  - Sentiment classification (LSTM/GRU)
- Language generation
  - Response generation (GPT-3)

Each topic and the reasons for the architecture distribution will be covered with further detail in the following Development chapter, but the latter essentially comes down to the available data and problem complexity.

B. Usage and Functionality

The usage of the digital sales assistant does not deviate significantly from that of a typical chat bot, with the user being able to talk with the assistant, either through text or voice, the assistant analysing and coming up with the best response and then delivering it to the user, repeating until the user's needs are fulfilled and the conversation ends.

It is worth noting that the complexity of this problem resides in the intermediate phase, which comprises of the utterance understanding, with the user's intention and sentiment being classified and slot tagging also being performed in this initial stage, and the response generation, based on the user's intention and respective action that is triggered, combined with the obtained results.

In figure 5, a brief visual overview of the solution's usage and functionality is represented through a flowchart.

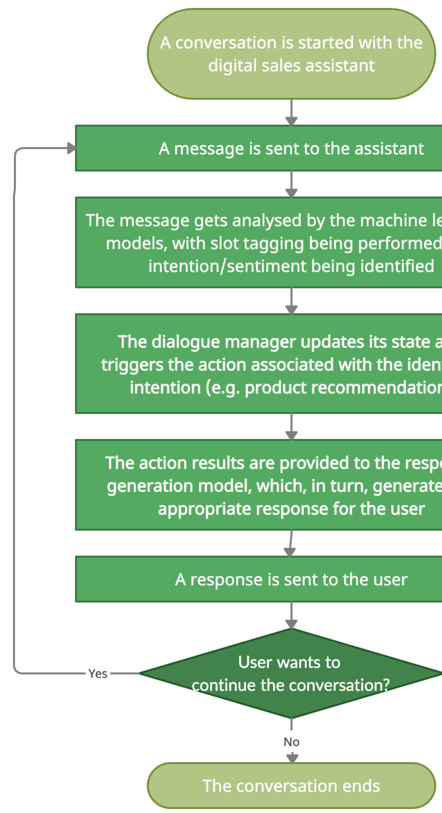

Fig. 5. Overview of the digital sales assistant usage and functionality

C. Benefits

When compared to other existing solutions, the solution proposed in this article contains a set of benefits that help it stand out on its own as an innovative approach, starting with the use of GPT-3 in many of the covered topics, which helps the assistant feel more natural to interact with, while also requiring very vew examples in order to work well.

On the other hand, the Dialogue Manager used in this solution is capable of maintaining a state/context of the entire conversation, all the way from the chat log to the last identified sentiment and current bot personality parameters, which further contributes to the natural conversations that can be had with the assistant.

Lastly, the recurring analysis of the user's sentiment allows the assistant to adapt its personality and expressiveness according to the obtained feedback, which helps with the personalization of the shopping experience, something that has been identified as currently lacking in E-commerce platforms and the current chat bot implementations.

D. Limitations

With the aforementioned benefits, there also come a few limitations when compared to other existing solutions.

On the one hand, with the increase of actions/intentions supported by the chat bot, the addition of new examples for the GPT-3 models also becomes necessary, which is currently done manually, in order to support more diverse scenarios and questions (e.g. allow the user to ask questions about the product, such as the type of fabric or the available colors).

On the other hand, the dialogue manager, in its current implementation, only supports a small set of actions, with the biggest priority for the proof of concept being identified by the team as the product recommendation. In the future, further actions would have to be implemented in order to make the assistant more appealing for retail companies to invest in and integrate with their E-commerce platforms.

IV. Development

In the following chapter, a summary of the various phases of development undertaken during this project's duration will be presented.

It's worth noting that the chapter's structure is loosely based on the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, starting with a business and data analysis, proceeding with data preparation and solution implementation, and finishing with its evaluation and deployment.

A. Business Understanding

In order to implement a digital sales assistant that aims to bring the level of personalization that a human sales assistant provides, it is necessary to analyse and understand how they work and how they accompany a client in his shopping.

Since one of the main objectives of this project is product recommendation, the team found that the best way to approach this analysis is to understand how a sales assistant talks with the client and what information is retained in order to guide the client to the product he wants.

Through this analysis, the team found that a sales assistant usually looks for two sources of information in a dialog with a client: (i) the conversation itself, in order to understand the type of clothes that the client wants, as well as its attributes (e.g. color, size, value); (ii) the client's feedback, based on the body language and tone of voice, for example, in order to understand the client's emotions and sentiment, which in turn allows the assistant to adapt its approach to each client (e.g. be more professional and direct, or be warmer and more expressive).

