Retail Platform with Natural Interaction: LOAD's Vision

Pedro Colarejo¹, Davide Ricardo¹, André Fernandes¹, Miguel Fonseca¹, Pedro Oliveira¹, Nidhal Cherni¹, João Abrantes¹, Hélio Guilherme² and António Teixeira²

¹LOAD Interactive, Lda., Aveiro, Portugal
²IEETA, Dep. Telecommunications and Information Technology Electronics, University of Aveiro, Aveiro, Portugal

{pedro.colarejo, davide.ricardo, andre.fernandes, miguel.fonseca, pedro.oliveira, nidhal.cherni, joao.abrantes,

Keywords: Retail, Information for Consumers, Virtual Assistants, Dialogue, Natural Language, Image Recognition, Blockchain.

Abstract: In a context of indirect sales channels, where the product reaches the final consumer through intermediaries, it is very difficult for consumers to know the origin of the product and identify the origin of problems. In this paper an innovative technological platform is proposed, oriented to the retail market, and capable of providing information from the entire product distribution chain to its various stakeholders. The platform is based on 3 pillars: a decentralized blockchain-based information network covering the path of a product from its origin to the end consumer; extraction of information about users/customers and products from images and video; interaction using natural language. A first instantiation of the platform is also presented as well as the first results. In its development, recent technologies were used in the areas of image recognition and dialogue systems.

1 INTRODUCTION

From the point of view of the global process of selling products/services by an organization to its customers, the customer service has always had a great importance in the ability of these organizations to generate revenue. This assistance to customers can consist of a set of services that help customers to know more information about a product/service they intend to purchase and help their choice. In this sense, these organizations have a higher cost to improve these services, investing in the training of employees, to understand by interacting with the clients their opinions and needs regarding the services provided. A bad provision of these services increases the probability of losing customers to the competition. Examples of poor practices in the provision of this type of services are well known, including excessive duration in problem solving, rude treatment, lack of transparency in resolution processes, not being able to give a concrete answer on the problem to be solved, promises that have not been fulfilled, etc.

In a context of indirect sales channels, where there is a value chain, and where the product reaches the final consumer from intermediaries, the customer service, during the sales process, becomes more complicated. The manufacturer who produces the product or provides the service does not reach the final consumer directly. There is a set of distributors and retailers that complete this value chain, where distributors buy the product and sell it to retailers, who will make the product available and sell to final consumers. There are advantages to this type of distribution channels, but there are also some limitations. On the one hand, the manufacturer loses control over the message that is transmitted to the end customer, as this is done by retailers and resellers, nor can it control the importance of the product facing its competitors in retail stores. Furthermore, in situations where there is a problem with the product (delivery delay, defects, etc.), it is very difficult for end consumers to understand the origin of the product and whether the problem is in the manufacturer, or elsewhere in the chain.

In recent years, organizations have invested in automated methods to provide these support services. Virtual Customer Assistants (VCA), which include chatbots, are the latest of these methods. VCAs are applications that are typically available online capable of establishing a conversation with customers via voice and/or text. VCA’s use has grown rapidly, as there is great potential in improving the provision of support services.
of customer service, both on the organization side and on the client side. Part of this is justified by the growing tendency among customers to use text and instant messaging as their first choices when communicating with an organization, rather than using the traditional voice channel (the phone). In addition, there is a growing tendency for customers to want to solve their own problems in a direct way, rather than asking someone else to do so for them.

From the point of view of organizations, the implementation of VCA, in addition to its greater efficiency, can result in a considerable reduction in costs, since the traditional voice channels, using the phone, are quite expensive, being much cheaper to keep this service fully available 24 hours a day, than the same availability with human resources.

However, there are several technological challenges to implement such assistants, capable of interacting as if they were human. The interaction platform has to be able to improve the way it interacts with customers, not only from the point of view of usability, but also from the user experience, that is, the perception with which customers are when interacting with the system, or the level of satisfaction/frustration that the customer experiences. This implementation tends to be sustained by technological advances in artificial intelligence, namely Machine Learning (ML), with applications in Natural Language Processing (NLP) and Image Recognition.

Thus, the main objective defined for this work, was the development of an innovative technological platform, oriented to the retail market, applicable in different business areas, to provide information of the entire product distribution chain to its various stakeholders. It must be based on a decentralized information network, covering the complete path of a product from its origin until it reaches the final consumer.

This article is structured as follows: in section 2, the work related to the areas most directly related to the implementation of the platform (Blockchain, image/video recognition and natural language dialogue system) is presented. In the third section is defined the main scenario to be considered, in order to realize the implementation of the prototype. In section 4 the general architecture of the system and its main components is described; in section 5 how the first prototype was implemented; and, in section 6, the first results of the integration of the various modules. The main conclusions from the work performed and some future paths are presented in section 7.

2 RELATED WORK

The development of a solution such as the one proposed in this paper requires the integration of results from various areas. In the following subsections is very briefly presented the work related to the implementation of a blockchain for the retail sector, object recognition technologies and technologies associated with conversational assistants (chatbots). No platforms or Assistants directly related to what is proposed in this paper were found.

2.2 Blockchain (for Retail)

Blockchain technology (Yaga, Mell, Roby, & Scarfone, 2019) has been attracting increasing attention in a wide range of industries due to its ability to reliably manage transaction-based applications without the need for a centralized authority. This was not possible before blockchain emerged. This technology ensures trust in transactions in a network between untrusted nodes, since all of these nodes can trade, even if they do not trust each other. Blockchain was initially developed to support secure digital currency transaction (Bitcoins) in order to prevent transactions from being improperly duplicated. However, its potential extends to several domains using peer-to-peer (P2P) architectures, where interconnected nodes make transactions with each other without using a centralized management system, such as electronic voting, purchase and sale of intellectual property, distribution of confidential medical information, or intelligent value chains applying Internet of Things (IoT), called Smart Logistics (Ückelmann, 2008; Kawa, 2012).

Great strides have been made in the use of Blockchain for application areas not related to cryptocurrencies, such as Slock.it, a company that works in smart electronic lock (Slocks), which can be unlocked with smartphones that have the appropriate token, and that can be purchased on Ethereum using Ethers. A Slock owner can rent his house or car at a price for timed access, as the door lock is automatic. Anyone interested in renting can identify the Slock, pay the requested amount and unlock the door. Another example are transactive energy companies with blockchain-based solutions, allowing you to buy and sell energy (generated by solar panels) automatically, according to the criteria defined by the user.

In relation to value chains, there is a huge potential for application of Blockchain technology. With a network based on Blockchain technology, product tracking and control is complete, since the
same information in the database is shared by all entities, where all updates are cryptographically verified and propagated over the network automatically, thus creating a record of the information history that can be validated by all entities in the network. There are several initiatives in the retail areas in pilot phases, some more advanced than others, as, for example, jewelry (e.g., De Beers¹ and TrustChain Jewelry²) and food industry (e.g., IBM Food Trust³ and Norwegian Sea Food⁴).

2.2 Object Recognition

A state-of-the-art survey of image recognition was carried out, and some technologies available on the market with the potential to integrate a first implementation of the platform were identified and characterized, such as:

- **Watson - Visual Recognition**⁵ - Developed by IBM, the Watson system offers - among many other things - the Visual Recognition API, which allows you to: sort content according to scenarios; recognize objects, faces, colors, food, text, and inappropriate content; train models for specific cases.

- **Amazon Rekognition**⁶ - The AWS cloud computing services platform has in its spectrum this Image Recognition service that allows you to identify objects, people, text, scenarios, activities, and explicit/inappropriate content in both image and video.

- **Google Mobile Vision**⁷ - Part of the Google ML Kit, in addition to identifying faces in both photographs and video, allows to find and observe facial landmarks (such as eyes, mouth, nose, etc.) providing information about whether, for example, the subject is laughing or has his eyes open. It also allows you to recognize text and present it according to its structure, thus maintaining paragraphs, lines, punctuation, etc.