A similar approach should be taken by the digital sales assistant, with the first source of information being covered by the slot tagging and the second being covered by the intention and sentiment classification. It is worth noting that these problems have been simplified for this project, taking into account its scope and time span, as they are more complex and require more tools in order to more closely resemble a real sales assistant (e.g. emotion classification, voice analysis).

In the following sections, a closer look into each of these topics and how each of them was covered/implemented will be presented to the reader.

B. Data Understanding and Preparation

For this project, many datasets were required in order to implement the several modules that compose the proposed solution, namely:

- **a product catalog**, for the Recommendation Engine;
- **a chat dataset**, in the retail domain, with tagged slots and labeled intentions, for the Slot Tagging, Intention Classification and Response Generation models;
- **a sentiment classification dataset**, containing user utterances labeled with the respective sentiment, for the Sentiment Classification model.

With each dataset, many challenges were found and many steps were taken in order to prepare the data to be used in the model training, which will be presented shortly.

- Product Catalog

In order to implement a proof of concept, the team decided to pursue a product catalog dataset to use as a knowledge base for the digital sales assistant the make recommendations on.

Given the familiarity with the domain, a clothing dataset was sought by the team, having ultimately found a french clothing store dataset, which required some cleaning and processing to be usable in the solution.

As such, the team implemented a ETL (Extract, Transform, Load) pipeline in order to ingest the raw dataset and produce a clean dataset with over 200 clothing categories, ready to be used as the solution's knowledge base.

The aforementioned ETL pipeline comprises of steps such as outlier removal (e.g. categories with few items), label translation (i.e. from french to english), column filtering and feature engineering (e.g. adding a *price range* column, which determines if an item is cheap, accessible or expensive).

- Chat Dataset

For the chat bot to be implemented, tasks such as slot tagging, intention classification and response generation require a dataset with various examples of interactions between a client and a sales assistant, in order to train models capable of performing these tasks as well as a human sales assistant.

Unfortunately, such as specific dataset is hard to obtain, more so in internet, as this kind of sensitive and private information is not usually published on public platforms such as Kaggle.

As such, the team found in GPT-3 the solution for this, since the model is known to perform well even when provided with only a few examples, given how generalized and complex the neural network is.

By building a small dataset, with a few dozens of examples of interactions, the team was able to train models capable of performing these tasks and generalizing to any given scenario with remarkable ease and effectiveness, which demonstrates the effectiveness of transfer learning when applied to NLP and opens the opportunity for the development of projects that previously did not have the necessary data to achieve a satisfactory degree of quality.

Below, an example taken from this dataset is presented to the reader:

> **User:** Hello, I am looking for a men's t-shirt.
> **Assistant:** Hello. What color would it be?
>
> **Intention:** Looking for a product
>
> **Slots**
> Product: T-shirt
> Gender: Men
> Color: ?
> Size: ?

It is worth noting that GPT-3 is a text generation model and, as such, the examples are provided with a structure similar to the one shown above, with no notable pre-processing being required.

When provided with the utterance *"User: I am looking for a blue t-shirt, please"*, the GPT-3 model generates the rest of the text based on the examples it was fine-tuned with, such as the one presented, with the response it finds appropriate, the intention of the user and the slots tagged accordingly (e.g. if this uttterance was provided as a follow-up to the presented example, the tag *Color* would be updated with the value *Blue*).

- Sentiment Classification Dataset

For the sentiment classification model, a dataset with utterances labelled with the respective sentiment (i.e. either positive or negative) was not hard to find, as it is a classic NLP problem.

Similarly to the chat dataset, the challenge was finding a public dataset in the domain of retail, particularly in related to clothes and fashion, the sub-domain the team decided to focus on.

To amplify the problem, the datasets found by the team did not meet the quality and size required to properly train a deep learning model, at least not to a satisfactory degree.

As such, the team decided to instead use one of the most used and mature datasets available publicly, a IMDB movie review dataset.

Given its maturity and successful applications in sentiment classification problems, while also not deviating too much from the field of retail, the team found it to be the best choice, given the resources and time-frame of this project.

The maturity of the dataset also meant that not much cleaning had to be performed, as it came with 50k examples, evenly distributed for both the positive and negative classes, and no missing or incomplete data being found during its analysis.