- **Google Cloud Vision API**⁸ - Is part of the Google Cloud AI platform, providing detection of objects, faces, emotions, scenarios, tags, explicit or inappropriate content, text and colors.

- **Vize.ai**⁹ - This company provides the cloud solution (Vize Custom Image Recognition API) and an On-device solution (Vize MobileModels) to recognize and classify images. Unlike most, this solution requires initial training according to categories defined by the user in order to cover specific problems.

- **Microsoft Azure Cognitive Services Vision**¹⁰ - It is one of the 5 sets of services that Microsoft provides through its Cognitive Services platform.

- **Cloudsight AI**¹¹ - Provides an API that allows you to filter and categorize images as well as monitor explicit/inappropriate content.

- **Clarifai**¹² - Clarifai provides a cloud solution that allows you to tag images and videos, create your own template that is applied to a specific use case, search content through visual or tag similarities, and moderate content by identifying whether the content is explicit/inappropriate or not.

- **Imagga**¹³ - It's a PaaS that provides image recognition APIs that allow companies to build and monetize image-intensive applications in the cloud.

- **Kairos**¹⁴ - It’s one of the leading companies in the field of Facial Recognition, providing an API and SDK that allow you to detect multiple faces, identify faces (answer the question, “who is this?”), check faces (answers the question, “this is...”), detect emotions, detect age groups, detect facial landmarks, detect levels of attention and finally detect ethnicities.

A comparison of these solutions is presented in Table 1. The most promising solutions considering the objectives were those of IBM and Microsoft. However, we wanted to ensure full flexibility to change and adapt to our idea, which was not guaranteed by these platforms. It was then considered beneficial to develop solutions tailored to the project, to gain more knowledge in the area and have more flexibility to adapt the functionalities. The libraries Tensorflow, OpenCV and PyZbar were adopted.

---

¹ https://www.jckonline.com/editorial-article/de-beers-blockchain-platform/
⁵ https://www.ibm.com/watson/services/visual-recognition/
⁶ https://aws.amazon.com/rekognition/
⁷ https://developers.google.com/vision
⁸ https://cloud.google.com/vision/
⁹ https://vize.ai/
¹⁰ https://azure.microsoft.com/en-us/services/cognitive-services/directory/vision/
¹¹ https://cloudsight.ai/
¹² https://www.clarifai.com/
¹³ https://imagga.com/
¹⁴ https://www.kairos.com/
2.3 Conversational Assistants

Dialogue systems, Chabot’s and Conversational Assistants are increasingly popular due to their usefulness: they provide easier and more versatile access to large and diverse sets of information. Conversation assistants perform chatbot-like interactions and allow speech as input and outputs. Google Assistant and Amazon Alexa are popular examples.

To allow interaction in a dialog format, a typical conversation system integrates several modules: (1) Automatic Speech Recognition (ASR) that converts speech into text; (2) Natural Language Understanding (NLU) processes ASR word sequences to identify important information, such as intentions and entities; (3) The Dialog Manager (DM) manages dialogue and context information, taking into account the intention, entities and previous conversations; (4) Natural Language Generation (NLG) generates phrases; (5) Speech synthesis uses text-to-speech (TTS) to produce synthetic speech.

In recent years, there has been impressive progress in NLP and NLU technologies, particularly accelerated by the increased popularity of technologies such as the Internet of Things (IoT) and artificial intelligence. As a result, several platforms have emerged providing sophisticated NLU capabilities, one of the main components of a dialog system, and tools aimed at developing dialog/chatbot systems, of which are representative examples: TrindiKit, OpenDial, IBM’s Watson Conversation Service, RASA, (Bocklisch, Faulkner, Pawlowski, & Nichol, 2017) PyDial, Dialog Flow, Alexa Skills, ICECAPS, Emora and Plato. One example to highlight is Plato (Papangelis, et al., 2020), a tool available by Uber in 2020 that can be used to create conversational agents, supporting interactions through speech, text, or dialogue. An application in Plato consists of four main components: the dialogue that defines and implements acts of dialogue and states; the domain that includes the ontology of the dialogue and the database that the dialogue system consults; the controller who orchestrates the conversations; and the agent that implements different components of each conversation agent.