Despite this, it must be noted that the user utterances themselves are raw data taken directly from IMDB, with no text pre-processing being applied. As such, the team saw this as a good opportunity to implement a pre-processing pipeline that cleans and normalizes the text, resorting to libraries such as SpaCy and Gensim, with the complete list of steps being presented below:

- Lowercase the text;
- Remove e-mails;
- Remove HTML tags;
- Remove URL's;
- Remove special characters and ponctuation;
- Remove accented characters;
- Convert words and verbs to their base form (e.g. **me** becomes **I**, **ran** becomes **run**);

It is worth noting that the team experimented with further pre-processing steps such as stop word removal and spelling correction, but decided to ultimately leave them out of the pipeline, as they were computationally demanding and, in some cases, actually resulted in worse models, since some of text became way too simplified, to the point of not having enough information to make a proper prediction (e.g. utterances such as *"These shoes do not fit me well"* were converted to *"shoes fit"*).

Another NLP technique that the team found interesting to apply and, ultimately, resulted in the most significant increase in model's performance, was Word2Vec.

The team decided to use a pre-trained model, containing over 400k word vectors with a dimension of 100 values, trained on a Wikipedia + Gigaword corpus with over 5.6B tokens [15], which increased the accuracy of the studied models by over 10%, significantly more than the 2% to 3% increase in accuracy gained from the pre-processing pipeline.

C. Implementation and Evaluation

In the following sub chapter, a brief overview of the implementation and evaluation process for each component of the proposed solution will be presented.

- Slot Tagging, Intention Classification and Response Generation

Since these tasks are all performed by GPT-3, it was decided that these modules should all be covered in the same section, as the following observations are common to all of them.

As stated previously, being a pre-trained model, GPT-3 only requires a small dataset in order to fine-tune a few layers and provide a state of the art model for the task at hand.

By using the API provided by OpenAI for its GPT-3 model, the fine-tuning process is further simplified, by allowing the provision of the dataset and fine-tuning of the model at runtime, instead of training it *a priori*.

Since the dataset is relatively small and the model runs on capable hardware, the fine-tuning process at runtime is seamless and the model predictions are instant, with the obtained model matching the expectations set by the team at the start of this project: a chat bot that talks in a very natural way and shows effectiveness in the other tasks at hand, with only a few examples being provided.

- Sentiment Classification

For the sentiment classification problem, the team decided use TensorFlow in order to implement and study the effectiveness of LSTM and GRU architectures and how the complexity of the entire model's architecture affects its performance, both in terms of time (e.g. training time) and prediction (e.g. accuracy).

Through some experimenting, the team settled on the architecture presented in table I for the base classification model, with either a LSTM or a GRU bidirectional layer at its core.

TABLE I
BASE MODEL ARCHITECTURE

Layer	Input Dimension	Output Dimension
Input	(300, 1)	(300, 1)
Embedding	(300, 1)	(300, 100)
Batch Normalization	(300, 100)	(300, 100)
Dropout	(300, 100)	(300, 100)
Bidirectional LSTM/GRU	(300, 100)	(100, 1)
Dense (Output)	(100, 1)	(2, 1)

With the intention of analysing the impact of increase in complexity of the model's architecture, the team decided to also implement the architecture presented in table II, where two LSTM/GRU layers are stacked on top of each other, with a Droput layer in the middle in order to mitigate the overfitting that results from the increase in model complexity.

TABLE II
STACKED MODEL ARCHITECTURE

Layer	Input Dimension	Output Dimension
Input	(300, 1)	(300, 1)
Embedding	(300, 1)	(300, 100)
Batch Normalization	(300, 100)	(300, 100)
Dropout	(300, 100)	(300, 100)
Bidirectional LSTM/GRU	(300, 100)	(300, 80)
Dropout	(300, 80)	(300, 80)
Bidirectional LSTM/GRU	(300, 80)	(40, 1)
Dense (Output)	(40, 1)	(2, 1)

It is worth noting that the inclusion of layers such as Dropout and Batch Normalization comes as a result of the experiments, as they helped mitigate the overfitting that was found in some of the tests.

In order to comprehend the impact of the increase in dimensionality of bidirectional layers, the team decided to train models with the same base architecture and increasing dimensionality, with the evaluation results being presented in table III.

TABLE III
DIMENSIONALITY IMPACT ANALYSIS

Dim.	LSTM (Acc.)	LSTM (AUC)	GRU (Acc.)	GRU (AUC)
15	0.65	0.66	0.77	0.80
30	0.75	0.77	0.78	0.80
50	**0.80**	**0.81**	0.78	0.80
100	0.78	0.79	**0.83**	**0.83**

From the results presented in table III, two conclusions can be taken.