3 SCENARIO

The following scenario was defined, depending on which the prototype was designed, developed:

- A consumer in a supermarket approaches a shelf with a particular product. The chatbot/Assistant detects the presence and starts a conversation, greeting him/her and asking what it can do to help (using the right gender context).
4 PROPOSED SYSTEM

Figure 1 schematically presents the architecture proposed for the system, consisting of 3 main modules:

- Interaction with the user.
- Image, Speech, and Language Processing.
- Blockchain Infrastructure.

The User interaction module is the frontend responsible for the user's information, from questions by voice, to his/her image, or even possibly text, but also for informing the user, through questions/answers by voice, image (e.g., map) and written information displayed on the screen.

The Image, Speech, and Language Processing Module is a system backend module, which includes the following components:
- Image recognition, responsible for image processing for facial and product codes recognition.

- NLP/NLU and Dialogue Management, supporting interaction by natural language through the identification of intentions and entities in user interactions (NLU), natural language generation and dialogue management.

The third main component of the system is dedicated to the entire Blockchain Infrastructure, implemented in EOS.IO. This is where all the information added to the product (transactions, state changes, value) is stored throughout its progression in the distribution chain.

5 PROTOTYPE DEVELOPMENT

5.1 Speech and Language Processing

The simplest way to communicate with the system is through natural language, whether the text is from the user or Speech Recognition.

To recognize speech and translate it into text a speech recognition service is used (from Google). To transform system’s response to speech a module was developed in C# using Microsoft Speech SDK.

To implement the determination of users' intentions, extraction of entities (e.g., topic, day), language generation and interaction management, the Plato Research Dialogue System, previously mentioned in this article, was used.
5.2 Image Processing

The integration of image processing with the user interface, by integrating the web interface through Websockets services, is a challenge, because it is where the harmony between the machine and the human needs to occur. For this prototype, image recognition focuses on three actions: face recognition, user gender recognition, and barcode scanner. Face recognition informs the assistant regarding the start of the conversation; gender recognition is needed to shape the conversation; the barcode scanner allows product identification, so that the system can obtain information from the blockchain and answer the questions posed by the customer.

Specific solutions have been created in Python using libraries such as Tensorflow, OpenCV and PyZbar. Tensorflow was used to detect client's gender and emotions; OpenCV was used to detect a person's face; and PyZbar was used to read barcodes (Abadi, et al., 2016).

5.2 Blockchain and Communications

The code base of the EOS blockchain has been identified as being the most suitable for the targeted prototype. We worked on the integration of external libraries (AI) in smart contracts. Focusing on applying proof of concept to a real solution to real problems (blockchain for intelligent logistics), the development of a DAPP (Decentralized App) was started, to store different states of a supply chain, from the creation of raw material to assembly, and until reaching the final consumer.

Support has been developed for a set of interactions with blockchain, such as:
- Deployment of the initial contracts for basic use of the blockchain.
- Creation of the default wallet.
- Creation of accounts to deploy contracts.
- Deployment of smart contracts.
- Interaction with the smart contract through transactions.

Automatic accounts were then created, multi-indexed tables and the available API (JSON RPC) were used to communicate with the blockchain and obtain data from the recorded tables. During this process, it was concluded that files (such as images) should be saved off-chain, and then their reference saved in the blockchain. The IPFS was selected as the most promising approach.

To integrate the blockchain with the dialog, an API was developed, which also intervenes in the interaction with the IoT component (temperature, humidity, and location sensors). In this way, it is possible to read and write data to a tag, using the RFID-RC522 shield, as well as allow the sending of images captured the assistant to be stored in IPFS (the distributed system for storing and accessing applications and data). The API between blockchain and dialog allows to:
- Register users, companies, devices, and products.
- Relate each device with different types of sensors and with a company.
- Relate each product to different IDs to be able to accept different barcodes and RFID.
- Send measurement messages from the IoT component, which will be stored on the blockchain (and it was possible to save location, time, temperature, and humidity for a specific product).
- Get all the values of each registered entity.
- Receive a request for characteristics of a product (for example: send barcode and receive general information about a product).
- Receive specific questions about a product for the different sensors (e.g., product x part at temperature y on day/hour z).