On the one hand, in the case of GRU, it is clear that the increase in dimensionality of its bidirectional layer had a significant and positive impact in its performance, with the most complex model performing the best out of all the studied models, with 83% accuracy and an AUC of 83%.

On the other hand, it is clear from the case of LSTM, that the increase in dimensionality may have a positive impact in the model's performance, but only until a specific point, at which the model may start to overfit and actually present worse results.

It is important to note that in the case of GRU, this would have eventually happened as well, in case the dimensionality kept being increased past 100 nodes in the bidirectional layer.

In order to keep benefiting from the increase in dimensionality, techniques and layers such as regularization and dropout can be employed, respectively, although at some point the only solution may be the increase in data volume.

After this experiment, the team decided to execute one more experiment, this time to understand the impact of the text pre-processing and Word2Vec technique, as well as the stacking of multiple LSTM/GRU layers, on the model's performance.

As such, four different models were trained during this experiment, varying between the base and stacked architecture and between LSTM and GRU bidirectional layers, with three separate datasets being used, each one having different degrees of pre-processing techniques employed to them:

- **Dataset 1:** Raw text;
- **Dataset 2:** Pre-processed text;
- **Dataset 3:** Pre-processed text and Word2Vec applied.

The results of the aforementioned experiment are presented in table IV.

TABLE IV
PRE-PROCESSING AND LAYER STACKING IMPACT ANALYSIS

Model	**No pre-processing (Acc.)**	**Pre-processing (Acc.)**	**Pre-processing + Word2Vec (Acc.)**
Base LSTM	0.67	0.70	0.79
Base GRU	0.64	0.66	**0.83**
Stacked LSTM	**0.69**	0.70	0.74
Stacked GRU	0.67	**0.71**	0.78

By verifying the results presented in table IV, it is clear that with an increase in pre-processing done to the dataset's text, significant gains in performance also come with it.

By applying the pre-processing pipeline to the dataset's text, the accuracy of the best model increases by 2%, while the application of Word2Vec further increases the performance by 12%.

On the other hand, it is interesting to note that without Word2Vec, and more so without any pre-processing, the models with stacked architectures performed best, with the application of Word2Vec significantly improving the performance of the base architecture models (i.e. Base GRU's accuracy increased by 17%) and even allowing them to surpass the performance of the stacked architecture.

As such, given the shorter training time, performance and computational resources necessary to execute the base architecture model, while applying Word2Vec and when compared to the stacked layers architecture, the team concludes that this is the preferred architecture for the problem at hand and settled on the Base GRU model obtained in the last experiment as the model to deploy in the proof of concept.

- Dialogue Manager

At the core of the solution sits the Dialogue Manager, the module responsible for receiving the user's utterances, interacting with the machine learning models in order to extract information from them and then executing the appropriate actions, such as coming up with product recommendations for the user, which are then sent back to the user.

Besides this, the module is also responsible for maintaining the conversation's state/context, which in turn allows the bot to adapt its personality and expressiveness based on the user's feedback and past utterances.

This is easily achieved with GPT-3 by dynamically updating the dataset with the personality that the bot should have and by adjusting hyperparameters such as temperature, which determines the expressiveness of the model.

- Search Engine

Bearing in mind that one of the main objectives of the digital sales assistant is product recommendation, the team worked on a module that not only offers excellent performance on this task, but is also easily integrated with the rest of the system.

At its core, the Recommendation Engine resorts to ElasticSearch, one of the most popular document-based databases, commonly used in the implementation of search engines, due to its maturity, performance and ease of use.

When provided with the slots tagged by the GPT-3 model (e.g. clothing category, color, size), the Recommendation Engine is able to query the ElasticSearch database and retrieve a list of recommended items to be presented to the user, based on the preferences and needs shown during the conversation.

D. Deployment

Being a proof of concept, the team found the idea of developing a web application interesting, in order to demonstrate a real use-case of the proposed solution.

As such, the team designed the web application represented by a component diagram in figure 6, which comprises of a single page application in the front-end and a REST API in the back-end, with the latter containing the chat bot and all its components, all the way from the Dialogue Manager and Recommendation Engine to the machine learning models, as presented in figure 4.

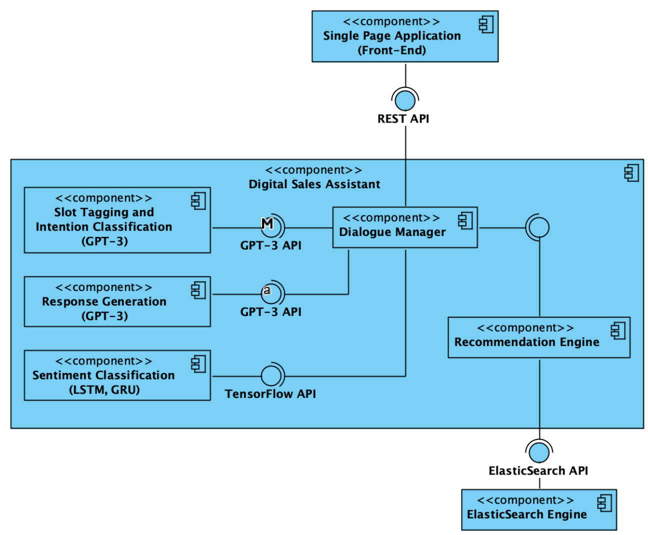

Fig. 6. Web application architecture

It must be noted that as of the delivery of this article, the implementation of the web application's front-end has not been finished, as such, a screenshot of its usage could not be included.

Nevertheless, the front-end essentially comprises of a typical chat interface with the digital sales assistant, closely following the usage illustrated in figure 5 and allowing both text and voice input, with the assistant's recommendations being dynamically presented to the user as the conversation goes on.

V. Conclusion

When looking towards the future of E-commerce, the demand for a more personalized experience and alternative interaction methods is becoming increasingly higher, as retail companies are continuously trying to close the gap still existant between physical and online retail.

With this project, the team hopes to bring a new solution to the table, a new way for clients to interact with their favorite E-commerce platforms and convert the skeptics, who have been avoiding these platforms for years, as they were never comfortable with its user experience and keep coming back to physical retail whenever possible.

As stated previously, most retail companies are still skeptical on the integration of alternative methods to interact with their platforms, as they still don't see the value in them. Despite the benefits presented by the proposed solution, many are the limitations still holding it back from being a must have in these companies' minds.

Currently, as a proof of concept, the team only implemented a small set of functionalities for the bot, essentially regarding product recommendation. Nevertheless, in order for the solution to catch retail companies' attention, a larger set of features would have to be implemented, building towards the main goal of the digital sales assistant: being able to accompany the entire shopping process of the client, all the way from the product recommendation to the checkout process.

Besides this, further improvements and fine-tunings could alrea dy be implemented for the existing functionalities, contributing to an increasingly more natural and smart chat bot, such as emotion recognition and analysis of the client's voice.

To conclude, the team hopes that with this article, a valuable contribution is made for the scientific community and more awareness is brought upon digital sales assistants and the field of conversational recommendation systems, as they are increasingly more important and relevant topics for the future of retail.

References

[1] A. Granov, "Customer loyalty, return and churn prediction through machine learning methods: for a Swedish fashion and e-commerce company," 2021.

[2] M. Rakhra, G. Gopinadh, N. S. Addepalli, G. Singh, S. Aliraja, V. S. G. Reddy, and M. N. Reddy, "E-Commerce Assistance with a Smart Chatbot using Artificial Intelligence," 2021.

[3] N. Mejia, "Natural Language Processing in Retail – Current Applications," 2020. [Online]. Available: https://emerj.com/ai-sector-overviews/nlp-retail

[4] Satisfi Labs, "Wicked Launches Broadway Ticketing — Case Study," 2019. [Online]. Available: https://satisfilabs.com/portfolio/wicked-4

[5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018.

[6] S. Ruder, M. E. Peters, S. Swayamdipta, and T. Wolf, "Transfer Learning in Natural Language Processing," 2019.

[7] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. Mccandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," 2020.

[8] A. Romero, "GitHub Copilot — A New Generation of AI Programmers," 2021. [Online]. Available: https://towardsdatascience.com/github-copilot-a-new-generation-of-ai-programmers-327e3c7ef3ae

[9] L. Floridi and M. Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," 2020.

[10] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," 2020.

[11] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan, and M. Zhou, "SuperAgent: A Customer Service Chatbot for E-commerce Websites," 2017.

[12] X. Cheng, Y. Bao, A. Zarifis, W. Gong, and J. Mou, "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure," 2021.

[13] C. Gao, W. Lei, X. He, M. de Rijke, and T.-S. Chua, "Advances and challenges in conversational recommender systems: A survey," 2021.

[14] D. Jannach, A. Manzoor, W. Cai, and L. Chen, "A Survey on Conversational Recommender Systems," 2021.

[15] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," 2014. [Online]. Available: https://nlp.stanford.edu/projects/glove