6 INITIAL RESULTS

This section presents information about results for the first prototype of the platform, starting with the Blockchain-based module and ending with the interaction with the system Assistant, visible face of the system and, in a certain way, integrator of the whole system.

Once the blockchain was implemented, creating the initial contracts to record the transactions to be recorded along the distribution chain, it was integrated with a set of sensors used to simulate temperature, humidity, and location data. After this definition, it was then possible to extract the information needed (topic, a date, and a location), making a call to the blockchain API in which the topic (temperature/humidity), date and location are passed as parameters. If the blockchain response is successful, a response is constructed by Plato with the entities received and the quantity returned. The following is an example of using blockchain:

```json
{"table":"product","scope":"loadtestaccount1","code": "loadtestaccount1", "json": "true"}
{"rows": [{"key":0,"user_id":0,"name":"lettuce","description":"organic lettuce","from_product":","sub_product
```

298
As mentioned in the previous section, the interaction with the user uses image to sense a human face approaching, its gender (male, female) and feeling (positive, negative, neutral). The same camera is also used for reading the product barcode. The barcode reading process allows to read most barcode formats, as shown in Figure 2.

Figure 2: Examples of bar code reading by the developed system.

Figure 3 presents an example of an interaction with the system, using the implemented interface.

In the example presented in Figure 3, we can observe the system locating a user who approaches the system, evaluating whether it should refer to verbal forms in the male or female, and his/her feeling at that moment, to better decide how to greet and answer the customer.

After the desired information collected, a map with a selection of the product route is displayed according to the question that was asked.

The reproduction of two other real interaction examples is presented next, following the internal flow of data from the user: NLU processing, request to the blockchain API, generation of the response, and transmission of the system's global response to the user.

**Interaction Example 1**

**User:**
What is the humidity in Aveiro on March 25, 2021?

**NLU:**
Topic = humidity
Day = 25
Month = March
Year = 2021
City = Aveiro

**Blockchain:**
API request: 
product/info/0?
sensors=humidity&time=1616630400.0 &location=Aveiro

**API Response:**
{'error': False, 'code': 9000, 'description': 'Product info sent with success', 'date': {'inTime': 0, 'date': [{'sensor': 'humidity', 'value': 4, 'time': 1610983085}]}}

**System Response:**
Humidity on March 25, 2021 in Aveiro was 4%
In the first interaction example the user asks a question regarding the storage temperature of the product that was recognized through its barcode, the language processing system identifies the topic (temperature), which is a question as well as date and location, uses this information to query the Blockchain and, based on the API response of this (value=4), generates a sentence to be transmitted as a response to the user. The second example is similar, changing the topic detected in the user's sentence, now "temperature". Although not integrating the examples presented, due to space limitations, the system is equipped with the ability to question the user if she/he does not immediately provide all the information necessary to have a complete query to the blockchain. For example, if you the customer doesn’t inform the local, the system will ask for it.

7 CONCLUSIONS

This article presents the vision for a platform that facilitates consumer access to blockchain-based product information by use of natural language dialogue supported by image, speech, and language processing technologies. As a first prototype, a system capable of interaction by speech with a customer that approaches the system was obtained, distinguishing gender and initial "mood" and capable of receiving requests for information by voice, distinguishing the different entities from the expression used, consulting the information on the blockchain, and responding to the user by voice, text, and image.

In its current state of development, several aspects require evolution, such as: greater flexibility in the extraction of users’ intentions and feelings; greater ability to convey intentions and emotions using speech (e.g., timbre adjustment); evolution of the blockchain to profit from the constant evolution of this technology. The system should also be subjected to user use and evaluation in environments gradually closer to the real, allowing its user-centric evolution.

ACKNOWLEDGEMENTS

This work was co-financed by project no. 38546 under the Incentive System R&D Projects-Individual Projects of Portugal 2020.

REFERENCES


